UNIVERSIDADE
EDUARDO
MONDLANE

FACULDADE DE CIÊNCIAS

Departamento de Matemática e Informática

Trabalho de Licenciatura em

Ciências de Informação Geográfica

Aplicação de Modelos de Regressão Espacial na Análise das Perdas Não Técnicas na Rede Elétrica de Distribuição de Baixa Tensão

Caso de Estudo: **Município da Cidade de Maputo**

Autor: Ivan Belchior Sarmento Bule

Maputo, Abril de 2025

UNIVERSIDADE
EDUARDO
MONDLANE

FACULDADE DE CIÊNCIAS

Departamento de Matemática e Informática

Trabalho de Licenciatura em

Ciências de Informação Geográfica

Aplicação de Modelos de Regressão Espacial na Análise das Perdas Não Técnicas na Rede Elétrica de Distribuição de Baixa Tensão

Caso de Estudo: **Município da Cidade de Maputo**

Autor: Ivan Belchior Sarmento Bule

Supervisor: Lic. Francesse Mauro Bacião, UEM

Maputo, Abril de 2025

Dedicatória

Aos meus avós em memória,

Chadreque Simione Bule e Elisa Arone Zango.

Que sempre foram o meu azimute e rumo de direção, fontes de sabedoria, felicidade, respeito e amor.

Declaração de Honra

Eu, **Ivan Belchior Sarmento Bule** declaro por minha honra que o presente Trabalho de Licenciatura é resultado da minha investigação e que o processo foi concebido para ser submetido apenas para a obtenção do grau de Licenciado em Ciências de Informação Geográfica, na faculdade de Ciências da Universidade Eduardo Mondlane.

Maputo, Abril de 2025

_____

Ivan Belchior Sarmento Bule

Agradecimentos

Primeiramente, agradeço a Deus, por me conceder força, saúde e determinação para concluir mais esta etapa da minha vida.

Ao meu supervisor, Francesse Mauro Bacião agradeço por ter sido modelo de inspiração durante a minha formação académica, pela paciência, dedicação e orientação ao longo deste trabalho. Suas valiosas contribuições, conselhos e críticas construtivas foram fundamentais para a realização deste projeto.

A Universidade Eduardo Mondlane, seu corpo docente, em particular afectos ao Curso de Licenciatura em Ciências de Informação e em especial ao docente Mário Azarias Chelengo, pela transmissão de conhecimentos e ensinamentos durante toda a minha formação acadêmica. Cada um de vós deixou uma marca significativa em minha jornada.

Agradecer à minha família, em particular à minha mãe Ana Linda Francisco Nhadevele, meu Irmão Emirbelde Flores Sarmento Bule e em especial aos Primos Clésio Bule e Edson Quive, meus tios Abílio Quive (em memória), Beatriz Bule, Júlio Munguambe e Zaida Bule, pela educação, apoio incondicional e encorajamento em todos os momentos, sem os quais nada disso teria sido possível. Sua confiança e amor me motivaram a buscar sempre o melhor.

A todos os meus amigos em especial, Cardoso Elias Bahule, Jeff Venâncio Quexoaio, Fernando Gomes e Francisco Mulima pelo incentivo, compreensão e apoio incondicional nos momentos mais desafiadores. Vocês são parte essencial desta conquista, o meu muito obrigado

A minha parceira Solange Cossa, pelo incentivo, compreensão e apoio incondicional nos momentos mais desafiadores e pelos momentos vividos durante esses anos.

Aos colegas de curso, em especial ao Cardo Baule, Fortunato Alberto Guirrugo pela camaradagem, apoio e momentos de troca de conhecimento e experiências e até as longas caminhadas, noites de estudos. A amizade e o espírito de colaboração de vocês tornaram essa caminhada mais leve e enriquecedora.

Aos meus colegas de estágio na EDM em especial ao meu supervisor Adérito Piloto, pela troca de experiências constantes, conselhos e motivação.

Por fim, a todos que, de alguma forma, contribuíram para que este trabalho fosse realizado, meu sincero agradecimento. Sem o apoio e a colaboração de cada um de vocês, este Trabalho não teria sido possível.

O MEU MUITO OBRIGADO!

Epígrafe

"Todo fenômeno está relacionado a todos os outros, mas coisas mais próximas tendem a ser mais relacionadas do que coisas distantes."

Waldo Tobler

Resumo

Neste estudo, as perdas não técnicas (PNTs) no Município da Cidade de Maputo (MCM) foram analisadas utilizando o modelo de Regressão Geograficamente Ponderada Multiescalar (MGWR), que integrou variáveis técnicas e socioeconômicas em um contexto espacial. Tradicionalmente, estudos sobre PNTs não consideram explicitamente a localização espacial, o que limita a compreensão das características dos locais onde as perdas ocorrem.

Os resultados do modelo MGWR demonstraram sua superioridade em relação ao modelo Regressão Geograficamente Ponderada (GWR), com um coeficiente de determinação elevado ($R^2 = 0,97$) e um critério de informação de Akaike significativamente menor (AIC = 0,173). Isso refletiu a capacidade do modelo de capturar padrões espaciais complexos em diferentes escalas e de ajustar variáveis explicativas de forma independente. A análise dos resíduos não detectou autocorrelação espacial significativa, indicando a adequação do modelo para os dados analisados.

A análise revelou que variáveis como os Contadores Integrados, Densidade Populacional e Contadores Split têm maior influência nas PNTs. Os Contadores Integrados, com 25% de influência, foram a variável mais impactante, destacando a vulnerabilidade desse tipo de contador em áreas urbanas densamente povoadas, onde o controle e monitoramento são mais difíceis. A Densidade Populacional, com 22% de influência, foi a segunda variável que mais teve impacto, indicando que uma fiscalização intensiva em áreas de alta densidade tem um impacto directo na redução das PNTs. Os Contadores Split, com 20% de influência, também se mostraram cruciais, dado que a instalação desses dispositivos dentro das residências aumenta sua suscetibilidade a fraudes.

Palavras-chave: Análise Espacial; Perdas Comerciais; Perdas Não Técnicas (PNTs), Regressão espacial.

Abstract

In this study, non-technical losses (NTLs) in the Maputo City Municipality (MCM) were analyzed using Multiscale Geographically Weighted Regression (MGWR), which integrated technical and socioeconomic variables in a spatial context. Traditionally, studies on NTLs do not explicitly consider spatial location, which limits the understanding of the characteristics of the locations where losses occur.

The results of the MGWR model demonstrated its superiority over the Geographically Weighted Regression (GWR) model, with a high coefficient of determination ($R^2 = 0.97$) and a significantly lower Akaike information criterion (AIC = 0.173). This reflected the model's ability to capture complex spatial patterns at different scales and to adjust explanatory variables independently. The analysis of the residuals did not detect significant spatial autocorrelation, indicating the adequacy of the model for the data analyzed.

The analysis revealed that variables such as Integrated Meters, Population Density and Split Meters have the greatest influence on NTPs. Integrated Meters, with 25% influence, were the most impactful variable, highlighting the vulnerability of this type of meter in densely populated urban areas, where control and monitoring are more difficult. Population Density, with 22% influence, was the second most important variable, indicating that intensive monitoring in high-density areas has a direct impact on reducing NTPs. Split Meters, with 20% influence, were also crucial, given that installing these devices inside homes increases their susceptibility to fraud.

Keywords: Spatial Analysis; Electricity Distribution; Commercial Losses; Non-Technical Losses (NTPs), Spatial Regression; Geographically Weighted Regression; GIS.

Abreviaturas

ASC	Área de Serviço ao Cliente
AIC	Critério de Informação de Akaike
FLV	Função de Log Verossimilhança
BT	Baixa Tensão
CRS	Complete Spatial Randomness
MCM	Município da Cidade de Maputo
EDM	Electricidade de Moçambique
DPRP	Direção de Proteção de Receitas e Controle de Perdas
GWR	Regressão Geograficamente Ponderada
GAM	Modelo Aditivo Generalizado
INE	Instituto Nacional de Estatística
MAUP	Problema da Unidade de Área Modificável
MGWR	Regressão Ponderada Geograficamente Multiescalar
MT	Média Tensão
PT	Perda Técnicas
PNT	Perdas Não Técnicas
SEP	Sistema Electrico de Potência
SIG	Sistema de Informação Geográfica

Índice

Dedicatória ..... **i**

Declaração de Honra ..... **ii**

Agradecimentos ..... **iii**

Epígrafe ..... **iii**

Resumo ..... **v**

Abstract ..... **vi**

Abreviaturas ..... **vii**

Lista de Figuras ..... **xi**

Lista de Tabelas ..... **xii**

Introdução **1**

1.1. Contextualização 1

1.2. Definição do problema 3

1.3. Justificativa 4

1.4. Relevância do Estudo 4

1.5. Objectivos 5

1.5.1. Objectivo Geral 5

1.5.2. Objectivos Específicos. 5

1.6. Área de Estudo 6

1.6.1. Localização Geográfica do Município da Cidade de Maputo 6

Revisão da Literatura **8**

2.1 Estado da Arte 8

2.2. Perdas de Energia Eléctrica 9

2.3 Perdas na rede de distribuição 9

2.3.1. Perdas Técnicas 10

2.3.2. Perdas Não Técnicas 11

2.4. Tipos de Perdas Não Técnicas 12

2.4.1. Perdas Não-Técnicas por falha de equipamentos 12

2.4.2. Erros de leitura ou diferenças no facturamento 12
2.4.3. Perdas por Ação do Consumidor – Fraude e furto de energia 13
2.5. Medidas de Prevenção e Combate a perdas não técnicas (PNT) 15
2.5.1. Inspeções 15
2.5.2. Denúncias 16
2.5.3. Combate 16
2.6. Uso de Sistemas de Informação Geográfica em Estudos de Perdas Não Técnicas de Energia Eléctrica 17
2.9. Modelos de Regressão Espacial 19
2.9.1. Modelos Globais *versus* Modelos Locais 21
2.9.2 Matrizes de Proximidade Espacial 22
Material e Metodologia **24**
3.1. Material 24
3.1.1. Aquisição de dados 25
3.1.2. Descrição dos dados usados 25
3.2. Métodos 26
3.2.1. Pré-processamento 28
3.2.2. Métodos de Transformação de Dados 29
3.2.2.1 Transformação Logarítmica 29
3.2.3 O Problema da Unidade de Área Modificável e o Problema de Escala 30
3.2.4. Validação dos Modelos 30
3.2.5. Análise Exploratória de Dados 31
3.2.5.1 Coeficiente de Correlação de Spearman rs 32
3.2.5.2 Função de Kernel não Paramétrico 33
3.2.5.3 Função de distribuição Cumulativa (G*) ou vizinho mais próximo 34
3.2.5.4 Índice Local de Moran I 35
3.3. Limitações de Estudo 36
3.4. Fluxograma Metodológico 38

Discussão de Resultados **39**

4.1. Análise Exploratória Geograficamente Ponderada 39

4.1.1 Avaliação do Agrupamento Espacial das PNTs 44

4.1.2. Avaliação da Dependência Espacial das Perdas não Técnicas (PNTs) 46

4.2. Aplicação dos Modelos de Regressão Espacial Local 48

4.2.1. Avaliação do Desempenho dos Modelos de Regressão Espacial Local 48

4.2.2. Avaliação Espacial do Desempenho dos Modelos de Regressão Espacial Local 51

4.3. Analise de Influência dos Factores Técnicos Socioeconómicos no Padrão de Distribuição Espacial das PNTs 53

Conclusões e Recomendações **58**

5.1. Conclusões 58

5.2 Recomendações 59

Referências Bibliográficas **60**

Anexos **66**

Anexo 1: Lista de Questionário para as Entrevista 66

Anexo 2. Teste de significância estatística do modelo GWR 67

Anexo 3. Teste de significância estatística do modelo MGWR 67

Anexo 4. Teste de Homocedasticidade do Modelo GWR 67

Anexo 5. Teste de Homocedasticidade do Modelo MGWR 68

Anexo 6. Relatório do Modelo GWR 68

Anexo 7. Relatório do Modelo MGWR 69

Anexo 8. Credencial de Pedido de Dados emitido a Electricidade de Moçambique **70**

Lista de Figuras

Figura 1. Localização do Município da Cidade de Maputo 6
Figura 2. Manipulação de contador integrado. Fonte: Adaptado de (Waldemar & Correa, 2007) 13
Figura 3. Ligação clandestina em um bairro na cidade de Maputo. 14
Figura 4. Contador Split, instalado na residência do cliente. 14
Figura 5. Matriz de ponderação espacial de primeira ordem normalizada Fonte: Adaptado de (Câmara & Correa, n.d.) 22
Figura 7. Fluxograma Metodológico 38
Figura 8. Distribuição espacial por bairro de variáveis Socioeconómicas: ***(a)*** Renda Media Mensal; ***(b)*** Índice de criminalidade. 40
Figura 9. Distribuição espacial por bairro de variáveis Tecnicas: ***(a)*** Densidade populacional; ***(b)*** Taxa de desemprego. 41
Figura 10. Distribuição espacial por bairro de variáveis Tecnicas: ***(a)*** Contadores Integrados; ***(b)*** Contadores Split 41
Figura 11. Matriz de correlações de Spearman. 42
Figura 12. Densidade de Kernel de Perdas não técnicas. 45
Figura 13. Função G(r) ou função de distribuição cumulativa da variável PNTs 46
Figura 14. Índice de Moran's I; **(a)** Gráfico de Dispersão; **(b)** Teste de significância 47
Figura 15. Análise gráfica dos modelos; **(a)** índice de Moran's I para o modelo GWR; **(b)** índice de Moran's I para o modelo MGWR 50
Figura 16. Análise gráfica dos modelos; **(a)** Histograma dos resíduos para o modelo GWR; **(b)** Histograma dos resíduos para o modelo MGWR 51
Figura 17. Avaliação espacial dos modelos; **(a)** Resíduos para o modelo GWR; **(b)** Resíduos para o modelo MGWR; 52
Figura 18. Coeficientes Estimados (β) para o modelo MGWR; **(a)** Renda Media Mensal; **(b)** Índice de Criminalidade; 54
Figura 19. Coeficientes Estimados (β) para o modelo MGWR; **(a)** Densidade Populacional; **(b)** Taxa de desemprego; 55
Figura 20. Coeficientes Estimados (β) para o modelo MGWR; (a) Contadores Integrados; (b) Contadores Split 56

Lista de Tabelas

Tabela 1. Resumo da eficiência das principais estratégias para localização de casas com Pontos. Fonte: Adaptado de Faria, 2016. 16

Tabela 2. Modelos globais versus Modelos locais. (Fonte: Adaptado de Druck et al., 2004) 21

Tabela 3. Programas utilizados 24

Tabela 4. Breve descrição dos Modelos de Regressão Espacial utilizados neste estudo..... 25

Tabela 5. Descrições das variáveis explicativas e dependente analisadas neste estudo. 26

Tabela 6. Constatações feitas durante as entrevistas com os 10 técnicos selecionados. 28

Tabela 7. Intervalos do coeficiente de correlação e magnitude da correlação. Fonte: Adaptado de (Faria, 2016)..... 33

Tabela 8. Análise de desepenho dos modelos. 48

Tabela 9. Percentagem de Influência dos factores técnicos e socioeconómicos no padrao de distribuição espacial das PNTs**.** 57

1

Introdução

Este capítulo tem como objectivo apresentar o contexto da pesquisa, o porquê de desenvolver o estudo, a justificativa, a relevância e os objectivos.

1.1. Contextualização

As perdas não técnicas (PNTs) representam uma ameaça crescente para a sustentabilidade das empresas de energia e para a confiabilidade no fornecimento eléctrico, especialmente em países em desenvolvimento, como Moçambique. Essas perdas, que podem ultrapassar 30% do total da energia distribuída em algumas regiões, resultam em prejuízos financeiros significativos e aumento nas tarifas para os consumidores (Botelho, 2022). Além de comprometer a receita das distribuidoras, as PNTs afectam directamente os consumidores, uma vez que a perda de receita das empresas muitas vezes é compensada pelo aumento das tarifas de electricidade, criando um ciclo vicioso onde os custos elevados podem incentivar ainda mais a prática de furtos, agravando o problema e dificultando o desenvolvimento de soluções eficazes (Jamil & Ahmad, 2010). Para assegurar a viabilidade financeira e operacional das empresas de energia, é crucial lidar com as PNTs de forma eficaz. Quando as perdas aumentam, a facturação das empresas diminui, limitando sua capacidade de investir em melhorias e expansão da infraestrutura eléctrica, o que impacta negativamente o crescimento do sector e a qualidade do fornecimento para a população (Ramos, 2014; Martins, 2021).

O relatório estatístico de contas de 2022 da Electricidade de Moçambique (EDM) estima as perdas em 28% da energia total injectada na rede, influenciadas por factores como redes precárias, falta de cadastro de clientes e a presença de contadores no interior das residências (Piloto, 2024). Entre as medidas adoptadas pela EDM, destacam-se a instalação de medidores inteligentes com leitura automatizada, técnica avançada para monitoramento e controle de perdas que permite o acompanhamento em tempo real do consumo e da carga dos clientes. Contudo, essa solução é considerada economicamente inviável para países em desenvolvimento como Moçambique, o que leva à busca por alternativas mais acessíveis (EDM, 2022). O relatório também menciona outras acções comumente empregadas, como a instalação de pontos de medição para balanço energético, regularização de ligações clandestinas, substituição de

contadores danificados ou antigos, reabilitação de redes precárias, realocação de contadores para postes e inspeções para verificar as condições das instalações eléctricas.

As pesquisas recentes sobre PNTs, têm evidenciado a importância da identificação de áreas vulneráveis, como locais de furtos e fraudes, por meio de modelos de regressão espacial (Cabral *et al.,* 2004; Angelos *et al.,* 2011; Faria, 2012). Entretanto, alguns estudos não têm considerado adequadamente as particularidades geográficas dessas áreas. A natureza imprevisível das PNTs, influenciada por factores socioeconômicos e técnicos que variam no tempo e no espaço, exige abordagens que levem em conta essas complexidades (Smith, 2004; Faria, 2016). Modelos de regressão espacial, como a regressão geograficamente ponderada (GWR), permitem incluir a dimensão geográfica na análise, estudando a distribuição espacial das perdas e simulando cenários dinâmicos para identificar as áreas mais vulneráveis e implementar medidas preventivas adequadas (Cardoso, 2002; Faria, 2016).

O presente estudo busca analisar as perdas não técnicas (PNTs) no contexto da rede eléctrica de baixa tensão, utilizando ferramentas de análise espacial e modelagem estatística. A partir dessa análise, espera-se contribuir para o desenvolvimento de estratégias mais eficazes e direcionadas ao combate dessas perdas, integrando metodologias inovadoras que considerem a localização espacial e a dinâmica dos factores envolvidos.

1.2. Definição do problema

De acordo com o relatório divulgado em abril de 2022 pela Electricidade de Moçambique (EDM), as Perdas Não Técnicas (PNT) resultaram em um prejuízo de aproximadamente 1,4 mil milhões de Meticais para a empresa no período de 2019 a 2022 (EDM, 2022). Essa situação tem dificultado a expansão da rede pública de distribuição, prejudicando o acesso à energia eléctrica no país, em um cenário onde apenas 39% da população nacional rural, equivalente a 12 milhões de pessoas, tem acesso ao serviço, deixando 71% dos moçambicanos sem electricidade (EDM, 2022). O relatório identifica ainda a cidade e província de Maputo como as principais áreas de ocorrência de fraudes e furto de energia, causando prejuízos comerciais significativos.

Segundo Técnicos da EDM, afectos na Área de Serviço ao Cliente (ASC) de KaMavota e Kampfumo, os casos de fraudes e furto de energia estão concentrados, especialmente em bairros com maior número de contadores analógicos e Split com o contador instalado no Postilete no interior das residências. Dentro desse cenário, os trabalhos de Smith (2004) e Faria (2016) ressaltam a importância de considerar a manifestação espacial de factores técnicos e socioeconômicos para compreender a distribuição espacial das PNTs em diferentes regiões. Esse enfoque é particularmente relevante para o Município da Cidade de Maputo, onde a configuração socioeconômica e a infraestrutura de medição eléctrica variam significativamente entre bairros.

Diante da complexidade dos elementos envolvidos, surge a necessidade de investigar em que medida os factores técnicos e socioeconômicos influenciam no padrão de distribuição espacial das PNTs.

1.3. Justificativa

A cidade de Maputo que contempla o Município da cidade de Maputo (MCM) é tido como o principal foco de ocorrência de casos de fraude e furto de energia eléctrica a nível nacional, contribuindo dessa forma para aumento significativo das perdas não técnicas (PNTs) (EDM, 2022). Ainda segundo a EDM, estima se que só do período 2019 a 2022, o MCM contribui com cerca de 50% das perdas não técnicas registadas a nível nacional, o que coloca em causa o projecto da companhia, de electrificação para os demais distritos e localidades numa altura em que apenas 39% da população rural possui energia eléctrica.

Visto que as Perdas não técnicas (PNTs) na rede eléctrica de baixa tensão, que compreende a distribuição até ao consumidor final, são influenciadas por uma variedade de factores técnicos e socioeconómicos, cientistas têm proposto metodologias científicas que analisam as PNTs com recurso a modelos de regressão espacial de modo a ter um conhecimento mais profundo no concernente ao padrão de distribuição espacial das mesmas.

Nesse contexto, torna se importante propor metodologias científicas que permitam uma análise da influência dos factores técnicos e socioeconómicos no padrão de distribuição espacial das PNTs no Município da Cidade de Maputo com recurso a modelos de regressão espacial combinado com os Sistemas de Informação Geográfica.

1.4. Relevância do Estudo

A metodologia apresentada neste estudo, permite analisar a influência dos factores técnicos e socioeconómicos no padrão de distribuição espacial das perdas não técnicas (PNTs) no Município da Cidade de Maputo. É relevante não só por fornecer uma compreensão mais abrangente das PNTs, como também permite uma visão mais detalhada das áreas mais vulneráveis, possibilitando a formulação de estratégias operacionais mais direcionadas e eficazes para o combate às PNTs. Ademais, estudo visa enriquecer o acervo nacional, visto que há escassa literatura que explora a influência combinada de factores técnicos e socioeconômicos no padrão de distribuição espacial das PNTs utilizando modelos de regressão espacial combinado com os Sistemas de Informação Geográfica.

A análise proposta poderá contribuir como uma ferramenta valiosa para o planeamento e a gestão da rede eléctrica, reforçando a importância de considerar o espaço geográfico nas estratégias de mitigação das PNTs, desse modo promovendo a eficiência financeira e operacional, e, a longo prazo, garantindo um fornecimento de energia mais confiável e sustentável.

1.5. Objectivos

1.5.1. Objectivo Geral

Analisar a influência dos factores técnicos e socioeconómicos no padrão de distribuição espacial das perdas não técnicas na rede eléctrica de distribuição de Baixa Tensão.

1.5.2. Objectivos Específicos.

- Identificar as áreas com maior concentração de perdas não técnicas no Município da Cidade de Maputo através do algoritmo de estimador de Kernel e Função Cumulativa G;
- Avaliar a dependência espacial das perdas não técnicas através do Índice de Moran I;
- Aplicar os modelos de regressão espacial geograficamente ponderados com vista a identificar o melhor modelo que se ajusta aos dados analisados; e
- Representar por meio de mapas temáticos a influência de factores técnicos e socioeconómicos no padrão de distribuição espacial das perdas não técnicas.

1.6. Área de Estudo

1.6.1. Localização Geográfica do Município da Cidade de Maputo

O Município da Cidade de Maputo (MCM) localiza-se aproximadamente entre as latitudes 25° 51'S e 26° 12'S e longitudes 32° 27'E e 32° 57'E. É delimitado a Norte pelo Município de Marracue, a Sul pelo Distrito de Matutuine, a Este pelo Canal de Moçambique e a Oeste pela Cidade da Matola e Distrito de Boane (CMM, 2024). No entanto, devido a disponibilidade dos dados, a área de estudo contemplou apenas 5 Distritos Municipais da Cidade de Maputo, nomeadamente: Distrito de KaMavota, KaMaxakeni, Ka Nlhamankulo, KaMubukwana e KaMpfumo, excluído os distritos Municipais como KaNyaka e Katembe conforme ilustra a figura 1.

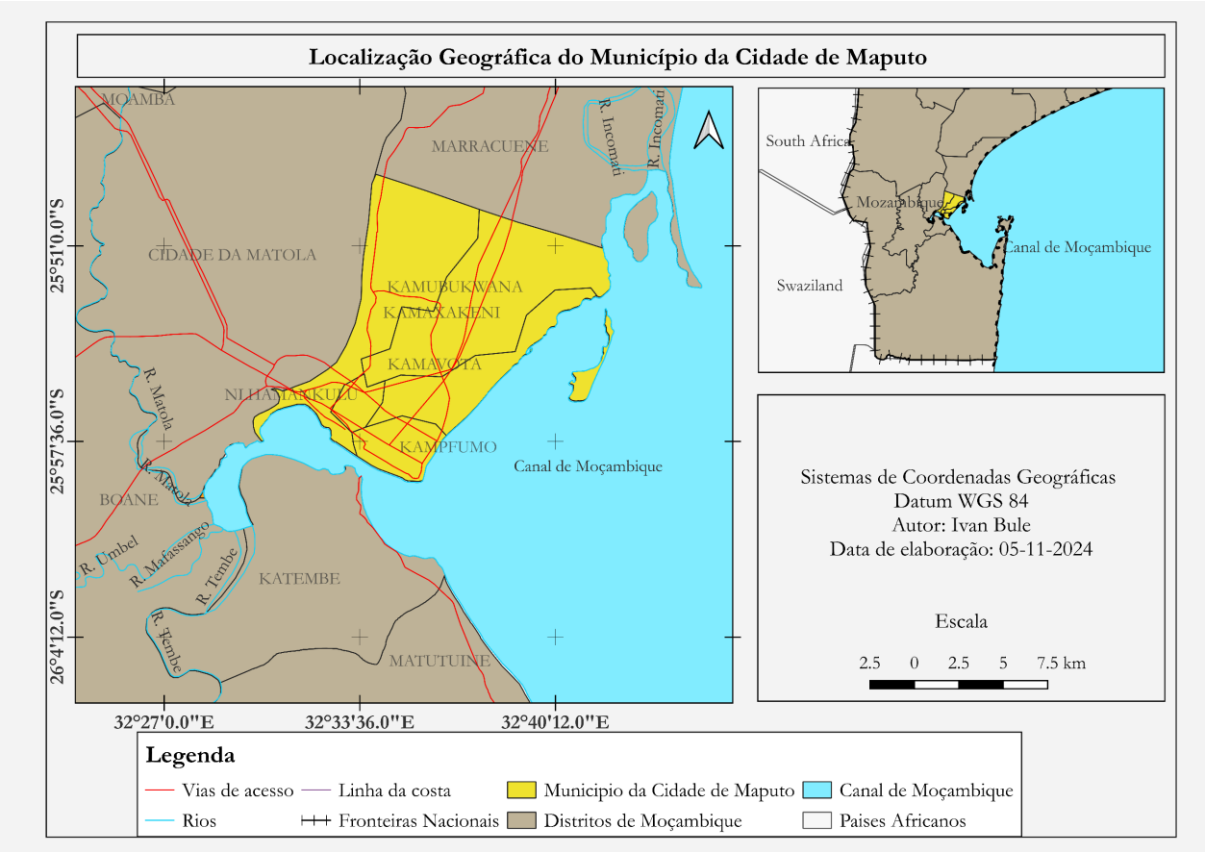


Figura 1. Localização do Município da Cidade de Maputo

Devido a sua importância política e económica e também a dimensão da sua população a cidade de Maputo tem estatuto de província, cuja jurisdição corresponde à área do Município. A área do Município é de 346,7 Km^2, comportando 64 bairros divididos por 7 Distritos Urbanos nomeadamente: KaMpfumu (Mafumo), KaMaxakeni (Maxaquene), Nlhamankulu (Chamanculo), KaMavota (Mavota), KaMubukwana (Mubukwana), KaTembe (Catembe) e KaNyaka (Inhaca). Em 2019, a população da cidade estava estimada em cerca de 1.170 mil habitantes (CMM, 2024).

O MCM é alimentado por aproximadamente 3201 transformadores (PTs) de variados níveis de tensão, dos quais aproximadamente 833 PTs alimentam o distrito urbano de KaMpfumu, 455 PTs para o distrito urbano de KaMaxakeni, 400 PTs para o distrito urbano de Nhlamankulu, 833 PTs para o distrito urbano de KaMavota, KaMubukwana com 511 PTs, KaTembe com 112 e KaNyaka com 57 PTs (EDM,2023).

No entanto, apesar de existir maior capacidade de fornecimento de energia, a infraestrutura da rede de distribuição, principalmente a de distribuição de baixa tensão deixa a desejar, pois, mais de 50% da infraestrutura da rede é precária, visto que em muitos bairros da periferia possui uma rede no qual vários moradores partilham a mesma baixada[1], oque contribui para a baixa qualidade de energia fornecida e a proliferação de acções do consumidor como fraude e roubo de energia. Ademais, mais de 60% da população do MCM possui os contadores de integrados, sendo apenas aproximadamente 40% da população que possui contadores inteligentes ou *split*[2].

[1] Baixada refere se ao ponto de toma (poste) que sustenta o fio de energia até a casa do cliente.
[2] Split refere se ao contador inteligente, ou seja, contador cuja o seu input (teclado) vem separado do mesmo.

2

Revisão da Literatura

Neste capítulo, foi realizada uma revisão da literatura com o objectivo de orientar a pesquisa sobre aplicação de modelos de regressão espacial na análise dos factores determinantes de Perdas Não Técnicas na rede eléctrica de baixa tensão. Nesse contexto, Perdas Não Tecnicas e Analise espacial com ênfase nos Modelos de Regressão Espacial são os principais conceitos essenciais discutidos extensivamente neste capítulo e durante todo o estudo.

2.1 Estado da Arte

As pesquisas em Perdas Não Técnicas (PNTs) como os trabalhos de (Jiang, 2002; Zhang, 2008; Nagi *et al.,* 2010; Nagi *et al.,* 2011; Angelos *et al.,* 2011; Faria, 2012; Ramos *et al.,* 2011; Ramos *et al.,* 2012; Ramos, 2014; Trevizan *et al.,* 2015; Piloto, 2024), têm, em sua maioria, focado na detecção de PNTs ou unidades irregulares. Esses estudos frequentemente empregam modelos baseados em técnicas de sistemas inteligentes, como redes neurais, lógica fuzzy, algoritmos de aprendizado de máquina e mineração de dados. No entanto, um ponto crucial que ainda não é amplamente abordado nesses trabalhos é a consideração das características locais (factores técnicos e socioeconómicos) que podem estar associados a ocorrência das PNTs (Smith, 2004).

A ausência dessa dimensão espacial limita a compreensão mais completa e precisa do problema, uma vez que as PNTs estão intimamente ligadas as características do espaço onde ocorrem (Faria, 2016). Para superar essa limitação, há uma necessidade crescente de desenvolvimento de novos métodos que levem em conta as características locais de ondem ocorrem as PNTs, como ressaltam Messinis & Hatziargyriou (2018) e Estevo Felix *et al.,* (2021).

Actualmente, autores tem vindo a dar ênfase na implementação de estudos de PNTs considerando as características locais de sua ocorrência.

Cancian (2013), em seu estudo, utilizou técnicas de agrupamento, como o algoritmo dos "k-Vizinhos Mais Próximos", aliado a análises gráficas e visuais, para investigar perdas comerciais em clientes de baixa tensão de concessionárias de energia eléctrica. Esse enfoque espacial melhorou o desempenho das inspeções, facilitando a identificação de unidades consumidoras com maior potencial de irregularidades e contribuindo para uma maior recuperação de receita.

Para além de Cancian, Faria (2016), em sua tese de mestrado, incorporou a dimensão geográfica para examinar as PNTs em um município do interior de São Paulo. Ao explorar a influência de factores técnicos e socioeconômicos na distribuição espacial das perdas, ele aplicou técnicas avançadas de análise espacial com enfoque na aplicação de modelos de regressão espacial, como o modelo aditivo generalizado (GAM) e a regressão geograficamente ponderada (GWR).

Essas duas abordagens comprovam de forma clara que o uso de dados espaciais é essencial para aumentar significativamente a precisão na identificação de perdas, corroborando a tese defendida nesta presente pesquisa de que as características geográficas não são apenas relevantes, mas são um elemento determinante na ocorrência das PNTs em várias regiões do Município da Cidade de Maputo. Ao integrar a dimensão espacial na análise, essa metodologia não só reforça a importância dessa abordagem, mas também tem potencial contributivo de melhorar a eficiência das inspeções e permitir uma compreensão mais profunda e direcionada das áreas com maior risco de irregularidades, tornando a redução das perdas uma meta mais tangível.

2.2. Perdas de Energia Eléctrica

Segundo Oliveira (2009), a definição da palavra "perda" muda dependendo do contexto em que é empregada. Em finanças, por exemplo, "perda" significa uma redução de valor, enquanto na área da engenharia e energia, refere-se ao desperdício de energia. Portanto, é fundamental entender o contexto em que a palavra é usada para compreender seu significado preciso e as consequências associadas.

No contexto da energia eléctrica, as perdas podem ser comparadas a um consumidor que nunca paga sua factura, simbolizando energia que é gerada e percorre a rede, mas que nunca chega ao consumidor final de maneira eficaz (Faria, 2016). Conforme Penin (2008), essas perdas globais representam a energia eléctrica que, apesar de gerada e distribuída através das redes de transmissão e distribuição, não é comercializada, seja por motivos técnicos ou comerciais.

2.3 Perdas na rede de distribuição

Segundo Oliveira (2022), compreender o nível das perdas de energia e suas características é crucial para avaliar a condição operacional da rede, a eficiência do fornecimento de energia e a eficácia das acções voltadas à redução dessas perdas. Essas perdas de energia eléctrica são identificadas por meio do balanço energético, realizado pelas concessionárias de energia eléctrica.

A rede de distribuição, sendo a última etapa no fornecimento de energia antes de chegar aos consumidores finais, é onde as perdas se tornam particularmente significativas. Em uma empresa de

distribuição de energia eléctrica, as perdas totais ocorrem na energia fornecida aos consumidores regulares, a outras distribuidoras e em diferentes níveis de tensão (Piloto, 2024). A maioria dos consumidores conectados directamente ao sistema de transmissão ou subtransmissão tem medições individuais feitas de hora em hora ou em intervalos ainda menores, de 10 ou 15 minutos. Com essas medições, é possível calcular as perdas nesses trechos de forma directa, pela discrepância entre a energia que entra e a que sai. Embora isso não elimine completamente as fraudes dos consumidores, ajuda a identificar os locais onde as perdas não técnicas estão ocorrendo (Oliveira, 2009).

2.3.1. Perdas Técnicas

As perdas técnicas no sistema eléctrico de potência (SEP) são um fenômeno inerente ao processo de geração, transmissão e distribuição de energia, causadas pelas características físicas dos dispositivos envolvidos (Faria, 2016). Segundo Costa (2018), essas perdas são geradas por processos físicos naturais, como o efeito Joule, que transforma parte da energia em calor ao passar pelos condutores e transformadores. A resistência eléctrica desses componentes é a principal causa das perdas durante as diferentes etapas do fornecimento de energia, e o uso de materiais mais eficientes é essencial para minimizar esses desperdícios, conforme destacado por (Costa,2018).

No entanto, a medição exata dessas perdas técnicas apresenta um desafio significativo, como aponta Faria (2016). A complexidade do sistema eléctrico, a necessidade de grandes volumes de dados, a aleatoriedade das cargas eléctricas e o crescimento constante dessas cargas tornam a análise das perdas técnicas uma tarefa complexa. Para aumentar a precisão dos resultados, a análise costuma ser feita por partes do sistema, permitindo uma abordagem mais detalhada.

Dantas (2006) e Oliveira (2009) classificam as perdas técnicas em três categorias principais: alta, média e baixa tensão. No sector de alta tensão, as discrepâncias na medição nas subestações são as principais responsáveis pelas perdas. Já no sistema de distribuição de média tensão, conforme CIRED (2017), as perdas são determinadas através do método de fluxo de carga. Na baixa tensão, o cálculo das perdas técnicas varia de acordo com a concessionária, que geralmente une os elementos da rede de distribuição (como transformadores, redes secundárias, ramais de serviço e medidores) e realiza os cálculos com base em curvas de carga típicas, sem depender dos dados de facturamento (Dantas, 2006; Oliveira, 2009).

Embora sejam classificadas como técnicas, Faria (2016) sugere que algumas dessas perdas variadas poderiam ser consideradas parte das Perdas Não Técnicas devido às incertezas envolvidas no processo de cálculo. A dificuldade em distinguir claramente entre perdas técnicas e não técnicas reflete a

complexidade do sistema eléctrico e a necessidade de abordagens mais rigorosas para a gestão e redução desses desperdícios.

2.3.2. Perdas Não Técnicas

As Perdas Não Técnicas (PNTs), ou perdas comerciais, representam um obstáculo significativo para o sector de energia eléctrica, impactando directamente a receita das empresas e a qualidade do serviço oferecido aos consumidores(Estevo Felix *et al.,* 2021). De acordo com Nunes (2000), essas perdas referem-se à energia gerada, transmitida e entregue, porém não faturada, o que resulta numa redução direta na receita.

Essas perdas diferem das Perdas Técnicas, que são inevitáveis e resultam de elementos físicos e operacionais do sistema eléctrico, como a resistência dos condutores e transformadores (Smith, 2004; Oliveira, 2018). As PNTs, por outro lado, incluem fraudes, furtos, imprecisões na medição e falhas de gestão, tornando-se uma questão crítica tanto para a eficiência energética quanto para a sustentabilidade econômica das empresas distribuidoras de energia.

Savian et *al.,* (2020) identificam diversos factores que levam os consumidores a praticarem o furto de energia eléctrica, com destaque para o aspecto socioeconômico. Isso sugere que as condições financeiras e sociais dos consumidores têm grande influência na decisão de cometer esse tipo de crimes. Além disso, Ferreira (2011) aponta que os medidores analógicos são particularmente suscetíveis à fraude, o que motivou as empresas de distribuição a substituí-los por medidores digitais, considerados mais seguros e econômicos.

No entanto, Oliveira (2013) ressalta que o estudo das PNTs deve considerar as características regionais onde essas perdas ocorrem. Em diferentes cidades, as motivações para fraudes variam, frequentemente ligadas a regiões menos favorecidas, onde os moradores têm acesso limitado a serviços públicos básicos, como água potável e saneamento. (Leite *et al.,* 2018) discutem a complexidade da relação entre criminalidade e contexto social, afirmando que, embora comunidades carentes geralmente apresentem altos índices de criminalidade, a dinâmica social é mais complexa do que uma simples correlação linear.

Embora a substituição de medidores analógicos por digitais seja uma medida importante para reduzir as PNTs, essa abordagem não é isenta de limitações. Por exemplo, a simples troca de tecnologia não resolve os problemas estruturais e socioeconômicos que incentivam as fraudes. Além disso, a eficácia dos medidores digitais depende de uma série de factores, como a manutenção adequada e a aceitação pelos consumidores.

Dessa forma, a análise das PNTs deve considerar não apenas as estratégias tecnológicas de mitigação, mas também as políticas sociais e econômicas que podem influenciar o comportamento dos consumidores. Somente uma abordagem integrada, que leve em conta tanto os aspectos técnicos quanto os contextos sociais, pode oferecer soluções mais eficazes e sustentáveis para o problema das PNTs.

2.4. Tipos de Perdas Não Técnicas

Smith citada por (Qadir & Jalal, 2017), desenvolveu uma pesquisa que foi conduzida em 102 nações ao longo de um extenso período de vinte anos. O relatório calcula a amplitude das fraudes relacionadas às PNTs e as divide em quatro tipos: Falhas dos contadores, Fraudes ou manipulação dos contadores, ligações clandestinas e problemas nas cobranças.

2.4.1. Perdas Não-Técnicas por falha de equipamentos

Para (Paulo, 2008), os medidores podem também ser fontes de perdas comerciais, pois um medidor com alguma anomalia poderá gerar um facturamento distorcido.

O facto de algumas empresas terem seu parque de medidores envelhecido ou não correctamente aferido, pode causar uma parcela das perdas comerciais, que pode ser importante quando for acumulado devido a grande volume de distorções de facturamento, tais distorções podem ser causadas no medidor pelos seguintes factores: desgaste devido à idade, as intempéries, ou mesmo devido à posição de instalação, podendo fazer seu mecanismo girar mais devagar, marcando valores menores (Penin, 2008).

2.4.2. Erros de leitura ou diferenças no facturamento

De acordo com Silva *et al.,* 2019, as falhas nos pagamentos também são um factor para o surgimento de problemas não traumáticos. Isso pode envolver erros propositais ou acidentais na etapa de cobrança, resultando em uma cobrança inferior ao verdadeiro consumo de energia dos consumidores. Santos *et al.,* (2020) enfatizam que esses erros podem decorrer de falhas no sistema de facturamento, questões administrativas ou conluio entre funcionários da empresa de serviços públicos e consumidores.

2.4.3. Perdas por Ação do Consumidor – Fraude e furto de energia

Segundo (Piloto, 2024), na área de fornecimento de energia eléctrica, fraude e furto têm definições distintas. A fraude de energia pode ser descrita como a acção intencional de alguém para evitar ou diminuir a energia cobrada (Fernando G. K. Guarda *et al.,* 2018). Isso ocorre quando o cliente realiza uma sobrecarga sem autorização da empresa de energia em sua instalação, muitas vezes de forma elaborada.

A fraude pode também envolver a manipulação do medidor de energia ou a instalação de um desvio na fiação de entrada antes do medidor (Penin, 2008). Nesses casos, o consumidor aumenta secretamente o consumo de electricidade em um sistema não aprovado, frequentemente de forma elaborada. Outra forma frequente de fraude acontece quando há interrupção no fornecimento regular de energia, devido à falta de pagamento, e o consumidor restabelece a ligação directamente à rede, sem autorização (Oliveira, 2009).

Ademais, segundo Botelho (2022), a conveniência de acesso ao contador pode permitir a execução de práticas como a interferência magnética, utilizando um imã para bloquear o movimento do disco magnético do contador analógico (modelos antigos), e a técnica conhecida como "by-pass", que envolve a conexão em paralelo com o equipamento, como ilustrado nas Figura 2.

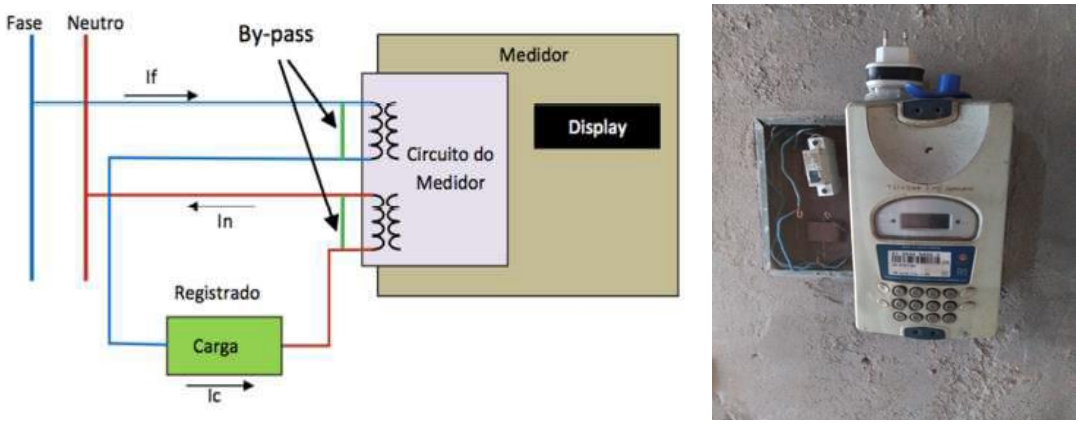


Figura 2. Manipulação de contador integrado. Fonte: Adaptado de (Waldemar & Correa, 2007)

De acordo com (Ferreira, 2015), as conexões clandestinas ou furtos consistem principalmente em ligações fora da lei feitas por consumidores que optam por não pagar pela electricidade utilizada. Em nações em desenvolvimento como Moçambique, essa modalidade de interações é comumente

observada em áreas carentes. As conexões ilegais são comumente feitas no ramal de Baixa Tensão (BT) e, nesse caso, estão visíveis, possibilitando sua identificação visual conforme mostrado na Figura 3.

Essas acções causam prejuízos financeiros e representam riscos de segurança para os fraudadores e para a comunidade em geral (Botelho, 2022).

Figura 3. Ligação clandestina em um bairro na cidade de Maputo.

Segundo Faria (2012), devido aos furtos, acidentes na rede eléctrica podem ocorrer devido à sobrecarga gerada nos fusíveis dos postos de transformação. Os incidentes afectam os perpetradores do roubo, residentes, vizinhos e qualquer pessoa próxima a essa actividade ilícita. Entre os principais acidentes estão os incêndios, choques eléctricos e curtos-circuitos. Adicionalmente, os cidadãos ainda podem correr o risco de sofrer acidentes envolvendo equipamentos e cortes frequentes de fornecimento de energia electrica (Piloto, 2024).

De acordo com Lima (2018), as formas mais comuns de furtos e fraudes no fornecimento de energia eléctrica incluem ligações clandestinas, auto-religação, práticas irregulares antes da medição e manipulações no processo de medição principalmente em dispositivos de medição instalados no interior das instalações do cliente conforme ilustra a figura 4.

Figura 4. Contador Split, instalado na residência do cliente.

Estas práticas englobam uma gama de acções intencionais para evitar ou reduzir o pagamento de energia consumida. Vieiralves citado por Lima (2018), argumenta que as perdas não técnicas, causadas tanto por fraudes quanto por furtos, estão profundamente enraizadas nas características socioeconômicas da área atendida e no comportamento dos consumidores. Esses factores são influenciados pelo desenvolvimento social, econômico e educacional da população, moldando o perfil mental, social e cultural da comunidade em questão.

Por outro lado, Piloto (2024), argumenta que as perdas PNTS não se restringem apenas a consumidores que cometem fraudes, mas também incluem aqueles que são inadimplentes, ou seja, aqueles que não pagam suas contas de energia pontualmente. Para Faria (2012), existe uma relação intrínseca entre as perdas de energia eléctrica e a inadimplência: um consumidor que, após um furto, tem sua ligação regularizada, pode eventualmente se tornar inadimplente, levando ao corte do fornecimento. Sem acesso à energia, há uma probabilidade de retorno ao furto.

Essas observações são corroboradas pelos achados da EDM, que identificou que a maioria das fraudes e furtos ocorre em áreas suburbanas com alta densidade de população de baixa renda, contadores integrados e contadores inteligentes split instalado no postilete na residência do cliente. Muitos desses consumidores alegam que o custo elevado da energia a força a recorrer a práticas fraudulentas ou de furto. O estudo que proponho visa não apenas mapear as áreas críticas, mas também analisar em que medida os factores socioeconómicos associadas a vulnerabilidades de cada tipo de contador afectam nos padrões da distribuição espacial das perdas não técnicas de modo a proporcionar uma abordagem abrangente e informada para melhorar a gestão das perdas e implementar políticas mais eficazes para combater fraudes e furtos de energia

2.5. Medidas de Prevenção e Combate a perdas não técnicas (PNT)

Diversas abordagens têm sido propostas para combater as PNTs, incluindo medidas tecnológicas, fiscalização direcionada e políticas sociais. Depuru *et al.* (2011) destacam a importância da combinação de ações preventivas, como o uso de contadores eletrônicos e inspeções regulares, com iniciativas educativas voltadas para a conscientização da população.

2.5.1. Inspeções

Segundo (Solteira, 2009), a técnica de inspeção é uma maneira menos eficiente de detectar consumidores com perdas não técnicas (PNTs) em uma rede de distribuição. Primeiramente, os

alimentadores com as maiores Perdas Não Técnicas são identificados através do cálculo das perdas globais desconsiderando as perdas técnicas. Depois, todos os clientes desses comedouros ou regiões suspeitas passam por uma inspeção. Contudo, devido ao elevado custo monetário e tempo exigido para monitorar todos os consumidores, essa abordagem é restrita a áreas com alto índice de fraudes(Paulo, 2008). Além do mais, os consumidores conseguem facilmente identificar as equipes de inspeções e esconder as evidências antes da fiscalização. A porcentagem de sucesso da varredura é baixa, aproximadamente de 10% a 12% (Roberto & Bastos, 2011).

2.5.2. Denúncias

Segundo Faria (2016), geralmente as inspeções ocorrem com base em denúncias, que têm alto índice de acerto, mas são poucas em comparação com à quantidade de inspeções que poderiam ser realizadas ao longo do tempo.

2.5.3. Combate

De acordo com Penin (2008), a redução de perdas no comércio inclui estratégias para identificar consumidores fraudulentos ou com padrões irregulares de consumo. Ao contrário da busca por furtos, que ocorre com base na localização do núcleo em regiões previamente mapeadas pelas concessionárias, a detecção de fraudes é desafiadora devido à sua natureza discreta e evasiva(Pr & Eng, 2018).

A Mineração de Dados tem sido amplamente adoptada pelas empresas em todo o mundo como uma estratégia para identificar clientes fraudulentos (Penin, 2008). Este método tem contribuído para aumentar as taxas de êxito nas inspeções no terreno, realçando estratégias como a análise do histórico de utilização, a georreferenciação dos clientes, a mineração de dados e a inteligência artificial conforme ilustra a tabela 1.

Tabela 1. Resumo da eficiência das principais estratégias para localização de casas com Pontos.
Fonte: Adaptado de Faria, 2016.

Estratégia	Retorno financeiro	Tempo de Planeamento	Área de Abrangência	Taxa de Sucesso
Inspeções	Baixo	Moderado	Por Área	10% a 12%
Denuncias	Alto	Baixo	Pontual	≈ 22%
Analise de Dados	Moderado	Alto	Pontual	> 20%
Software dedicados	Alto	Alto	Pontual/Por Área	> 20%

2.6. Uso de Sistemas de Informação Geográfica em Estudos de Perdas Não Técnicas de Energia Eléctrica

Os Sistemas de Informações Geográficas são formas especializadas de lidar com dados de localização usando tecnologia específica. Esses sistemas incluem desde a colecta e armazenamento de informações até sua análise, criação e exibição (Sánchez *et al.,* 2013). Actualmente, os SIGs têm sido bastante usados para o estudo das perdas não técnicas. Segundo Gomez *et al.,* (2016), é essencial utilizar os SIG para reconhecer padrões espaciais relacionados às perdas não técnicas, combinando informações de consumo de energia com dados geográficos a fim de mapear regiões com altas taxas de perdas. Druck *et al.,* (2004) ressaltam múltiplas vantagens importantes do emprego de Sistemas de Informação Geográfica (SIG) em pesquisas sobre Perdas Não Técnicas (PNT) em redes de distribuição de electricidade. Notavelmente, uma das principais vantagens dos Sistemas de Informações Geográficas é sua habilidade em apresentar de forma clara e compreensível os dados geoespaciais. Isso possibilita a rápida identificação das áreas problemáticas e dos padrões espaciais das PNTs.

2.7. Análise Espacial em Perdas Não Técnicas

Segundo Estevo Félix *et al.,* (2021), a análise espacial possibilita integrar o factor espacial e identificar as características das subáreas afectadas pelas PNTs. Ela utiliza ferramentas para explorar e modelar processos que se manifestam através de padrões espaciais, conhecidos como fenômenos geográficos. Nessa análise, são avaliadas as propriedades e relações, levando em conta de forma clara a localização espacial do fenômeno estudado.

De acordo com Faria (2016), essa técnica é frequentemente empregada em diversas disciplinas, como epidemiologia, criminologia, geologia, agronomia e ecologia. Segundo Druck *et al.,* (2004), os desafios na análise espacial são identificados como: pontos de referência, regiões com taxas agregadas e distribuições contínuas.

Os padrões pontuais referem-se a fenômenos representados por pontos específicos no espaço, como a localização de crimes, casos de doenças ou espécies vegetais. Essas informações são cruciais para análises que exigem uma visão detalhada e precisa de eventos individuais(Estevo Felix *et al.,* 2021).

Por outro lado, áreas com taxas agregadas tratam de dados populacionais que são reunidos em unidades delimitadas, como sectores censitários, para proteger a privacidade e evitar a exposição de informações

individuais. Esses dados agregados oferecem uma visão mais ampla, porém menos detalhada (Faria, 2016). Por fim, superfícies contínuas são estimadas a partir de amostras distribuídas de forma regular ou irregular, refletindo características geográficas contínuas, como aquelas encontradas em mapas geológicos ou topográficos. Esses dados são fundamentais para entender as variações espaciais de fenômenos naturais e suas implicações (Faria, 2016; Estevo Felix *et al.,* 2021).

2.8. Análise Espacial de Dados Agregados por Áreas

De acordo com Fotheringham *et al.,* (2000), a importância da análise espacial está em levar em conta a dimensão espacial dos dados, pois a falta de atenção a isso pode resultar em interpretações erradas. A investigação espacial de dados de regiões envolve diferentes técnicas para estudar dados espaciais cuja posição é determinada por regiões delimitadas por formas geométricas (Faria, 2016). Esta abordagem é especialmente benéfica ao lidar com eventos agrupados em locais geográficos como municípios, bairros. Nestas situações, a localização exata dos eventos não pode ser determinada; em vez disso, há apenas uma estimativa para cada região. Seguindo a mesma linha, Faria (2016) destaca a importância de levar em conta tanto a autocorrelação espacial quanto a heterogeneidade espacial. A correlação espacial indica o quanto um valor de uma variável em determinado local se assemelha aos valores em locais próximos, ao passo que a heterogeneidade espacial envolve a diversidade nos padrões espaciais devido a vários factores subjacentes.

Wong & Lee (2005) discutem a utilização de métodos de clusterização espacial, como quadrantes e o índice de Moran, que são eficazes na identificação de agrupamentos e *outliers*. A técnica de dividir em quartos divide a área de estudo em partes menores, permitindo uma análise mais detalhada da distribuição espacial dos dados, mostrando como os pontos se agrupam ou se distanciam. Além disso, Wong & Lee (2005), ressaltam a importância de levar em conta a questão da escala na análise espacial. A forma como os dados são combinados pode ter um grande impacto nos resultados e significados das análises. Ao juntar informações em vários níveis, como bairros, distritos ou cidades, é possível observar padrões espaciais únicos que não seriam percebidos em outras escalas. Selecionar a escala correcta é fundamental para assegurar que os resultados da análise sejam precisos e pertinentes ao contexto específico da pesquisa Druck *et al.,* (2004).

Faria (2016) complementa esse ponto de vista, ao destacar que é importância ter-se uma abordagem cuidadosa na seleção de escalas e métodos de análise para prevenir distorções nos resultados devido ao "Problema da Unidade de Área Modificável" (MAUP). Este problema mostra como os resultados

podem variar com diferentes formas de agregação espacial, impactando as conclusões tiradas dos dados.

2.9. Modelos de Regressão Espacial

De acordo com (Waldemar & Correa, 2007), os modelos de regressão espacial são fundamentais para a análise de dados em nível agregado. Para Faria (2016), os modelos de regressão desempenham um papel crucial na investigação de dados espaciais, uma vez que possibilitam a investigação e medição das conexões entre variáveis em determinado cenário. São métodos estatísticos que visam compreender como uma variável dependente (y) pode ser explicada ou prevista por uma ou mais variáveis independentes (x), também chamadas de variáveis explicativas geográficas. O modelo de regressão espacial é semelhante a uma regressão linear, mas inclui a componente espacial na variável dependente Y, conforme demonstrado por (Câmara & Correa, n.d.)

$$Y = \rho WY + X\beta + \varepsilon, \ \ \varepsilon \sim N(0, \sigma^2) \tag{1}$$

Onde,

Y é um vector de observações dependente com n áreas, X é uma matriz de variáveis explicativas com k variáveis das n áreas, $\boldsymbol{\beta}$ é um vetor de coeficientes de regressão e $\boldsymbol{\varepsilon}$ é um vetor de erros aleatórios com distribuição normal, média zero e variância constante σ^2.

De cordo com Bailey & Gatrell (1995) citado por (Lise, 2011), na análise de regressão objetiva-se:

- encontrar um bom ajuste entre valores preditos pelo modelo e os valores observados da variável dependente;
- descobrir quais variáveis explicativas contribuem de forma significativa para este relacionamento linear.

Para alcançar os objetivos mencionados, a premissa padrão pressupõe que as observações são não correlacionadas, o que implica que os resíduos ($\boldsymbol{\varepsilon}$) do modelo são igualmente independentes e não correlacionados com a variável dependente (Y) (Waldemar & Correa, 2007).

No entanto, em informações espaciais com presença de dependência espacial, a suposição comum de que as observações são não correlacionadas não é provável de ser verdadeira (Cressie, 1993). Os vestígios da regressão frequentemente exibem autocorrelação espacial, resultando em variações regionais significativas nas relações do modelo. A análise dos resíduos da regressão para detectar pistas da organização espacial é fundamental na regressão espacial. Através de testes de autocorrelação espacial, como o índice de Moran, é possível perceber que as observações apresentam uma correlação maior do que a esperada se fossem independentes (Waldemar & Correa, 2007). Quando há

autocorrelação, é preciso escolher um modelo que leve em conta as interferências provocadas por essa dependência.

As categorias dos modelos de regressão espacial são globais e locais. Anselin (1988) citado por (Estevo Felix *et al.,* 2021) afirma que modelos globais usam um parâmetro único para representar a estrutura espacial, enquanto modelos locais realizam regressões espaciais diferentes em sub-regiões, conforme Fotheringham, Brunsdon & Charlton (2002). Essa abordagem é útil em estudos de análise espacial, onde é importante entender como diferentes áreas estão relacionadas entre si e como essas relações podem influenciar variáveis de interesse.

Na análise espacial de áreas, os dados geralmente provêm de levantamentos populacionais, como censos demográficos, estatísticas de saúde e registros criminais Druck *et al.,* (2004). Parte-se do pressuposto de que há homogeneidade interna nas unidades de levantamento, ou seja, mudanças significativas ocorrem apenas nos limites dessas áreas (Faria, 2016). Druck *et al.,* (2004) citado por Faria (2016), entende que essa suposição nem sempre é verdadeira visto que as unidades de levantamento são frequentemente definidas por critérios políticos, como Estados e Municípios, ou por critérios operacionais, como sectores censitários. Não há garantia de que a distribuição dos eventos seja homogênea dentro dessas unidades ou áreas de estudo.

Segundo Faria (2016), em locais com desigualdades sociais significativas, é frequente a agregação de diferentes grupos sociais em espaços de colecta como subúrbios e áreas mais abastadas. Isso leva a indicadores distorcidos que refletem a média de grupos populacionais bastante variados. Na primeira etapa crucial da análise de dados espaciais agregados por áreas, é fundamental escolher um modelo inferencial adequado, como enfatizado por Druck *et al.,* (2004). Nesta aboradgem, cada região é considerada ter uma identidade única com uma distribuição de probabilidade exclusiva, o que o define como um modelo discreto de estatística espacial.

Os mapas coropléticos, também chamados de mapas temáticos, são comumente empregados para visualizar os padrões espaciais do fenômeno sendo analisado (Faria, 2016). Além de apresentar de forma clara as diferenças geográficas, esses mapas possibilitam abordar questões essenciais, como a existência de grupos significativos ou a aleatoriedade dos padrões vistos na região estudada.

Essa metodologia não só auxilia na compreensão dos dados agrupados por regiões, como também auxilia na identificação de regiões com características únicas e possíveis valores discrepantes, gerando assim uma análise mais minuciosa e embasada das relações espaciais dos fenômenos estudados (Faria, 2016)

2.9.1. Modelos Globais *versus* Modelos Locais

Os modelos de regressão espacial globais são essenciais para detectar padrões amplos e tendências gerais em uma determinada região (Cressie, 1993). No entanto, esses modelos podem ignorar variações significativas dentro de uma região, restringindo a capacidade de compreender heterogeneidades espaciais importantes. Em contrapartida, (Estevo Felix *et al.,* 2021)argumentam que os modelos locais são cruciais para investigar como os impactos das variáveis explicativas que mudam em diversas regiões, possibilitando interpretações mais precisas e contextualizadas. Anselin (1988) citado por (Estevo Felix et al., 2021), considera que esses modelos fornecem uma maneira eficaz de lidar com a diversidade espacial nos dados.

Os modelos locais são especialmente eficientes ao identificar padrões espaciais específicos, como agrupamentos de casos de doenças ou o efeito localizado de medidas governamentais (Druck *et al.,* (2004). Getis & Ord (1992) citado por (Waldemar & Correa, 2007) ressaltam que, ao contrário dos modelos globais, que geralmente generalizam efeitos em uma grande extensão, os modelos locais são capazes de capturar nuances espaciais significativas que podem ser negligenciadas em escalas maiores (Faria 2016). A tabela 2 abaixo apresenta as disparidades entre os modelos.

Tabela 2. Modelos globais versus Modelos locais. (Fonte: Adaptado de Druck et al., 2004)

Global	Local
Estatísticas dizem respeito a região como um todo (1 valor)	Desagregação locais das estatísticas globais (muitos valores)
Estatísticas globais e não mapeáveis	Estatísticas locais e mapeáveis
Ênfase nas similaridades da região	Ênfase nas diferenças ao longo do espaço
Procura regularidades ou leis	Procura por exceções ou hotspots locais
Ex: Regressão Clássica, Spatial Lag, Spatial Error, …	Ex: GWR, MGWR, Regimes Espaciais, …

Portanto, a tabela 1 mostra que, enquanto os modelos globais são mais generalistas, os modelos locais fornecem uma análise mais detalhada e contextualizada, essencial para estudos que lidam com a variabilidade espacial, como a análise das perdas não técnicas. Nesse estudo serão empregues os

modelos de regressão espacial local nomeadamente o modelo GWR e MGWR para a análise das PNTs, com fundamento na literatura revisada que demostra que os modelos de regressão espacial locais são mais eficazes em relação aos modelos globais pois permitem capturar a variação local do fenómeno em análise, oque se pressupõe nesse estudo.

2.9.2 Matrizes de Proximidade Espacial

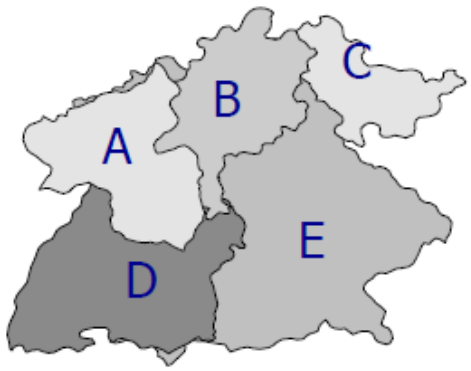
Para estimar a variabilidade espacial de dados de área, uma ferramenta básica é a matriz de proximidade espacial, também chamada matriz de vizinhança(Câmara & Correa, n.d.). Dado um conjunto de n áreas {A1,..,An}, construímos a matriz W(1) (n x n), onde cada um dos elementos wij representa uma medida de proximidade entre Ai e Aj. Esta medida de proximidade pode ser calculada a partir de um dos seguintes critérios (Faria, 2016):

$W_{ij} = 1$, se o centroide de A_i está a uma determinada distância de A_j; caso contrário $w_{ij} = 0$;

$W_{ij} = 1$, se Ai compartilha um lado comum com A_j, caso contrário $w_{ij} = 0$;

$W_{ij} = L_{ij}/L_i$, onde L_{ij} é o comprimento da fronteira entre A_i e A_j e L_i é o perímetro de A_i

Como a matriz de proximidade é utilizada em cálculos de indicadores na fase de análise exploratória, é muito útil normalizar suas linhas, para que a soma dos pesos de cada linha seja igual a 1. Isto simplifica muito vários cálculos de índices de autocorrelação espacial, como se verá a seguir (Câmara & Correa, n.d; Oliveira, 2013) A Figura 4 ilustra um exemplo simples de matriz de proximidade espacial, em que os valores dos elementos da matriz refletem o critério de adjacência e foram normalizados.



	A	B	C	D	E
A	0	0,5	0	0,5	0
B	0,25	0	0,25	0,25	0,25
C	0	0,5	0	0	0,5
D	0,33	0,33	0	0	0,33
E	0	0,33	0,33	0,33	0

Figura 5. Matriz de ponderação espacial de primeira ordem normalizada Fonte: Adaptado de (Câmara & Correa, n.d.)

Segundo (Zaidan, 2017), a ideia da matriz de proximidade espacial pode ser generalizada para vizinhos de maior ordem (vizinhos dos vizinhos).

Segundo (Estevo Felix *et al.,* 2021), a lei de Tobler, afirma que "tudo está relacionado com tudo, mas coisas mais próximas estão mais relacionadas do que coisas distantes". Isso implica que a dependência espacial está presente em todas as direções, mas enfraquece à medida que aumenta a distância entre as localizações dos dados. Essa lei é amplamente utilizada na construção da matriz de ponderação $W_{(n\times n)}$. De acordo (Faria, 2012), os critérios mencionados anteriormente e outros diversos critérios abordados na literatura, os elementos da matriz de ponderação w_{ij} tendem a ser inversamente proporcionais à distância entre os centroides das áreas A_i e A_j . Esse princípio é particularmente relevante em áreas como a epidemiologia, onde faz sentido que regiões mais próximas compartilhem características mais similares do que regiões mais distantes (Faria 2016).

No entanto, (Estevo Felix *et al.,* 2021) destaca que no caso das perdas não técnicas (PNTs), a Lei de Tobler nem sempre se aplica de forma tão direta. Embora a proximidade espacial possa influenciar o comportamento de determinadas áreas, as PNTs são fortemente afetadas por fatores socioeconômicos, estruturais e culturais, que nem sempre seguem o padrão de maior similaridade em áreas adjacentes.

Assim, enquanto a matriz de ponderação espacial baseada na distância entre áreas pode ser útil, no contexto das PNTs é necessário incorporar outros factores na análise, como a tipologia dos consumidores e o tipo de infraestrutura, que podem alterar o padrão de dependência espacial observado (Faria, 2016). Isso mostra que, em certos casos, a relação entre áreas distantes pode ser mais forte do que entre áreas vizinhas, desafiando a aplicabilidade universal da Lei de Tobler nesse contexto.

Deste modo, nesse estudo pretende aplicar como critério para construção da matriz de vizinhança, a vizinhança semântica proposta por Faria (2016), no qual a proximidade entre as áreas é determinada pelo nível de semelhança em relação às suas variáveis socioeconômicas e características da rede de distribuição. Assim, áreas próximas não precisam obrigatoriamente compartilhar fronteiras; além disso, áreas distantes podem ter um valor de proximidade w_{ij} maior do que áreas geograficamente mais próximas.

3

Material e Metodologia

Neste capítulo tem se a descrição detalhada sobre tudo o que foi realizado para o alcance dos resultados na pesquisa.

3.1. Material

Neste subcapítulo, são apresentados os materiais utilizados no desenvolvimento deste estudo, incluindo os dados, ferramentas e recursos metodológicos empregados para a análise das perdas não técnicas (PNTs).

A tabela 3 apresenta os programas utilizadas para a concretização deste trabalho.

Tabela 3. Programas utilizados

Software	**Versão**	**Descrição**
Quantum GIS	QGIS 3.34 (Prizren)	Software de Sistema de Informação Geográfica (SIG) de código aberto utilizado para visualização, edição e análise de dados espaciais.
Geoda	Geoda 1.22	Ferramenta estatística voltada para análise espacial, especialmente útil para explorar dependência espacial e autocorrelação geográfica.
Jamovi	2.3.28	Plataforma de estatística gratuita baseada em R, com interface gráfica amigável, voltada para análises estatísticas clássicas e avançadas.
MGWR	2.2	Software especializado em Regressão Geograficamente Ponderada (GWR) e Multiescalar (Multiscale GWR), utilizado para modelar relações espaciais complexas.
Rstudio	R-4.4.3	Ambiente de desenvolvimento integrado (IDE) para a linguagem R, utilizado para programação estatística, visualização de dados e modelagem.

3.1.1. Aquisição de dados

Para o alcance dos objectivos traçados neste trabalho, foram utilizados dados de diferentes fontes, referentes à área de estudo.

O Município da Cidade de Maputo é composto por 6 Áreas de Serviço ao Cliente (ASC), no entanto, devido a disponibilidade de dados, foram usados 5 ASC para este estudo.

Nesse contexto, foram usados dados das inspeções georreferenciadas, realizadas nas 5 ASC que compõem o Município de Maputo, correspondentes ao período 2022 a 2023. Estes dados foram adquiridos na Electricidade de Moçambique, concretamente no Departamento de Proteção de Receitas e Controle de Perdas (DPRP).

Foram igualmente adquiridos dados que reflectem a realidade socioeconómica da área de estudo a partir de dados disponibilizados no Censo 2017, Inquérito sobre Orcamneto Familiar (IOF) 2022 e no Anuário Estatístico 2020 disponibilizados pelo Instituto Nacional de Estatística (INE).

A Tabela 4 apresenta uma breve descrição dos modelos de regressão espacial usados para este estudo.

Tabela 4. Breve descrição dos Modelos de Regressão Espacial utilizados neste estudo

Nome do Modelo	**Instituição (Desenvolvedora)**	**País**
GWR (Geographically Weighted Regression)	University of Newcastle, UK	Reino Unido
MGWR (Multiscale Geographically Weighted Regression)	Arizona State University (Center for Spatial Analysis and Research)	EUA

3.1.2. Descrição dos dados usados

A Tabela 5 apresenta as variáveis selecionadas para a análise das perdas não técnicas (PNTs), com foco em indicadores que caracterizam as áreas de estudo. As variáveis são descritas de forma a captar aspectos importantes sobre a distribuição espacial e as características socioeconômicas e técnicas dos bairros, oferecendo uma base robusta para a modelagem das PNTs.

Tabela 5. Descrições das variáveis explicativas e dependente analisadas neste estudo.

	Variáveis	Sigla	Descrição
Variáveis Socioeconómicas	Densidade populacional	DENS_P	Medida demográfica que expressa a quantidade de indivíduos residentes em uma determinada área geográfica.
	Índice de Criminalidade (%)	IND_CR	Percentual de índice criminalidade (Consumo de drogas, roubo e assaltos).
	Renda Media dos agregados familiares em (MT)	RND	Rendimento médio mensal dos agregados familiares.
	Taxa de Desemprego	TAXA_D	indicador econômico que representa a proporção da população economicamente activa
Variáveis Técnicas	Casas em situação de irregularidade com contadores integrados	CONT_INT	Número de casas em situação irregular com contadores integrados.
	Casas em situação de irregularidade com contadores SPLIT	CONT_SPLIT	Número de casas em situação irregular com contadores SPLIT, mas com o contador instalado no postilete
Variável dependente	Casas em situação de irregularidade (Furto e Fraude)	PNT	Número de casas em situação irregular

3.2. Métodos

De acordo com Faria (2016), não existe uma abordagem universal para lidar com o desafio das PNTs. Cada município tem sua própria realidade socioeconômica e sociocultural. Assim, a compreensão das perdas é facilitada ao conhecer o perfil do local estudado e as subáreas onde elas ocorrem.

Com vista a obter informações detalhadas e qualitativas adicionais a literatura revisada, que pudessem auxiliar na definição das variáveis do estudo que melhor caracterizam a área em causa, foram conduzidas entrevistas com especialistas e profissionais do sector elétrico. Usou-se da entrevista que é uma das principais técnicas de colectas de dados e pode ser definida como "conversa realizada face a

face pelo pesquisador junto ao entrevistado, seguindo um método para se obter informações sobre determinado assunto" (Cervo & Bervian, 2002, p.65).

Cada entrevista teve uma duração média de 30 minutos e foi conduzida pessoalmente, conforme a disponibilidade dos participantes. Importa referir que, foi usada a entrevista do tipo semiestruturada[3], cujas questões constam e podem ser conferidas no (Apêndice I**)** da presente pesquisa.

O processo de seleção dos entrevistados seguiu a amostragem aleatória simples, visando garantir a diversidade e a representatividade das respostas. De acordo com Mattar (2001) este tipo de amostragem é um tipo em que cada constituinte da população pode ser nomeada para compor a amostra e gozar das mesmas chances dos seus constituídos. Isto significa que, o pesquisador tem a gestão sobre o erro amostral (a diferença entre um resultado amostral e o verdadeiro resultado populacional) na pesquisa.

Salientar que, no presente estudo, as unidades amostrais são escolhidas por acaso, isto é, das entrevistas aplicadas aos participantes da pesquisa, foram dadas as mesmas probabilidades de inclusão a todos entrevistados sem olhar para as suas especificidades.

Para Lakatos & Marconi, (2001), a técnica de colecta diz respeito a parte prática da colecta de dados, é, portanto, o momento em que se reúnem regras ou processos para praticar a ciência. Assim, durante este momento, diferentes técnicas podem ser utilizadas, sendo as mais empregues: a entrevista, o questionário, a observação e a pesquisa documental.

Os técnicos submetidos as entrevistas são provenientes de várias Áreas de Serviço ao Cliente (ASC) da Cidade de Maputo composta por, nomeadamente:

- ASC de KaMavota;
- ASC de KaMaxaqueni;
- ASC de Kampfumo;
- ASC de KaMubukana;
- ASC de Nlhamankulu.

[3] As entrevistas semiestruturadas podem ser definidas como "uma lista das informações que se deseja de cada entrevistado, mas a forma de perguntar (a estrutura da pergunta) e a ordem em que as questões são feitas irão variar de acordo com as características de cada entrevistado" (Gil, 2002, p.25).

Durante a entrevista foram selecionados 10 técnicos do total de 30 entrevistados, com base na sua experiência profissional (mínimo de 5 anos no setor das inspeções), envolvimento em projetos relacionados à detecção e prevenção de fraudes elétricas, e disponibilidade para participar da pesquisa. O perfil d dos técnicos escolhido deveu-se ao facto de estes estarem melhor abalizados com conhecimento prático sobre perdas não técnicas, o que lhes permite fornecer dados com maior propriedade.

A tabela 6 ilustra as constatações feitas nas entrevistas as quais foram de grande importância para a definição das variáveis considerando a realidade geográfica da área de estudo:

Tabela 6. Constatações feitas durante as entrevistas com os 10 técnicos selecionados.

Constatações
A maioria dos clientes fraudulentos possuem contadores integrados dentro das suas residências
Cerca de 75% dos clientes fraudulentos são os que tem instalações monofásicas
Cerca de 40% dos clientes fraudulentos possuem contador *split* com o mesmo instalado no postilete de suas residências.
Clientes com o nível de renda acima de 3 salários mínimos, geralmente tem vários eletrodomésticos e por razão disso, são os mais propensos a acções ilegais como a fraude visto que os mesmos eletrodomésticos acarretam custos elevados de consumo de energia.
Os bairros como Maxaquene, Mahotas que apresentam infraestrutura urbana desordenada "assentamentos informais" são as que mais tem clientes fraudulentos, ademais, a rede nesses bairros é obsoleta, razão pela qual vários clientes partilham a mesma baixada devido a falta de arruamento que permita a colocação das baixadas para cada cliente.
Os bairros com alto índice de criminalidade como e o caso de Maxaquene apresentam altos níveis de clientes fraudulentos, pois realizar actividades de inspeções nesses bairros chega a ser um calcanhar de Aquiles para os técnicos, visto que nesses bairros muitas das vezes os técnicos são agredidos violentamente, o que coloca em causa a sua integridade física.

3.2.1. Pré-processamento

Conforme exposto no capítulo 4 sobre a aquisição de dados, este estudo concentrou-se na análise das perdas não técnicas (PNTs) no Município da Cidade de Maputo (MCM). Durante a etapa inicial de

preparação dos dados, identificou-se a necessidade de realizar transformações estatísticas para melhorar a adequação dos dados às análises subsequentes.

Dessa forma, foi adoptada a transformação logarítmica para algumas variáveis, uma abordagem amplamente utilizada em estudos que lidam com dados que apresentam grande amplitude de variação. Essa transformação é eficaz para estabilizar a variância, reduzir o impacto de valores extremos e aproximar os dados de uma distribuição normal, como recomendado em análises espaciais e regressões ponderadas geograficamente (GWR) (Druck *et al.,* 2004; Sinha, 2018).

3.2.2. Métodos de Transformação de Dados

Em geral uma transformação nos dados envolve a aplicação de alguma fórmula matemática aos valores de um variável, buscando obter os dados em uma forma mais apropriada para a posterior modelagem, maximizando a informação, satisfazendo premissas de modelos ou simplesmente prevenindo erros(Mel, 2021). Entre as transformações mais realizadas e importantes está a normalização ou padronização dos dados(LÓPEZ FLORES, 2017), feita com o objetivo de homogeneizar a variabilidade das variáveis de uma base de dados, criando um intervalo de amplitude similar onde todas as variáveis irão residir. Em geral a normalização é necessária no caso de variáveis com unidades diferentes ou dispersões muito heterogêneas.

3.2.2.1 Transformação Logarítmica

A transformação logarítmica é uma técnica estatística amplamente utilizada em analise de dados e ciência de dados para transformar variáveis que apresentam uma distribuição assimétrica ou que possuem uma grande variação em suas magnitudes(LÓPEZ FLORES, 2017). Essa transformação é especialmente útil quando se trabalha com dados que seguem uma distribuição exponencial ou que contêm outliers, pois ajuda a estabilizar a variância e a tornar os dados mais adequados para análise estatística. Ao aplicar a transformação logarítmica, os dados são convertidos em uma escala logarítmica, o que pode facilitar a interpretação e a visualização dos resultados.

Sendundo (LÓPEZ FLORES, 2017), a transformação logarítmica é realizada aplicando a função logarítmica a cada valor de uma variável. A forma mais comum de transformação logarítmica é o logaritmo natural, que utiliza a base 'e' (aproximadamente 2,718). Outra opção é o logaritmo na base 10, que é frequentemente utilizado em contextos onde se deseja uma interpretação mais intuitiva. A fórmula básica para a transformação logarítmica é dada por (equação 2):

$$(y' = \log(y) \quad (2)$$

Onde,

(y) é o valor original e (y') é o valor transformado. É importante ressaltar que a transformação logarítmica só pode ser aplicada a valores positivos, pois o logaritmo de zero ou de um número negativo não é definido.

3.2.3 O Problema da Unidade de Área Modificável e o Problema de Escala

Neste estudo, foi utilizado o método de agregação espacial descrito por Wong (2004), que busca representar a superfície terrestre, um meio contínuo, por meio de áreas com fronteiras bem definidas. A definição dessas áreas baseia-se na agregação de regiões com características semelhantes relacionadas ao fenômeno em análise. Este processo está diretamente associado ao Problema da Unidade de Área Modificável (PUAM), ou Modifiable Areal Unit Problem (MAUP), que se refere às inconsistências estatísticas geradas pelas diferentes formas de agregação espacial aplicadas a uma observação. Segundo Faria (2016), além da influência das fronteiras das zonas na análise, como mencionado, também existe a influência da escala que está ligada à dimensão da unidade de área considerada. Conforme Druck *et al.* (2004), a diminuição da escala (zonas mais extensas) geralmente homogeneíza os dados, diminui a variação aleatória e intensifica correlações que parecem ser mais robustas do que em áreas menores (que têm escala mais alta).

A ampliação da escala para áreas menores considera a diversidade dos dados, aumentando a variação aleatória e diminuindo correlações. Com maior "zoom", o fenômeno torna-se mais aleatório. Não existe uma escala perfeita; modelos devem atender ao que se deseja elucidar. A escala é um problema inerente aos dados agrupados (Faria, 2016). No caso deste estudo, os dados disponibilizados pelo INE estão agregados por áreas, mas fazem referência a características e relacionamentos individuais. Para possibilitar a aplicação dessas informações na análise espacial, foi necessário adaptar as unidades de análise, considerando as características específicas de cada residência, apesar da agregação das áreas.

3.2.4. Validação dos Modelos

Após a correção de todas as simulações, o desempenho dos modelos foi avaliado utilizando métricas estatísticas como o coeficiente de determinação (R^2), o Critério de Informação de Akaike (AIC) e o Índice de Moran dos resíduos. O R^2 mede a força da associação linear entre os valores observados e

ajustados, variando de 0 a 1, onde valores mais próximos de 1 indicam maior capacidade do modelo em explicar a variância dos dados (Rhymee et al., 2022).

$$R^2 = \frac{n\Sigma Q_i Q_i' - Q_i Q_i'}{\left(\sqrt{n(\Sigma Q_i^2) - (\Sigma Q_i)^2}\right) - \left(\sqrt{n(\Sigma Q_i'^2) - (\Sigma Q_i^1)^2}\right)} \quad (3)$$

Onde Qi descreve os valores observados do modelo, Qi' descreve os valores simulados do modelo e, n representa o número de valores de dados no período de estudo.

O AIC, por sua vez, avalia a qualidade do ajuste do modelo, penalizando a complexidade. Modelos com menor AIC são preferíveis, indicando um equilíbrio entre qualidade do ajuste e simplicidade do modelo (Burnham & Anderson, 2002). O Critério de Informação de Akaike AIC foi desenvolvido a partir a distância de Kullback-Leibler (K-L), a qual é uma distância entre o modelo verdadeiro, que geralmente é uma abstração conforme a equação 3, e o modelo candidato.

$$AIC = -2\sum_{1=1}^{n} \ln L(\hat{u}_i, y_i) + 2 * n \quad (3)$$

onde y_i é o i-ésimo valor da resposta e $\hat{u}_i$ é a estimativa de y_i ,quando se ajusta um modelo de *p* parâmetros por meio da maximização da Função de Log Verossimilhança (FLV). O termo que se adiciona à FLV, chamado de função de penalidade, tem a finalidade de corrigir um viés proveniente da comparação de modelos de diferentes números de parâmetros. Entre vários modelos candidatos, deve ser escolhido aquele que apresentar o menor valor de AIC.

Por fim, o Índice de Moran é usado para verificar a autocorrelação espacial nos resíduos, fundamental para avaliar se o modelo capturou adequadamente os padrões espaciais nos dados. Valores significativos do Índice de Moran indicam a presença de autocorrelação espacial residual, sugerindo que o modelo precisa ser ajustado para incorporar dependências espaciais (Anselin, 1995) citado por (Fernando G. K. Guarda *et al.,* 2018). A combinação dessas métricas fornece uma avaliação robusta da adequação e desempenho dos modelos ajustados.

3.2.5. Análise Exploratória de Dados

Para (Pazderin *et al.,* 2023), uma fase essencial na análise exploratória de regiões é a identificação da estrutura de correlação espacial que mais adequadamente representa os dados. Esta análise possibilita mensurar o grau de autocorrelação espacial entre as áreas, oferecendo uma melhor compreensão dos padrões espaciais que estão presentes. O resultado alcançado é diretamente afetado pela definição da vizinhança escolhida na matriz de ponderação, conforme abordado na seção anterior(Faria, 2016).

A definição adequada do nível de dependência entre variáveis é fundamental para sustentar hipóteses ligadas à presença de dependência espacial no fenômeno que está sendo analisado(Waldemar & Correa, 2007). Esse método não só simplifica a detecção de padrões, mas também oferece suporte para a criação de relações causais e a compreensão dos mecanismos subjacentes que regulam a distribuição espacial das variáveis examinadas(Faria, 2016). Dessa forma, a estrutura de correlação espacial se apresenta como um recurso essencial no procedimento de pesquisa, guiando escolhas metodológicas e ajudando na formação de análises mais sólidas e embasadas.

3.2.5.1 Coeficiente de Correlação de Spearman rs

Segundo Faria (2016), o coeficiente de correlação de Spearman rs é uma alternativa eficaz para determinar a relação entre duas variáveis, baseando-se na ordenação ou ranking dos valores das variáveis em questão. O cálculo de rs envolve a organização dos valores das duas variáveis em ordem crescente (ou decrescente), conforme a equação 4, substituindo cada valor original da série de dados por um número inteiro positivo que representa sua posição hierárquica ou ordinal dentro da série.

$$rs = 1 - \frac{6\Sigma\, d^2}{n(n+1)(n-1)} \tag{4}$$

O coeficiente de Spearman é preferível ao de Pearson por ser menos sensível a assimetrias e *outliers*, não requerendo distribuição normal nos dados. Ele identifica melhor associações entre variáveis em um conjunto de dados, segundo Ferreira (2014, p. 104).

De acordo com Faria (2016), uma alternativa para avaliar o nível de significância do coeficiente de correlação é através do teste *t* de *Student*. Formulam-se duas hipóteses sobre rs:

- H0: o valor do coeficiente de correlação ocorreu por mero acaso.
- H1: o valor do coeficiente de correlação é maior do que se poderia esperar, caso acontecesse por acaso.

Para o teste das hipóteses supracitadas, utiliza-se a distribuição *t* de *Student* com $n-2$ graus de liberdade. A equação (5) estabelece a relação entre os parâmetros envolvidos no cálculo da função *t*:

$$t = \frac{r_s\sqrt{n-2}}{\sqrt{(1-r_s^2)}} \tag{5}$$

sendo n o número de amostras.

Em seguida, compara-se o valor t calculado com os valores críticos da distribuição t de Student para diversos graus de liberdade para obter o valor do nível de significância. Depois de obter o valor de t, ele é comparado com os valores críticos da distribuição t de Student para determinar a significância estatística. Os coeficientes de correlação variam de -1 a 1, onde -1 representa uma correlação negativa perfeita e 1 uma correlação positiva perfeita conforme mostra a tabela 7.

Tabela 7. Intervalos do coeficiente de correlação e magnitude da correlação. Fonte: Adaptado de (Faria, 2016).

Intervalos do Coeficiente de Correlação de Spearman r_s	***Magnitude de Correlação***
$0 < r_s < 0.3$	Fracamente positiva
$0.3 < r_s < 0.8$	Moderadamente positiva
$0.8 < r_s < 1.0$	Altamente positiva
$- 0.3 < r_s < 0$	Fracamente negativa
$- 0.8 < r_s < - 0.3$	Moderadamente negativa
$- 1.0 < r_s < -0.8$	Altamente negativa
$r_s = 0$	Não há correlação

3.2.5.2 Função de Kernel não Paramétrico

Diversas abordagens estatísticas foram desenvolvidas para a estimativa de intensidade em processos de pontos (Diggle, 1983; Druck *et al.,* 2004). Cada método apresenta suas próprias vantagens e desvantagens em diferentes contextos (Li *et al.,* 2020). Neste estudo, a estimativa não paramétrica por meio de uma função de kernel suavizado foi empregada para modelar a intensidade de eventos de perdas não técnicas de energia elétrica no MCM. Este método é reconhecido por sua flexibilidade na captura de variações locais na intensidade dos eventos (Wang *et al.,* 2021; Chen *et al.,* 2019).

No entanto é preciso levar em consideração que quando se estima a intensidade via função de *kernel*, a escolha chave não é a escolha da função de *kernel* paramétrico ou não paramétrico, mas sim da largura de banda, *b*. Bivand, Pebesma e Gómez-Rubio (2008) citado por Faria (2016) demonstraram empiricamente que a intensidade do processo de pontos para diferentes funções de *kernel* produzem estimativas semelhantes para uma mesma largura de banda; no entanto, para uma mesma função de *kernel*, diferentes valores de largura de banda produzem resultados totalmente distintos. A intensidade estimada $\hat{\lambda}(s)$ foi calculada usando a fórmula descrita na equação (6):

$$\hat{\lambda}(s) = \frac{1}{b^2}\sum_{i=1}^{n} K\left(\frac{\|s-s_i\|}{b}\right) / \, q(\|s\|) \tag{6}$$

onde $k(a)$é uma função de kernel simétrica e $q(\|s\|)$ é uma função de correção de borda para compensar a perda de dados próxima aos limites da região de estudo conforme ilustra a equação 7.

$$k(a) = \begin{cases} \frac{3}{\pi}(1 - \|a\|^2)^2 \; se \; a \in (-1,1) \\ 0, caso \; contrário \end{cases} \tag{7}$$

A seleção da largura de banda *b* em estimativas de intensidade via função de kernel é complexa, e não há um método universalmente aplicável (Faria, 2016). Geralmente, múltiplos valores são testados para determinar o mais adequado ao processo em estudo. Fernando *at al.,* (2018) sugerem um critério que minimiza o erro quadrático médio em processos estacionários. A escolha da largura de banda é mais crítica do que a função de kernel em si, uma vez que diferentes larguras podem levar a resultados distintos para a mesma função.

3.2.5.3 Função de distribuição Cumulativa (G*) ou vizinho mais próximo

A seleção da função estatística para a análise espacial em processos de pontos varia de acordo com o objetivo da pesquisa(Eduardo *et al.,* 2006). A função **G** é apropriada para examinar distâncias entre eventos próximos, capturando interações locais, como padrões de repulsão ou agregação (Baddeley, Rubak & Turner, 2015).

Na análise das perdas não técnicas de energia elétrica, a função G se destaca por ser mais eficiente na identificação de pontos críticos em regiões específicas, essenciais para detectar padrões de roubo e desperdício (Diggle, 2014). Dessa forma, a sua utilização é explicada neste trabalho, uma vez que fornece uma avaliação minuciosa das interações próximas, fundamental para a compreensão da disposição das perdas no espaço.

A função **G** mensura a distribuição de distâncias a partir de um evento arbitrário i para o evento mais próximo j. Tais distâncias são definidas como $d_i = min_j \{d_{ij}, \forall j \neq i\}$ sendo i=1, …, n. A função G é definida em (8).

$$\hat{G}(s) = \frac{\#\{d_i : d_i \leq r \forall i\}}{n} \tag{8}$$

O numerador de (8) representa o número de eventos (#) inscritos em um círculo com centro em i e raio r; n é o número total de pontos inscritos no círculo. Sob Complete Spatial Randomness (CSR), a função G teórica é obtida via (9):

$$G = 1 - e^{\lambda \pi r^2} \tag{9}$$

onde λ é a *intensidade* (propriedade de primeira ordem) e representa o número médio de eventos por unidade de área.

De acordo com (Faria 2016), a avaliação da semelhança com o padrão de (CSR) pode ser feita ao comparar a função $\hat{G}(\mathrm{r})$ empírica, com a função G(r), teórica, em diferentes distâncias (r). Para isso, utiliza-se a simulação de Monte Carlo para gerar um envelope de curvas que representam o padrão CSR, com intensidade estimada λ para a área em estudo. Caso a função empírica descrita na equação (8) esteja dentro do envelope, os pontos são considerados aleatoriamente distribuídos, indicando que a análise espacial não contribui em nada para a compreensão do problema. Se estiver acima, aponta para uma distribuição em clusters; se abaixo, para dispersão espacial (Baddeley *et al.,* 2015).

3.2.5.4 Índice Local de Moran I

O Índice Local de Moran, originalmente proposto por Luc Anselin (1994), é amplamente utilizado em análises espaciais para identificar clusters e outliers locais Eduardo *et al.,* 2006. Segundo (Faria, 2016; Fernando G. K. Guarda *et al.,* 2018). O Índice de Moran, também conhecido como Moran's I, é uma medida estatística amplamente utilizada na análise espacial para avaliar a autocorrelação espacial de um conjunto de dados(Pazderin *et al.,* 2023). Essa ferramenta é fundamental para entender como as variáveis estão distribuídas geograficamente e se há padrões significativos que indicam a presença de agrupamentos ou dispersões. O Índice de Moran varia de -1 a +1, onde valores próximos a +1 indicam uma forte autocorrelação positiva, ou seja, áreas similares tendem a se agrupar, enquanto valores próximos a -1 indicam autocorrelação negativa, sugerindo que áreas semelhantes estão distantes umas das outras.

No contexto das PNTs, o Índice Local de Moran (ILM) é empregue para identificar áreas de maior concentração de fraudes ou irregularidades, oferecendo insights sobre os padrões espaciais associados a fatores socioeconômicos, densidade populacional e características dos contadores (Fernando G. K. Guarda *et al.,* 2018). Por exemplo, áreas com maior densidade de contadores do tipo SPLIT ou regiões com rendas significativamente diferentes podem apresentar padrões de PNTs que se destacam em relação às suas vizinhanças (Eduardo *et al.,* 2006).

O cálculo do Índice de Moran envolve a comparação de um valor observado em uma localização específica com os valores observados em suas vizinhanças(Eduardo *et al.,* 2006). Para calcular o índice, é necessário definir uma matriz de pesos espaciais que determina a relação entre as diferentes unidades de análise. Essa matriz pode ser baseada em distâncias, adjacências ou outras definições de proximidade(Faria, 2016). Segundo (Eduardo *et al.,* 2006), o Índice de Moran é então calculado utilizando a fórmula que considera a soma dos produtos das diferenças entre os valores observados e a média, ponderados pela matriz de pesos conforme a equação 10.

$$\boldsymbol{I} = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_i} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_i(y_i-\bar{y})(y_j-\bar{y})}{\sum_{i=1}^{n}(y-\bar{y})^2} \tag{10}$$

Onde,

n: número de áreas ou observações; $\mathbf{y_i}$: valor do atributo na área i; $\boldsymbol{\bar{y}}$**:** valor medio do atributo na região de estudo; $\mathbf{W_{ij}}$**:** pesos atribuídos conforme a conexão entre áreas.

No caso de da sua aplicação em perdas não tecnicas (PNTs), isso significa que, para cada unidade consumidora ou área, é possível avaliar a contribuição individual para o padrão espacial geral, destacando regiões prioritárias para inspeções ou intervenções. Dessa forma, o Índice de Moran não apenas identifica hotspots de perdas, mas também orienta estratégias mais eficientes de monitoramento e mitigação.

3.3. Limitações de Estudo

Apesar dos avanços alcançados com a aplicação de modelos de regressão espacial na análise das perdas não técnicas (PNT) na rede eléctrica de distribuição de baixa tensão no Município da Cidade de Maputo, o presente estudo apresenta algumas limitações que devem ser consideradas.

Em primeiro lugar, a disponibilidade e a qualidade dos dados representaram um desafio significativo. Muitos dos dados utilizados foram provenientes de fontes secundárias, podendo conter inconsistências ou lacunas temporais e espaciais, o que pode afectar significativamente a precisão dos resultados obtidos nos modelos espaciais.

Além disso, o estudo se concentrou apenas em um recorte geográfico específico de 5 Áreas de Serviço ao Cliente no Município da Cidade de Maputo o que limita a generalização dos resultados para outros contextos urbanos ou regiões com características socioeconômicas e técnicas diferentes.

Outra limitação relevante diz respeito à natureza estática da análise. A aplicação dos modelos de regressão espacial foi feita com base em dados de um único período, o que impossibilita observar variações temporais e dinâmicas que poderiam enriquecer a compreensão do fenômeno ao longo do tempo.

O estudo também se restringiu a variáveis socioeconômicas e tecnicas selecionadas, como densidade populacional, taxa de desemprego, índice de criminalidade e condições de infraestrutura da rede, não abrangendo outras dimensões que poderiam influenciar as PNT, como factores institucionais, normativos, culturais ou tecnológicos.

Por fim, embora os modelos de regressão espacial, como GWR e MGWR, tenham proporcionado uma análise mais refinada da variabilidade espacial dos fatores que explicam as perdas não técnicas, eles ainda apresentam limitações quanto à escolha de parâmetros, sensibilidade a outliers e pressupostos estatísticos que, se violados, podem comprometer a robustez dos resultados.

3.4. Fluxograma Metodológico

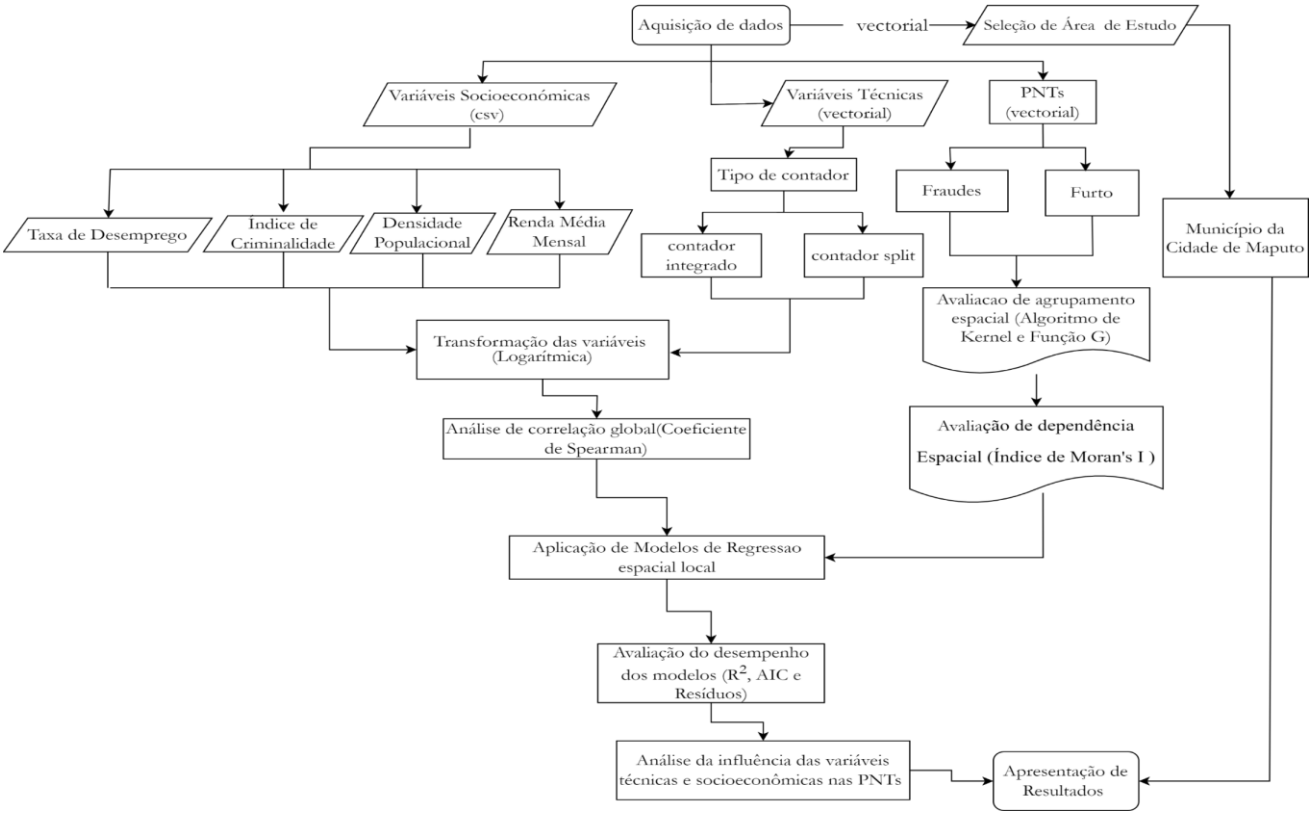


Figura 6. Fluxograma Metodológico

4

Discussão de Resultados

Nesta seção, os resultados obtidos com a aplicação da metodologia proposta são apresentados por meio de tabelas, gráficos e mapas, com o objetivo de descrever de maneira didática a análise realizada.

4.1. Análise Exploratória Geograficamente Ponderada

De acordo com Faria (2016), antes da modelagem espacial, é realizada uma fase de análise exploratória que envolve a visualização dos dados em mapas, além da identificação de padrões de dependência espacial do fenômeno em questão. Associa-se cada cor a um intervalo de classe. A segmentação por intervalos de valores para a construção das legendas dos mapas temáticos é feita com base no critério de intervalos de desvio padrão. Segundo Ferreira (2014), esta técnica é a que melhor considera a distribuição estatística das variáveis a serem mapeadas.

As Figuras 7 – 9 apresentam mapas temáticos com a distribuição espacial das variáveis explicativas analisadas. Cada cor corresponde a um intervalo de classe, e as legendas foram elaboradas com base no critério de intervalos de desvio padrão, conforme a técnica sugerida por Ferreira (2014), que considera a distribuição estatística das variáveis.

Neste contexto, alguns comentários são feitos a partir da análise das Figuras:

- Conforme Fig. 7a, a variável Renda Média Mensal (MZN) indica que as áreas situadas na porção norte do município apresentam rendas mais elevadas, logo, trata-se de uma região de alto poder aquisitivo, enquanto as regiões ao sul e leste da cidade são aquelas com menor poder aquisitivo.
- Conforme Fig. 7b, os índices de criminalidade apresentam distribuição elevada na porção central e leste do município, refletindo maior vulnerabilidade socioeconômica, enquanto as áreas ao norte apresentam níveis mais baixos.
- Conforme Fig. 8a, as regiões centrais e sudeste do município concentram uma maior quantidade de contadores integrados, trate se de uma porção do MCM que a presenta uma certa precariedade em termos de infraestrutura da rede de distribuição, o que constitui um atrativo para as práticas ilegais (furtos e fraudes).

- Conforme Fig. 8b, as áreas a este e sudeste apresentam uma menor concentração de contadores Split, logo trata se de bairros do MCM que apresentam maior número de assentamentos informais, visto que muitas vezes o técnico se vê obrigados a atribuir a mesma baixa a vários clientes devido a falta de espaço para a implantação de devida de Postes de apoio.
- Conforme Fig. 9a, a densidade populacional é maior nas áreas centrais e norte do município, enquanto as áreas periféricas a leste apresentam densidades mais baixas.
- Conforme Fig. 9b, os índices de criminalidade permanecem elevados nas regiões leste e sudeste, reforçando o impacto das condições socioeconômicas, que por conseguinte, essas áreas são as de difícil acesso para os técnicos nas suas campanhas de inspeções, propiciando assim o aumento das PNTs.

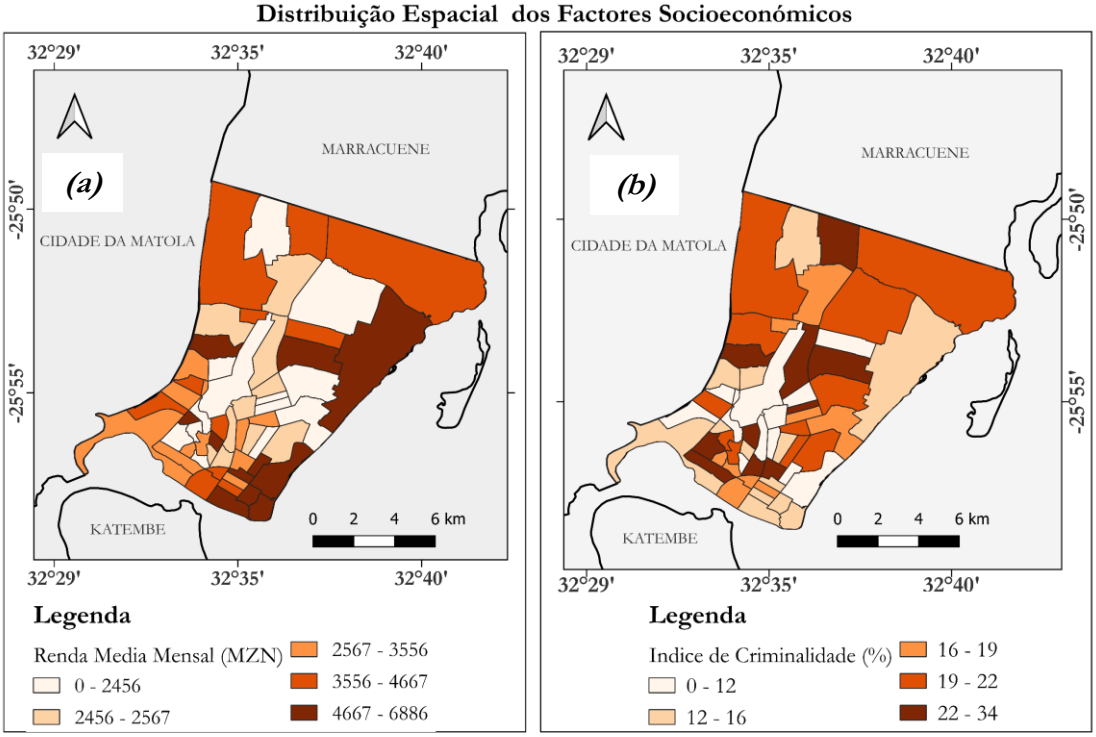


Figura 7. Distribuição espacial por bairro de variáveis Socioeconómicas: ***(a)*** Renda Media Mensal; ***(b)*** Índice de criminalidade.

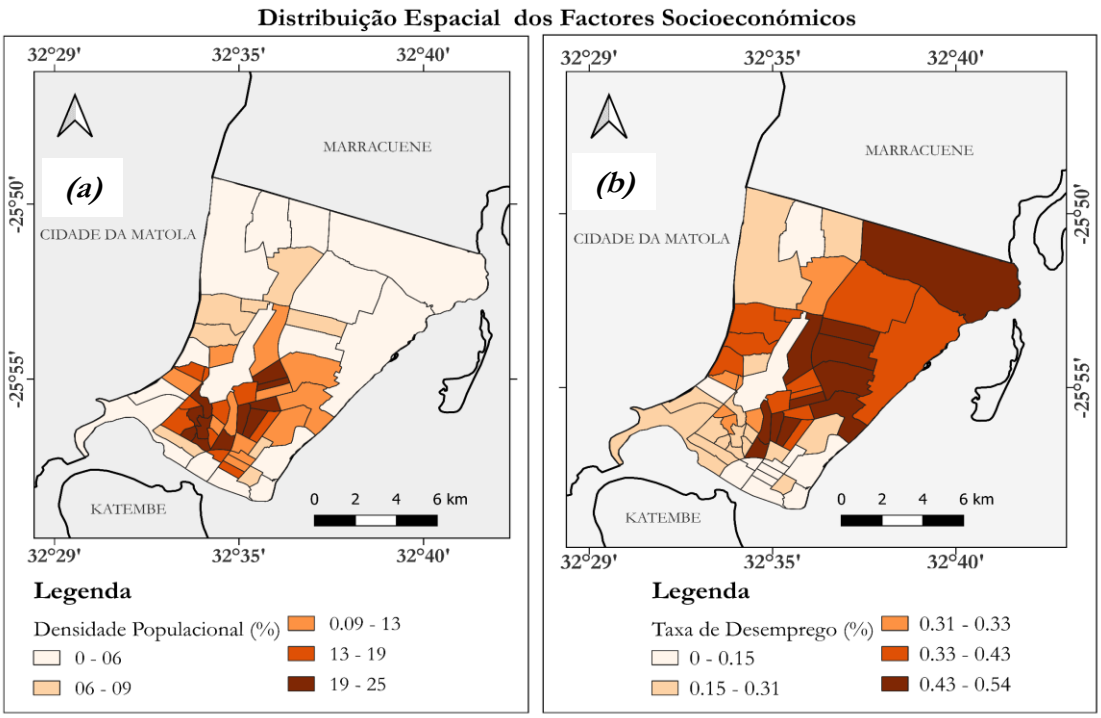


Figura 8. Distribuição espacial por bairro de variáveis Tecnicas: ***(a)*** Densidade populacional; ***(b)*** Taxa de desemprego.

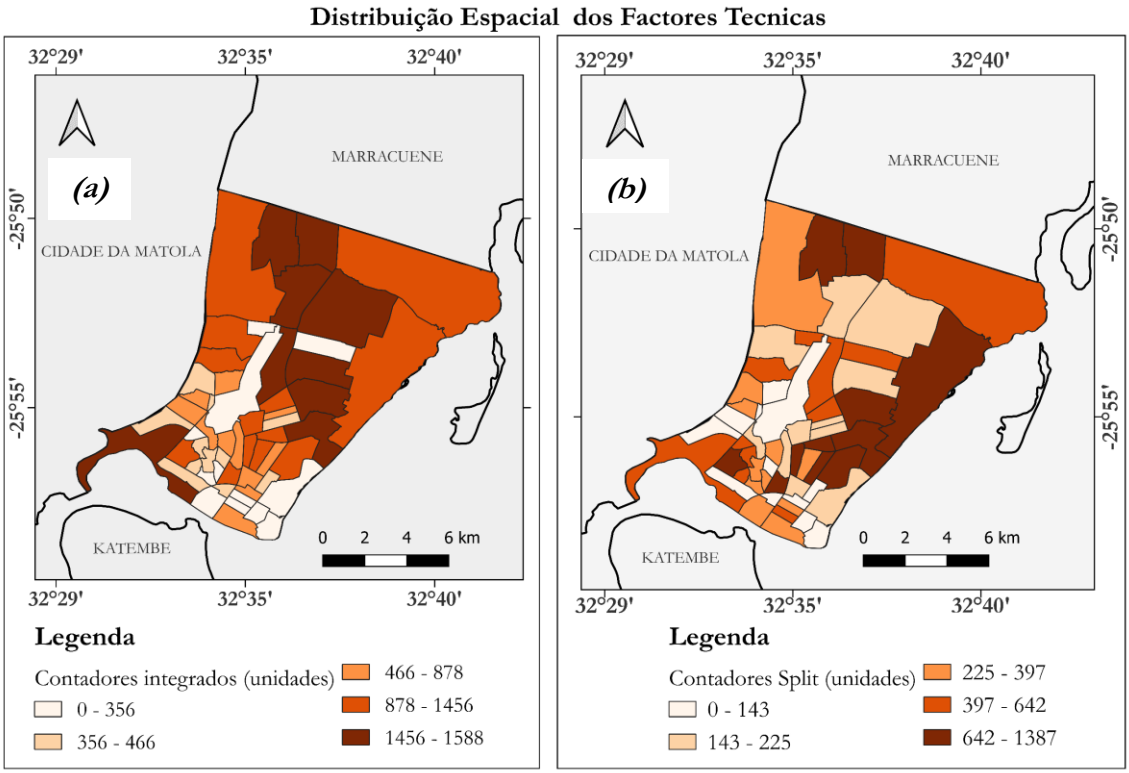


Figura 9. Distribuição espacial por bairro de variáveis Tecnicas: ***(a)*** Contadores Integrados; ***(b)*** Contadores Split

Para além de analisar as variáveis visualmente por meio de mapas temáticos como proposto por Faria (2016), é fundamental também considerar a correlação entre as variáveis em análise. Nesse cenário, é exibida na figura 10 a matriz simétrica dos coeficientes de correlação de Spearman rs.

Matriz de Correlações

		PNT	RND	IND_CR	DENS_P	TAXA_D	CONT_INT	CONT_SPLIT
PNT	Rho de Spearman	—						
	gl	—						
	p-value	—						
RND	Rho de Spearman	−0.498***	—					
	gl	54	—					
	p-value	< .001	—					
IND_CR	Rho de Spearman	0.448***	−0.187	—				
	gl	54	54	—				
	p-value	< .001	0.167	—				
DENS_P	Rho de Spearman	0.140	−0.371**	0.094	—			
	gl	54	54	54	—			
	p-value	0.303	0.005	0.493	—			
TAXA_D	Rho de Spearman	0.532***	−0.317*	0.230	0.322*	—		
	gl	54	54	54	54	—		
	p-value	< .001	0.017	0.088	0.015	—		
CONT_INT	Rho de Spearman	0.877***	−0.351**	0.372**	−0.035	0.434***	—	
	gl	54	54	54	54	54	—	
	p-value	< .001	0.008	0.005	0.798	< .001	—	
CONT_SPLIT	Rho de Spearman	0.656***	−0.230	0.352**	0.273*	0.452***	0.422**	—
	gl	54	54	54	54	54	54	—
	p-value	< .001	0.088	0.008	0.042	< .001	0.001	—

Nota. * p < .05, ** p < .01, *** p < .001

Figura 10. Matriz de correlações de Spearman.

Com base na Figura 10, seguem algumas análises:

- A correlação global entre as variáveis PNT e RND é moderadamente negativa de -0,498 (p < 0,001). Isso indica uma tendência de que áreas com maior densidade de redes (RND) tenham menos perdas não técnicas (PNT), o que pode ser atribuído a melhores condições de monitoramento e manutenção em redes mais densas.
- A correlação global entre PNT e IND_CR é moderadamente positiva de **0,448** (p < 0,001). Isso sugere que subáreas com maior índice de criminalidade (IND_CR) tendem a registrar mais perdas não técnicas, possivelmente devido a um ambiente propício para práticas ilegais, como furtos de energia.

- A correlação global entre PNT e DENS_P é fracamente positiva de 0,140 e não significativa (p = 0,303). Embora essa relação seja fraca, sugere-se que subáreas mais densamente povoadas podem ter maior potencial de perdas, mas outros fatores podem estar influenciando essa relação.
- A correlação global entre PNT e TAXA_D é moderadamente positiva de **0,532** (p < 0,001). Isso indica que uma maior taxa de desemprego (TAXA_D) está associada a um aumento nas perdas não técnicas, refletindo um potencial impacto socioeconômico, como a dificuldade em pagar contas de energia.
- A variável CONT_INT apresenta uma forte correlação positiva de **0,877** com PNT (p < 0,001). Isso sugere que contadores integrados (CONT_INT) estão significativamente associados a áreas com maiores perdas não técnicas, possivelmente devido à facilidade de manipulação desses dispositivos.
- A correlação global entre PNT e CONT_SPLIT é moderadamente positiva de **0,656** (p < 0,001). Isso indica que contadores SPLIT também estão associados a perdas não técnicas, embora em menor intensidade que os contadores integrados.

Com bases nos resultados observados, em que 54% das relações significativas apresentam correlação positiva e 46% apresentam correlação negativa, podemos observar uma distribuição quase equilibrada entre as duas direções de correlação. Essa combinação de correlações positivas e negativas sugere que os fatores analisados influenciam as perdas não técnicas (PNTs) de maneira distinta em diferentes em diferentes áreas do Município da Cidade de Maputo.

A correlação positiva indica que, em algumas áreas, o aumento de certos fatores (como densidade populacional, tipo de contador ou renda) pode estar relacionado ao aumento das PNTs, enquanto a correlação negativa sugere que, em outras áreas, o aumento desses fatores pode estar associado à redução das PNTs. Essa diversidade de padrões de correlação reforça a necessidade de aplicar métodos de análise local, para explorar a variabilidade espacial desses efeitos. Os métodos locais permitiriam identificar áreas específicas onde as relações entre as variáveis podem ser mais complexas e variar de forma significativa, permitindo um entendimento mais aprofundado do impacto espacial dos fatores socioeconômicos e técnicos nas PNTs(Zaidan, 2017).

A ausência de uma forte correlação entre as variáveis em escala global não significa necessariamente uma ausência de correlação forte a nível local. Além disso, pode haver variáveis que possuam uma fraca correlação com as perdas, porém são relevantes para uma compreensão mais aprofundada do problema qualitativo e vice-versa (Faria, 2012). De acordo com Rodrigues (2012), o índice global destaca as

semelhanças, assumindo que cada subárea pode ser expressa por um único valor. Contudo, a existência de disparidades regionais torna essa suposição inválida (Pearl, 2009).

Em algumas situações, pode haver resultados distintos ao analisar os dados de forma global e local. Pearl, (2009) & Faria, (2016) ressaltam que a relação não significa causa e efeito, já que diversos modelos não levam em consideração variáveis essenciais. Ainda assim, é possível observar uma alta correlação espacial, sendo necessário ter cautela ao interpretar os dados para evitar conclusões errôneas (Fernando G. K. Guarda et al., 2018).

4.1.1 Avaliação do Agrupamento Espacial das PNTs

Segundo Qadir e Jalal (2017), a análise da dependência espacial de fenômenos tornou-se nos últimos tempos uma estratégia fundamental para entender os padrões de distribuição em informações geoespaciais. O padrão de pontos é descrito pela densidade espacial estatística dos dados observados. A intensidade $\lambda(s)$, que mede a distribuição espacial, é proporcional à densidade. Para processos com a mesma densidade, o número de eventos varia com a intensidade, sendo maior onde esta é maior (BIVAND; PEBESMA; GÓMEZ-RUBIO, 2008).

O mapa apresentado na Figura 12 ilustra, de forma detalhada, o padrão de agrupamento espacial das perdas não técnicas (PNTs) na cidade de Maputo, no período de 2022 a 2023. A representação por meio de uma escala cromática, que varia do verde ao vermelho, oferece uma visualização clara da densidade dessas perdas. As áreas em verde, indicativas de menor incidência de PNTs, contrastam com as áreas em vermelho, que sinalizam concentrações significativamente elevadas dessas perdas.

A análise espacial do mapa revelou que os distritos de KaMavota e KaMaxakeni constituem os principais focos de PNTs, conforme demonstrado pelas intensas manchas vermelhas, mostrando que essas regiões estão particularmente vulneráveis a furtos e fraudes. Por outro lado, distritos como Kampfumlo, Nlhamankulu apresentam um perfil menos crítico, evidenciado pelos tons mais claros e verdes, o que sugere uma menor concentração de PNTs conforme se pode ver na figura 11 abaixo.

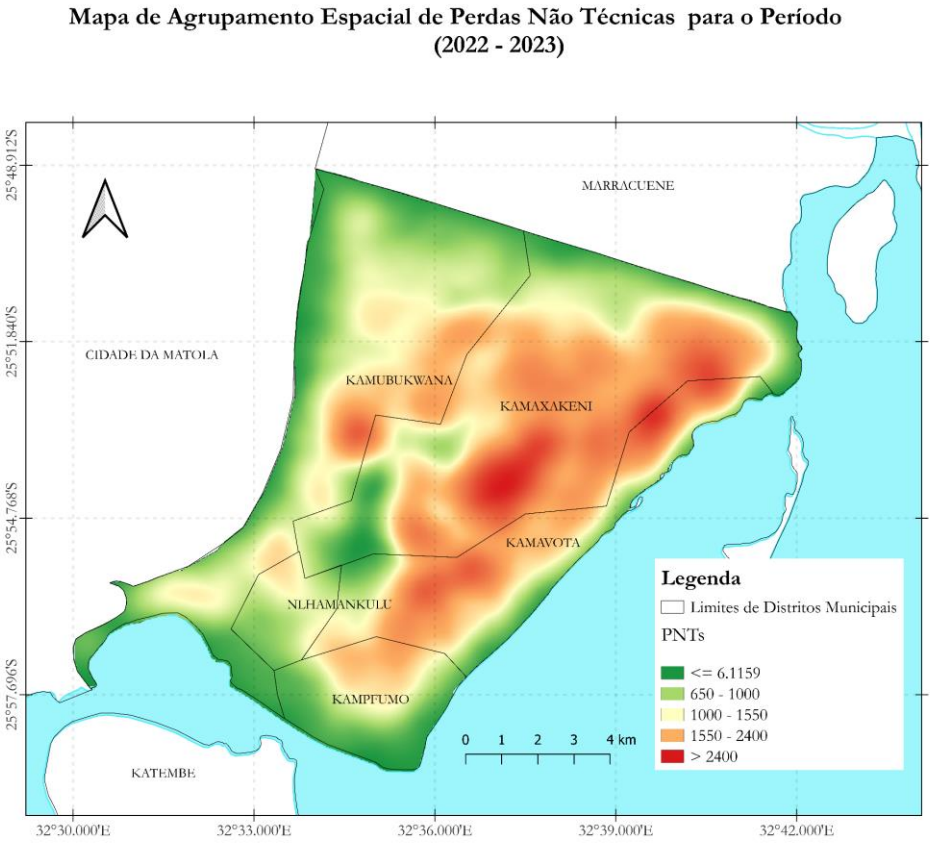


Figura 11. Densidade de Kernel de Perdas não técnicas.

Adicionalmente, com o intuito de medir o grau de similaridade com o padrão de Aleatoriadade Espacial Completa (CST) tarduzida da sigla em inglês (Complete Spatial Randomness), a função G foi implementada conforme a ilustra a (Figura 12). É apresentada nessa figura a função G estimada, também chamada de função empírica $G(r)$, identificada na figura como $Gobs(r)$.

Onde:

Gkm(r): A linha sólida representa a função acumulada G(r) considerando uma estimativa baseada no método de Kernel (km).

Gbord(r): A linha tracejada pode indicar uma correção de borda, onde é levada em consideração a proximidade de eventos às bordas da região estudada.

Ghan(r): A linha pontilhada pode representar o estimador de Hanisch, utilizado para correções em dados espacialmente correlacionados.

Gpois(r) A linha tracejada e pontilhada pode estar mostrando a função G teórica sob um padrão de ponto aleatório (Poisson), indicando como os eventos seriam distribuídos se ocorressem aleatoriamente.

O gráfico patente na figura 12, mostra que, à medida que a distância r aumenta, a função G acumulada também aumenta, o que é esperado, pois eventos mais distantes começam a ser incluídos na contagem. Segundo (Faria, 2016), o fato de a função Gpois(r) (padrão aleatório) estar abaixo das outras funções indica que os eventos observados (PNTs) não seguem um padrão completamente aleatório. Esse comportamento torna-se ainda mais evidente a partir de distâncias superiores a 300 metros, onde se observa que, para cada casa com PNT, há pelo menos uma casa nas proximidades na mesma situação a similaridade do que foi observado por Faria (2016).

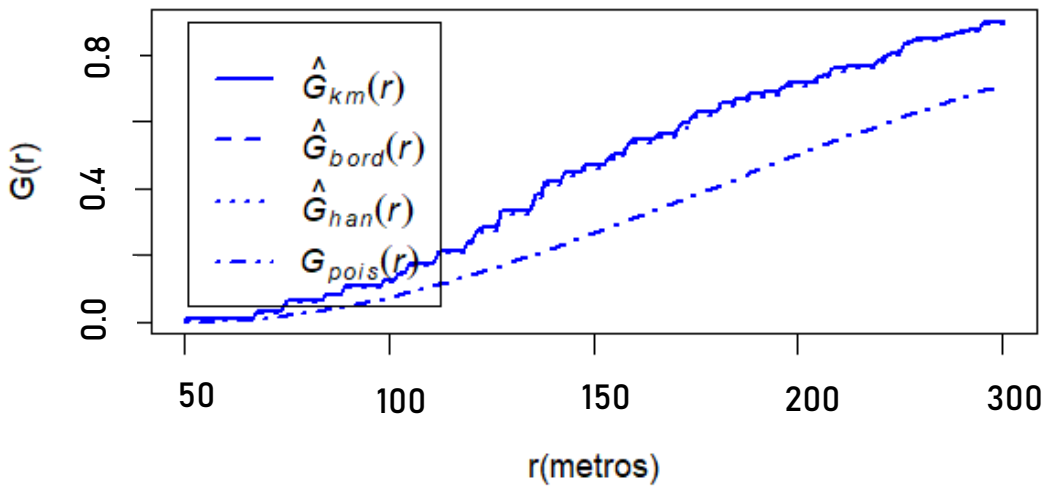


Figura 12. Função G(r) ou função de distribuição cumulativa da variável PNTs

4.1.2. Avaliação da Dependência Espacial das Perdas não Técnicas (PNTs)

Como referido no subcapítulo 4.8, o Índice de Moran, também conhecido como Moran's I, é uma medida estatística amplamente utilizada na análise espacial para avaliar a autocorrelação espacial de um conjunto de dados(Waldemar & Correa, 2007). Essa ferramenta é fundamental para entender como as variáveis estão distribuídas geograficamente e se há padrões significativos que indicam a presença de agrupamentos ou dispersões(Câmara & Correa, n.d.).

Para a devida análise nas perdas não técnicas (PNTs), foram considerados os dados do período 2022 a 2023, dada a disponibilidade dos dados das inspeções georreferenciadas. No gráfico à esquerda (Figura 13a), o índice de Moran (Moran's I) apresentou um valor de 0,392, indicando uma autocorrelação espacial positiva fraca a moderada. O que significa que regiões com altos índices de PNTs tendem a estar próximas umas das outras, assim como regiões com baixos índices de PNTs. Esse padrão espacial sugere a existência de características geográficas ou sociais comuns em áreas vizinhas, contribuindo para a formação de clusters espaciais de PNTs.

No gráfico à direita (Figura 13b), observa-se a distribuição da estatística de permutações, com um pseudo p-valor de 0,001. Este resultado reforça a significância estatística da autocorrelação espacial identificada, indicando que ela não é aleatória. Com base em 999 permutações, a análise permite rejeitar com segurança a hipótese nula de ausência de autocorrelação espacial, conforme sugerido por Anselin (1995) citado por Faria (2016). A linha verde no gráfico destaca o valor observado de Moran's I, que se distancia significativamente da distribuição esperada sob a hipótese nula. Esse resultado sublinha a importância da autocorrelação espacial positiva, evidenciando padrões espaciais relevantes nas PNTs.

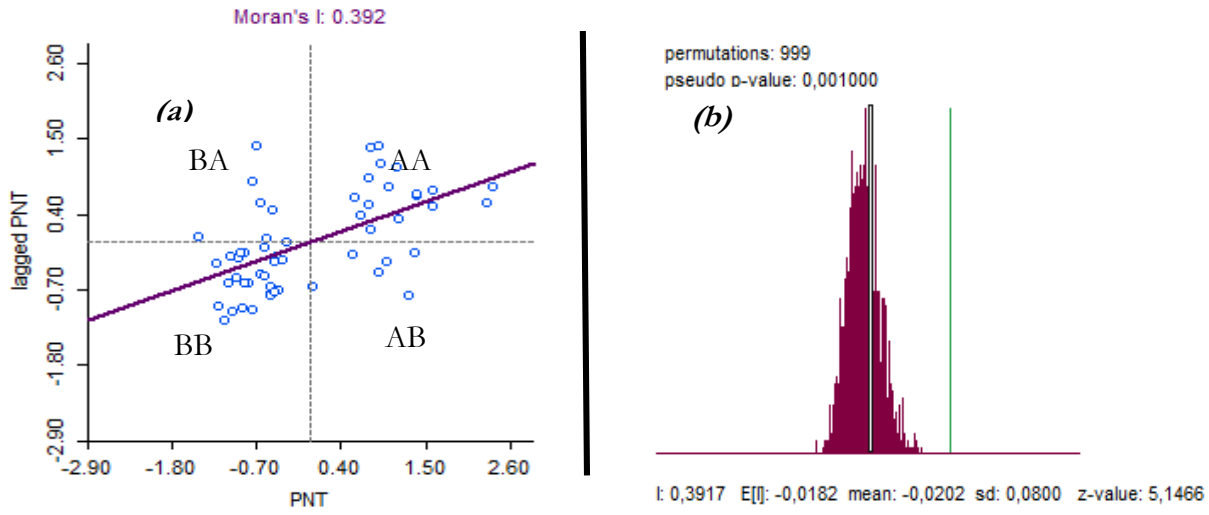


Figura 13. Índice de Moran's I; ***(a)*** Gráfico de Dispersão; ***(b)*** Teste de significância

Os resultados obtidos indicam a existência de uma tendência geográfica significativa no contexto das Perdas Não Técnicas (PNTs), com certas áreas registrando níveis consistentemente mais altos ou mais baixos. Essa distribuição pode ser atribuída a fatores socioeconômicos e técnicos que influenciam diretamente a ocorrência das perdas.

Conforme destacado por Fotheringham *et al.,* (2002), a escolha adequada da matriz de pesos espaciais desempenha um papel crucial na captura precisa das interações espaciais. Neste estudo, a utilização de uma matriz de pesos baseada na distância semântica permitiu explicar de maneira consistente as relações

de proximidade entre as áreas analisadas, fornecendo uma base robusta para a identificação da influência espacial nas PNTs.

4.2. Aplicação dos Modelos de Regressão Espacial Local

De forma a analisar a influência dos fatores técnicos e socioeconômicos no padrão de distribuição das perdas não técnicas (PNTs) na cidade de Maputo, foram consideradas as variáveis explicativas, nomeadamente: renda média mensal, índice de criminalidade, densidade populacional, taxa de desemprego, e o tipo de contador existente nas instalações de clientes da empresa Electricidade de Moçambique (integrado e split). Como variável dependente, foram analisadas as perdas não técnicas, compostas por fraudes e furtos. Foram usados os modelos de regressão espacial local nomeadamente: o modelo de regressão geograficamente ponderado (GWR) e o modelo de regressão geograficamente ponderado Multi escalar (MGWR).

4.2.1. Avaliação do Desempenho dos Modelos de Regressão Espacial Local

Os resultados apresentados na Tabela 8 indicam que o modelo de regressão geograficamente ponderado multiescalar (MGWR) apresentou o melhor desempenho entre os modelos avaliados. Com um índice de R^2 de 0,97 contra 0,96 do modelo Geograficamente Ponderado (GWR), este modelo demonstrou uma elevada capacidade de explicar a variabilidade observada nos dados, superando o modelo GWR. De acordo com Fotheringham *et al.,* (2017), o MGWR é especialmente eficaz na modelagem de relações que variam em diferentes escalas espaciais, permitindo capturar padrões espaciais complexos de forma mais precisa.

Adicionalmente, o modelo MGWR apresentou o menor valor de AIC (0,173), evidenciando sua maior parcimônia. Assim, o desempenho superior do MGWR reforça sua adequação para a análise das dinâmicas espaciais das PNTs em relação ao modelo GWR.

Tabela 8. Análise de desepenho dos modelos.

Modelo de Regressão	R^2	AIC	Log-likelihood	Moran's I	
				I	p-value
GWR	0,96	5,192	10,389	-0,079	0,038
MGWR	0,97	0,173	14,053	-0,113	0,228

Adicionalmente, a log-verossimilhança máxima alcançada pelo modelo MGWR, de 14,053, reafirma sua eficácia no ajuste dos dados. Este resultado está em linha com as observações de LeSage e Pace (2009), que destacam que uma log-verossimilhança mais elevada está associada a explicações mais satisfatórias para os dados analisados.

A análise dos resíduos foi realizada por meio de dois gráficos principais: a Figura 14 apresenta o plot dos resíduos, utilizado para avaliar a presença de autocorrelação espacial por meio do Índice de Moran, contribuindo para validar a adequação do modelo ajustado; a Figura 15 exibe o histograma dos resíduos, utilizado para verificar sua normalidade, sendo desejável que a distribuição apresente forma aproximadamente simétrica e próxima à curva normal.

No modelo GWR, os resíduos apresentaram um índice de Moran de -0,079, conforme ilustrado na (Figura 14a), com p-valor de 0,023 e z-score de -0.813, de acordo com o **Anexo 2**. Esses resultados indicam que os resíduos seguem uma distribuição essencialmente aleatória, não havendo indícios significativos de autocorrelação espacial. Isso reforça a adequação do modelo GWR, evidenciando que a dependência espacial não comprometeu a análise.

Para o modelo MGWR, como mostrado na (Figura 14b), o índice de Moran foi de -0,113, com p-valor de 0,089 e z-score de -1,269, conforme o **Anexo 3**. Embora esse índice seja ligeiramente mais alto do que no GWR, os resíduos ainda apresentam um padrão próximo ao acaso, sem relevância estatística. O leve aumento no z-score pode indicar uma tendência sutil para padrões espaciais nos resíduos, mas sem impacto significativo, o que ressalta a robustez do modelo MGWR.

Ao comparar os dois métodos, observa-se que o MGWR apresenta uma pequena elevação na autocorrelação dos resíduos em relação ao GWR. No entanto, ambos os modelos permanecem dentro de um intervalo que indica aleatoriedade nos resíduos, validando sua capacidade de representar as variáveis analisadas com confiabilidade.

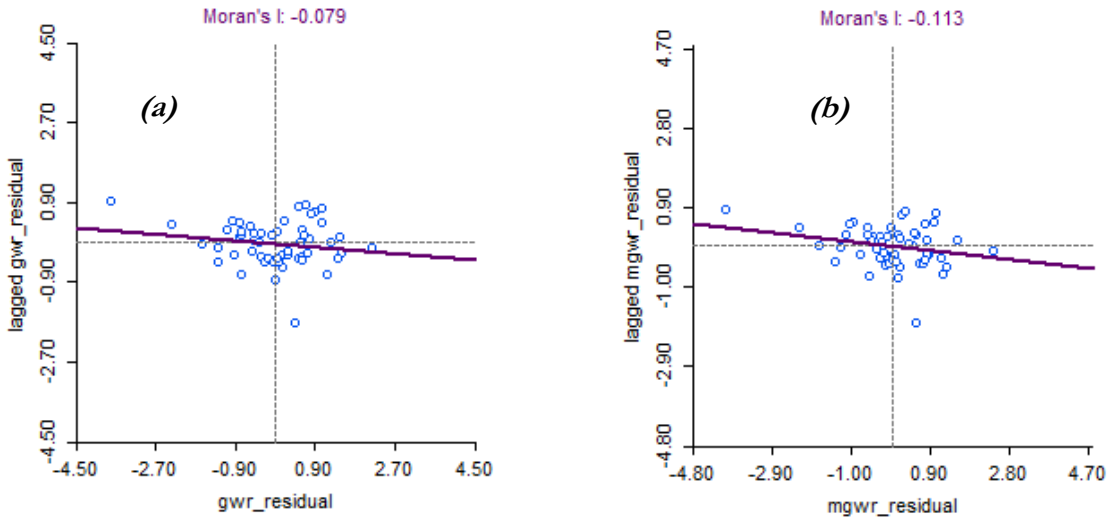


Figura 14. Análise gráfica dos modelos; ***(a)*** índice de Moran's I para o modelo GWR; ***(b)*** índice de Moran's I para o modelo MGWR

A distribuição dos resíduos nos modelos GWR e MGWR, conforme apresentado nas Figuras 15a e 15b, é aproximadamente simétrica, sugerindo uma distribuição normal, o que é desejável para a validade dos modelos de regressão (Osborne, 2000; Anselin, 2005). A amplitude dos resíduos é menor na Figura 15a (GWR), variando entre -0,8 e 0,6, em comparação com a Figura 15b (MGWR), que exibe uma amplitude ligeiramente maior, de -0,8 a 0,7. Isso sugere que o MGWR pode capturar mais variações espaciais locais (Fotheringham *et al.,* 2002). Em termos de forma, a curva da Figura 15b (MGWR) é mais homogênea, enquanto a da Figura 15a (GWR) apresenta maior concentração perto do valor central, refletindo a maior flexibilidade do MGWR (Fotheringham *et al.,* 2002; Brunsdon *et al.,* 1998). A dispersão mais uniforme dos resíduos na Figura 15b (MGWR) sugere um ajuste mais aprimorado, corroborando a superioridade do MGWR para dados com variações espaciais complexas (Yang *et al.,* 2016).

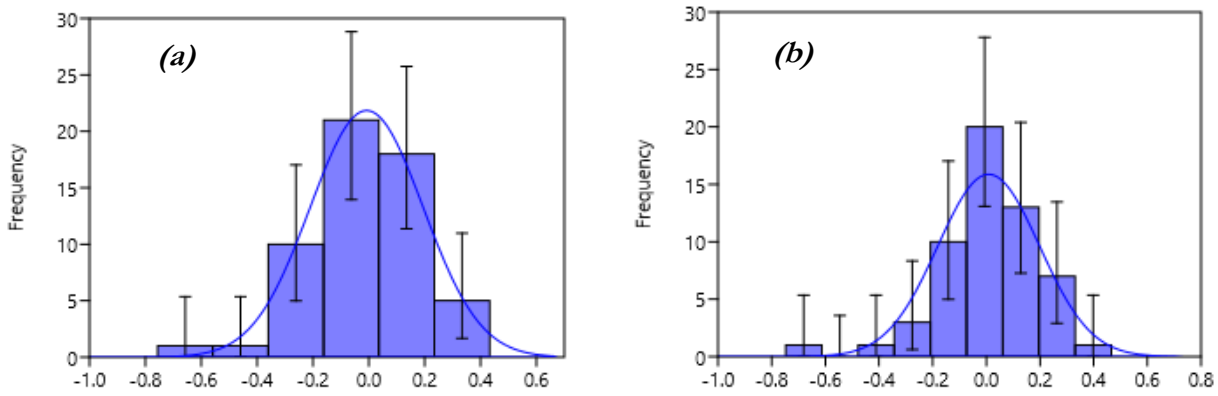


Figura 15. Análise gráfica dos modelos; ***(a)*** Histograma dos resíduos para o modelo GWR; ***(b)*** Histograma dos resíduos para o modelo MGWR

4.2.2. Avaliação Espacial do Desempenho dos Modelos de Regressão Espacial Local

A avaliação espacial dos modelos foi realizada por meio do mapeamento dos resíduos (Figura 16), considerado uma das abordagens mais relevantes para identificar padrões espaciais em resultados de modelos de regressão, conforme destacado por Anselin (2005). Essa análise permite identificar regiões subestimadas ou superestimadas, que podem evidenciar padrões espaciais importantes para interpretação.

No modelo GWR, os resíduos apresentaram uma variação de -0,164 a 0,432 (Figura 16a), com destaque para áreas críticas localizadas principalmente nas regiões centrais e costeiras. Em contraste, o modelo MGWR exibiu resíduos variando entre -0,094 e 0,466 (Figura 16b), com menor amplitude e uma distribuição mais homogênea. Essa redução na magnitude dos resíduos e a uniformidade espacial destacam a capacidade superior do MGWR em minimizar discrepâncias locais.

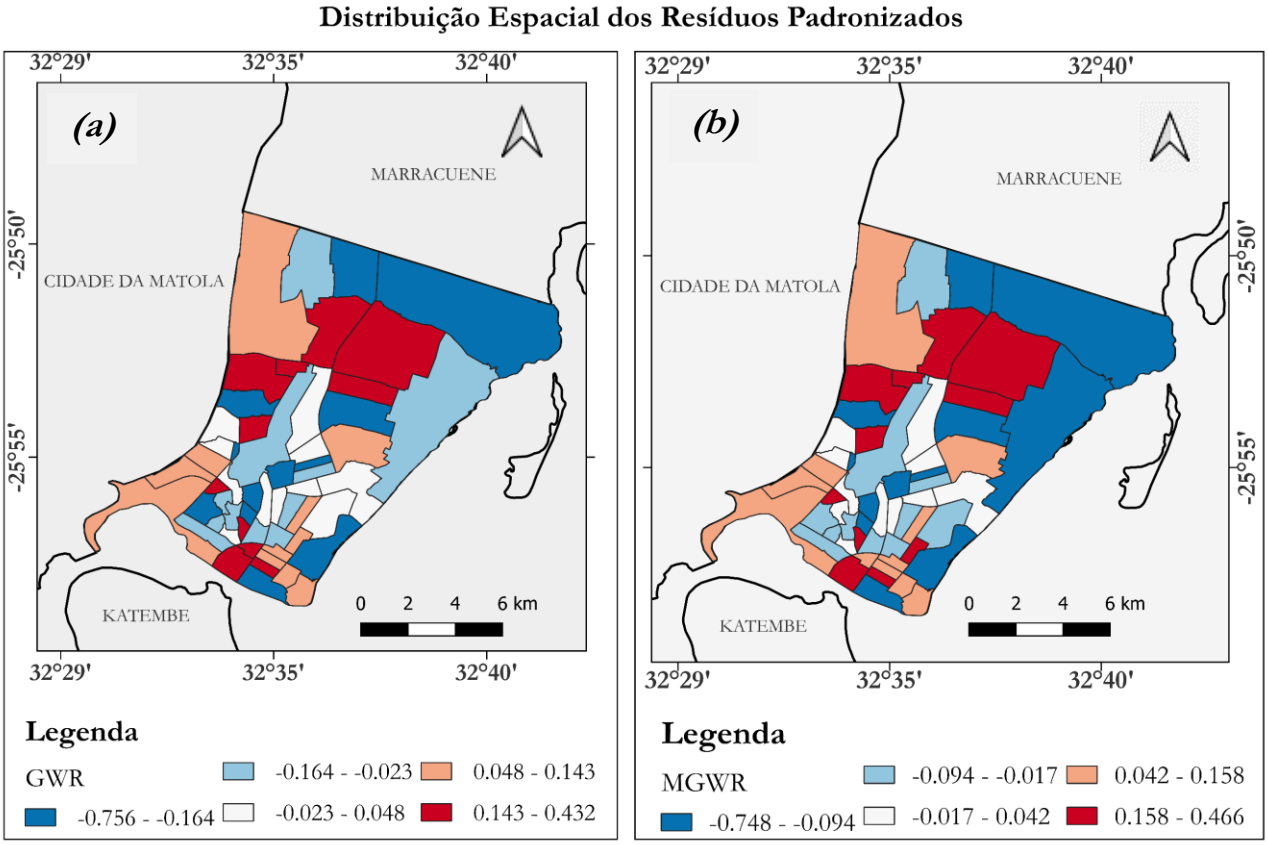


Figura 16. Avaliação espacial dos modelos; ***(a)*** Resíduos para o modelo GWR; ***(b)*** Resíduos para o modelo MGWR;

Com base nos resultados obtidos, o modelo MGWR (Regressão Geograficamente Ponderada Multiescalar) demonstrou desempenho superior ao modelo GWR (Regressão Geograficamente Ponderada). O MGWR alcançou um R^2 de 0,97, ligeiramente superior ao R^2 de 0,96 do GWR, e obteve um AIC significativamente mais baixo (0,173 contra 5,192), indicando um melhor equilíbrio entre qualidade de ajuste e complexidade. Ambos os modelos apresentaram resíduos sem autocorrelação espacial significativa, conforme os testes de Moran's I, evidenciando que a estrutura espacial dos dados foi adequadamente capturada. Embora o GWR tenha apresentado resíduos ligeiramente menos correlacionados espacialmente, as diferenças não foram estatisticamente significativas. Portanto, considerando o maior R^2, o menor AIC e a capacidade do MGWR de lidar com variações em múltiplas escalas espaciais, esse modelo foi selecionado como o mais adequado para representar os dados analisados.

4.3. Analise de Influência dos Factores Técnicos Socioeconómicos no Padrão de Distribuição Espacial das PNTs

A análise dos coeficientes estimados para as variáveis independentes foi realizada utilizando o modelo MGWR, com os resultados categorizados em cinco classes de intensidade nomedamente: Muito Baixo, Baixo, Moderado, Alto e Muito Alto. Esta classificação permite uma visualização clara da distribuição espacial das variáveis, facilitando a identificação de padrões de influência de cada factor nas perdas não técnicas (PNTs) no Município da Cidade de Maputo.

Para (Estevo Felix *et al.,* 2021), uma das grandes vantagens do MGWR é a sua habilidade de captar padrões espaciais locais, possibilitando uma análise mais exata da distribuição das PNTs. Embora a regressão GWR convencional possam detectar tendências locais, o MGWR proporciona uma compreensão mais aprofundada, considerando que os efeitos de factores como densidade populacional, tipo de contador ou características socioeconômicas podem divergir consideravelmente de um lugar para outro associado a dimensão da área analisada. Por exemplo, a influência do tipo de contador no nível de PNT pode ser mais intensa em localidades com alta densidade populacional, sendo um bairro em termos territoriais pequeno e apresentar uma relação distinta em áreas com menor densidade.

Nas figuras 17 e 19, as estimativas dos coeficientes da regressão espacial MGWR, para variáveis nomeadamente: Renda média mensal, Índice de criminalidade, Densidade populacional, taxa de desemprego, Contadores Integrados e Split, foram classificadas em cinco categorias: Muito Baixo, Baixo, Moderado, Alto e Muito Alto.

Os coeficientes da renda média mostram uma associação mais forte nas regiões do norte do MCM (Figura 17a), com coeficientes classificados como "Alto" e "Muito Alto", sugerindo que áreas com maior poder aquisitivo podem estar mais expostas a PNTs. Isso pode ser explicado pelo maior consumo de energia e a possibilidade de maior evasão no controle devido ao aumento da demanda (Jain et al., 2022). Por outro lado, os coeficientes negativos ou classificados como "Baixo" nas áreas centrais e sul (Figura 18b) indicam uma relação inversa, onde as regiões com menor renda podem ter menor incidência de PNTs, possivelmente devido à restrição de consumo e à limitação de recursos para infrações.

O índice de criminalidade, por sua vez, apresenta coeficientes mais elevados nas regiões do centro-sul, com valores classificados como "Alto" e "Muito Alto" na Figura 17b, corroborando a ideia de que a criminalidade tem uma associação directa com as PNTs, uma vez que práticas ilegais, como o furto de energia, tendem a ser mais comuns em áreas com maior vulnerabilidade social (Zhang *et al.,* 2018). Em

contraste, as regiões do Norte, com índices de criminalidade mais baixos, apresentam coeficientes classificados como "Baixo" na Figura 18b, demonstrando menos impacto da criminalidade nas PNTs.

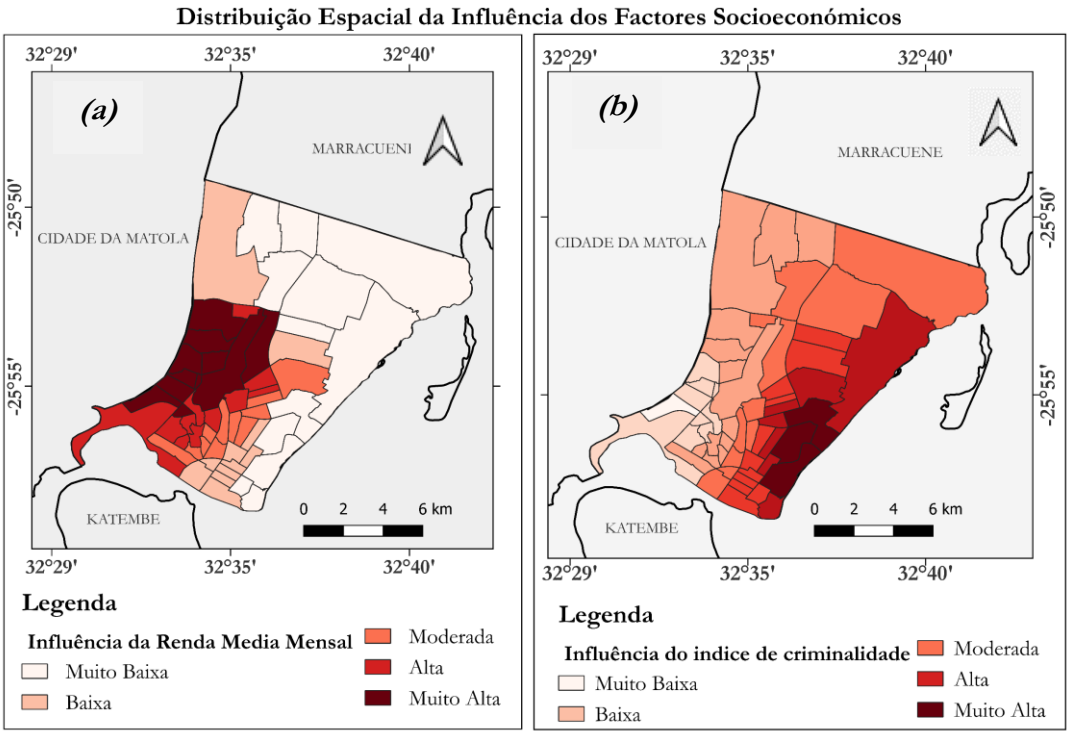


Figura 17. Coeficientes Estimados (β); ***(a)*** Renda Media Mensal; ***(b)*** Índice de Criminalidade;

Em relação à densidade populacional, na figura 18a, observa-se que nas áreas de alta densidade populacional no sul, os coeficientes negativos indicam que ações de fiscalização mais intensas e uma infraestrutura de distribuição mais robusta contribuem para o controle das PNTs, conforme aponta Almeida et al. (2021). No norte, as áreas com menor densidade populacional apresentam coeficientes mais elevados, sugerindo que a dispersão populacional dificulta a fiscalização e facilita a ocorrência de PNTs.

Por fim, a taxa de desemprego, com coeficientes negativos no sul (Figura 18b) e positivos no norte, reflete a influência do contexto social na prática de furtos de energia. Nas regiões periféricas com altos índices de desemprego, como o norte, a falta de recursos pode levar a um aumento na ocorrência de PNTs, como observado por Silva et al. (2020), enquanto políticas públicas de apoio no sul podem mitigar esses efeitos.

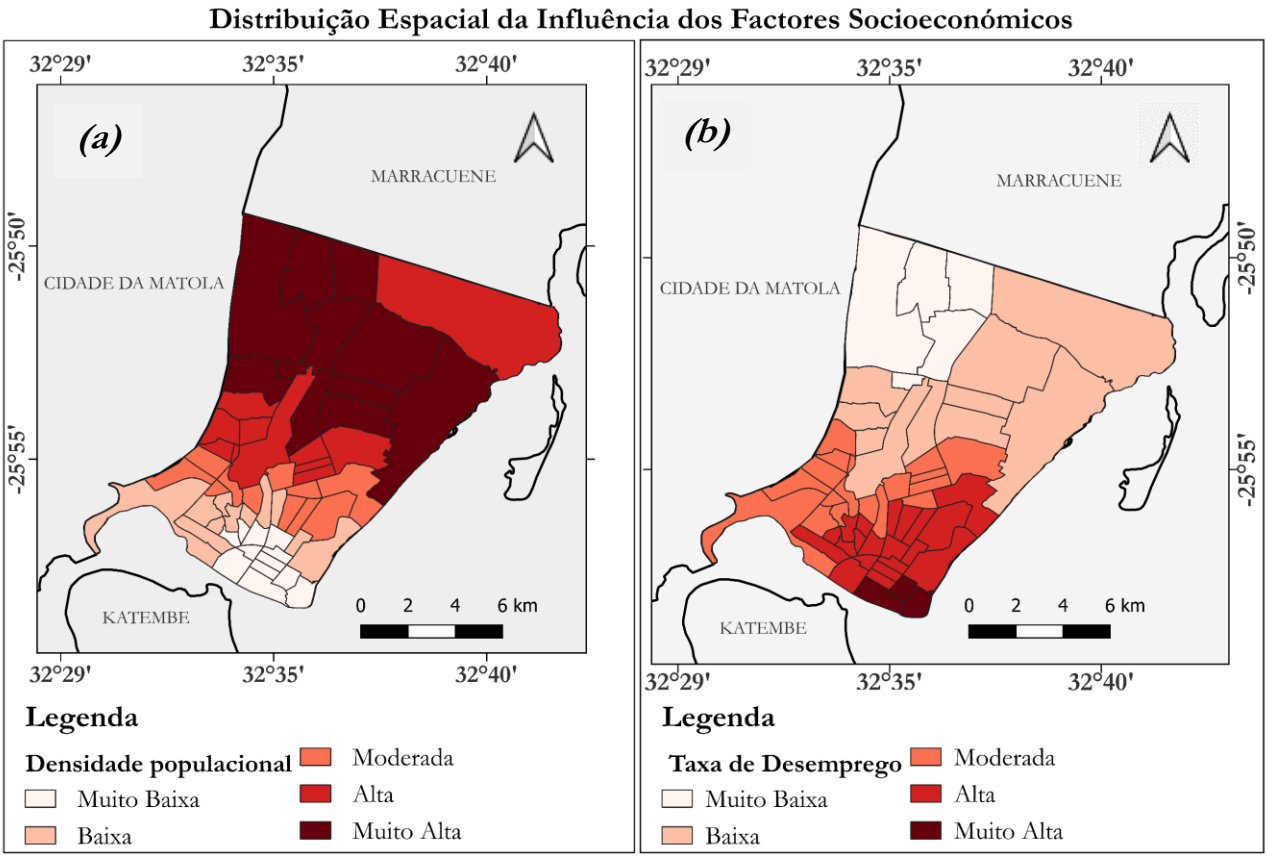


Figura 18. Coeficientes Estimados (β); ***(a)*** Densidade Populacional; ***(b)*** Taxa de desemprego;

Os contadores split, conforme a (figura 19a), apresentam coeficientes positivos mais significativos no sul e no centro-leste, indicando que a vulnerabilidade desse tipo de contador está concentrada em regiões urbanas com alta densidade populacional. Esse padrão está de acordo com o estudo de Chakravarthy *et al.* (2021), que aponta deficiências no planeamento e na implementação de contadores divididos em áreas urbanas. A localização comum desses contadores dentro das residências em áreas densamente povoadas os torna mais suscetíveis a interferências ilegais, uma vez que é possível manipular e fraudar o equipamento sem a necessidade de intervenções externas. No norte do MCM, os índices mais baixos indicam uma menor quantidade desses dispositivos, refletindo o impacto reduzido em áreas menos povoadas e mais afastadas.

A análise espacial da influência dos factores técnicos sobre as perdas não técnicas, no que se refere especificamente aos contadores integrados, revela uma distribuição marcadamente desigual no território do Município da Cidade de Maputo. Conforme ilustrado no mapa da (figura 19b) a direita, observa-se que a influência é muito alta em grande parte das zonas localizadas no norte e nordeste do município, incluindo áreas adjacentes ao distrito de Marracuene. Essas regiões apresentam-se com

coloração mais escura, indicando que os fatores técnicos exercem um peso considerável na ocorrência de perdas não técnicas nessas áreas.

Por outro lado, nas regiões sul e sudoeste, a influência dos contadores integrados tende a ser moderada ou baixa, sugerindo que, nesses locais, os contadores integrados têm menor relação com a ocorrência de perdas não técnicas. Essa variação espacial pode estar associada a diversos factores, como a infraestrutura eléctrica instalada, níveis de manutenção, frequência de inspeções técnicas, além das condições socioeconômicas dos bairros, que impactam diretamente na forma como os equipamentos são utilizados e mantidos(Faria, 2016).

A elevada influência observada nas zonas mais periféricas pode indicar desafios importantes, como a baixa qualidade da rede eléctrica, dificuldades de acesso por parte das equipas técnicas e limitações na fiscalização, factores que favorecem a manipulação ou mau funcionamento dos contadores integrados. Além disso, é possível que, nessas áreas, exista um número maior de instalações antigas ou com menor controle automatizado, o que eleva a vulnerabilidade técnica desses equipamentos.

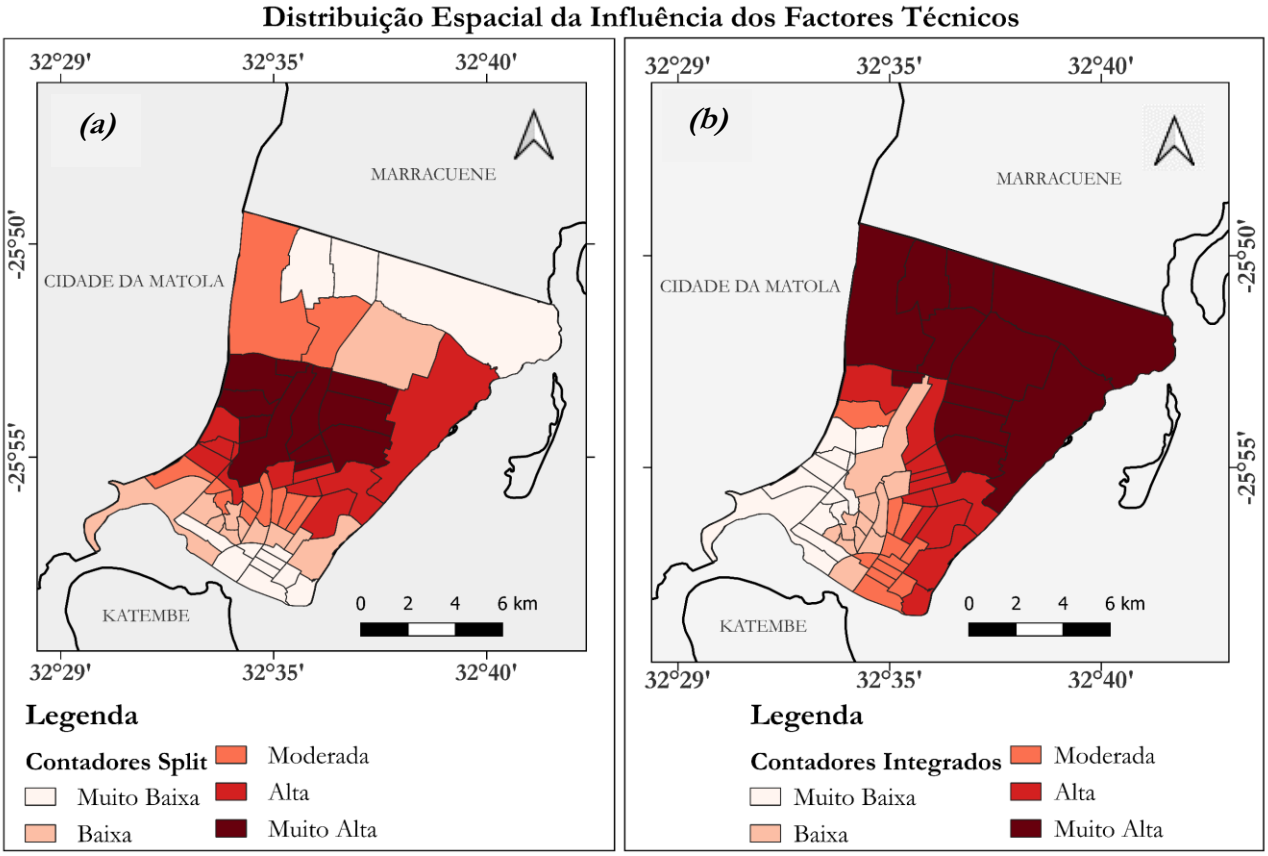


Figura 19. Coeficientes Estimados (β); *(a)* Contadores Integrados; *(b)* Contadores Split

Com base na tabela 9, os Contadores Integrados são a variável mais impactante, com uma influência de 25% e uma amplitude de variação de 0,53, o que sugere uma associação forte e significativa com as PNTs.

Tabela 9. Percentagem de Influência dos factores técnicos e socioeconómicos no padrao de distribuição espacial das PNTs**.**

Variável	Intervalo de Variação	Amplitude de Variação	Percentagem de Influência (%)
Contadores Integrados	0,12 a 0,65	0,53	25%
Densidade Populacional	-0,34 a 0,41	0,75	22%
Contador Split	-0,08 a 0,52	0,60	20%
Indice de Criminalidade	-0,22 a 0,48	0,70	18%
Renda Media	-0,18 a 0,34	0,52	8%
Taxa de Desemprego	-0,19 a 0,29	0,48	7%

Esse resultado reflecte a vulnerabilidade desse tipo de contador, especialmente em áreas urbanas densamente povoadas, onde o controle e monitoramento são mais difíceis.

A Densidade Populacional vem em segundo lugar, com 22% de influência e uma amplitude de variação de 0,75, indicando que a fiscalização intensiva em regiões de alta densidade tem um impacto direto na redução das PNTs. A variável Contadores Split segue com 20% de influência e uma amplitude de 0,60, evidenciando que a localização e o tipo de contador têm um papel crucial, pois esses dispositivos são suscetíveis a fraudes, especialmente quando instalados dentro das casas dos consumidores.

O Índice de Criminalidade com (18%) tem um impacto moderado, reflectindo uma associação mais forte com as PNTs em áreas de maior vulnerabilidade social. A criminalidade, associada à falta de controle, parece ser mais prevalente em regiões com alta incidência de práticas ilegais, como o furto de energia. Já as variáveis Renda Média (8%) e Taxa de Desemprego (7%) apresentam influências mais limitadas. Embora desempenhem um papel secundário, esses factores econômicos ainda podem impactar a ocorrência de PNTs, especialmente nas regiões mais periféricas, onde as condições socioeconômicas favorecem a evasão do controle energético.

5

Conclusões e Recomendações

5.1. Conclusões

Neste estudo, foi utilizado o modelo de regressão espacial MGWR (Multiscale Geographically Weighted Regression) para integrar o espaço geográfico na análise das perdas não técnicas (PNTs), permitindo examinar a distribuição das variáveis estudadas. O uso deste modelo possibilitou a identificação de padrões espaciais complexos e das interações locais entre variáveis técnicas e socioeconômicas, representadas em mapas temáticos detalhados. Após a comparação com o modelo GWR tradicional, o MGWR apresentou um melhor ajuste, refletindo sua capacidade superior de capturar as variações espaciais em diferentes escalas, o que melhorou a precisão da análise.

A avaliação indicou que o Sul e o Centro-este do Município da Cidade de Maputo são regiões particularmente vulneráveis, devido à alta concentração populacional e ao uso predominante de contadores Integrados e split instalados nas residências, o que as torna mais propensas a perdas. Observou-se que, nas áreas densamente povoadas, os factores técnicos predominam e têm maior influência sobre as PNTs. Esses achados ressaltam a importância de se adoptar medidas específicas para cada área, considerando tanto as características técnicas quanto as socioeconômicas, a fim de reduzir as PNTs.

Os contadores integrados mostraram a maior taxa de influência de 25% nas PNTs, evidenciando que essa variável desempenha um papel fundamental nas perdas. Regiões onde há predominância desse tipo de contador costumam apresentar maior frequência de PNTs. Seguinda da densidade populacional exerceu uma significativa influência de 22% nas PNTs. Em regiões com maior densidade populacional, a interação entre elementos técnicos e socioeconômicos resulta em uma probabilidade aumentada de perdas não técnicas.

Contador split, apesar de ter uma influência um pouco inferior à dos contadores integrados, com uma percentagem de 20%, aindaassim, exerceu um impacto considerável nas PNTs. Esse tipo de contador também está associado a um incremento nas perdas em regiões onde sua prevalência é maior. Por outro lado, A taxa de criminalidade revelou uma contribuição significativa, com a conexão entre a criminalidade e as PNTs destacando a relação entre aspectos sociais e o roubo de energia.

5.2 Recomendações

Com base nos resultados obtidos, recomenda-se que a empresa de distribuição de energia invista na modernização de tipo de contadores, especialmente no que diz respeito à substituição gradual dos contadores integrados por dispositivos mais avançados como os contadores inteligentes, priorizando as áreas identificados como sendo as que mais são afectados por perdas por influência dos contadores integrados. Apesar de serem uma alternativa, os contadores split ainda apresentam fragilidades, o que torna necessário um fortalecimento de suas tecnologias e monitoramento, aliando manutenções regulares e estratégias educativas voltadas à conscientização dos consumidores. Em áreas de alta densidade populacional, o uso de Sistemas de Informação Geográfica (SIG) deve ser priorizado para mapear e identificar padrões espaciais das perdas não técnicas, facilitando a fiscalização e a alocação mais eficiente de recursos. Além disso, a colaboração com órgãos de segurança pública torna-se fundamental em regiões com maiores índices de criminalidade, onde o furto de energia é mais frequente.

Para os estudos futuros, recomenda a aplicação dos modelos de regressão local com a inclusão de variáveis que não foram integrados nesse estudo, afim de analisar a sua influência nas PNTs. Assim como integrar abordagens temporais, a partir da análise de séries históricas, para possibilitar o acompanhamento da evolução das perdas não técnicas ao longo do tempo, oferecendo subsídios para o planeamento de estratégias mais eficazes e sustentáveis.

Referências Bibliográficas

Aguero, J.R. (2012). "Improving the efficiency of power distribution systems through technical and non-technical losses reduction". In: *Proceedings of the IEEE PEStransmission and distribution conference and exposition.*

Assad, E. D. & Sano, E. E. (1998). Sistema de Informações Geográficas: aplicações na agricultura. 2ªed. Brasília: Embrapa-SPI.

ALMEIDA, M., SILVA, P., & GOMES, L. (2021). *Dinâmicas Socioeconômicas e Impacto nas Perdas de Energia*. Revista Brasileira de Geografia.

Anselin, L. (2013). *Spatial Econometrics: Methods and Models*. Springer Science & Business Media.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Springer Netherlands.

Bailey, T. C., & Gatrell, A. C. (1995). *Interactive Spatial Data Analysis*. Longman Scientific & Technical.

Batty, M. (2005). *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals*. MIT Press.

Botelho, G. L. (2022*). Perdas Não Técnicas de Energia Eléctrica: Um Desafio Comportamental*. Rio de Janeiro, RJ – Brasil

Câmara, G., & Correa, V. (n.d.). *Análise espacial de áreas*.

Cabral, J.E., Pinto, J.O., Gontijo, E.M., Filho, J.R. (2004). Fraud detection in electrical energy consumers using rough sets. *IEEE International Conference on Systems, Man and Cybernetics*.

Chiaverini, G. (2018). *Perdas de energia elétrica no Brasil: desafios e perspectivas*. Rio de Janeiro: Editora FGV.

CHAKRAVARTHY, R., JAIN, M., & GUPTA, A. (2021). *Spatial Patterns of Energy Loss in Developing Nations*. Energy Policy Journal.

Cardoso, M. J. S. (2022). ESTUDO E ANÁLISE DAS PERDAS NÃO TÉCNICAS DE ENERGIA ELÉTRICA NO TERRITÓRIO PARAIBANO. João Pessoa – Brasil

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. Wiley.

CCM. (2024). Caracterização Geográfica e História da Cidade de Maputo. Disponível em Sobre Nós - CMM (cmmaputo.gov.mz) acesso em 20 de Maio de 2024.

Câmara, G., & Correa, V. (n.d.). *Análise espacial de áreas. 0*.

Eduardo, L., Carvalho, X., Silva, H. N., Felipe, C., Loureiro, G., & Meneses, H. B. (2006). *Regressão linear geograficamente ponderada em ambiente SIG*. 18–26.

Estevo Felix, G., Costa da Silva, N., & Teles Faria, L. (2021). *Análise Espacial Exploratória das Perdas Não Técnicas em Sistemas de Distribuição de Energia Elétrica*. https://doi.org/10.48011/sbse.v1i1.2317

Eléctricidade de Moçambique. (2022). Relatório de Contas.

Eléctricidade de Moçambique. (2021). Manual de Procedimentos Técnico - Comerciais. Ordem de Serviço N° 028/CA/2021 de 05 de outubro.

Energia, P. D. E., & Distribui, N. A. (2021). Perdas de Energia Elétrica na Distribuição. 2021, 1–17.

FARIA, L. T. de. (2012). Sistema Inteligente Híbrido Intercomunicativo para Detecção de Perdas Comerciais. (Dissertação de Mestrado em Engenharia Eléctrica). São Paulo: Universidade Estadual Paulista, 2012. 112 p.

Faria, L. T. D. (2016). *Estimação EspaçoTemporal das Perdas Não Técnicas Estimação Espaço- Temporal das Perdas Não Técnicas*. Pós-Graduação em Engenharia Elétrica. Ilha Solteira – Brasil

Ferreira, F. J. M. (2011). Perdas de energia elétrica em sistemas de distribuição. 2. ed. Rio de Janeiro: Editora Blucher.

Ferreira, J. T. V. (2013). Redes de Distribuição de Energia Eléctrica de Media e Baixa Tensão. (Dissertação de Mestrado em Engenharia Eletrotécnica). Coimbra: Instituto Superior de Engenharia de Coimbra.

Fernando G. K. Guarda, Alzenira Da Rosa Abaide, Nelson Knak Neto, Laura L. C. dos Santos, Marcelo Bruno Capeletti, Bruno Knevitz Hammerschmitt, & Lucio Rene Prade. (2018). Avaliação técnica de métodos para detecção de perdas não-técnicas em sistemas de distribuição de energia elétrica. *Proceedings IX Simpósio Brasileiro de Sistemas Elétricos*, *January*. https://doi.org/10.20906/sbse.v2i1.3030

Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics*. Sage.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2000). *Quantitative Geography: Perspectives on Spatial Data Analysis*. Sage Publications.

FOTHERINGHAM, A. S., CHARLTON, M. E., & BRUNSDON, C. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley.

Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis*, 24(3), 189-206.

Gujarati, D. N., & Porter, D. C. (2009). Basic Econometrics. McGraw-Hill.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2018). Multivariate Data Analysis. Cengage Learning.

GÓMEZ, L., MARTÍNEZ, J., & RODRÍGUEZ, A. (2019). *Technical Vulnerabilities in Integrated Meters*. Energy Systems Research.

Gomes, A. R., Santos, L. P., & Oliveira, M. J. (2021). Policy choices and their impact on technical efficiency: A spatial analysis of energy systems**.** *Energy Systems and Policy Journal*

Kale, S., Singh, R., & Gupta, N. (2020). Demographic density and technical challenges in electricity distribution: Insights from rural and urban analysis. *Journal of Energy Polic.*

Kutner, M. H., Nachtsheim, C. J., & Neter, J. (2005). Applied Linear Regression Models. McGraw-Hill/Irwin.

Heppenstall, A. J., Sarkhosh, K. S., & Rezaie, K. (2012). The relationship between electricity theft and social deprivation: A spatial analysis. *Energy Policy*, 40(4), 406-414.

IRENA (2021). Renewable Energy Losses in Power Systems: *A Guide to Mitigation and Measurement. Abu Dhabi*

IEEE (2019). IEEE Guide for Electric Energy Loss Measurement. New York.

Jamil, F., & Ahmad, E. (2010). "The relationship between electricity consumption, electricity prices and GDP in Pakistan." *Energy Policy*

JAIN, M., ZHANG, Y., & SINGH, A. (2022). *Socioeconomic Determinants of Energy Theft*. International Energy Studies.

Lawson, A. B. (2001). *Statistical Methods in Spatial Epidemiology*. Wiley.

Leite, J. B., Roberto, J., & Mantovani, S. (2018). *Detecting and Locating Non-Technical Losses in Modern Distribution Networks. 9*(2), 1023–1032. https://doi.org/10.1109/TSG.2016.2574714

Lise, Á. E. (2011). *Universidade federal de santa Maria centro de tecnologia programa de pós-graduação em engenharia elétrica.*

Miranda, J. I. (2005). Fundamentos de Sistemas de Informações Geográficas. Brasília: Embrapa

Informação Tecnológica.

Medeiros, G. L. M. (2023). *PERDAS NÃO TÉCNICAS NO SISTEMA DE DISTRIBUIÇÃO DE ENERGIA ELÉTRICA BRASILEIRO: PREVENÇÃO E TRATAMENTO REGULATÓRIO*. Rio de Janeiro, RJ – Brasil.

Messinis, G.M., Hatziargyriou, N.D. (2018). Review of non-technical loss detection methods, *Electrical Power Systems Research*, v. 158, p. 250-266.

Mel, R. (2021). *Mestrado em Engenharia da Energia Solar*.

Monteiro, D. A. (2012). Perdas não técnicas de energia elétrica: causas, impactos e medidas de combate. São Paulo: Editora Atlas.

Montgomery, D. C., & Runger, G. C. (2014). Applied Statistics and Probability for Engineers. Wiley.

Nunes, F. A. C. (2000). Perdas não técnicas de energia elétrica: causas, impactos e medidas de combate. Rio de Janeiro: Editora FGV.

Openshaw, S. (1984). *The Modifiable Areal Unit Problem*. Geo Books.

Oliveira, J. T. (2013). *Geoprocessamento aplicado no estudo de identificação das zonas inundáveis no município de Belo Vale – MG*. Universidade Federal de Minas Gerais. Programa de Pós-Graduação do Departamento de Cartografia.

Paulo, S. (2008). *Combate , Prevenção e Otimização das Perdas Comerciais de Energia Elétrica Combate , Prevenção e Otimização das Perdas Comerciais de Energia Elétrica*.

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.

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

Pr, Q. E. L., & Eng, T. (2018). *The Impact of Non-Technical losses: A South African perspective compared to global trends*.

Qadir, J., & Jalal, F. (2017). *Using Geographic Information Systems to Develop a Robust Electricity Utility Network Using Geographic Information Systems to Develop a Robust Electricity Utility Network. October*. https://doi.org/10.23953/cloud.ijarsg.320

Roberto, P., & Bastos, F. D. M. (2011). *Tese Diagnóstico de perdas comerciais de energia elétrica na distribuição* 57 *usando redes Bayesianas Diagnóstico de perdas comerciais de energia elétrica na distribuição usando redes Bayesianas*.

Rodrigues, T.C.V. (2012). Regressão binomial negativa geograficamente ponderada: modelando super dispersão espacial. Dissertação (Mestrado em Estatística) – Instituto de Ciências Exatas, Universidade de Brasília.

Smith, T. B. (2004). Electricity theft: a comparative analysis. *Energy Policy*. v. 32, p. 2067-2076.

Sánchez, L.; Juan M. et al. (2013*). GEOGRAPHICAL INFORMATION SYSTEMS (GIS) AND MULTI-CRITERIA DECISION MAKING (MCDM) METHODS FOR THE EVALUATION OF SOLAR FARMS LOCATIONS: CASE STUDY IN SOUTH-EASTERN SPAIN. Renewable and Sustainable Energy Reviews*, v. 24, p. 544-556.

Savian, F., et al. (2020). Fatores socioeconômicos e suas influências nas perdas não técnicas de energia elétrica.

Santos, A. M., Silva, J. M., & Silva, J. M. (2019). Efeitos da Corrupção sobre as Perdas Não Técnicas em Sistemas de Distribuição de Energia Elétrica. Revista de Energia Elétrica, 12(2), 1-12.

Silva, J. M., Santos, A. M., & Silva, J. M. (2020). Análise Espacial de Perdas Não Técnicas em Sistemas de Distribuição de Energia Elétrica. Revista Brasileira de Energia

Silva, J. (2001). Geoprocessamento para Análise Ambiental. Rio de Janeiro.

Sinha, E. (2018). Multivariate Statistical Analysis in Climate and Environmental Research. Springer. *(Este autor fornece fundamentos sobre transformações de dados, incluindo a transformação logarítmica, especialmente no contexto de estudos ambientais e climáticos.)*

Situação da Eletricidade em Moçambique - energypedia

SILVA, J., ALMEIDA, P., & BARROS, F. (2020). *Desemprego e Perdas Não Técnicas em Regiões Urbanas*. Estudos Econômicos.

Solteira, C. D. I. (2009). *Avaliação de Metodologias de Cálculo de Perdas Técnicas em Sistemas de Distribuição de Energia Elétrica Avaliação de Metodologias de Cálculo de Perdas Técnicas em Sistemas de Distribuição de Energia Elétrica*.

VASCONCELOS, F. M. (2017). Geração, Transmissão e Distribuição de Energia Elétrica. (Desenvolvimento de material didático ou instrucional). Londrina Editora e Distribuidora Educacional.

Vicari, M. B. (2012). *Uso de SIG e Analise Multicritério para Levantamento do Potencial de Implantação de Usinas Eólicas e Solar no Rio Grande do Sul.* Trabalho de conclusão de curso. Engenharia Ambiental – Faculdade de Engenharia e Arquitetura da Universidade de Passo Fundo/RS

Waldemar, J., & Correa, M. (2007). *APLICAÇÃO DO GEOPROCESSAMENTO NA AVALIAÇÃO E ESPACIALIZAÇÃO DAS PERDAS FÍSICAS DE ÁGUA DO SISTEMA DE ABASTECIMENTO PÚBLICO DE RONDONÓPOLIS - MT Application of the Geographic Information Systmems in the evalution and spatial distribuition of the physical l. 19*(2), 51–67.

Waller, L. A., & Gotway, C. A. (2004). Applied Spatial Statistics for Public Health Data. Wiley.

Wong, D. W., & Lee, J. (2005). *Statistical Analysis of Geographic Information with ArcView GIS and ArcGIS*. Wiley.

ZHANG, Y., LI, J., & WANG, X. (2018). *Crime and Energy Theft: A Geospatial Analysis*. Urban Studies.

Zhou, Q., Wang, J., & Fan, Z. (2019). Spatial inequality and technical vulnerability: A case study of energy losses in urban areas**.** *Energy Research and Social Science*.

Anexos

Anexo 1: Lista de Questionário para as Entrevista

1. Quais são os principais tipos de fraudes e ligações clandestinas que vocês identificam com mais frequência durante as inspeções?

2. Em quais regiões ou bairros vocês tendem a encontrar mais casos de furtos e ligações clandestinas?

3. Quais são os métodos mais comuns utilizados pelos consumidores para fraudar o sistema?

4. Que fatores você acredita que levam os consumidores a recorrer a esses comportamentos ilegais?

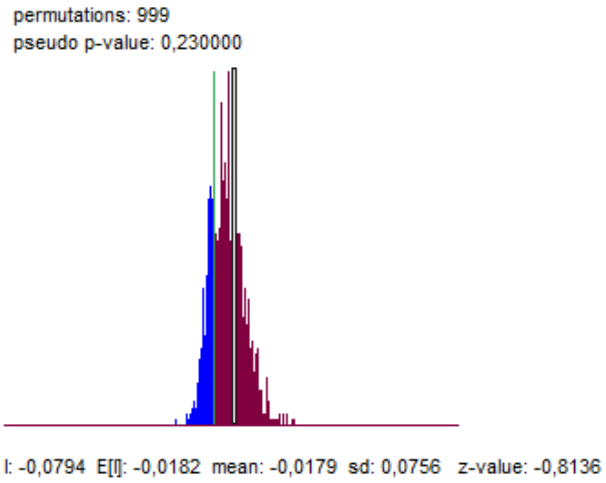
5. Quais são as principais dificuldades que vocês enfrentam durante as inspeções e no combate a esses problemas?

6. Que estratégias e tecnologias a empresa tem adotado para prevenir e detectar fraudes e Furtos?

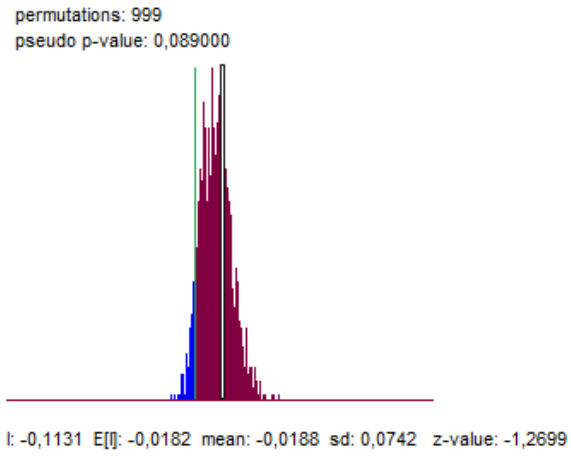
7. Há diferenças significativas no perfil dos consumidores que cometem esses atos ilegais?

8. Que tipo de ações ou políticas a empresa tem implementadas para reduzir a incidência de furtos, fraudes?

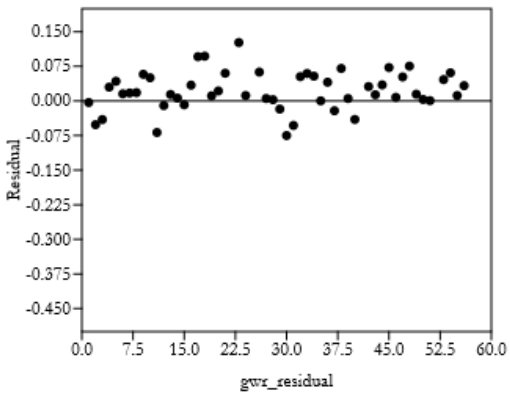
Anexo 2. Teste de significância estatística do modelo GWR



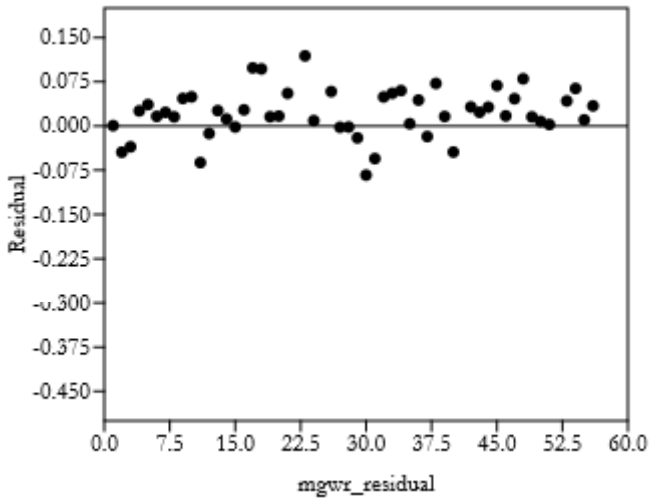
Anexo 3. Teste de significância estatística do modelo MGWR



Anexo 4. Teste de Homocedasticidade do Modelo GWR



Anexo 5. Teste de Homocedasticidade do Modelo MGWR



Anexo 6. Relatório do Modelo GWR

```
Geographically Weighted Regression (GWR) Results
------------------------------------------------------------------------------
Coordinates type:                                             Projected
Spatial kernel:                                     Adaptive bisquare
Criterion for optimal bandwidth:                                 AICc
Bandwidth used:                                                56.000

Diagnostic Information
------------------------------------------------------------------------------
Residual sum of squares:                                        2.262
Effective number of parameters (trace(S)):                     11.985
Degree of freedom (n - trace(S)):                              44.015
Sigma estimate:                                                 0.227
Log-likelihood:                                                10.389
Degree of Dependency (DoD):                                     0.900
AIC:                                                            5.192
AICc:                                                          13.837
BIC:                                                           31.492
R2:                                                             0.960
Adj. R2:                                                        0.948
Adj. alpha (95%):                                               0.033
Adj. critical t value (95%):                                    2.182

Summary Statistics For GWR Parameter Estimates
------------------------------------------------------------------------------
Variable                     Mean        STD        Min     Median        Max
--------------------  ----------  ---------- ---------- ---------- ----------
Intercept                  -0.016      0.024     -0.050     -0.024      0.033
database_R                  0.057      0.013      0.024      0.060      0.075
database_I                  0.071      0.015      0.034      0.074      0.095
database_1                 -0.036      0.010     -0.059     -0.032     -0.025
database_C                  0.682      0.040      0.628      0.668      0.770
database_2                  0.252      0.018      0.206      0.249      0.308
database_T                  0.041      0.008      0.021      0.042      0.055
database_D                  0.015      0.043     -0.071      0.031      0.072
```

Anexo 7. Relatório do Modelo MGWR

```
Multiscale Geographically Weighted Regression (MGWR) Results
------------------------------------------------------------------------------
Coordinates type:                                                    Projected
Spatial kernel:                                              Adaptive bisquare
Criterion for optimal bandwidth:                                         AICc
Score of change (SOC) type:                                     Smoothing f
Termination criterion for MGWR:                                     1.0e-05
Number of iterations used:                                               34

MGWR bandwidths
------------------------------------------------------------------------------
Variable                  Bandwidth      ENP_j   Adj t-val(95%)           DoD_j
Intercept                    51.000      1.780            2.256           0.857
database_R                   55.000      1.042            2.023           0.990
database_I_C                 47.000      1.931            2.290           0.837
database_I_A                 55.000      1.355            2.139           0.925
database_C_I                 43.000      1.745            2.247           0.862
database_C_S                 55.000      1.301            2.121           0.935
database_T_D                 55.000      1.561            2.200           0.889
database_D_P                 45.000      2.425            2.384           0.780

Diagnostic Information
------------------------------------------------------------------------------
Residual sum of squares:                                                 1.985
Effective number of parameters (trace(S)):                              13.139
Degree of freedom (n - trace(S)):                                       42.861
Sigma estimate:                                                          0.215
Log-likelihood:                                                         14.053
Degree of Dependency (DoD):                                              0.877
AIC:                                                                     0.173
AICc:                                                                   10.651
BIC:                                                                    28.810
R2:                                                                      0.965
Adj. R2:                                                                 0.953
```

Anexo 8. Credencial de Pedido de Dados emitido a Electricidade de Moçambique

UNIVERSIDADE EDUARDO MONDLANE

Faculdade de Ciências

Á
Empresa Electricidade de Moçambique
EDM

CREDENCIAL

No âmbito de elaboração do Trabalho Final do Curso de Licenciatura em Ciências de Informação Geográfica, com o Tema: **"Geoprocessamento Aplicado a análise de Factores Determinantes de Perdas não técnicas na rede eléctrica de baixa tensão"**, a Direcção da Faculdade de Ciências da Universidade Eduardo Mondlane, credencia o estudante **Ivan Belchior Sarmento Bule** para junto da Empresa Electricidade de Moçambique, efectuar a recolha de dados sobre perdas não técnicas de energia eléctrica para alimentar o seu trabalho de culminação de estudos.

Antecipadamente agradecemos o vosso apoio e colaboração.

Cordiais Saudações.

UNIVERSIDADE EDUARDO MONDLANE
SECRETARIA
FACULDADE DE CIÊNCIAS

O Director

Prof. Doutor Daúd Liace Jamal
(Professor Associado)

Maputo, 13 de Junho de 2024

ELECTRICIDADE DE MOÇAMBIQUE, E.P.
02-
20-06-24

Av. Julius Nyerere, nº 3453, Campus Principal,C. Postal 257, Tel.: (+258) 21 493377, Fax.: (+258) 21 493377,
Maputo – Moçambique

Digitalizada com CamScanner