



Disciplina SER 301 – Análise Espacial
Professores: Eduardo G. Camargo e Antônio Miguel V. Monteiro

“Análise Espacial de Padrões de Consumo de Energia Elétrica em Áreas Urbanas”

JOEL DAVID MELO TRUJILLO

São José dos Campos – SP
Dezembro/2011

SUMÁRIO

Capítulo 1 - Introdução	3
Capítulo 2 -Análise Espacial da Demanda Elétrica em Áreas Urbanas	5
2.1 ANÁLISES DE PADRÕES DE DISTRIBUIÇÃO DE PONTOS	5
2.1.1 Caracterização de Distribuição de Pontos	5
2.1.2 Estimadores de Intensidade (<i>Kernel Estimation</i>)	8
2.1.2.1 Conclusões do Estimador de Intensidade	10
2.1.3 Estimadores de Dependência Espacial	10
2.1.3.1 Método do Vizinho Mais Próximo	11
2.1.3.2 Função K	13
2.1.3.3 Conclusões dos Estimadores de Dependência Espacial	15
2.2 ANÁLISES DE PADRÕES DE ÁREAS	15
2.2.1 Modelos de Distribuição de Dados em Áreas	16
2.2.1 Análise de Autocorrelação Espacial	16
2.2.2 Matrizes de Proximidade Espacial	17
2.2.3 Indicador Global de Autocorrelação Espacial	18
2.2.4 Indicadores Locais de Associação Espacial	21
2.2.5 Conclusões da Análise de Padrões de Área	22
Capítulo 3 - Conclusões Finais	24
3.1 TRABALHOS FUTUROS	24
REFERÊNCIAS	25

Capítulo 1 - Introdução

O elevado nível de expansão das cidades, o envelhecimento das redes elétricas e as práticas do mercado de eletricidade têm deixado as concessionárias de distribuição com grandes desafios no planejamento da expansão das redes. As empresas elétricas têm como objetivo oferecer um serviço confiável e econômico na área de serviço para respeitar os padrões impostos pela lei vigente. A confiabilidade, eficiência e economia do sistema de distribuição dependem da localização dos elementos da rede dentro da área de serviço e da sua capacidade para poder abastecer a demanda elétrica. Portanto, para realizar um planejamento eficiente da expansão da rede elétrica com uma viabilidade econômica e técnica, e manter os consumidores satisfeitos pela qualidade do serviço prestado, as empresas precisam identificar a como a demanda elétrica encontra-se distribuída na área de concessão, quais são os padrões que existem dentro da área de serviço e verificar a existência de aglomerados que tenham uma dinâmica de crescimento de demanda elétrica diferenciada.

O supracitado motivou a realizar uma análise exploratória da variável demanda elétrica com o fim de encontrar padrões de consumo de energia em áreas urbanas que possa ter alguma relação com o espaço geográfico e que possam direcionar a expansão das redes elétricas.

Para realizar a análise exploratória utilizou-se como variável de entrada a base de dados de instalação de transformadores da empresa elétrica concessionária na cidade de estudo. Nesta base tem-se a coordenada geográfica da localização de cada transformador e sua capacidade nominal em KVA. Com o fim de aproveitar as características dos dados foram realizadas análises de padrões de distribuição de pontos e análise de padrões de áreas, os detalhes destas análises serão mostrados no Capítulo 2.

Na área de planejamento e no setor elétrico, em geral, é muito utilizada por profissionais e pesquisadores a palavra *carga* para referir-se à *demanda de energia* e o *consumo de energia elétrica*. Para a Agência Nacional de Engenharia Elétrica (AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA - ANEEL), a *carga* é a caracterização da demanda do sistema, em um determinado ponto de interesse, definida por uma ou mais das seguintes grandezas: potência ativa, demanda de energia ativa e demanda de energia reativa (ANEEL, 2009). Neste trabalho utilizou-se o termo “*carga*” que deverá ser entendido como demanda de energia elétrica.

Por outro lado, os usuários da rede elétrica estão familiarizados com a palavra *energia elétrica*, mas as palavras *energia elétrica* e *demanda elétrica* são termos diferentes do ponto de vista da Engenharia Elétrica. Segundo a ANEEL, a *energia elétrica fornecida* é a quantidade de energia elétrica (kWh) fornecida e medida pela distribuidora: aos usuários para o consumo próprio. Porém, a *demanda elétrica* é a média das potências elétricas ativas ou reativas, solicitadas ao sistema elétrico pela parcela dos aparelhos elétricos instalados e em operação, durante um intervalo de tempo especificado, as unidades utilizadas são kW ou KVA (ANEEL, 2009). Neste trabalho a variável demanda elétrica é a variável de interesse que se deseja estudar a fim de realizar uma expansão das redes elétricas eficiente, econômica e confiável.

Os resultados das análises realizadas podem ser de ajuda para mostrar a evidencia de dependência espacial no processo de instalação de transformadores e existência autocorrelação espacial entre a densidade de carga (consumo de demanda por unidade de área) de uma subárea e seus vizinhos. Tais informações irão auxiliar a tomada de decisão no planejamento da expansão das redes elétricas

As análises realizadas mostram como se podem utilizar técnicas de análise espacial para explorar o consumo de energia a fim de poder compreender fatores da área de estudo que podem direcionar a expansão das redes elétricas na cidade. As duas aplicações deste trabalho foram realizadas utilizando o programa R Development Core Team (2008).

O trabalho desenvolvido não procura realizar uma previsão de demanda, tarefa que é importante no planejamento de expansão. A determinação da demanda em cada localidade dentro da área de serviço está fora do foco deste trabalho. Este trabalho procura ajudar no processo exploratório antes de aplicar técnicas de simulação, ou outras técnicas, que permitam determinar a quantidade de demanda em cada localidade da área de serviço.

No capítulo seguinte se mostraram as análises de padrões de pontos e áreas. Logo, no Capítulo 3 serão mostradas as conclusões e futuros trabalhos a desenvolver com o fim de criar ferramentas de auxílio para expansão das redes elétricas.

Capítulo 2 -Análise Espacial da Demanda Elétrica em Áreas Urbanas

Neste capítulo serão apresentadas as análises de padrões de distribuição de pontos e padrões de áreas. Os conceitos teóricos dessas análises podem ser encontrados em Bailey e Gatrell (1995). Também foram consultadas a publicação de Fucks, Sá Carvalho, Câmara e Monteiro (2004) e Bivand, Pebesma e Gómez-Rubio (2008).

Um breve resumo teórico de cada uma dessas técnicas serão mostradas antes de apresentar os resultados obtidos.

2.1 ANÁLISES DE PADRÕES DE DISTRIBUIÇÃO DE PONTOS

2.1.1 Caracterização de Distribuição de Pontos

O interesse primário ao analisar padrões de distribuição de pontos é determinar se os eventos observados, que para a aplicação do trabalho é a instalação de transformadores, exibem algum padrão sistemático, em oposição à uma distribuição aleatória. Busca-se detectar a existência de padrão de conglomerados espaciais (*cluster*), através da constatação de um número acima do esperado de casos excessivamente próximos, considerando uma distribuição estocástica, usualmente um processo de Poisson. Se um padrão de eventos pontuais apresenta desvios significativos do comportamento esperado para uma distribuição de Poisson, isto indicará a existência de uma distribuição espacial diferente da completa aleatoriedade, que merece ser objeto de maior análise.

Desde o ponto de vista da estatística, os processos pontuais são definidos como um conjunto de pontos irregularmente distribuídos em um terreno, cuja localização foi gerada por um mecanismo estocástico. Para sua caracterização, este processo estocástico pode ser descrito em termos dos efeitos de primeira ordem e efeitos de segunda ordem.

Os efeitos de primeira ordem, também conhecidos como globais ou de larga escala, correspondem a variações no valor médio do processo no espaço. Neste caso, o interesse é a intensidade do processo, isto é, o número de eventos por unidade de área. Para a primeira aplicação do trabalho será números de transformadores por unidade de área. Por outro lado, os

efeitos de segunda ordem, também denominados locais ou de pequena escala, representam a dependência espacial no processo, proveniente da estrutura de correlação espacial. Para medir a dependência espacial, se procurará estimar o relacionamento entre pares de eventos por unidade de área no espaço, o que corresponde a uma aproximação do cálculo da covariância entre as variáveis aleatórias que representam cada evento, os quais são um processo estocástico, na qual cada ocorrência é uma realização de uma variável aleatória distinta.

Considera-se um conjunto de pontos (u_1, u_2, \dots) numa determinada região A onde ocorreram estes eventos. O processo pontual é modelado considerando sub-regiões S em A através de sua esperança $E[N(S)]$ e a covariância $C[N(S_i)N(S_j)]$, na qual $N(S)$ denota o número de eventos em S . Considerando que o objetivo da análise de pontos é estimar as prováveis localizações de ocorrência de determinados eventos, essas estatísticas devem ser inferidas considerando o valor limite da quantidade de eventos por área.

Este valor limite corresponde à esperança de $N(S)$ para uma pequena região du em torno do ponto u , quando essa tende a zero. Essa esperança é denominada intensidade (propriedade de primeira ordem), sendo definida como:

$$\lambda(u) = \lim_{|du| \rightarrow 0} \left\{ \frac{E[N(du)]}{|du|} \right\} \quad 2.1$$

Propriedades de segunda ordem podem ser definidas da mesma forma, considerando a intensidade conjunta $\lambda(u_i, u_j)$ entre duas regiões infinitesimais $|du_i|$ e $|du_j|$ que contém os pontos u_i e u_j .

$$\lambda(d(u_i), d(u_j)) = \lim_{du_i, du_j \rightarrow 0} \left\{ \frac{C[N(S_i)N(S_j)]}{du_i, du_j} \right\} \quad 2.2$$

Quando o processo é estacionário, $\lambda(u)$ é uma constante, ou $\lambda(u) = \lambda$; se também é isotrópico, $\lambda(u_i, u_j)$ se reduz à $\lambda(|r|)$, sendo $|r|$ a distância entre os dois pontos. Quando o processo é não estacionário, ou seja, a intensidade média varia na região A , a modelagem da estrutura de dependência $\lambda(u_i, u_j)$ deve incorporar a variação de $\lambda(u)$. A maior parte das técnicas de análise de distribuição de pontos supõe, explícita ou implicitamente, um comportamento estacionário e isotrópico do processo aleatório subjacente aos eventos analisados.

Para a primeira aplicação do trabalho, a ocorrência de instalação dos transformadores está condicionada pela distribuição dos usuários. Além disso, características individuais dos

usuários como consumo residencial, comercial e industrial alto ou baixo podem condicionar a instalação da capacidade dos transformadores a ser instalado em uma determinada localidade. É possível, entretanto, modelar estes eventos e detectar áreas de maior densidade de carga, verificando a variação da intensidade do evento na região e a estrutura de correlação local.

A análise estatística dos padrões de distribuições de pontos requer um modelo teórico de referência, base para o desenvolvimento de métodos formais que checam a significância dos resultados exploratórios. O modelo teórico mais simples (e bastante aplicado na prática) é conhecido como aleatoriedade espacial completa (*Complete Spatial Randomness* - CSR). Este modelo divide a região de estudo A em subáreas S_i e modela a distribuição de eventos pontuais como um processo aleatório

$$\{Z_i(u_i), u_i \in S_i : i = 1, \dots, n\} \quad 2.3$$

Neste caso, considera-se que $Z_i(u_i)$ é o número de eventos que ocorrem na subárea S_i . No modelo CSR, as ocorrências em cada subárea são não correlacionadas e homogêneas, e estão associadas à mesma distribuição de probabilidade de Poisson. Numa visão intuitiva, pode-se considerar que a posição dos eventos é independente e de que os eventos tem igual probabilidade de ocorrência em toda a região A .

Esta formulação permite estabelecer uma base de comparação entre uma distribuição completamente aleatória (que seria gerada por um processo de Poisson) e os dados coletados em campo. O procedimento mais usual para estimar a probabilidade associada ao padrão encontrado será produzir uma simulação do processo aleatório na região de estudo. Dado um número fixo de eventos medidos em campo (denotado por n), determinamos o retângulo envolvente da região A (seja $\{(x, y) : x_1 \leq x \leq x_2, y_1 \leq y \leq y_2\}$). Os eventos são gerados a partir de abscissas x , obtidas de uma distribuição uniforme em (x_1, x_2) e de ordenadas y , obtidas de uma distribuição uniforme em (y_1, y_2) . Pontos que caem fora da região são rejeitados. Este processo é repetido até que n eventos tenham sido obtidos na região.

Podemos gerar um conjunto de simulações, para que possamos obter uma base de comparação entre o comportamento de um processo aleatório e a distribuição dos eventos medidos. Os conceitos de CSR são utilizados para caracterizar os efeitos de segunda ordem em distribuição de pontos, utilizando os métodos do vizinho mais próximo e da função K ,

descritos na seção 2.1.3. Também esses conceitos são utilizados para avaliação em vários métodos de detecção de aglomerados (*clusters*).

2.1.2 Estimadores de Intensidade (*Kernel Estimation*)

Uma alternativa para analisar o comportamento de padrões de pontos é estimar a intensidade pontual do processo em toda a região de estudo. Para isto, pode-se ajustar uma função bidimensional sobre os eventos considerados, composto de uma superfície cujo valor será proporcional à intensidade de amostras por unidade de área. Esta função realiza uma contagem de todos os pontos dentro de uma região de influência, ponderando-os pela distância de cada um à localização de interesse, como mostrado na Figura 2.1.

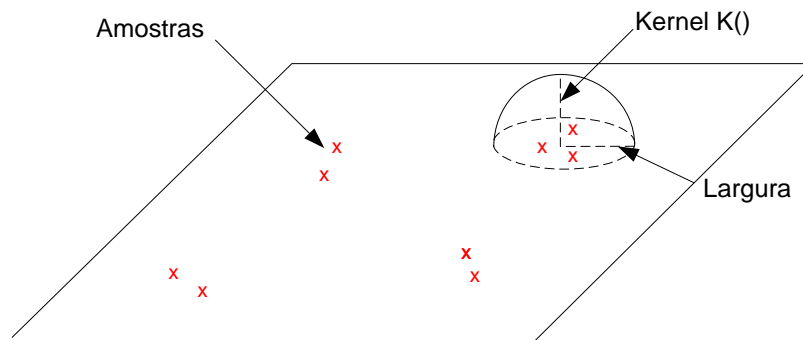


Figura 2.1. Estimador de intensidade de distribuição de pontos.

Considerando que u_1, \dots, u_n são localizações de n eventos observados em uma região A e que u represente uma localização genérica cujo valor se deseja estimar. O estimador de intensidade é computado a partir dos m eventos $\{u_i, \dots, u_{i+m-1}\}$ contidos num raio de tamanho τ em torno de u e da distância d entre a posição e a i -ésima amostra, a partir de funções cuja forma geral é:

$$\hat{\lambda}_\tau = \frac{1}{\tau^2} \sum_{i=1}^n k\left(\frac{d(u_i, u)}{\tau}\right), d(u_i, u) \leq \tau \quad 2.4$$

O estimador mostrado em 2.4 é chamado *kernel estimator* e seus parâmetros básicos são: (a) um raio de influência ($\tau \geq 0$) que define a vizinhança do ponto a ser interpolado e controla o "alisamento" da superfície gerada; (b) uma função de estimação com propriedades de suavização do fenômeno. O raio de influência define a área centrada no ponto de estimação u que indica quantos eventos u_i contribuem para a estimativa da função intensidade λ . Um raio muito pequeno irá gerar uma superfície muito descontínua; se for grande demais, a

superfície poderá ficar muito amaciada. No caso da função de interpolação $k()$, é comum usar funções de terceira ou quarta ordem, como o *kernel gaussiano*.

$$K(r) = \frac{1}{2\pi\tau} e^{-\left(\frac{r^2}{2\tau^2}\right)} \quad 2.5$$

Neste estimador, r representa a distância entre a localização na qual se deseja calcular a função e o evento observado.

O estimador de intensidade é muito útil para fornecer uma visão geral da distribuição de primeira ordem dos eventos.

A Figura 2.2 ilustra a aplicação do estimador de intensidade para a instalação de transformadores no ano 2001. Esta figura foi obtida considerando o método de Sheather e Jones (1991), o qual se encontra disponível em R Development Core Team (2008). Por outro lado, a função de estimação utilizada é a função do *kernel gaussiano*.

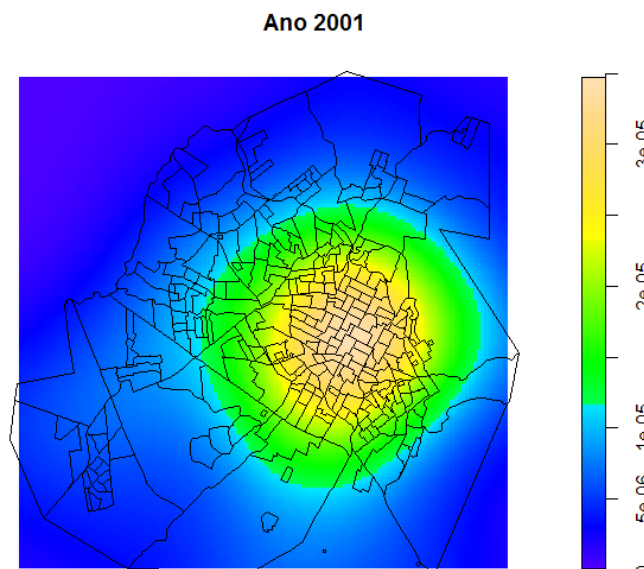


Figura 2.2 Estimador de intensidade para a distribuição de transformadores instalados no ano 2001.

Da Figura 2.2 pode-se observar que se tem uma alta intensidade no centro de atividades da cidade e que esta intensidade diminui na medida em que se afasta do centro de atividades.

A Figura 2.3 ilustra a aplicação do estimador de intensidade para a instalação de transformadores no ano 2002. Os parâmetros utilizados são os mesmos da Figura 2.2.

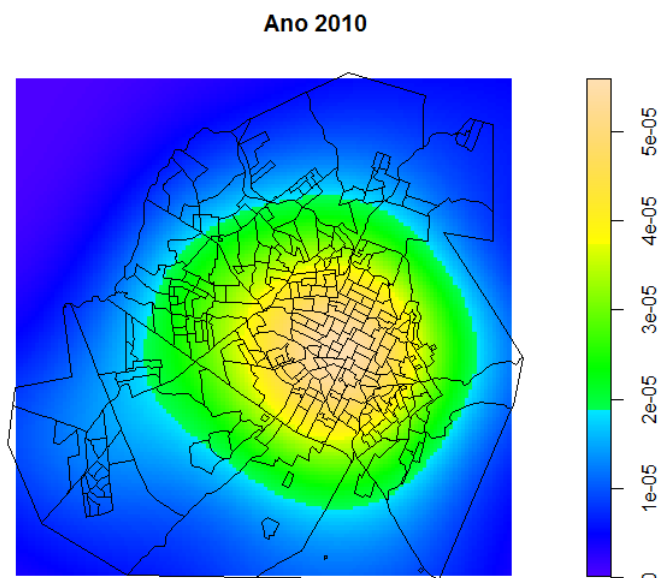


Figura 2.3 Estimador de intensidade para a distribuição de transformadores instalados no ano 2010.

Comparando as Figuras 2.2 e 2.3 pode-se observar que a intensidade no centro de atividades da cidade aumentou, mantêm-se a característica de crescimento centro-periferia, a intensidade diminui na medida que se afasta do centro de atividades, e que a intensidade em zonas da periferia da cidade apresenta um aumento considerável.

2.1.2.1 Conclusões do Estimador de Intensidade

Considerando que o maior número de transformadores corresponde às zonas de maior demanda elétrica e observando as Figuras 2.2 e 2.3, pode-se afirmar que a cidade tem um centro de maior consumo, que para as aplicações do trabalho coincidem com o centro de atividades da cidade.

Também, pode-se concluir que a cidade apresenta um comportamento de demanda centro-periferia. Maior crescimento de demanda elétrica no centro da cidade e diminui na medida em que se afasta do centro.

2.1.3 Estimadores de Dependência Espacial

Para a estimação de propriedades de segunda ordem do processo pontual, as técnicas mais utilizadas são o vizinho mais próximo e a função K, descritos a seguir.

2.1.3.1 Método do Vizinho Mais Próximo

O método do vizinho mais próximo estima a função de distribuição cumulativa $\hat{G}(r)$ baseado nas distâncias r entre eventos em uma região de análise.

Esta função de distribuição pode ser estimada empiricamente da seguinte forma:

$$\hat{G}(r) = \frac{\#(d(u_i, u_j) \leq r)}{n} \quad 2.6$$

Sendo que o valor normalizado acumulado para uma distância r corresponde à soma dos vizinhos mais próximos de cada evento cuja distância é menor ou igual a r , dividido pelo número de eventos na região.

O gráfico desta função de distribuição cumulativa empírica $\hat{G}(r)$ pode ser usada como um método exploratório para verificar a existência de evidência de interação entre os eventos. Se este gráfico apresenta um crescimento rápido para pequenos valores de distância, esta situação aponta para interação entre os eventos caracterizando agrupamentos nestas escalas. Se este gráfico apresentar valores pequenos no seu início, e só crescer rapidamente para valores maiores de distância, esta situação aponta para uma distribuição mais regular.

A Figura 2.4 mostra a função $G_{pois}(r)$, $\hat{G}_{han}(r)$, $\hat{G}_{bord}(r)$ e $\hat{G}_{km}(r)$ que representam a função de distribuição teórica de Poisson, função de distribuição estimada com correção de Hanish (1984), função de distribuição estimada com correção de borda e a função de distribuição obtida através do estimador Kaplan-Meier (BADLEY; GILL, 1997) respectivamente, para os dados da instalação de transformadores para o ano 2001, com distância mínima de 0 m e distância máxima de 200 m. Verifica-se que a curva mostra um crescimento acentuado até distâncias no entorno de 150 m para depois se estabilizar, o que caracteriza agrupamento nesta faixa de distâncias.

A análise de vizinhança pode ser usada como método formal para se comparar estatisticamente a distribuição dos eventos observados com o que se esperaria na hipótese da aleatoriedade espacial completa (CSR). Esta metodologia consiste em se criar funções de distribuição estimadas para a distribuição CSR, a fim de se acessar a significância dos desvios. Na hipótese de CSR, a função de distribuição $G(r)$ seria dada por um processo de Poisson.

$$G(r) = 1 - e^{-\lambda\pi r^2} \quad 2.7$$

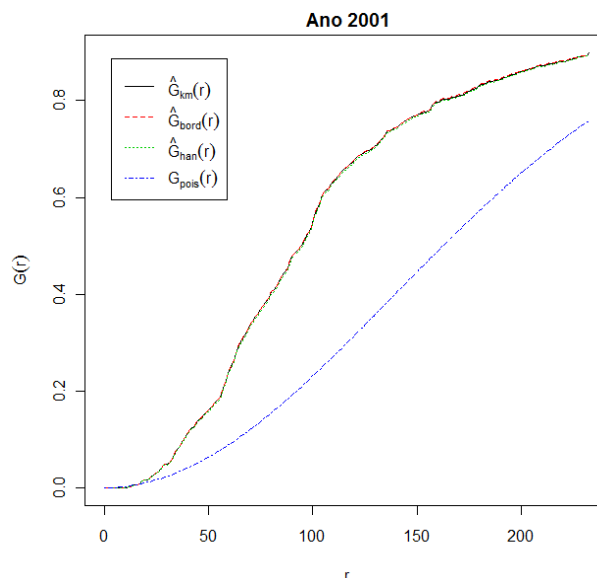


Figura 2.4 Função de distribuição do vizinho mais próximo para instalação de transformadores do ano 2001.

O gráfico da distribuição estimada ($G(r)$) versus a distribuições estimadas ($\hat{G}_{han}(r)$, $\hat{G}_{bord}(r)$ e $\hat{G}_{km}(r)$), permite medir a significância dos desvios relativo a aleatoriedade. Se a condição CSR for válida para os dados observados, os gráficos das curvas de $\hat{G}_{han}(r)$, $\hat{G}_{bord}(r)$ e $\hat{G}_{km}(r)$ versus $G(r)$ deve ser praticamente linear com um ângulo de 45 graus. Se o dado apresenta tendências para agrupamentos, os traçados no gráfico estarão acima da linha de 45 graus, ao passo que para padrões de regularidade os traçados ficarão abaixo da linha de 45 graus.

A Figura 2.5 mostra um exemplo do gráfico da distribuição da função $G(r)$ mostrando o posicionamento da distribuição e das distribuições estimadas, para os dados referentes aos dados de instalação de transformadores para o ano 2011. Neste caso percebe-se a posição das funções estimadas acima da linha de 45 graus, o que caracteriza agrupamento para as distâncias em análise.

Embora o método do vizinho mais próximo forneça uma indicação inicial da distribuição espacial, ele considera apenas escalas pequenas. Para se ter informação mais efetiva para o padrão espacial em escalas maiores, o melhor método a ser utilizado é o da função K.

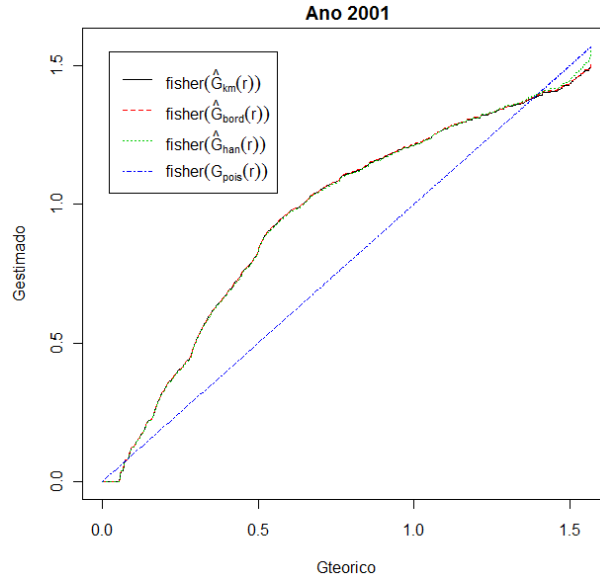


Figura 2.5 Gráfico de $\hat{G}_{han}(r)$, $\hat{G}_{bord}(r)$ e $\hat{G}_{km}(r)$ versus $G(r)$ para os dados de instalação de transformadores do ano 2001.

2.1.3.2 Função K

A função K, também denominada medida de momento de segunda ordem reduzido, é definida para o processo univariado como:

$$\lambda K(r) = E(\# \text{ eventos contidos a uma distância } r \text{ de um evento arbitrário}) \quad 2.8$$

Na qual, # está associado ao número de eventos, $E()$ é o operador de estimativa, e λ é a intensidade ou número médio de eventos por unidade de área, assumida constante na região. Uma estimativa de $K(r)$ é:

$$\hat{K}(r) = \frac{A}{n^2} \sum_i^n \sum_{j, i \neq j}^n \frac{I_r(d_{ij})}{w_{ij}} \quad 2.9$$

Na qual A é a área da região, n é o número de eventos observados, $I_r(d_{ij})$ é uma função indicação cujo valor é 1 se $(d_{ij}) \leq r$ e 0 em caso contrário, e w_{ij} é a proporção da circunferência do círculo centrado no evento i que está dentro da região (correção devido ao efeito de borda).

A função K é usada como ferramenta exploratória na comparação entre estimativa empírica ($\hat{K}(r)$) e a resultante de um processo de padrão de pontos espacial aleatório ($K(r)$). Para um processo aleatório $K(r)$ seria πr^2 . Portanto, uma forma de comparar a estimativa K

de um conjunto de dados observados com πr^2 seria realizar o gráfico da função $\hat{L}(h)$ definida como:

$$\hat{L}(r) = \sqrt{\frac{\hat{K}(r)}{\pi}} - r \quad 2.10$$

O gráfico de $\hat{L}(r)$ em função da distância r indica atração espacial entre eventos ou agrupamentos para valores positivos, sendo o agrupamento mais forte em picos positivos, e indica repulsão espacial ou regularidade em pontos de valores negativos. Uma abordagem similar à do vizinho mais próximo pode ser feita para se estimar a significância dos desvios da distribuição $\hat{L}(r)$ em relação à condição de aleatoriedade (CSR). Desta forma, se todos os valores estimados da função $\hat{L}(r)$ estiverem acima do valor teórico da função $L(r)$, teremos agrupamentos em todas as escalas. Depressões negativas na função estimada $\hat{L}(r)$ que estiverem abaixo dos valores estimados evidenciam regularidade nessa escala, portanto, se todos os valores de $\hat{L}(r)$ estiverem abaixo da função $L(r)$ e com valores negativos, tem-se regularidade em todas as escalas.

A Figura 2.6 mostra o gráfico da função $\hat{L}(r)$ para os dados em análise (Figura 2.1). Verificam-se valores que existem agrupamentos dentro de uma faixa de distâncias entre 2,8 km e 3,0 km.

A função L foi obtida para todo o período disponível, ou seja, desde o ano 2001 até o ano 2010, Na Tabela 2.1 mostra a evolução do valor máximo em km da Função L para cada ano. Pode-se observar que a distância máxima durante os 10 anos encontra-se em torno de 3,4 km. Considerando simetria entorno desse valor máximo com um desvio de 0,1 km, nesse caso pode-se obter que para distâncias entre 3,3 a 3,5 km há evidências de agrupamento.

Tabela 2.1 Valor Máximo em km da Função L desde 2001 até o 2010.

Ano	Distância Máxima
2001	2948
2002	3029
2003	3431
2004	3430
2005	3430
2006	3430
2007	3430
2008	3436
2009	3434
2010	3600

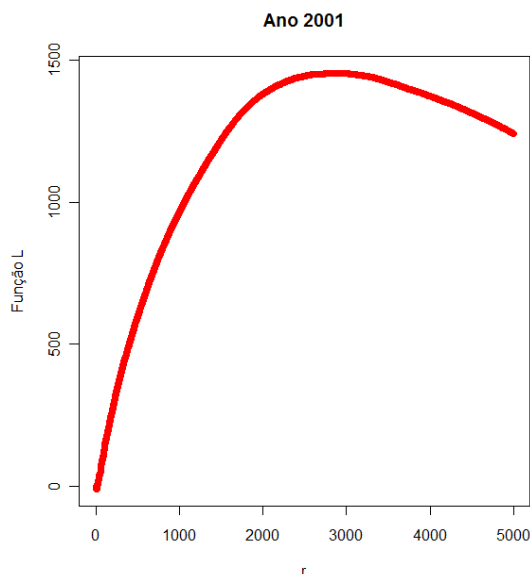


Figura 2.6 Função L para a distribuição de transformadores do ano 2001.

2.1.3.3 Conclusões dos Estimadores de Dependência Espacial

Os resultados obtidos das técnicas aplicadas demonstram a evidência de agrupamentos. Estes agrupamentos são importantes para os planejadores, devido a que dentro da área de serviço existem consumidores que atraem outros consumidores incrementando a demanda elétrica em seu entorno. Por exemplo, geralmente, a construção de um centro comercial facilita o crescimento da demanda residencial próxima da sua área de influência, mas a construção de uma rodovia facilita o crescimento da demanda industrial na área.

As técnicas mostradas nas seções anteriores permitem encontrar o raio da área de influência dos agrupamentos presentes na área urbana. O valor do raio encontrado permitirá aos engenheiros de planejamento determinar a abrangência dessas construções, desta forma, pode-se analisar a capacidade da infraestrutura elétrica para abastecer as demandas elétricas destas construções.

2.2 ANÁLISES DE PADRÕES DE ÁREAS

2.2.1 Modelos de Distribuição de Dados em Áreas

O modelo de distribuição mais utilizado para dados de área é o modelo de variação espacial discreta. Considere-se a existência de um processo estocástico Z_i , $i = 1, \dots, n$, onde Z_i é a realização do processo espacial na área i e n é o total de áreas A_i . O objetivo principal da análise é construir uma aproximação para a distribuição conjunta de variáveis aleatórias $Z = \{Z_1, \dots, Z_n\}$, estimando sua distribuição.

De forma semelhante ao modelo de eventos pontuais discutido na seção 2.1, considere-se Z_i como a variável aleatória que descreve a contagem, indicador ou taxa associada à área A_i . O valor observado Z_i , correspondente à contagem na i -ésima área. A hipótese mais comum é supor que a variável aleatória Z_i , que descreve o número de ocorrências em cada área pode ser associada a uma distribuição de probabilidade de Poisson. Tal hipótese justifica-se por ser esta a distribuição estatística mais adequada a fenômenos que envolvem contagens de eventos, como é o caso na maioria dos dados agregados por áreas. Evidentemente outras distribuições podem ser mais adequadas, dependendo da variável a ser analisada. Taxas podem ser modeladas utilizando a distribuição normal, pois ainda que esta admita valores negativos, evidentemente impossíveis neste tipo de indicador, as propriedades da distribuição normal podem ser adequadas.

A alternativa à hipótese de variação espacial discreta é supor que os dados apresentam variação espacial contínua. Considera-se um processo estocástico $\{Z(x), x \in A, A \subset \mathfrak{R}^2\}$, cujos valores podem ser conhecidos em todos os pontos da área de estudo. Neste caso, as contagens agregadas devem ser transformadas em taxas ou indicadores, pois o que varia continuamente no espaço são as taxas e não as contagens. Para o desenvolvimento da segunda aplicação será considerado como taxa a densidade de carga instalada, isto é, os KVA totais instalados por unidade de área.

2.2.1 Análise de Autocorrelação Espacial

Para estimar a magnitude da autocorrelação espacial entre as áreas são utilizadas o índice global de Moran, o índice de Geary e o variograma. Quando se dispõe de grande número de áreas, resultantes, por exemplo, de escalas espaciais detalhadas, a natureza dos processos envolvidos é tal que é muito provável a existência de diferentes regimes de correlação espacial em diferentes sub-regiões. Para evidenciar estes regimes espaciais,

podem-se utilizar os indicadores locais de autocorrelação espacial e o mapa de espalhamento de Moran. Todas estas estatísticas dependem da definição de vizinhança adotada, como se mostrará na seção 2.2.2.

Por outro lado, o variograma é a base das análises realizada pela geoestatística. A geoestatística não pode ser aplicada com dados de densidade de carga, devido a que esta densidade não se encontra distribuída em forma uniforme em toda a área urbana e não apresentam uma correlação direta com a distância. Portanto, prosseguiu-se a explorar a técnica de Moran utilizando os resíduos de uma regressão linear para contorna a instabilidade das taxas ou indicadores, como será mostrada na seção 2.4.

2.2.2 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. Dado um conjunto de n áreas $\{A_1, \dots, A_n\}$, construímos a matriz $W^{(1)}(n \times n)$, onde cada um dos elementos w_{ij} representa uma medida de proximidade entre A_i e A_j . Esta medida de proximidade pode ser calculada a partir de um dos seguintes critérios:

- $w_{ij} = 1$, se o centróide de A_i está a uma determinada distância de A_j ; caso contrário $w_{ij} = 0$.
- $w_{ij} = 1$, se A_i compartilha um lado comum com A_j , caso contrário $w_{ij} = 0$.
- $w_{ij} = \frac{l_{ij}}{l_i}$, 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 vários cálculos de índices de autocorrelação espacial.

A ideia da matriz de proximidade espacial pode ser generalizada para vizinhos de maior ordem (vizinhos dos vizinhos). Com critério análogo ao adotado para a matriz de vizinhança de primeira ordem, pode-se construir as matrizes $W^{(2)}, \dots, W^{(n)}$.

Na Figura 2.7 ilustra-se o diagrama de vizinhança adotada para a aplicação deste trabalho. O critério seguido para obter este diagrama é a se contiguidade que o segundo critério dos antes mencionados.

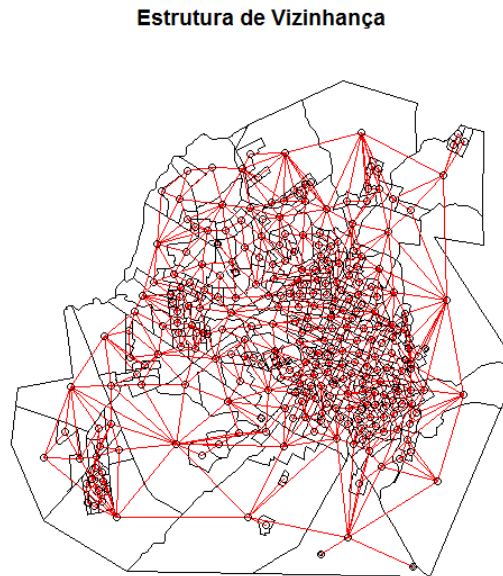


Figura 2.7 Estrutura de vizinhança para aplicação da análise de padrões por área.

2.2.3 Indicador Global de Autocorrelação Espacial

Autocorrelação espacial mostra como os valores estão correlacionados no espaço. Neste contexto, as funções de autocorrelação espacial e o variograma são utilizados para estimar quanto o valor observado de um atributo numa região é dependente dos valores desta mesma variável nas localizações vizinhas. O índice global de Moran I é a expressão da autocorrelação considerando apenas o primeiro vizinho, como se mostra:

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (z_i - \bar{z})(z_j - \bar{z})}{\sum_{i=1}^n (z_i - \bar{z})^2} \quad 2.11$$

Na qual, n é o número de áreas, z_i é o valor do atributo considerado na área i , \bar{z} é o valor médio do atributo na região de estudo e w_{ij} são os elementos da matriz normalizada de proximidade espacial. Neste caso a correlação será computada apenas para os vizinhos de primeira ordem no espaço, conforme estabelecido pelos pesos w_{ij} . No caso que se deseje utilizar uma matriz de proximidade de maior ordem deve utilizar-se w_{ij}^k em lugar w_{ij} . Desta forma, pode-se a função de autocorrelação considerando matrizes de maior ordem.

De uma forma geral, o índice de Moran presta-se a um teste cuja hipótese nula é de independência espacial; neste caso, seu valor seria zero. Valores positivos (entre 0 e +1) indicam para correlação direta e negativos (entre 0 e -1) para correlação inversa. Uma vez calculado, é importante estabelecer sua validade estatística. Para estimar a significância do índice, será preciso associar a este uma distribuição estatística, sendo mais usual relacionar a estatística de teste à distribuição normal. Outra possibilidade, sem pressupostos em relação à distribuição, e abordagem mais comum é um teste de pseudo-significância. Neste caso, são geradas diferentes permutações dos valores de atributos associados às regiões; cada permutação produz um novo arranjo espacial, onde os valores estão redistribuídos entre as áreas. Como apenas um dos arranjos corresponde à situação observada, pode-se construir uma distribuição empírica de I. Se o valor do índice I medido originalmente corresponder a um “extremo” da distribuição simulada, então se trata de valor com significância estatística.

Algumas variações do modelo de Moran são o teste C de Geary e o teste Ipop. O primeiro (C de Geary) difere do teste I de Moran por utilizar a diferença entre os pares, enquanto que Moran utiliza a diferença entre cada ponto e a média global. Assim, o indicador C de Geary assemelha-se ao variograma, e o I de Moran ao correlograma.

A hipótese implícita do cálculo do índice de Moran é a estacionariedade de primeira e segunda ordem, e o índice perde sua validade ao ser calculado para dados não estacionários. Quando existir não-estacionariedade de primeira ordem (tendência), os vizinhos tenderão a ter valores mais parecidos que áreas distantes, pois cada valor é comparado à média global, inflacionando o índice. Da mesma forma, se a variância não é constante, nos locais de maior variância o índice será mais baixo, e vice-versa. Quando o dado é não-estacionário, a função de autocorrelação continua decaindo mesmo após ultrapassar a distância onde há influências locais. Como as populações são diferentes entre as subáreas, como é de praxe, então a densidade de carga possuem variâncias diferentes e a hipótese subjacente ao teste baseado no índice I de Moran de que as variáveis sejam variáveis aleatórias independentes e identicamente distribuídas é violada. Desta forma, áreas com populações pequenas possuem densidade de carga com maior variância e mais propensas a assumir os valores mais extremos entre os observados. Uma forma de visualizar isto é realizando um gráfico da densidade de carga versus a área, como se mostra na Figura 2.8, na qual se pode visualizar o efeito funil, existe pequenas áreas com uma grande densidade de carga. Para aplicação desta seção, a extensão das subáreas urbanas e o número de habitantes tem uma relação direta, isto permite

realizar a análise a respeito da extensão de área é preferida, pois é mais comum esse dado no planejamento de expansão das redes elétricas.

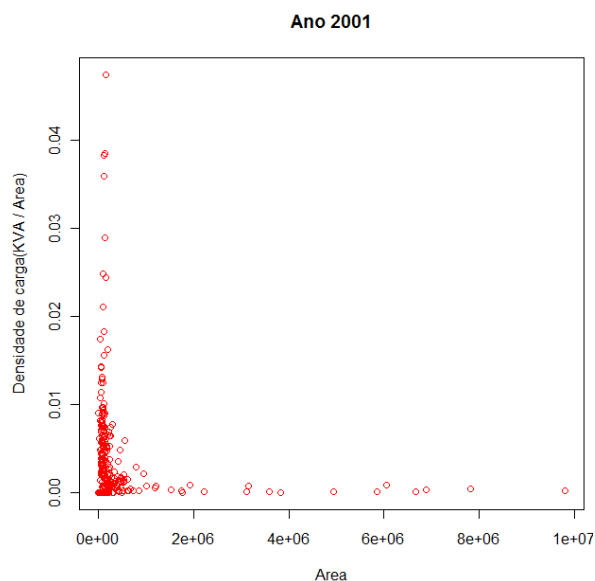


Figura 2.8 Gráfico que mostra o instabilidade da densidade de carga, efeito funil.

Para poder contornar este problema foi seguido o procedimento apresentado por Bivand, Pebesma e Gómez-Rubio (2008), na qual apresentam uma regressão linear de termo constante. Logo, realiza-se a análise de padrões de área com respeito aos resíduos dessa regressão linear. Este processo é válido assumindo que os coeficientes de Moran tem uma distribuição normal. A seguir se mostram os resultados da aplicação desta análise.

Na Figura 2.9 mostra-se o diagrama de espalhamento de Moran. Este diagrama é uma maneira adicional de visualizar a dependência espacial. Construído com base nos valores normalizados (valores de atributos subtraídos de sua média e divididos pelo desvio padrão), permite analisar o comportamento da variabilidade espacial. A ideia é comparar os valores normalizados do atributo numa área com a média dos seus vizinhos, construindo um gráfico bidimensional de z (valores normalizados) por wz (média dos vizinhos), que é dividido em quatro quadrantes, como mostrado nessa figura para os resíduos da regressão linear supracitada. Os quadrantes podem ser interpretados como: Q1 (valores positivos, médias positivas) e Q2 (valores negativos, médias negativas): indicam pontos de associação espacial positiva, no sentido que uma localização possui vizinhos com valores semelhantes. Q3 (valores positivos, médias negativas) e Q4 (valores negativos, médias positivas): indicam pontos de associação espacial negativa, no sentido que uma localização possui vizinhos com valores distintos.

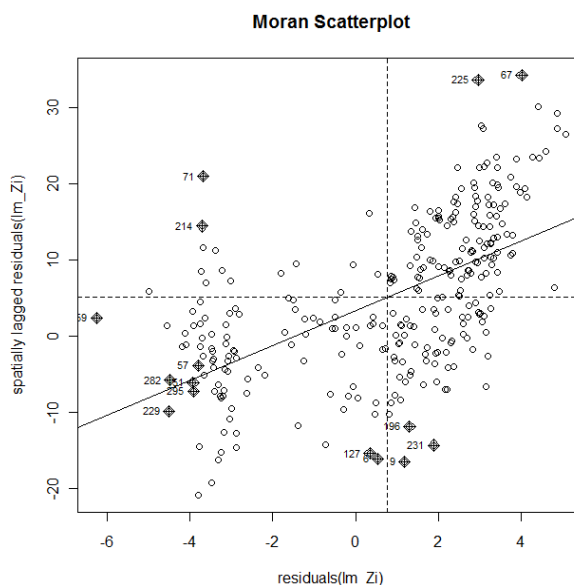


Figura 2.9 Diagrama de espalhamento de Moran.

O diagrama de espalhamento de Moran corrobora o processo de dependência espacial. Na Figura 2.9, a maior parte das subáreas se encontra localizadas nos quadrantes Q1 e Q2, que apresentam associação espacial positiva. Os pontos localizados nos quadrantes Q3 e Q4 podem ser vistos como regiões que não seguem o mesmo processo de dependência espacial das demais observações. Evidentemente, o diagrama reflete a estrutura espacial nas duas escalas de análise: vizinhança e tendência.

O índice de Moran I é equivalente ao coeficiente de regressão linear que indica a inclinação da reta de regressão (α) de wz em z . Para o caso dos dados apresentados na Figura 2.9, este coeficiente é igual a 0,54, o mesmo valor calculado aplicando-se a fórmula da equação 2.11.

2.2.4 Indicadores Locais de Associação Espacial

Os indicadores globais de autocorrelação espacial, como o índice de Moran, fornecem um único valor que é a medida da associação espacial para todo o conjunto de dados, que é útil na caracterização da região de estudo como um todo. Quando lidamos com grande número de áreas, é muito provável que ocorram diferentes regimes de associação espacial e que apareçam máximos locais de autocorrelação espacial, onde a dependência espacial é ainda mais pronunciada. Assim, muitas vezes é desejável examinar padrões em maior detalhe. Portanto, é preciso utilizar indicadores de associação espacial que possam ser associados às diferentes localizações de uma variável distribuída espacialmente. Os indicadores locais

produzem um valor específico para cada área, permitindo assim a identificação de agrupamentos. O índice local de Moran pode ser expresso para cada área i a partir dos valores normalizados z_i do atributo como:

$$I_i = \frac{Z_i \sum_{j=1}^n w_{ij} Z_j}{\sum_{j=1}^n Z_j^2} \quad 2.12$$

A significância estatística do uso do índice de Moran local é computada de forma similar ao caso do índice global. Para cada área, calcula-se o índice local, e depois se permuta aleatoriamente o valor das demais áreas, até obter uma pseudo-distribuição para a qual podemos computar os parâmetros de significância. Uma vez determinada a significância estatística do índice local de Moran, é útil gerar um mapa indicando as regiões que apresentam correlação local significativamente diferente do resto dos dados. Estas regiões podem ser vistas como "bolsões" de não-estacionariedade, pois são áreas com dinâmica espacial própria e que merecem análise detalhada. Na Figura 2.10 mostra-se a significância dos índices de Moran locais. Logo, foi aplicado o teste de Shapiro-Wilk para conferir se os índices encontrados tem uma distribuição normal, o teste deu como valor 0.376, portanto, é aceita a hipótese nula de distribuição normal.

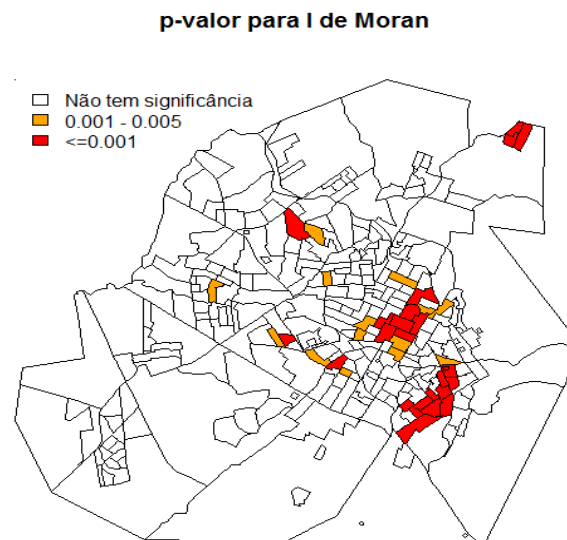


Figura 2.10 Indicador de autocorrelação espacial para a instalação de transformadores do ano 2001.

Apenas os valores cm significância maior de 95%

2.2.5 Conclusões da Análise de Padrões de Área

Os resultados evidenciam que existem algumas subáreas com autocorrelação espacial com uma significância estatística considerável. Estas subáreas tem dinâmica espacial própria

e merecem uma análise detalhada. Esta informação pode ajudar aos planejadores da empresa elétrica para encontrar lugares que podem se saturar mais devido a chegada de novas cargas. Para a rede elétrica os lugares com autocorrelação com significância considerável podem apresentar um efeito considerável a diferença de subáreas que não apresentam tal autocorrelação. Por exemplo, Instalar uma nova carga em uma subárea que apresenta uma autocorrelação com significância igual a 0,05 produzirá um efeito sobre os vizinhos considerável, possivelmente os vizinhos terão um incremento de densidade de carga, desta forma, nessa área se poderá obter uma saturação que dever ser analisada para se determinar se a rede elétrica nessa região pode abastecer a nova carga. No entanto, a mesma carga instalada em subáreas sem significância estatística considerável poderia não apresentar nenhum efeito nos vizinhos.

Capítulo 3 - Conclusões Finais

Neste trabalho foram apresentadas análises espaciais de padrões de pontos e áreas com o fim de ajudar aos planejadores na expansão das redes elétricas. A aplicação das análises realizadas consideram a instalação de transformadores em um sistema elétrico real, informação que se encontra disponível na maioria de empresas elétricas de distribuição.

A análise de padrões de pontos permitiu determinar que a área de serviço tem um centro de atividades que coincide com o centro de atividade da cidade. Em outras cidades pode encontrar-se mais de um centro de atividades ou pode não coincidir com o lugar com maior densidade de carga. Esta classe de análise permite explorar a área de serviço da empresa elétrica.

A análise de padrões de áreas pode ajudar aos planejadores na tomada de decisão da instalação de novas cargas e fornecer diretrizes de expansão em lugares com autocorrelação espacial. Enfatiza-se que se deve ter cuidado no momento de análise da densidade de carga, devido a que se apresenta o efeito funil, aplicações de técnicas de bayes empírico global e local devem ser exploradas.

Estas classes técnicas espaciais podem ser aplicadas em qualquer área de serviço, mas deve considerar-se um adequado uso das técnicas. Também os resultados obtidos neste trabalho não devem ser aplicados com muito critério devido a que embora em alguns lugares não apresente significância estatística podem ser diferentes no mundo real. É necessário considerar na aplicação dos resultados a experiência e o conhecimento dos planejadores.

3.1 TRABALHOS FUTUROS

O passo seguinte dentro da análise exploratório é pesquisar se existe alguma variável que pode ajudar explicar o comportamento da densidade de carga na área urbana, este passo será realizado em um trabalho futuro com informações do censo 2010 e com a base de consumo de energia.

REFERÊNCIAS

AGÊNCIA NACIONAL DE ENGENHARIA ELÉTRICA - ANEEL – PRODIST. **Procedimentos de distribuição de energia elétrica no sistema elétrico nacional**. Brasília: Agência Nacional de Energia Elétrica, 2009. Disponível em: <<http://www.aneel.gov.br>>. Acesso em: 01 fev. 2009.

BAILEY, T. C.; GATRELL, A. C. **Interactive spatial data analysis**. Essex, UK: Longman, 1996. 413p.

BIVAND, R. S.; PEBESMA, E. J.; GOMEZ-RUBIO, V.. Applied spatial data analysis with R. Springer, NY. <http://www.asdar-book.org/>. 2008.

FUCKS, S.; SÁ CARVALHO, M.; CÂMARA, G.; MONTEIRO, A. M. V. **Análise espacial de dados geográficos**. São José dos Campos: EMBRAPA, 2004.

BADDELEY, A.J; GILL, R.D. Kaplan-Meier estimators of interpoint distance distributions for spatial point processes. **Annals of Statistics** **25** (1997) 263-292.

R Development Core Team (2008). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>. 2008.

SHEATHER, S. J.; JONES M. C. A reliable data-based bandwidth selection method for kernel density estimation. **J. Roy. Statist. Soc. B**, 683–690. 1991.