



Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Estatística

**ANÁLISE DA LOCALIZAÇÃO DAS
EMPRESAS DE TRANSPORTE RODOVIÁRIO
DE CARGA NO DISTRITO FEDERAL**

BERNARDO NOGUEIRA SCHLEMPER 07/44506
BRUNO OURIQUES DE LIMA NOGUEIRA NUNES 07/54251

Brasília

2011

Bernardo Nogueira Schlemper 07/44506
Bruno Ouriques de Lima Nogueira Nunes 07/54251

ANÁLISE DA LOCALIZAÇÃO DAS EMPRESAS DE TRANSPORTE RODOVIÁRIO DE CARGA NO DISTRITO FEDERAL

Relatório apresentado à disciplina Estágio Supervisionado II do curso de graduação em Estatística, Departamento de Estatística, Instituto de Exatas, Universidade de Brasília, como parte dos requisitos necessários para o grau de Bacharel em Estatística.

Orientador: Prof. Dr. Alan Ricardo da Silva

Brasília

2011

Dedico este trabalho a meus pais, pelo amor, apoio e pelos valores por eles passados a mim, a meu irmão e meus familiares; minha namorada Janaina, e ao meu colega de trabalho e amigo Bruno e a todos os amigos que encontrei durante a graduação.

Bernardo Nogueira Schlemper

Dedico este trabalho a meus pais , por tudo que eles representam na minha vida, a meus irmãos e a todos os meus familiares que me acompanharam nesta longa jornada; a todos os meus verdadeiros amigos, em especial o Bernardo, grande companheiro e colega de trabalho.

Bruno Ouriques de Lima Nogueira Nunes

Agradecimentos

A Deus, pela força, perseverança, e por ter iluminado nossos caminhos para que pudéssemos concluir mais uma etapa de nossas vidas.

Ao professor orientador Alan por todo auxílio não somente nas reuniões semanais, mas principalmente quando enfrentávamos uma situação de dificuldade, pela cobrança fundamental para que o trabalho seguisse adiante, pelo incentivo para melhorarmos sempre, pela motivação e dedicação durante todo o trabalho, e por sua empolgação que nos contagiou e ajudou a seguir em frente.

Aos nossos pais, por acreditarem em nossa capacidade, por todo suporte dado, pelo incentivo a seguir em frente, pela amizade e pelo carinho.

Aos nossos familiares pelas orações e torcida.

A todos os professores que passaram por nosso caminho, por quem tivemos a sorte e a honra de ter seus conhecimentos transmitidos a nós.

Aos nossos amigos e colegas pelo incentivo e apoio constantes.

Resumo

O Transporte Rodoviário de Carga(TRC) representa atualmente o modo de transporte mais utilizado no Brasil, sendo sua parcela na matriz de transportes igual a 61,1%. Isso pode ser explicado por suas vantagens quando comparado com outros tipos de transporte, tal como a possibilidade de entrega na porta do comprador e a facilidade de embarque e desembarque de mercadorias. Sendo assim, informações sobre as características do TRC, assim como identificar quais fatores influenciam sua distribuição espacial são de fundamental importância para os interessados nessa forma de transporte.

A distribuição espacial da frota de veículos responsável por esse modo de transporte normalmente é influenciada por características regionais, como a existência de portos e rodovias, ou locais onde são produzidos determinados produtos, o que nos permite afirmar que o espaço geográfico é um fator determinante para tal distribuição. Assim, modelos estatísticos convencionais não são os mais adequados para a avaliação de tais eventos, visto que não permitem a utilização desse espaço associado a outras variáveis aleatórias na análise.

O presente trabalho apresenta uma análise referente a localização das empresas de TRC no Distrito Federal através do processo de análise pontual, visando assim observar se a distribuição espacial dessas empresas apresenta algum padrão definido ou se está disposta de maneira aleatória. Para tal fim, avaliou-se não somente a distribuição das empresas mas também como estão dispostos cada tipo de veículo de TRC no território do Distrito Federal.

Lista de Tabelas

3.1	Tabela com as distâncias e desvios padrão	19
5.1	Tabela de frequências dos tipos de veículo	38

Lista de Figuras

3.1	Mapa da região de <i>Soho</i> elaborado por John Snow	12
3.2	Tipos de Distribuição Espacial dos Pontos	16
3.3	Demonstração do cálculo da distância euclidiana	18
3.4	Vizinho mais próximo	21
3.5	Vizinho mais próximo com simulação	22
3.6	Função K	24
3.7	Criação da superfície	27
4.1	Funcionamento do Sistema Lat/Long	30
4.2	Utilização do plano cartesiano	31
4.3	Linha do Equador em UTM	31
4.4	Meridiano Central em UTM	32
4.5	Fusos existentes no Brasil	32
5.1	Mapa do Distrito Federal com representação das empresas de TRC	34
5.2	Estimadores de intensidade com diferentes valores de τ	35
5.3	Método do vizinho mais próximo para o modelo geral	36
5.4	Envelope de simulação do modelo geral	36
5.5	Função K do modelo geral	37

5.6	Distribuição dos veículos do tipo Caminhão Simples	39
5.7	Método do vizinho mais próximo para Caminhão Simples	39
5.8	Envelope de simulação para Caminhão Simples	40
5.9	Função de intensidade para Caminhão Simples com $\tau=5$	41
5.10	Função K para Caminhão Simples	41
5.11	Distribuição dos veículos do tipo Caminhão Leve	42
5.12	Método do vizinho mais próximo para Caminhão Leve	43
5.13	Envelope de simulação para Caminhão Simples	43
5.14	Função de intensidade para Caminhão Leve com $\tau=5$	44
5.15	Função K para Caminhão Leve	45
5.16	Distribuição dos veículos do tipo Semi-reboque	45
5.17	Método do vizinho mais próximo para Semi-reboque	46
5.18	Envelope de simulação para Semi-reboque	47
5.19	Função de intensidade para Semi-reboque com $\tau=5$	47
5.20	Função K para Semi-Reboque	48
5.21	Distribuição dos veículos do tipo Reboque	49
5.22	Método do vizinho mais próximo para Reboque	49
5.23	Envelope de simulação para Reboque	50
5.24	Função de intensidade para Reboque com $\tau=10$	51
5.25	Função K para Reboque	51

Sumário

RESUMO	iv
1 INTRODUÇÃO	1
1.1 OBJETIVOS	2
2 TRANSPORTE RODOVIÁRIO DE CARGA	4
2.1 INTRODUÇÃO	4
2.2 LEGISLAÇÃO BÁSICA	5
2.2.1 RNTRC	6
2.3 CUSTOS DE TRANSPORTE	8
3 ANÁLISE PONTUAL DE DADOS GEOGRÁFICOS	9
3.1 INTRODUÇÃO	9
3.2 PRIMEIRA UTILIZAÇÃO DA ANÁLISE ESPACIAL	10
3.3 ANÁLISE ESPACIAL DE EVENTOS	12
3.4 ANÁLISE DE PADRÕES DE DADOS COM REPRESENTAÇÃO POR PONTOS	14
3.5 ANÁLISES COM BASE NA DISTÂNCIA ENTRE PONTOS	18
3.6 ANÁLISE DA DENSIDADE DE PONTOS	24

3.7	INTERPOLAÇÃO DE DADOS PONTUAIS	27
4	MATERIAL E MÉTODOS	28
4.1	INTRODUÇÃO	28
4.2	SISTEMA Lat/long	29
4.3	SISTEMA UTM	30
5	ANÁLISE DOS RESULTADOS	33
5.1	INTRODUÇÃO	33
5.2	MODELO GERAL	34
5.3	ANÁLISE POR TIPO DE VEÍCULO	38
6	Conclusão	53
	Conclusão	53
	Referências Bibliográficas	55

Capítulo 1

INTRODUÇÃO

A Agência Nacional de Transportes Terrestres (ANTT) foi instituída através do Artigo 21 da Lei N° 10.233 de 5 de junho de 2001, e sua esfera de atuação é constituída por: I - o transporte ferroviário de passageiros e cargas ao longo do Sistema Nacional de Viação; II - a exploração da infra-estrutura ferroviária e o arrendamento dos ativos operacionais correspondentes; III - o transporte rodoviário interestadual e internacional de passageiros; IV - o transporte rodoviário de cargas; V - a exploração da infra-estrutura rodoviária federal; VI - o transporte multimodal; VII - o transporte de cargas especiais e perigosas em rodovias e ferrovias (ANTT, 2001).

O Sistema Nacional de Viação - SNV é constituído pela infra-estrutura viária e pela estrutura operacional dos diferentes meios de transporte (ANTT, 2001). O SNV possui alguns objetivos tidos como essenciais, tais como: I- Dotar o país de infra-estrutura viária adequada; II- Garantir a operação racional e segura dos transportes de pessoas e bens; III- promover o desenvolvimento social e econômico e a integração nacional. Dentro deste contexto, uma infra-estrutura viária adequada é aquela que torna o custo total do transporte mínimo, sendo este dado pela soma dos custos

de investimentos, manutenção e operação dos sistemas. Seguindo a mesma lógica, uma operação racional e segura é aquela que possui um gerenciamento eficiente das vias, dos terminais, dos equipamentos e dos veículos, com o objetivo de minimizar os custos operacionais. Tal ação impacta diretamente no preço dos fretes e tarifas, além de garantir a segurança e credibilidade do transporte. Ainda sim, mesmo com essa estrutura, um fator que pode influenciar o preço dos fretes é a localização espacial das empresas, ou seja, quanto mais distante da malha viária, maiores os custos de transporte.

Com base nas informações supracitadas, este trabalho buscará verificar qual a distribuição espacial das empresas do transporte rodoviário de carga no Distrito Federal, utilizando o banco de dados do RNTRC (Registro Nacional dos Transportadores Rodoviários de Carga), gentilmente cedido pela ANTT. Para tal, utilizará técnicas de Estatística Espacial como a análise pontual. Dessa forma, cada uma dessas empresas será representada por um ponto em um mapa, onde será analisado se eles estão aleatoriamente distribuídos, agrupados em aglomerados ou se estão regularmente distribuídos ao redor da malha viária.

1.1 OBJETIVOS

O objetivo geral do trabalho é analisar a localização das empresas de transporte rodoviário de carga para verificar se existe algum padrão de distribuição espacial dessas empresas no Distrito Federal.

Os objetivos específicos são:

- Utilizar estatística espacial, mais especificamente análise pontual de dados, para estudar a disposição das empresas no DF.
- Realizar as análises utilizando o *software* SAS 9.2 com a proc IML;
- Verificar se há relação entre a localização das empresas e o tipo de veículo utilizado para o transporte das cargas.

Capítulo 2

TRANSPORTE RODOVIÁRIO DE CARGA

2.1 INTRODUÇÃO

É entendido por Transporte Rodoviário, aquele feito através de estradas, rodovias e ruas. Na matriz do transporte de carga brasileira, o Transporte Rodoviário de Carga (TRC) representa 61,1% do total, enquanto o transporte ferroviário compõe 20,7%, o aquaviário 13,6%, o dutoviário 4,2% e o aéreo 0,4% (ANTT, 2010). Essa grande disparidade entre a percentagem do transporte rodoviário e os demais modos se deve a algumas vantagens apresentadas por esse meio de transporte, entre as quais pode-se citar a facilidade de embarque e desembarque da mercadoria no veículo transportador e o fato dele ser favorecido pelas políticas públicas.

O mercado do transporte rodoviário de carga é formado por usuários, montadoras, operadores e transportadores, sendo esse último grupo dividido em transportadores autônomos, empresas e cooperativas. Dessa forma, é fundamental que se possua informações referentes aos veículos que realizam o transporte, assim como sua distribuição espacial, pois essas informações podem ser usadas tanto por esse

mercado, para definir a estrutura de preços a ser utilizada, quanto pelo governo, na elaboração de uma política de regulação do setor.

Sendo o Transporte Rodoviário a forma mais utilizada para o transporte de cargas no Brasil, e conseqüentemente, um fator importante no desenvolvimento econômico e social do país, é interessante que se tenha conhecimento sobre a localização das empresas e dos clientes, não somente sua distribuição espacial, mas quais os fatores que influenciam as empresas a estarem onde estão. Outra informação fundamental é a localização das rodovias pelas quais os veículos dessas empresas trafegam, pois com tais dados é possível analisar se existe a possibilidade de serem feitas alterações nas rotas realizadas pelos veículos transportadores, ou realocação das empresas de tal forma que elas sejam beneficiadas com essas mudanças de alguma forma, como pela redução dos custos com o transporte ou gastando menos tempo no deslocamento.

2.2 LEGISLAÇÃO BÁSICA

Segundo a ANTT (2001), a lei que dispõe sobre o Transporte Rodoviário de Cargas realizado em vias públicas, no território nacional, por conta de terceiros e mediante remuneração, além dos mecanismos de sua operação e a responsabilidade do transportador é a lei N° 11.442, de 5 de Janeiro de 2007. A atividade econômica tratada por esta lei é de natureza comercial e pode ser exercida tanto por pessoas físicas, quanto por pessoas jurídicas em um regime de livre concorrência. Mas para essa atividade acontecer é necessário a inscrição no Registro Nacional de Transportadores Rodoviários de Carga (RNTRC) em alguma das duas categorias disponíveis : a de Transportador Autônomo de Carga (TAC) e a Empresa de Trans-

porte Rodoviário de Cargas (ETC). Considera-se por TAC a pessoa física que tenha no transporte rodoviário de cargas a sua atividade profissional; e por ETC a pessoa jurídica constituída por qualquer forma prevista em lei que tenha no transporte rodoviário de cargas a sua atividade principal.

O TAC deverá comprovar ser proprietário, co-proprietário ou arrendatário de, pelo menos, um veículo automotor de carga, registrado em seu nome no órgão de trânsito, como veículo de aluguel e deverá comprovar ter experiência de, pelo menos, três anos na atividade, ou ter sido aprovado em curso específico. Já a ETC deverá ter sede no Brasil, comprovar ser proprietária ou arrendatária de, pelo menos, um veículo automotor de carga, registrado no País, indicar e promover a substituição do Responsável Técnico, que deverá ter, pelo menos, três anos de atividade ou ter sido aprovado em curso específico além de demonstrar capacidade financeira para o exercício da atividade e idoneidade de seus sócios e de seu responsável técnico. Um ressalvo importante dentro da lei é que em todos os veículos automotores de carga, deverá constar o número de registro no RNTRC na forma a ser regulamentada pela (ANTT, 2001).

2.2.1 RNTRC

O Registro Nacional de Transportadores Rodoviários de Carga (RNTRC) foi instituído pela Lei 10.233, de 5 de junho de 2001, Arts. 14-A e 26, item IV, pela Lei nº 11.442, de 5 de janeiro de 2007 e pela Resolução nº 3056, de 12 de março de 2009, da ANTT, que dispõe sobre o exercício da atividade de transporte rodoviário de cargas por conta de terceiros e mediante remuneração, estabelecendo procedimentos

para inscrição e manutenção no RNTRC (ANTT, 2001).

Para se registrar no RNTRC, todas as Empresas de Transporte Rodoviário de Cargas, as Cooperativas de Transporte de Cargas e os Transportadores Autônomos de Cargas, que praticam atividade econômica de transporte rodoviário de cargas no Brasil, por conta de terceiros e mediante remuneração devem atender aos seus requisitos da Resolução nº 3056. O registro no RNTRC traz diversos benefícios não somente aos transportadores, ao regularizar o exercício de suas atividades por meio da habilitação formal, identificando parâmetros de participação no mercado e inibindo a participação de atravessadores não qualificados; mas também aos usuários, que passam a ter maior segurança ao contratar o transportador, maior informação sobre a oferta de transporte e redução de custos do seguro; e ao país, já que após o registro se torna mais fácil conhecer a oferta do transporte rodoviário de carga, identificar a distribuição espacial, composição e idade média da frota, delimitar as áreas de atuação dos transportadores e fiscalizar a atividade.

Devem estar registrados no RNTRC quaisquer veículos de carga que executem transporte rodoviário de carga mediante remuneração com capacidade de carga útil superior a 500kg, exclusive aqueles que sejam utilizados somente para transportar carga própria (classificados na categoria “particular”). Nas rodovias federais e nas concedidas à iniciativa privada, a Polícia Rodoviária Federal e fiscais da ANTT, respectivamente, serão responsáveis pela fiscalização do porte obrigatório do CRNTRC (Certificado de Registro Nacional de Transportador Rodoviário de Cargas).

2.3 CUSTOS DE TRANSPORTE

Os custos de transporte podem ser descritos como as despesas realizadas no transporte de cargas, desde a sua origem, até o seu destino final. Como em qualquer outra operação, os custos são divididos em fixos e variáveis. Esta classificação pode ser feita com relação a distância percorrida. Os custos variáveis são aqueles que podem variar de acordo com a quilometragem (distância) percorrida no transporte da mercadoria, enquanto os custos fixos, já são pré-definidos e não variam de acordo com a distância. São exemplos de custos fixos : depreciação, motorista e seguros. Já exemplos de custos variáveis são: Pneus, Combustível, Lubrificantes, Manutenção, Pedágio entre outros.

Algumas empresas de transporte de carga, podem não estar tão bem localizadas quanto outras, isso provavelmente implicaria em um maior tempo no embarque e desembarque da mercadoria e possivelmente acarretaria em um aumento no preço do frete. Se essas empresas estivessem localizadas mais próximas de uma Rodovia, o escoamento do frete seria mais veloz e talvez houvesse uma diminuição nos custos de transporte. Para identificar quais empresas podem estar “mal-localizadas” podemos usar técnicas de estatística espacial. Consideramos cada empresa como um ponto no mapa, e a partir disso, é possível a visualização de padrões, ou não, seguido pelos pontos no mapa. Assim identificaríamos quais empresas podem estar perdendo com a sua localização espacial.

Capítulo 3

ANÁLISE PONTUAL DE DADOS GEOGRÁFICOS

3.1 INTRODUÇÃO

A estatística espacial é um ramo da estatística que permite utilizar o espaço geográfico associado ao conceito de variáveis aleatórias, e assim, gerar um novo conceito de funções aleatórias, pois permite além de identificar, localizar e visualizar a ocorrência de fenômenos no espaço, modelar a ocorrência de tais fenômenos, adicionando, por exemplo, os fatores determinantes, a estrutura de distribuição espacial ou a identificação de padrões, ou seja, uma análise diferente da proporcionada pela estatística clássica. Na abordagem tradicional para dados sociais, econômicos e ambientais, utilizam-se técnicas estatísticas bem estabelecidas como teste de hipóteses, modelos lineares e análise de variância, que apesar de úteis e importantes, não levam em conta o espaço geográfico.

Segundo (Bailey e Gatrell, 1995), a análise espacial de dados é uma ferramenta que possibilita manipular dados espaciais de diferentes formas e extrair como resposta um conhecimento que não se tinha anteriormente, tendo como objetivo apro-

fundar a compreensão do processo, avaliar evidências de hipóteses a ele relacionadas, ou ainda, tentar prever valores em áreas onde as observações não estão disponíveis.

Dentro da análise espacial, a taxonomia mais comum para descrever problemas de análise espacial considera três tipos de dados. O primeiro tipo são os padrões pontuais, onde os fenômenos são apresentados através de ocorrências identificadas como pontos localizados no espaço, chamados de processos pontuais. O segundo tipo são as superfícies contínuas, que são estimadas a partir de um conjunto de amostras de campo, podendo estar regularmente ou irregularmente distribuídas, sendo que normalmente, esse tipo de dados decorre de levantamento de recursos naturais. E o terceiro tipo são as áreas, que estão associadas a levantamentos populacionais, e que originalmente, fazem referência a indivíduos localizados em pontos específicos do espaço, mas que por algum motivo não é possível obter os dados específicos desses pontos. No caso, o trabalho irá utilizar apenas a análise de Padrões Pontuais.

3.2 PRIMEIRA UTILIZAÇÃO DA ANÁLISE ESPACIAL

Em 1854, no trabalho “On the Mode of Communication of Cholera”, o médico britânico John Snow observou que os casos de cólera na cidade de Londres aconteciam em certas localizações com maior incidência que em outras, e resolveu estudar esse fenômeno. Para a realização desse estudo, ele decidiu representar o local onde aconteciam os casos por pontos no mapa da cidade. Snow localizou as residências em que haviam ocorrido casos de cólera através do registro de óbitos, utilizando o endereço desses locais e os poços de provisão de água presentes naquelas ruas do

bairro de Soho, localizado na zona sul de Londres (Figura 3.1).

Até então, a teoria dominante para a explicação da disseminação da cólera era a teoria miasmática, tese que afirmava que doenças como a cólera e a peste negra eram transmitidas pela poluição do ar ou da água, teoria à qual o próprio Snow era favorável. No entanto, após seu estudo, Snow identificou evidências que o levavam a crer que a transmissão da cólera não ocorria devido à poluição como se acreditava.

Ao conversar com moradores da região, identificou a fonte do surto como uma bomba de água em *Broad Street*. Após essa descoberta, foram feitos exames microscópicos de uma amostra de água dessa bomba, e embora não houvessem evidências que levassem a acreditar que essa água era a fonte da contaminação, o estudo realizado por Snow utilizando a análise pontual de dados foi suficiente para as autoridades locais, que ordenaram o fechamento da bomba. Em consequência disso, e também do fato de muitas pessoas terem saído da área com medo da contaminação, a mortalidade por cólera na região foi reduzida drasticamente.

Mais tarde, John Snow divulgou esse estudo e mostrou em um mapa como as residências em que haviam ocorrido os casos de cólera se agrupavam ao redor dessa bomba. Ele também usou a análise pontual para elaborar um mapa mais amplo abrangendo toda a cidade de Londres, ilustrando a relação entre a qualidade da fonte de água e a incidência de casos de cólera, que era maior em localidades próximas a fontes de água de qualidade inferior.

Esse estudo foi inovador não somente nos campos da medicina, biologia e epidemiologia, mas também na área da estatística, pois proporcionava uma nova abordagem

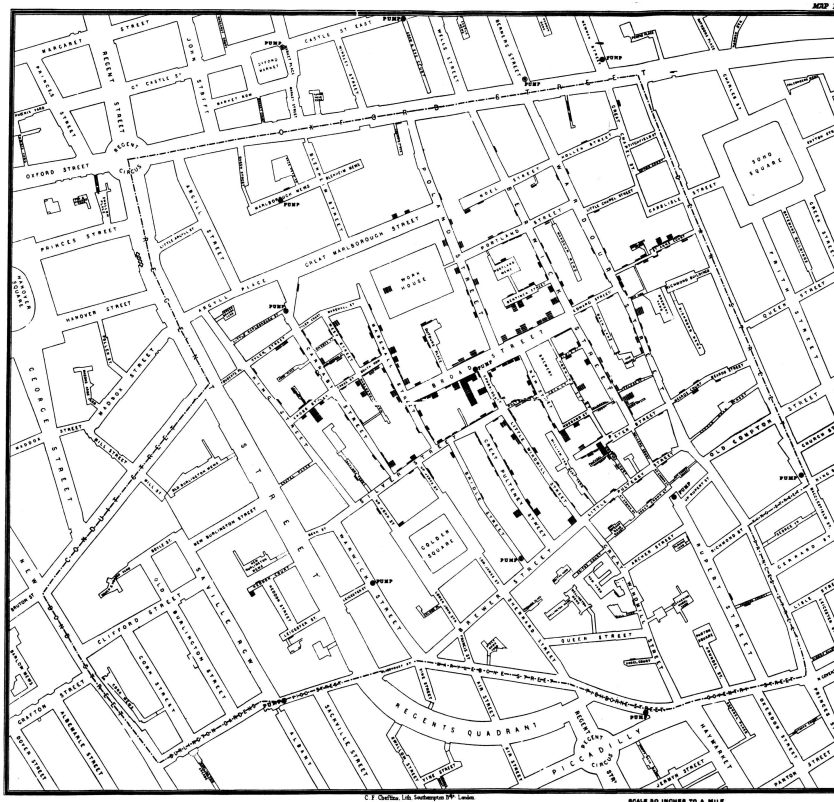


Figura 3.1: Mapa da região de *Soho* elaborado por John Snow

que permitia utilizar variáveis geográficas para realizar análises estatísticas, sendo considerado o primeiro caso de utilização da análise espacial.

3.3 ANÁLISE ESPACIAL DE EVENTOS

A análise espacial de eventos, segundo Bailey e Gatrell(1995), pode-se distinguir em três grupos de métodos. São eles:

Visualização de Dados Espaciais: A visualização gráfica é fundamental na análise espacial, por isso seu objetivo principal é mostrar graficamente os dados espaciais para que assim seja possível identificar padrões espaciais nos dados, gerando hipóteses testáveis, avaliando o ajuste de modelos propostos ou então a validade das previsões resultantes. A forma mais usual de tais dados serem apresentados é

através de mapas. Com utilização de ferramentas de SIG (Sistema de Informação Geográfica), a elaboração desses mapas se tornou mais simples e rápida. Exemplos disso são os dados pontuais, que apresentam uma fácil interpretação quando representados em um mapa, ou dados de área representados em um mapa coroplético, onde uma determinada superfície estatística é representada por uma área gráfica primitiva no mapa, semelhante a estrutura de dados “polígono” utilizada no SIG.

Análise Exploratória de Dados: Esse método busca uma descrição dos eventos, tentando gerar hipóteses e modelos apropriados através da aplicação de métodos exploratórios de dados espaciais. Para tal finalidade, são usadas ferramentas do SIG para medir a distância entre os pontos ou calcular densidades e assim descrever de forma mais clara os padrões espaciais. Com essa análise pode-se obter resultados na forma de um mapa ou gráficos mais convencionais, mas geralmente, busca-se como resultado a visualização gráfica, pois dessa forma é possível identificar padrões, inter-relações e observações incomuns.

Modelagem de Dados Espaciais: O objetivo da modelagem é especificar modelos estatísticos e estimar parâmetros que permitam realizar uma análise precisa dos dados obtidos. Nesse método há grande preocupação com fenômenos estocásticos, que são influenciados pelas leis de probabilidade. Dessa forma, são utilizadas ferramentas que consideram a probabilidade de um evento ocorrer em relação a um conjunto de outros eventos existentes, criando-se assim uma nuvem de pontos.

Esses três conceitos são diferentes, e essa diferenciação é fundamental. No entanto, devido à inter-relação existente entre esses métodos que estão ligados em um

processo interativo, eles não possuem delimitações bem definidas. Essas abordagens estão acompanhadas de interpretações intuitivas, e são basicamente uma maneira de determinar uma primeira impressão da área e de algum padrão que possa ser observado. Entretanto, é necessário atenção pois essa interpretação intuitiva do que pode ser um padrão aleatório pode ser ambígua, como foi explicado por (Bailey e Gatrell, 1995).

Atualmente, é comum a utilização de métodos de visualização e exploratórios em conjunto, o que é denominado pelo termo visualização-exploratória. Essa combinação é muito usada devido à semelhança entre esses métodos, e nela são empregados procedimentos da visualização que mapeam dados espacialmente georreferenciados, com os procedimentos exploratórios, analisando os dados e buscando observar algum padrão. Esse método da visualização-exploratória é o mais utilizado, porém, deve-se sempre utilizar na análise o método mais apropriado ao caso que está sendo avaliado.

3.4 ANÁLISE DE PADRÕES DE DADOS COM REPRESENTAÇÃO POR PONTOS

Os processos pontuais são definidos como uma base de dados representados por um conjunto de pontos irregularmente distribuídos em um terreno, ou seja, representam uma série de localizações em um mapa. Segundo (Bailey e Gatrell, 1995), são a forma mais simples de representar dados espaciais, e em sua forma primária, esses dados espaciais representam apenas as coordenadas do evento estudado.

Ao se realizar a análise pontual, o que se deseja saber é se existe algum padrão

espacial de distribuição desses pontos. Após obter o mapa com a localização dos pontos em seu território, existem três tipos de distribuição esperados: aleatório, regular e aglomerado.

No padrão regular, a distância média entre os pontos tende a ser constante, e ao se analisar o mapa de forma geral, é possível notar uma regularidade no espaçamento entre os pontos, como se eles estivessem espalhados propositadamente para que não houvesse uma grande diferença na distância entre eles. Esse tipo de padrão é muito comum ao se analisar, por exemplo, a localização de pontos de ônibus em uma cidade, que tendem a ser igualmente espaçados para que moradores de todas as áreas tenham fácil acesso à rede de transporte público, ou a presença de unidades básicas de saúde em um município.

Outro tipo de padrão é o aglomerado, onde se observa a concentração dos pontos em uma determinada área em detrimento de outras. Ao observar-se um mapa que apresente esse padrão, também chamado de agrupado ou clusterizado, estudos podem ser feitos para identificar a causa da aglomeração no local. Dependendo do estudo realizado, é possível também controlar os fatores causadores da aglomeração, como foi o caso do estudo de John Snow, que observou uma aglomeração nos casos de cólera em uma região de Londres e assim pôde tomar providências para a diminuição de incidência da doença.

O padrão aleatório é observado quando os pontos estão localizados no território de tal maneira que não apresentem nenhum modelo de distribuição. Eles estão distribuídos ao acaso e não são observados locais que apresentem uma maior concen-

tração que possa justificar o padrão aglomerado, nem uma uniformidade na distância entre esses pontos que justifiquem o padrão regular.

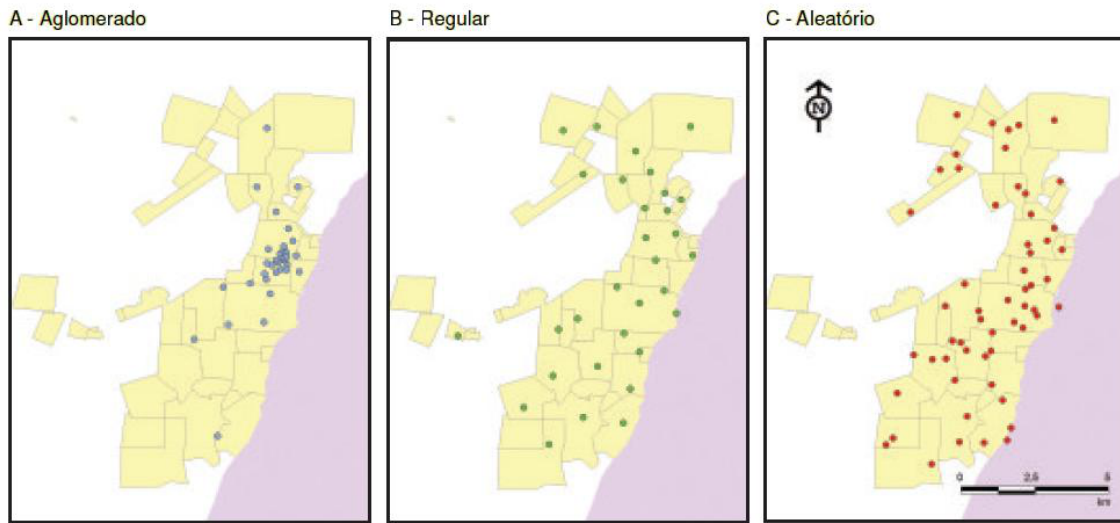


Figura 3.2: Tipos de Distribuição Espacial dos Pontos

Segundo (Druck et al, 2004) para caracterizar os processos pontuais, pode-se descrevê-lo em termos dos efeitos de primeira ordem e dos efeitos de segunda ordem. Os efeitos de primeira ordem, também chamados de efeitos globais, são as variações no valor médio do processo no espaço, ou seja, o interesse aqui é a intensidade do processo, a quantidade de eventos por unidade de área. Já os efeitos de segunda ordem, também conhecidos por efeitos locais, correspondem a dependência espacial no processo, que é oriundo da correlação espacial. A dependência espacial é medida através da estimação do relacionamento entre pares de eventos no espaço. Em uma região A onde ocorreram eventos, considera-se um conjunto de pontos (u_1, u_2, \dots) . O processo pontual é modelado considerando sub-regiões S em A através de sua esperança $E[N(S)]$ e a covariância $C[N(S_i), N(S_j)]$, onde $N(S)$ denota o número de eventos em S .

Como o objetivo da análise é estimar as prováveis localizações de ocorrência de eventos, as estatísticas devem ser inferidas considerando o valor limite da quantidade de eventos por área. O valor limite corresponde à esperança de $N(S)$ para uma região d_u em torno do ponto u , quando ela tende a zero. A esperança é denominada intensidade (Propriedade de Primeira Ordem), sendo definida por :

$$\lambda(u) = \lim_{|du| \rightarrow 0} \left\{ \frac{E[N(du)]}{|du|} \right\} \quad (3.1)$$

Já as propriedades dos efeitos locais (Propriedades de Segunda Ordem), são definidas da mesma forma do que as de Primeira Ordem, mas agora considera-se a intensidade conjunta $\lambda(u_i, u_j)$ entre duas regiões infinitesimais $|du_i|, |du_j|$ que contém os pontos u_i e u_j .

$$\lambda(d(u_i), d(u_j)) = \lim_{du_i, du_j \rightarrow 0} \left\{ \frac{C[N(du_i), N(du_j)]}{du_i, du_j} \right\} \quad (3.2)$$

Para analisar estatisticamente os padrões de distribuição de pontos, é necessário tomar como referência, um modelo teórico que servirá de base para o desenvolvimento de métodos que irão checar os resultados apurados. O modelo teórico que é mais aplicado é o de aleatoriedade espacial completa (CSR). O modelo em questão, divide a região de estudo A em subáreas S_i e modela a distribuição de eventos de pontos como um processo aleatório

$$\{Z_i(u_i), u_i \in S_i : i = 1, \dots, n\} \quad (3.3)$$

sendo que $Z_i(u_i)$ é o número de casos que ocorreram na subárea S_i .

Neste modelo, considera-se que a posição de cada caso ocorrido é independente e os casos tem probabilidade de ocorrência igual, em toda a região de estudo. Dessa

forma podemos comparar uma distribuição com aleatoriedade total com os dados apurados.

3.5 ANÁLISES COM BASE NA DISTÂNCIA ENTRE PONTOS

Uma das formas de avaliar os padrões espaciais de pontos e notar a presença de aglomerados é medindo a média da distância entre os pontos. No padrão aglomerado, a distância média entre os pontos é pequena e nos outros padrões (Aleatório e Regular) a distância entre pontos é grande. Essas distâncias entre os pontos, são facilmente calculadas em um SIG, já que este sistema guarda as coordenadas de cada ponto. Pode-se calcular a distância euclidianda utilizando a equação:

$$d = \sqrt{((x_1 - x_2)^2 + (y_1 - y_2)^2)} \quad (3.4)$$

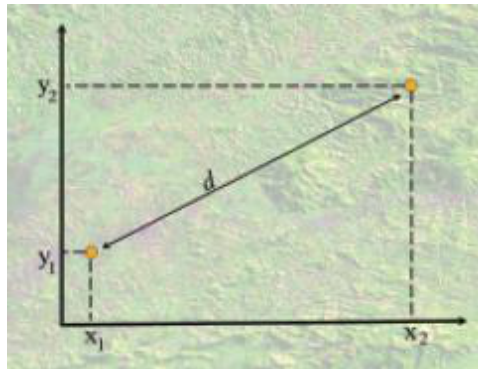


Figura 3.3: Demonstração do cálculo da distância euclidiana

Pode-se observar as estatísticas de distância entre os pontos da 3.2 na Tabela 3.1. Nota-se que o padrão regular é o que apresenta a maior distância média e o menor desvio padrão da distância, já que espera-se que a distribuição desse padrão seja mais homogênea dentre os demais. A menor distância média é encontrada

Tabela 3.1: Tabela com as distâncias e desvios padrão

	DISTÂNCIA MÉDIA	DESVIO PADRÃO
Aglomerado	358	688
Regular	1094	321
Aletório	890	643

no padrão aglomerado, já que seus pontos estão mais concentrados na região de estudo, por sua vez apresenta maior desvio padrão entre os três padrões analisados. No padrão aleatório encontra-se alto desvio padrão e alta distância média entre os pontos.

As estatísticas de distância entre os pontos servem para uma série de técnicas utilizadas para a detecção de aglomerados de pontos. De modo geral a análise de aglomerados é um método de classificação que utiliza a característica de certos objetos com o intuito de agrupá-los (Bailey e Gatrell, 1995).

Uma das formas de análise de aglomerados é usando as técnicas de aglomerados hierárquicos. Nela, a detecção de aglomerados começa com os eventos separados e procede-se a união das observações mais semelhantes em grupos, com base em critérios pré-definidos (Santos, 2007). Na análise hierárquica de aglomerados, inicialmente a menor distância entre dois pontos é identificada e esses pontos são ajuntados em um novo aglomerado. Após isso, as distâncias entre os aglomerados também são calculadas e os mesmos podem ser reagrupados em um novo aglomerado e assim por diante.

São vários métodos hierárquicos que podem ser utilizados. O método mais usual é o de vizinhos mais próximos, que agrupa os pontos que estão mais próximos na área de estudo. Identifica-se aglomerados vizinhos de primeira ordem, que são os

que representam os pontos mais próximos do que a distância limiar entre os pontos proposta e que tenham no mínimo a quantidade mínima de pontos pré-estabelecida. Estes aglomerados são tidos como pontos individuais e podem ser reagrupados em aglomerados de segunda ordem e assim por diante até atingir-se um aglomerado singular ou não atender mais as condições pré-estabelecidas.

Este método estima a função de distribuição cumulativa $\hat{G}(h)$ baseado nas distâncias h entre os eventos que ocorreram na região de estudo. A função de distribuição pode ser estimada empiricamente por :

$$\hat{G}(h) = \frac{\#(d(u_i, u_j) \leq h)}{n} \quad (3.5)$$

onde o valor normalizado acumulado para uma distância h , corresponde à soma dos vizinhos mais próximos de cada evento cuja distância é menor ou igual a h , dividido pelo número de eventos naquela região, representado por n (Bailey e Gatrell, 1995).

Ao colocar em um gráfico os resultados desta função de distribuição cumulativa empírica $\hat{G}(h)$ pode-se usá-los como um método exploratório para verificar a evidência de interação dos eventos em estudo. Se o gráfico apresentar um rápido crescimento para valores pequenos de distância, esta situação mostra que há interação entre os eventos caracterizando agrupamentos nestas faixas de distâncias. Se o gráfico apresentar valores pequenos no começo e só crescer aceleradamente para valores maiores de distância, esta situação mostra uma distribuição mais regular dos eventos.

A análise de vizinhança é usada para comparar estatisticamente a distribuição dos eventos observados com a hipótese de aleatoriedade espacial completa (CSR).

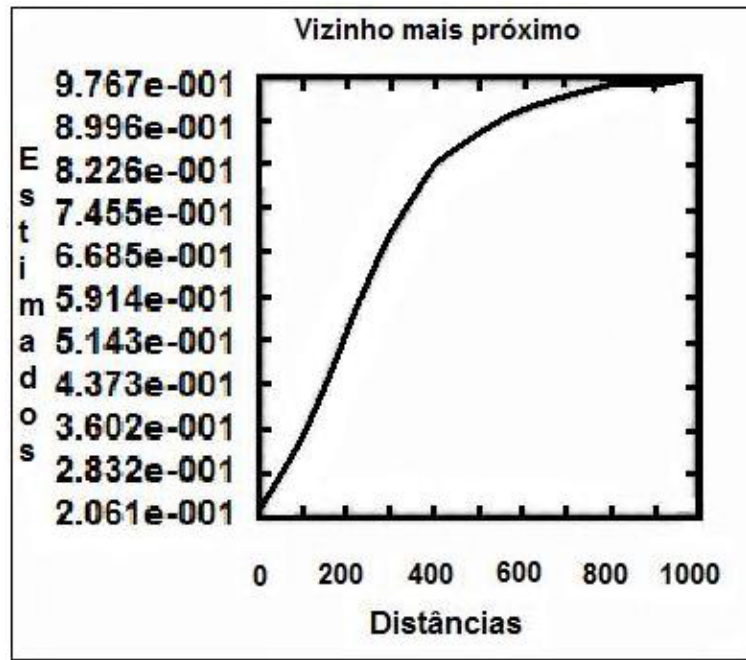


Figura 3.4: Vizinho mais próximo

Para isso, cria-se envelopes de simulação para a distribuição CSR a fim de se acessar a significância dos desvios (Druck et al, 2004). Na hipótese de aleatoriedade espacial completa a função de distribuição $G(h)$ seria dada por um processo de Poisson:

$$G(h) = 1 - e^{-\lambda h^2}, h \geq 0 \quad (3.6)$$

A estimação simulada para a distribuição $G(h)$ assumindo aleatoriedade espacial completa é dada por:

$$\bar{G}(h) = \frac{\sum_i^k \hat{G}_i(h)}{k} \quad (3.7)$$

onde $G_i(h), i = 1, 2, \dots, k$, são funções de distribuição empíricas estimadas a partir de k simulações independentes dos n eventos, na hipótese de CSR (Druck et al, 2004).

Para testar se existe ou não condição de aleatoriedade, calcula-se ainda os en-

velopes de simulação superiores e inferiores, definidos por:

$$U(h) = \max\{\hat{G}_i(h)\}, i = 1, \dots, k$$

$$L(h) = \min\{\hat{G}_i(h)\}, i = 1, \dots, k$$

O gráfico da distribuição estimada $\bar{G}(h)$, da distribuição simulada $G(h)$ e dos envelopes de simulação superior e inferior, torna possível medir a significância dos desvios referentes a aleatoriedade. Se a condição CSR for válida para os eventos estudados, o gráfico da curva de $G(h)$ versus $\bar{G}(h)$ deve ser praticamente linear com um ângulo de 45° . Se os eventos mostram-se tendenciosos ao agrupamento, as linhas do gráfico estarão acima da linha de 45° , e se mostrarem uma tendência ao regularismo, estarão abaixo.

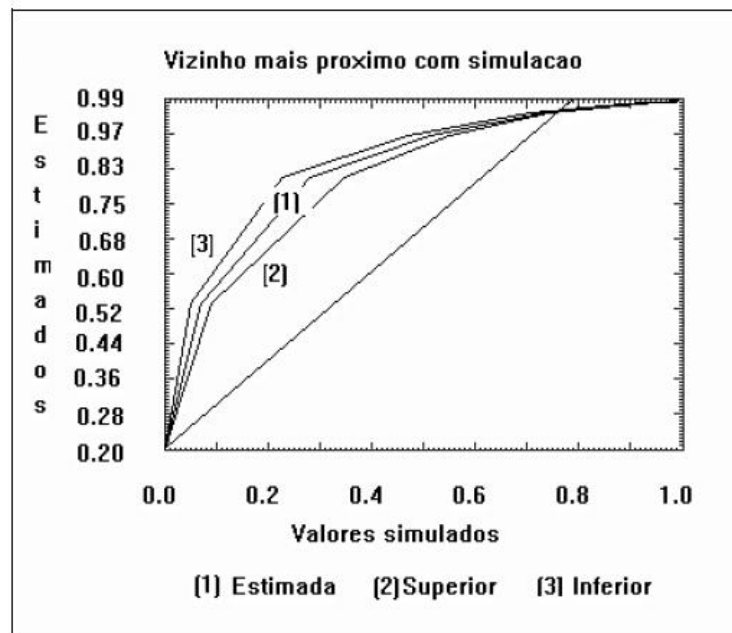


Figura 3.5: Vizinho mais próximo com simulação

Apesar de ser uma bom método para fornecer indicações iniciais da distribuição espacial, o método de vizinho mais próximo considera apenas escalas de distância

pequenas. Se o caso em questão possuir escalas de distância maiores, é melhor trabalhar como a Função K .

A função K é definida para o processo univariado como:

$$\lambda K(h) = E(\# \text{ eventos contidos a uma distância } h \text{ de um evento arbitrário}) \quad (3.8)$$

onde $E()$ é o operador da esperança e λ é o número médio de eventos por unidade de área, assumida constante na região. Uma estimativa de $K(h)$ é :

$$\hat{K}(h) = \frac{A}{n^2} \sum_i^n \sum_{j, i \neq j}^n \frac{I_h(d_{ij})}{w_{ij}} \quad (3.9)$$

onde A é a área da região, n é o número de eventos observados, $I_h(d_{ij})$ é uma função indicatriz cujo valor é 1 se $(d_{ij}) \leq h$ 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 utilizada de forma semelhante ao método do vizinho mais próximo, ou seja, compara-se uma estimativa empírica $\hat{K}(h)$ e a resultante de um processo de padrão de pontos espacial aleatório $\bar{K}(h)$. Para um processo que segue aleatoriedade $K(h)$ seria πh^2 . Dessa forma, pode-se comparar a estimativa K de um conjunto de dados apurados com πh^2 representando no gráfico a função $\hat{L}(h)$ que é definida como:

$$\hat{L}(h) = \sqrt{\frac{\hat{K}(h)}{\pi}} - h \quad (3.10)$$

Segundo (Druck et al, 2004) a plotagem da função $\hat{L}(h)$ em função da distância h , indica atração espacial entre eventos ou agrupamentos para valores positivos e indica repulsão espacial ou regularidade para valores negativos. Da mesma forma que no método do vizinho mais próximo, pode-se estimar a significância dos desvios

da distribuição $\hat{L}(h)$ em relação à CSR. Os envelopes superior e inferior também são calculados da mesma forma do que a do método do vizinho mais próximo, ou seja, a partir de k simulações de n eventos na região A. A partir daí a análise dos envelopes e do gráfico da função é a de que a presença de picos positivos que se situam acima do envelope superior evidencia ocorrência de agrupamentos e a presença de depressões negativas que situam-se abaixo do envelope inferior evidencia ocorrência de regularidade.

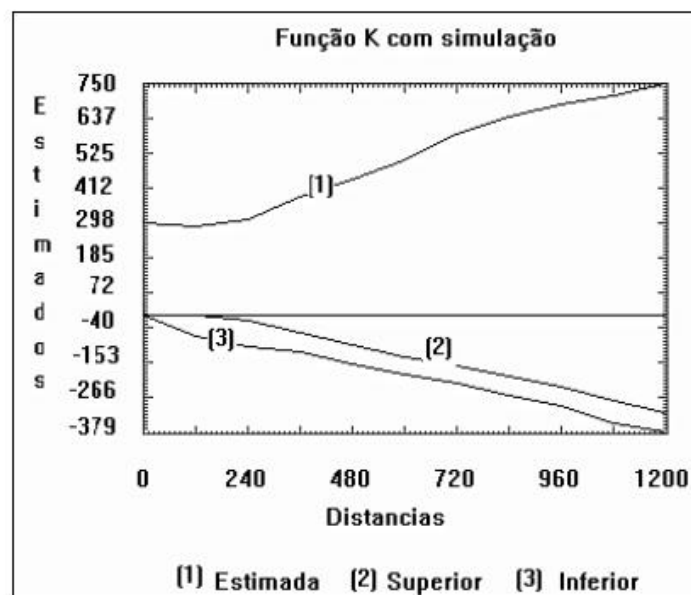


Figura 3.6: Função K

3.6 ANÁLISE DA DENSIDADE DE PONTOS

Existem vários procedimentos de estimação da densidade de eventos, sendo a estimativa Kernel a mais conhecida e utilizada. Essa estimativa cria uma superfície de densidade que permite a identificação visual de áreas quentes, através de uma técnica de interpolação exploratória. As áreas quentes são definidas como regiões onde há

concentração de eventos indicando que de alguma maneira exista aglomeração na distribuição espacial.

A estimativa de Kernel é uma técnica estatística, não paramétrica, onde uma determinada distribuição de pontos em um mapa é transformada em uma superfície contínua de risco para sua ocorrência. Dessa forma, é possível filtrar a variabilidade de um conjunto de dados sem alterar suas características locais, como foi explicado por (Bailey e Gatrell, 1995). Esse estimador não é apenas uma forma de identificar aglomerados, mas um método que permite a exploração e demonstração do padrão de pontos apresentado, o que se mostra útil após a criação da superfície contínua baseada nos dados pontuais.

A superfície criada pela estimativa de Kernel possui densidades calculadas para todas as localizações, o que facilita o trabalho de quem a estiver utilizado, pois não é necessário que se defina de forma aleatória onde dividir a área quente para que nela esteja presente o aglomerado. Outro motivo pelo qual o estimador de Kernel é utilizado é sua utilidade ao fornecer uma visão geral da distribuição de primeira ordem dos eventos, já que permite uma visualização das áreas que necessitam mais atenção e não é afetado pelas divisões político-administrativas.

A estimativa de Kernel foi desenvolvida inicialmente com o intuito de se obter uma estimativa suavizada de uma probabilidade univariada ou multivariada baseado na amostra de uma observação, ou um histograma suavizado. As estimativas de intensidade de uma distribuição de pontos e de probabilidade de densidade bivariada são muito semelhantes.

É necessário que se defina dois parâmetros básicos para a aplicação da estimativa de Kernel. O primeiro é o raio de influência, representado por τ . Esse parâmetro define a vizinhança do ponto que será interpolado e controla a suavidade da superfície que será criada. Esse raio define um disco centrado em S , que pe uma localização na região R , na qual os ponto s_i vão contribuir para a estimativa da função de intensidade. O outro parâmetro é a função de estimação K (Kernel), utilizada para suavizar o fenômeno. Os Kernel mais utilizados são o normal e o quártico. Com base nesses parâmetros, se S representa uma região em R e (s_1, \dots, s_n) representam as localizações das n observações, então a intensidade estimada λ em S é dada por:

$$\lambda_\tau(S) = \sum_{i=1}^n \frac{1}{\tau^2} K \left[\frac{(S - S_i)}{\tau} \right] \quad (3.11)$$

Nessa equação, K é a função de alisamento, uma função de densidade bivariada, conhecida como Kernel, previamente escolhida. O parâmetro τ como dito previamente, é o raio de influência, fator que determinará o grau de suavização da superfície obtida. A superfície será criada a partir do ajustamento da função bidimensional sobre os eventos em questão, sendo que ela apresentará um valor proporcional à intensidade dos eventos por unidade de área . Enquanto a função Kernel (normal, quadrática, uniforme ou triangular) não é um ponto crítico, o raio de influência escolhido pode causar alterações significativas na estimativa final (Bailey e Gatrell, 1995).

A função apresentada tem por objetivo contar os eventos dentro de uma região de influência, ponderando para cada ponto, a distância entre o ponto e uma localização de interesse. Ao se aplicar o estimador de intensidade em uma grade de localizações

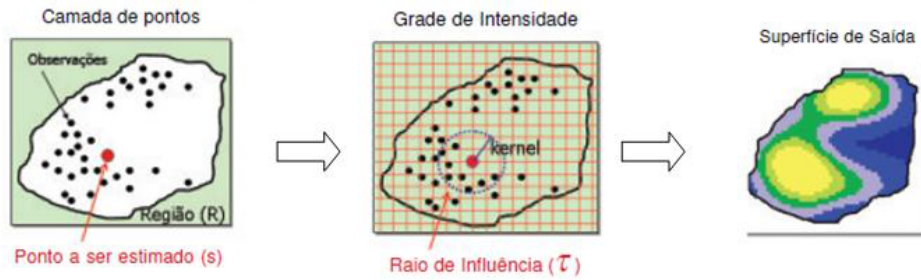


Figura 3.7: Criação da superfície

presente em uma determinada região R , é como se cada ponto s da grade fosse representado numa função tridimensional. A partir disso, são medidas as distâncias para cada ponto s_i presente na área de influência, e elas contribuem para a estimativa de intensidade em S de acordo com a sua distância a S . A criação da superfície de intensidade é ilustrada na 3.7

3.7 INTERPOLAÇÃO DE DADOS PONTUAIS

A técnica de interpolação de dados pontuais é comumente aplicada quando se quer comparar dois processos espaciais. Geralmente ela é utilizada quando os dados de população não estão disponíveis, ou quando se trabalha com amostras. No caso de trabalho com amostras, é possível comparar a distribuição espacial dos casos com o total das amostras. As interpolações de dados pontuais são importantes para ponderar a distribuição espacial de um processo contínuo no espaço, baseadas em amostras pontuais deste processo (Bailey e Gatrell, 1995). Isso ocorre principalmente em processos ambientais, onde espera-se que eles se alterem no espaço e possuam uma forte vinculação espacial.

Capítulo 4

MATERIAL E MÉTODOS

4.1 INTRODUÇÃO

Para a realização do estudo acerca da localização das empresas de transporte rodoviário no Distrito Federal, e posterior verificação da existência de algum padrão de distribuição espacial, é necessário que se possua alguma fonte que indique onde estas empresas estão instaladas. Tais informações foram obtidas através do banco de dados do RNTRC (Registro Nacional dos Transportadores Rodoviários de Carga), gentilmente cedido pela ANTT.

No banco de dados do RNTRC, existem variáveis indicando qual o tipo de transportador, a Unidade da Federação (no caso, apenas o Distrito Federal), o CEP da empresa, a quantidade de transportadores, o número do transportador, categoria e a quantidade de veículos.

A variável tipo de transportador mostra se o responsável pelo transporte é uma pessoa física, uma empresa ou cooperativa. Nesse trabalho, foram desconsideradas as pessoas físicas, sendo utilizadas apenas as pessoas jurídicas (empresas e cooperativas). O número do transportador é o registro do transportador junto à ANTT. A

categoria indica em que grupo o veículo se enquadra, de acordo com o peso máximo que ele transmite ao pavimento, e a quantidade de veículos mostra qual o tamanho da frota de cada empresa.

No entanto, a variável de maior interesse inicialmente nesse banco de dados é o CEP, já ele que é de fundamental importância para a obtenção da localização das empresas estudadas. Através de uma MACRO implementada no *software* SAS 9.2, o valor presente no campo CEP será utilizado e juntamente com informações geográficas de latitude e longitude, será obtida a localização das empresas no mapa do DF. Essas informações geográficas serão obtidas através do sistema Lat/Long.

4.2 SISTEMA Lat/long

O sistema de coordenadas geográficas (lat/long), é o sistema referencial de localização terrestre. Esse sistema baseia-se em valores angulares que são expressos em graus, minutos e segundos de latitude ou de longitude. As latitudes (Paralelos) são as linhas imaginárias E-W (leia-se Leste-Oeste) paralelas a linha imaginária do Equador e as Longitudes (Meridianos) são as linhas imaginárias N-S (leia-se Norte - Sul) , passando pelos polos, que correspondem a intersecção da superfície terrestre com planos hipotéticos contendo o eixo de rotações terrestre (IBGE, 2001).

No sistema de latitudes, o referencial zero usado é a linha imaginária do Equador, e a partir dele os valores angulares crescem para o Norte e Sul até 90° , sendo que cada grau se divide em 60 Minutos e cada minuto se divide em 60 segundos. Já no sistema de longitudes o referencial zero utilizado é o meridiano de Greenwich e a partir dele os valores crescem para Leste e Oeste até 180° . Os valores dos graus, minuto e

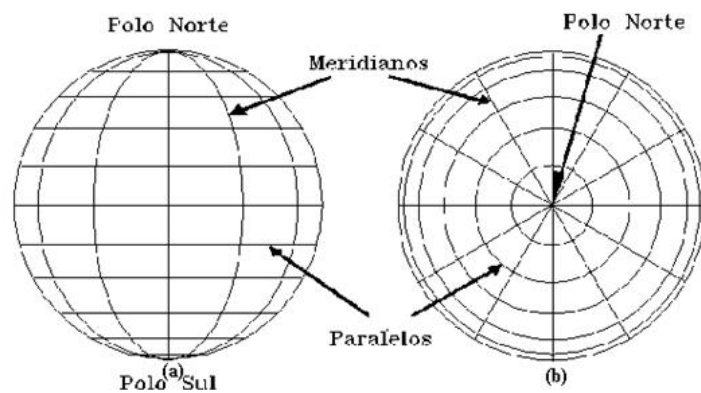


Figura 4.1: Funcionamento do Sistema Lat/Long

segundos das latitude e longitudes são diferenciados colocando-se as letras N, S, E ou W a medida que estas variam para Norte,Sul, Leste ou Oeste, respectivamente.

4.3 SISTEMA UTM

O Sistema de Coordenadas UTM (Universal Transversa de Mercator) baseia-se no plano cartesiano e utiliza o metro como unidade elementar para medir as distâncias e dessa forma definir a posição de um objeto no espaço. A diferença básica entre os Sistemas UTM e o Lat/Long (Sistema de Coordenadas Geográficas) é que o primeiro não acompanha a curvatura da terra, dessa forma podemos nos referenciar a esse sistema como Sistema de Coordenadas Planas. O sistema UTM foi adotado no Brasil como o padrão pelo Exército e pelo IBGE.

O Sistema é formado por 60 fusos, cada fuso possui 6° de longitude e são numerados a partir do fuso 180° a 174° Gr. e segue para leste até o encontro com o meridiano central de Greenwich. Já a extensão latitudinal do sistema fica entre 80° Sul e 84° Norte. As origens do Sistema UTM são as mesmas do Sistema Lat/Long: Meridiano Central e Linha do Equador.



Figura 4.2: Utilização do plano cartesiano

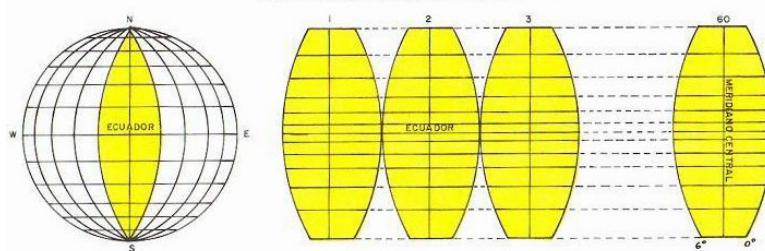


Figura 4.3: Linha do Equador em UTM

Cada fuso no Sistema UTM possui um sistema de coordenadas plano e único, com valores repetidos ao longo de todos os fusos, ou seja, para saber a localização de um ponto no Sistema UTM, é preciso saber quais são as suas coordenadas, e a qual fuso elas pertencem. No sistema UTM não encontra-se coordenadas negativas porque atribui-se valores de 500 Km para a contagem das coordenadas que variam ao longo do Equador e 10.000 Km para a contagem das coordenadas que variam ao longo do meridiano Central (Figura 4.4).

Por ser um país com grandes extensões longitudinais, o Brasil possui 8 fusos

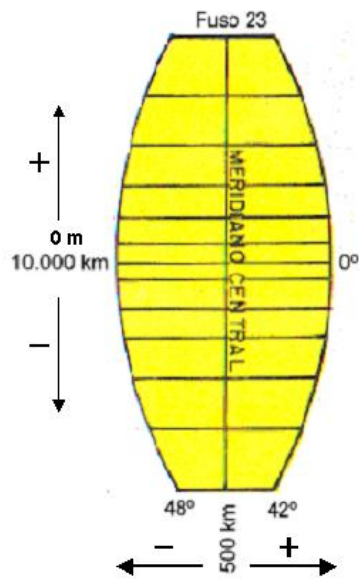


Figura 4.4: Meridiano Central em UTM

UTM, que vão do fuso 18 ao fuso 25(Figura 4.5).



Figura 4.5: Fusos existentes no Brasil

Capítulo 5

ANÁLISE DOS RESULTADOS

5.1 INTRODUÇÃO

Este capítulo tem por base analisar os resultados obtidos através de um programa implementado no *software* SAS que permite obter a coordenada (latitude e longitude) de cada empresa a partir do CEP. Dessa forma é possível obter a localização exata desse ponto e pode-se representá-lo no mapa do DF. Após a representação desses pontos, será possível realizar uma série de análises espaciais, dentre elas, o método do vizinho mais próximo, a análise de vizinhança, a função K de Ripley e a superfície de intensidade para observar como a distribuição desses pontos se comporta. As análises serão úteis para se chegar a uma conclusão a respeito da distribuição espacial das empresas de transporte de carga do Distrito Federal e do tipo de caminhão utilizado pelas empresas.

Tendo por base o mapa com a localização das empresas de TRC no Distrito Federal, representado na Figura 5.1, deseja-se agora saber se a distribuição desses pontos no território é aleatória, regular, ou se há existência de aglomerados, além de verificar se as distribuições por tipo de veículo apresentam alguma relação com a

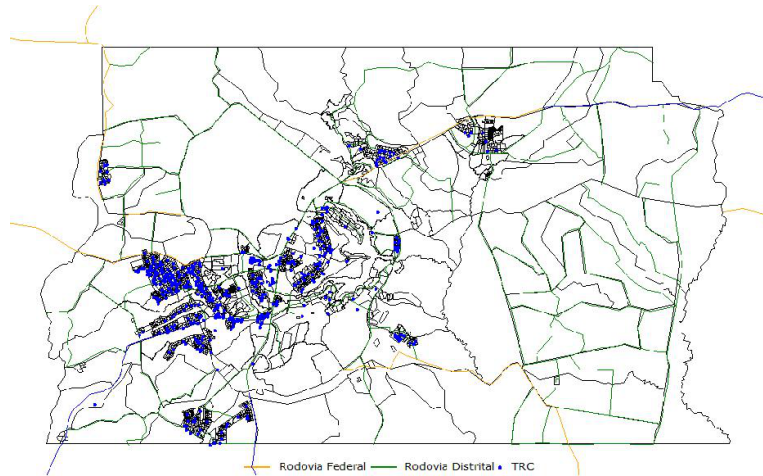


Figura 5.1: Mapa do Distrito Federal com representação das empresas de TRC

localização das Rodovias Brasileiras.

5.2 MODELO GERAL

O modelo geral é aquele em que não houve diferenciação dos tipos de veículo, ou seja, as análises foram feitas considerando todas as empresas, independente de qual a categoria dos veículos que ela possuía. Dessa forma, foram utilizadas todas as 1002 empresas, que possuem um total de 1646 veículos.

Aplicando a esses pontos o estimador Kernel para diferentes valores de τ (5, 10 e 20 quilômetros) obtém-se os mapas da Figura 5.2.

Verifica-se que existe uma concentração das empresas nas regiões de Taguatinga e Ceilândia, mas não se pode afirmar de maneira inequívoca que exista aglomeração. Para isso pode-se utilizar o método do vizinho mais próximo. Ao representar graficamente os resultados do método do vizinho mais próximo, pode-se utilizá-lo para verificar a existência de interação dos eventos em questão da seguinte forma: quando o gráfico apresentar um crescimento rápido pra valores pequenos de distância, mostra

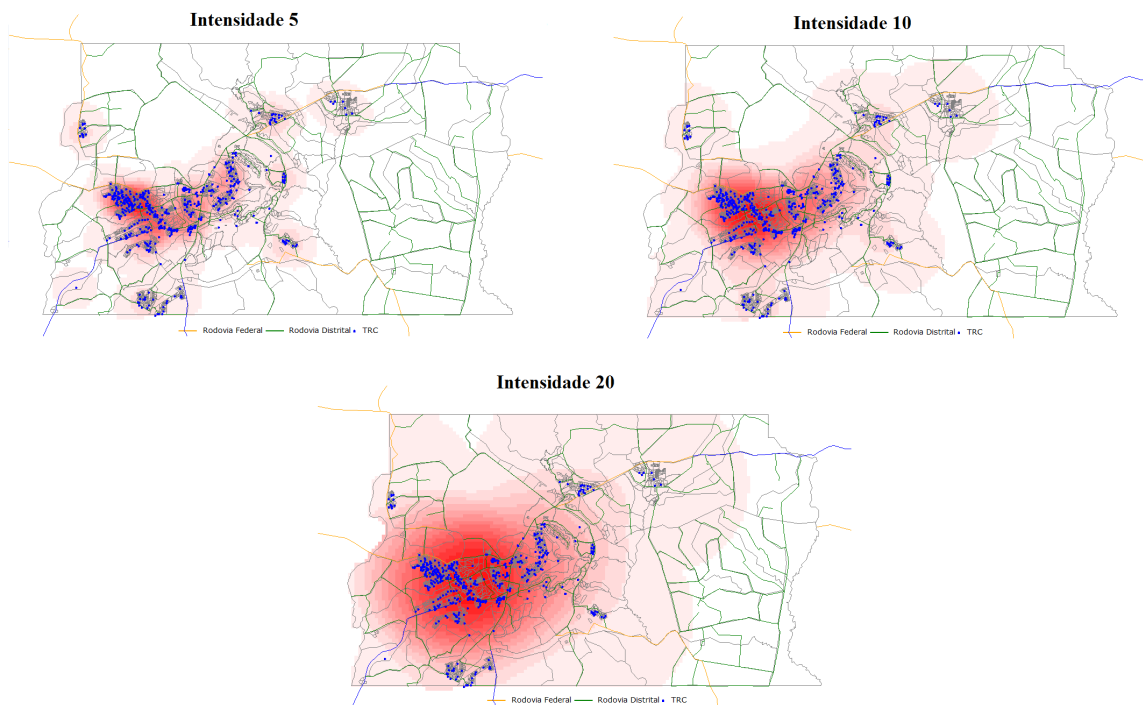


Figura 5.2: Estimadores de intensidade com diferentes valores de τ

interação entre os eventos, caracterizando agrupamentos nessa faixa de distância. Esse método está representado na Figura 5.3.

O gráfico do método do vizinho mais próximo mostra uma aglomeração a partir de distâncias h pequenas, em torno de 5 quilômetros, estabilizando-se em 10 quilômetros. Isso significa dizer que para essa faixa de distância, é possível afirmar que existe aglomeração. Para verificar se a distribuição dos eventos em questão se dá de forma aleatória, ou não, pode-se utilizar a análise de vizinhança. O gráfico torna possível medir a significância dos desvios referentes a aleatoriedade. Se os eventos forem aleatórios o gráfico deverá ser praticamente linear com um ângulo de 45° . Se o evento for tendencioso ao agrupamento, as linhas do gráfico estarão acima da linha de 45° e se forem tendenciosos a regularidade, estarão abaixo. A análise de

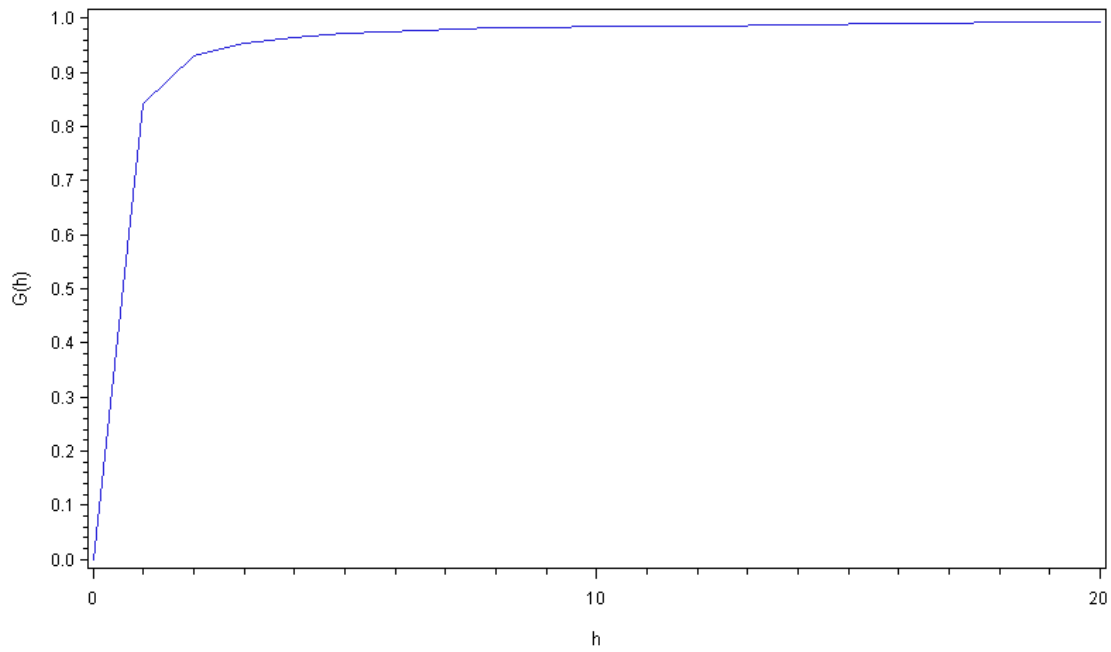


Figura 5.3: Método do vizinho mais próximo para o modelo geral

vizinhança para o modelo geral está representada na Figura 5.4.

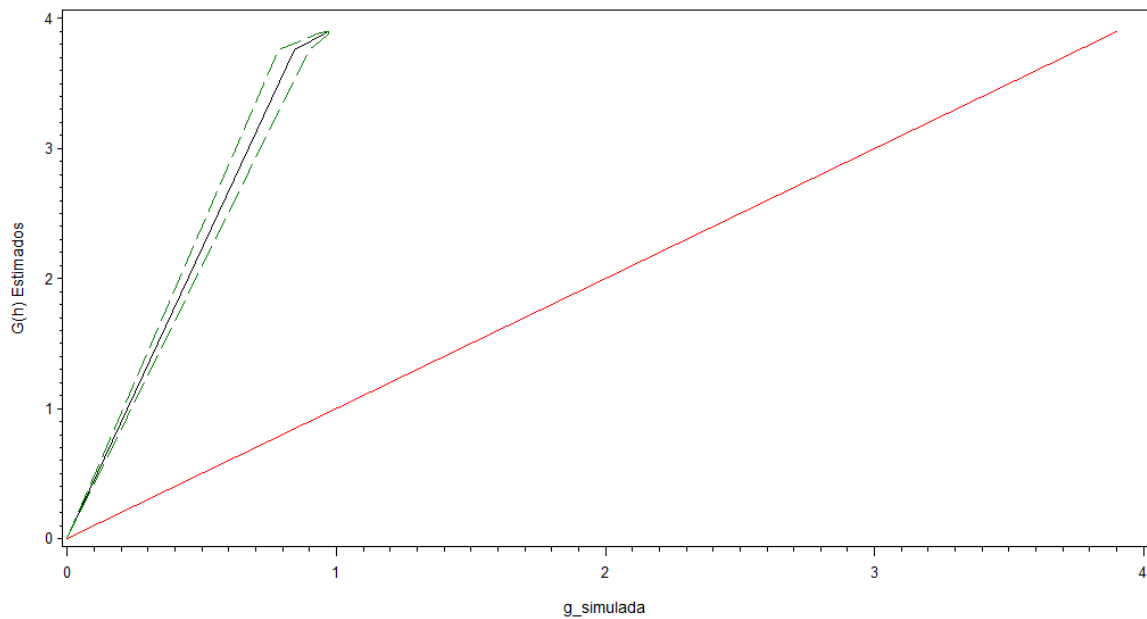


Figura 5.4: Envelope de simulação do modelo geral

Como é possível observar, as linhas do gráfico encontram-se acima da linha de 45° ,

corroborando a afirmação de que existe tendência ao agrupamento na distribuição das empresas transporte rodoviário de carga do DF. O gráfico da função K , na Figura 5.5 reitera a interpretação fornecida pelo método do vizinho mais próximo.

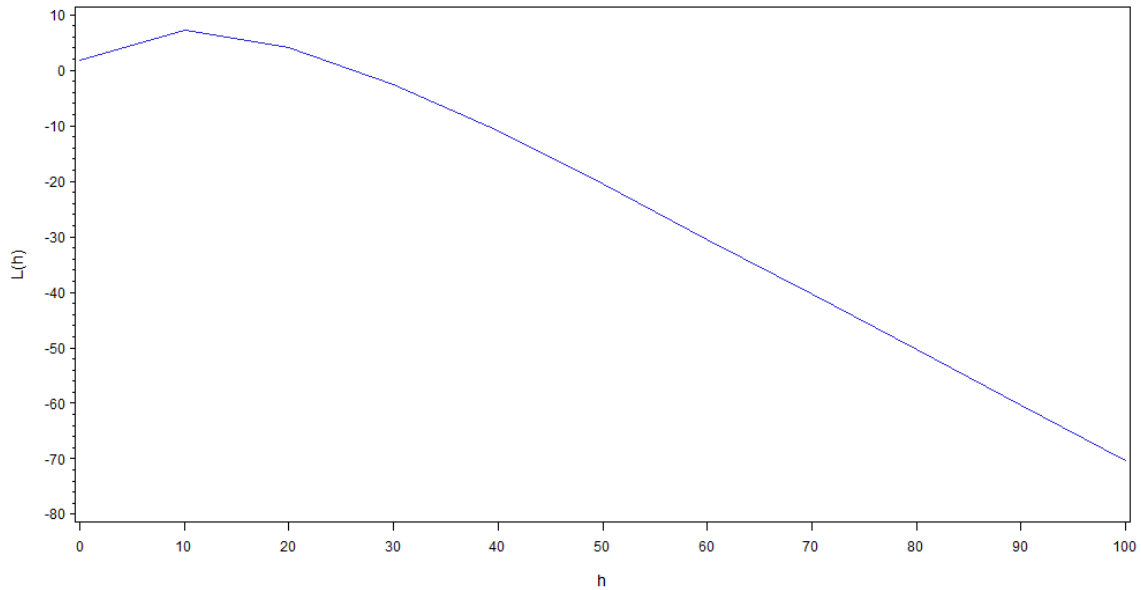


Figura 5.5: Função K do modelo geral

Este pico no gráfico para valores até 10 quilômetros, é entendido como um indício de tendência a aglomeração nessa faixa de distância.

Tais análises são úteis para a avaliação da localização dessas empresas no território do Distrito Federal, mas existem tipos diferentes de caminhões nessas empresas, o que permite fazer um estudo mais aprofundado, dividindo esses veículos em categorias, e refazendo essas análises do modelo geral para cada categoria de veículo separadamente.

Ao todo, são 5 tipos de caminhões utilizados pelas empresas em questão. São eles: Caminhão Simples, Caminhão Leve, Reboque, Semi-Reboque e Bi-Trem. Como o

Tabela 5.1: Tabela de frequências dos tipos de veículo

Categoria	Frequência	Porcentagem
Caminhão Leve	338	20,53%
Caminhão Simples	993	60,33%
Reboque	60	3,65%
Semi-reboque	253	15,37%
Semi-reboque com 5 ^a roda (Bitrem)	2	0,12%

número total de caminhões Bi-trem das empresas analisadas é muito pequeno (dois caminhões), este não constará na análise. Os 1646 caminhões estão representados em suas respectivas categorias na Tabela 5.1.

5.3 ANÁLISE POR TIPO DE VEÍCULO

Para a primeira parte da análise dos resultados, serão considerados apenas os caminhões simples das Empresas de TRC (Transporte Rodoviário de Carga). Primeiramente é necessário visualizar os dados em questão, para se ter uma idéia da distribuição espacial dos pontos. Tal representação stá na Figura 5.6.

Ao todo são 993 pontos que aparentemente parecem aglomerados em certas partes do mapa. Para verificar a existência, ou não, dessa tendência a aglomeração, pode-se utilizar o método do vizinho mais próximo. Na Figura 5.7, apresenta-se o gráfico do método do Vizinho mais próximo para o Caminhão do tipo Simples.

É possível observar que para uma distância h de 10 quilômetros, o gráfico apresenta um crescimento rápido e se estabiliza em torno de 2,5 km. Ou seja, nessa faixa de distância temos agrupamentos. Na análise de vizinhança para os Caminhões do tipo Simples obteve-se o seguinte gráfico representado na 5.8.

Como as linhas do gráfico estão todas acima da linha de 45°, os eventos em

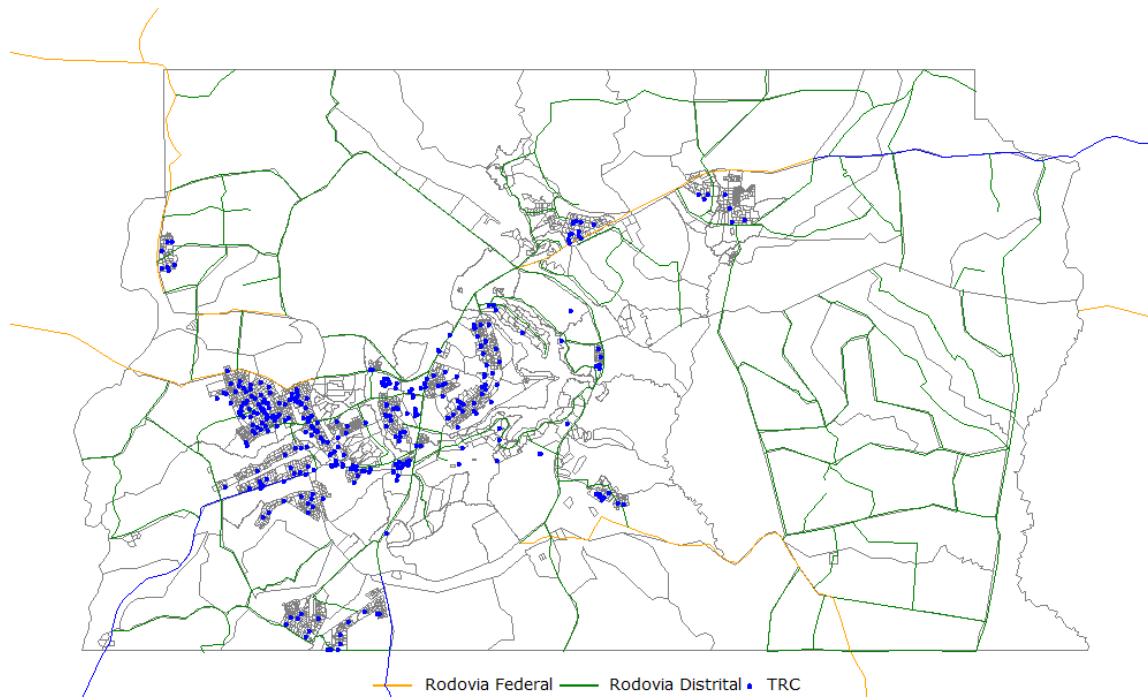


Figura 5.6: Distribuição dos veículos do tipo Caminhão Simples

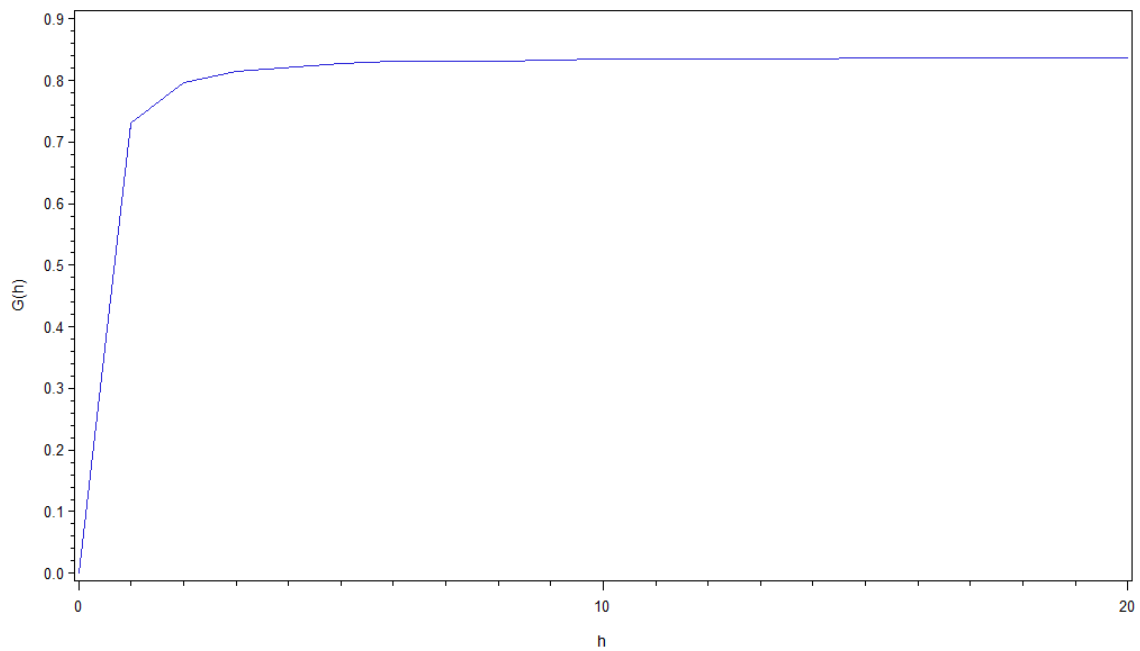


Figura 5.7: Método do vizinho mais próximo para Caminhão Simples

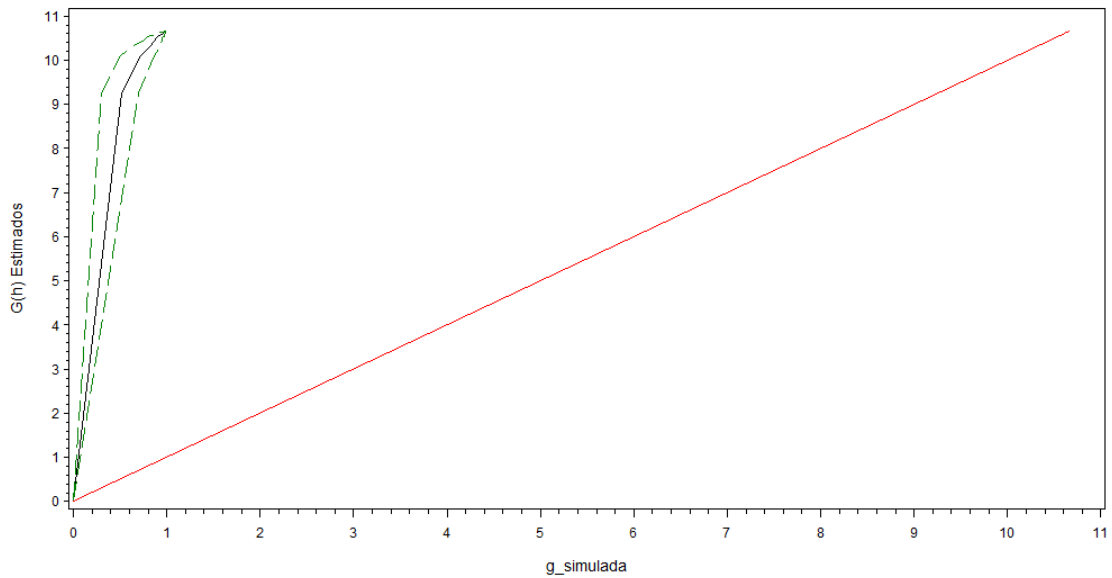


Figura 5.8: Envelope de simulação para Caminhão Simples

questão apresentam uma tendência ao agrupamento. Por fim, pode-se criar uma superfície de densidade que permite enxergar “áreas quentes” no mapa estudado. As áreas quentes são aquelas regiões onde há uma concentração de eventos indicando uma aglomeração na distribuição espacial. Na Figura 5.9, tem-se a superfície de densidade para o Caminhão do tipo Simples.

Observa-se que na área onde situam-se as Regiões Administrativas de Taguatinga, Ceilândia e Guará, há uma “área quente” que indica uma aglomeração na nossa distribuição espacial. No gráfico da função K (Figura 5.10) para a distribuição dos caminhões do tipo Simples é possível observar um pico para valores até 10 quilômetros, indicando que nessa faixa de distância há tendências a aglomerados, reafirmando o que foi encontrado no método do vizinho mais próximo.

A análise dos caminhões do tipo leve, se dará da mesma forma. Primeiramente a visualização dos eventos, com cada ponto da Figura 5.11 sendo um caminhão do

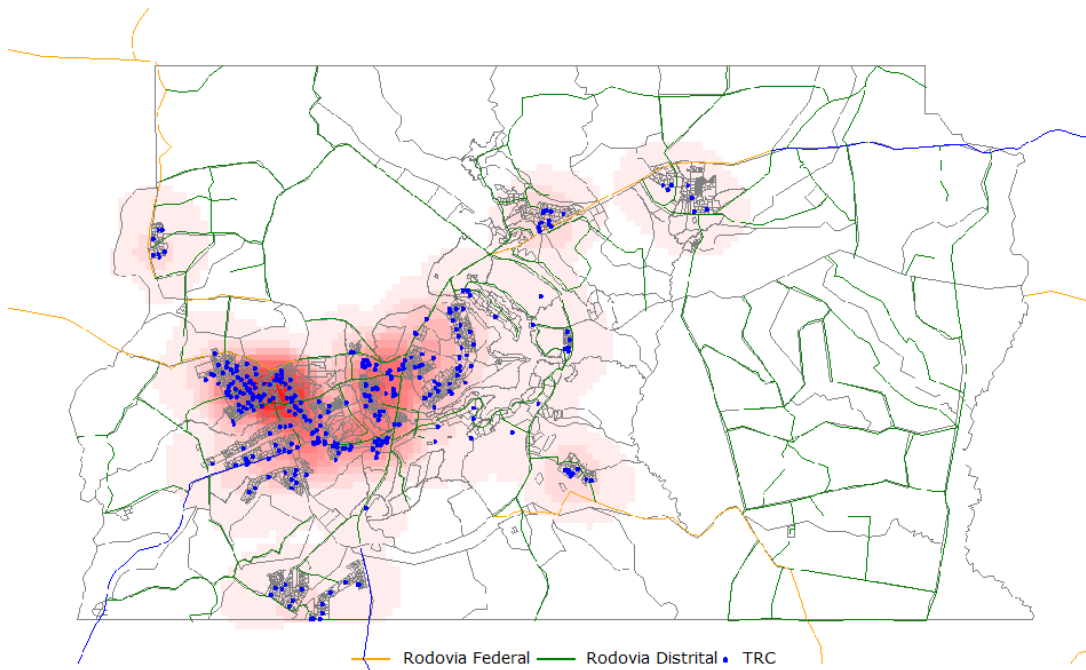


Figura 5.9: Função de intensidade para Caminhão Simples com $\tau=5$

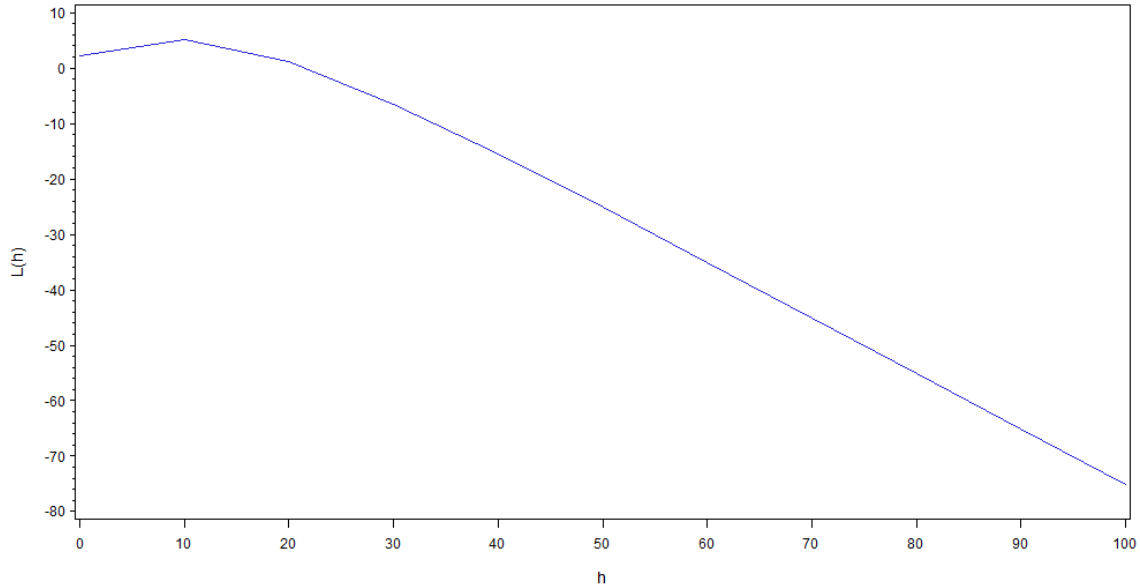


Figura 5.10: Função K para Caminhão Simples

tipo leve das Empresas de TRC do DF.

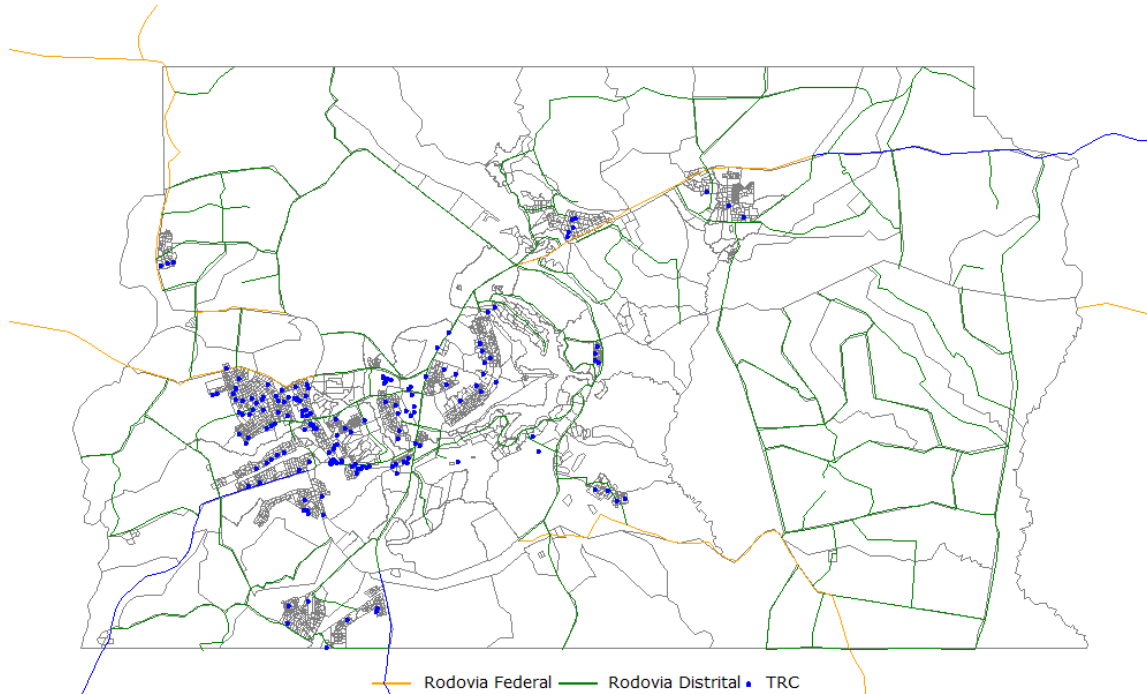


Figura 5.11: Distribuição dos veículos do tipo Caminhão Leve

Ao todo são 338 caminhões do tipo leve que novamente parecem ter uma tendência de aglomerados na região de Taguatinga e Ceilândia. Para verificar a existência de interação dos eventos estudados, utiliza-se o método do vizinho mais próximo. A Figura 5.12 mostra o gráfico do método do vizinho mais próximo para o Caminhão do tipo Leve.

O gráfico apresenta um crescimento relativamente rápido para uma distância h de 10 km, já que este se estabiliza em torno de 4 km. Conclui-se que nessa faixa de distância, os eventos tem interação caracterizando um agrupamento. É interessante verificar se a distribuição dos caminhões do tipo leve é aleatória ou possui alguma tendência a regularidade ou a aglomeração. Na Figura 5.13 temos o gráfico da análise de vizinhança para este tipo de caminhão

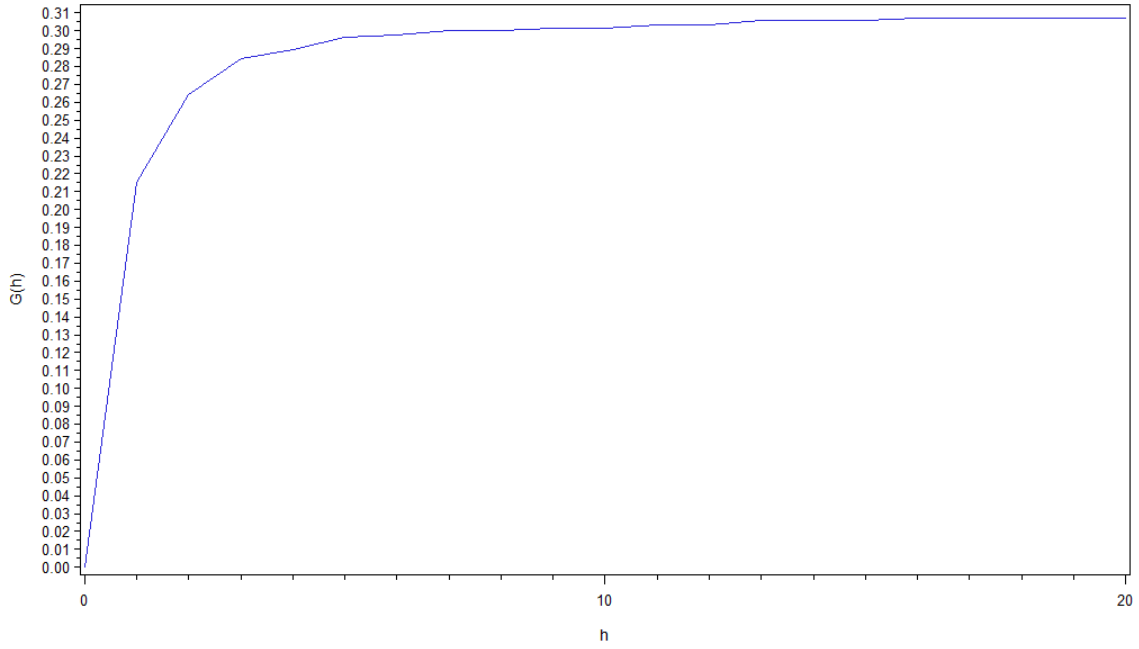


Figura 5.12: Método do vizinho mais próximo para Caminhão Leve

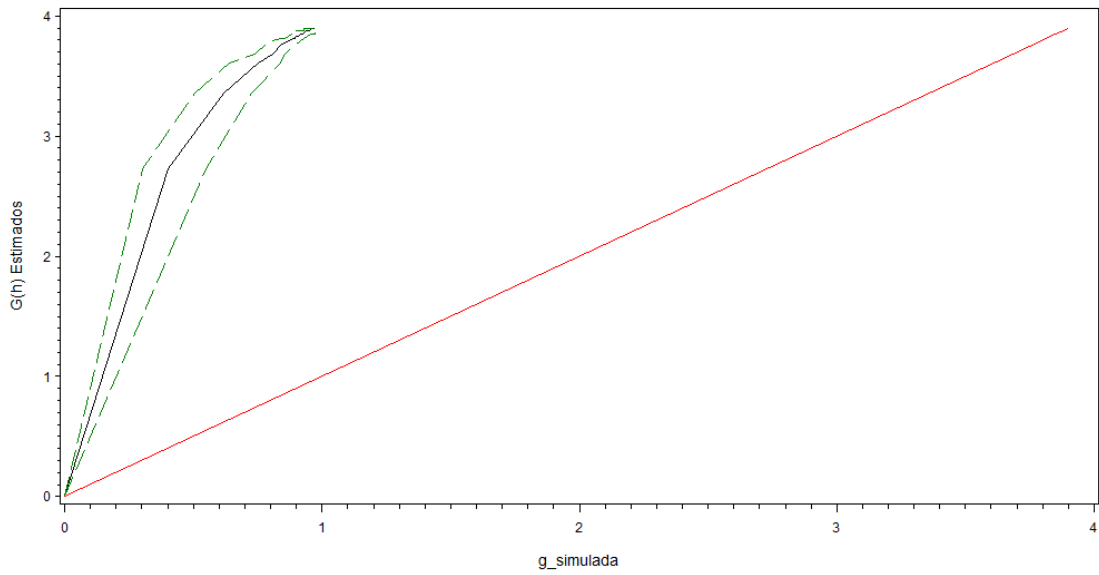


Figura 5.13: Envelope de simulação para Caminhão Simples

As linhas do gráfico estão bem acima da linha de 45° , indicando que a distribuição dos caminhões do tipo leve tem uma certa tendência ao agrupamento. Para visualizar aonde se encontram estas aglomerações de pontos, basta construir a superfície de densidade. Na Figura 5.14 observa-se que a “área quente” dessa densidade, cai, novamente, em cima das R.A’s de Taguatinga e Ceilândia.

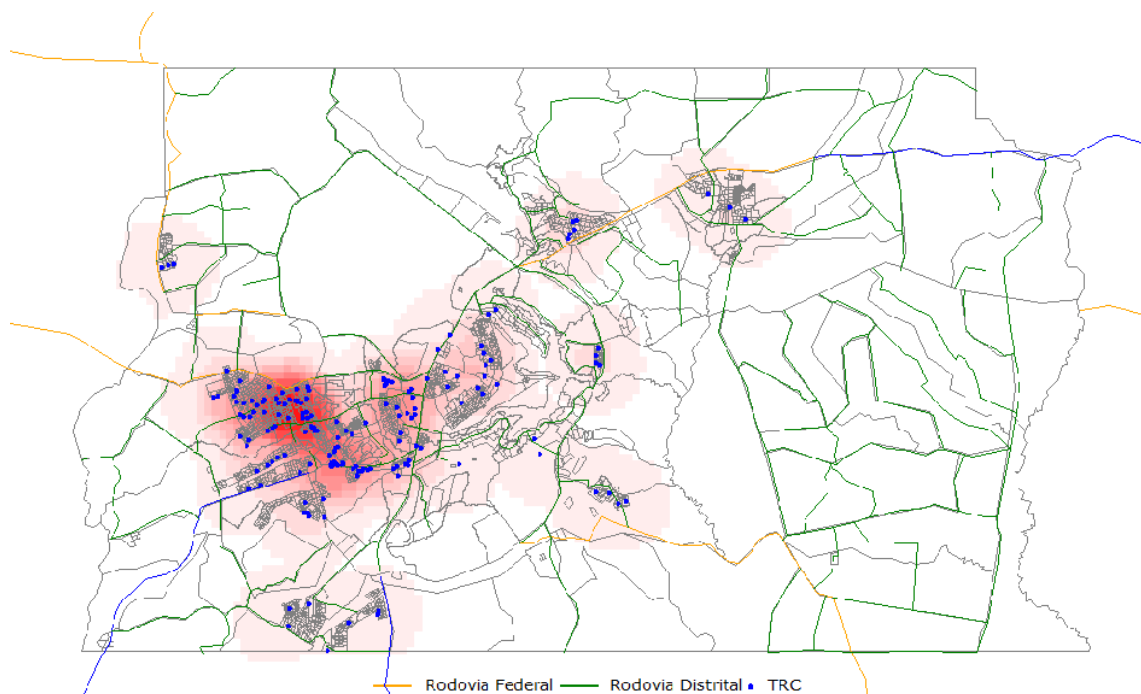


Figura 5.14: Função de intensidade para Caminhão Leve com $\tau=5$

Analisando o gráfico da função K para os caminhões do tipo leve, presente na Figura 5.15, encontra-se que os mesmos possuem tendências de aglomeração para valores compreendidos na faixa de 0 a 10 quilômetros, já que houve um pico no gráfico para esta faixa de distância. A função K só ratifica o que foi observado no método do vizinho mais próximo.

Outro caminhão estudado é o do tipo Semi-Reboque. Na Figura 5.16, é possível visualizar como os caminhões desse tipo estão distribuídos pelo Distrito Federal.

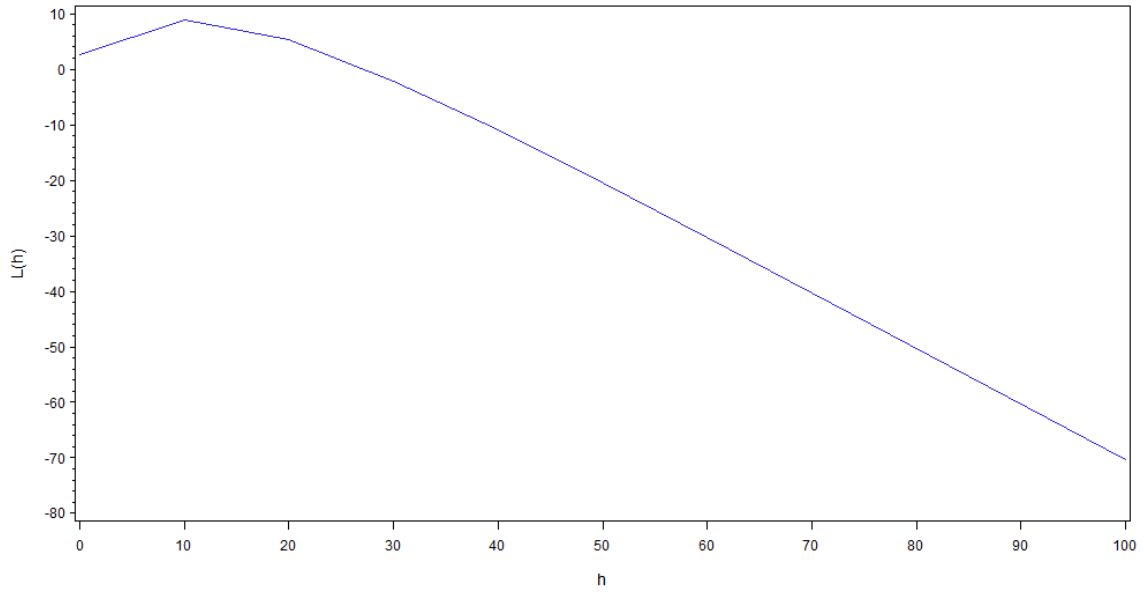


Figura 5.15: Função K para Caminhão Leve

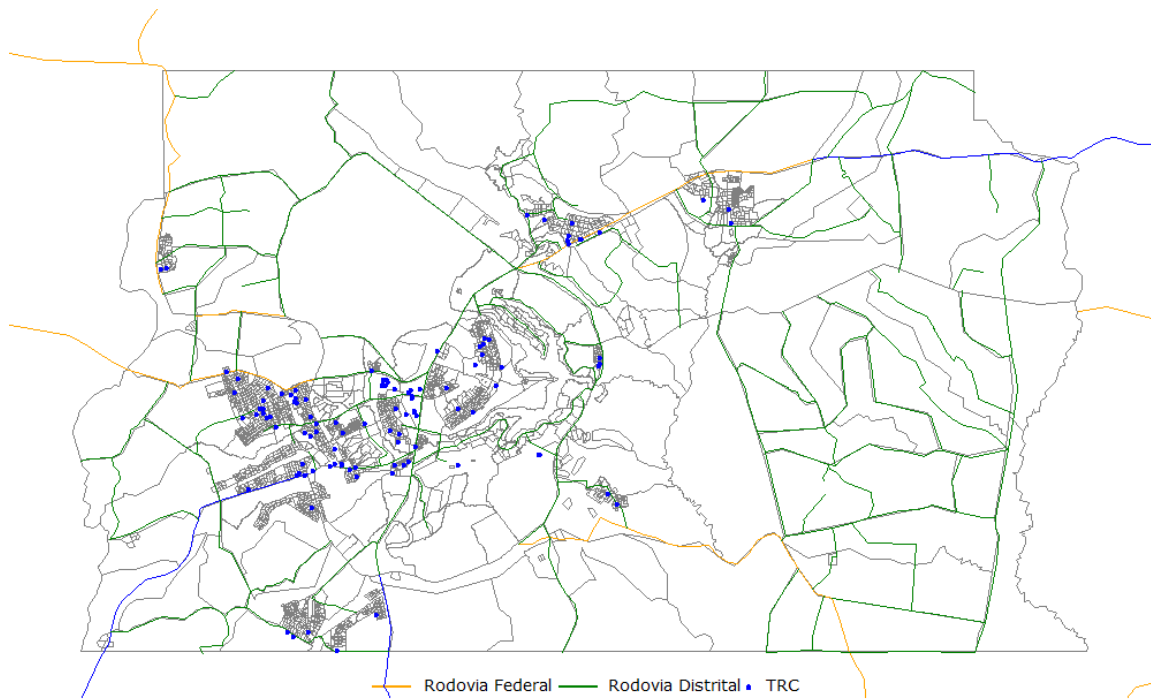


Figura 5.16: Distribuição dos veículos do tipo Semi-reboque

Ao todo são 253 caminhões desse tipo que aparentemente estão agrupados na mesma região dos caminhões simples e leves. Pode-se verificar a interação entre os eventos em estudo através do método do vizinho mais próximo, que tem como gráfico para este tipo de caminhão a Figura 5.17.

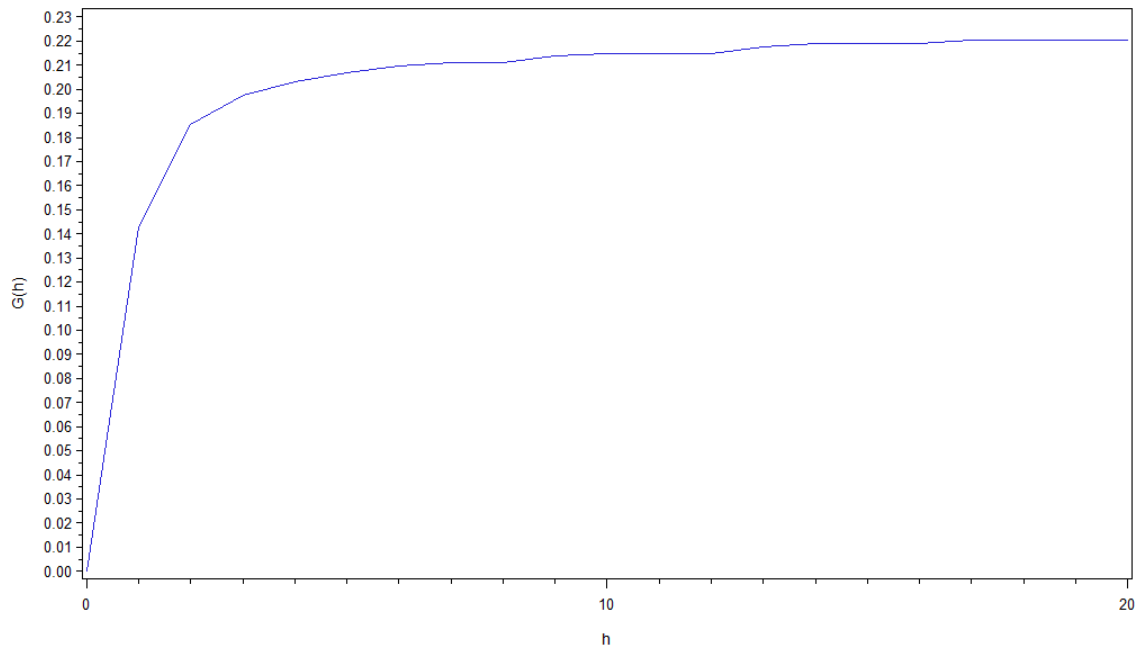


Figura 5.17: Método do vizinho mais próximo para Semi-reboque

O gráfico para este tipo de caminhão também apresenta um crescimento acelerado para uma distância h de 10km, estabilizando-se em torno de 4 km. Para testar se a distribuição dos pontos segue um modelo aleatório, a análise de vizinhança será utilizada. Dela temos o gráfico representado na Figura 5.18.

Assim como os outros dois tipos de caminhões analisados anteriormente, o envelope de simulação dos caminhões do tipo Semi-Reboque mostra suas linhas acima da linha de 45° indicando uma tendência ao agrupamento. Esse agrupamento é

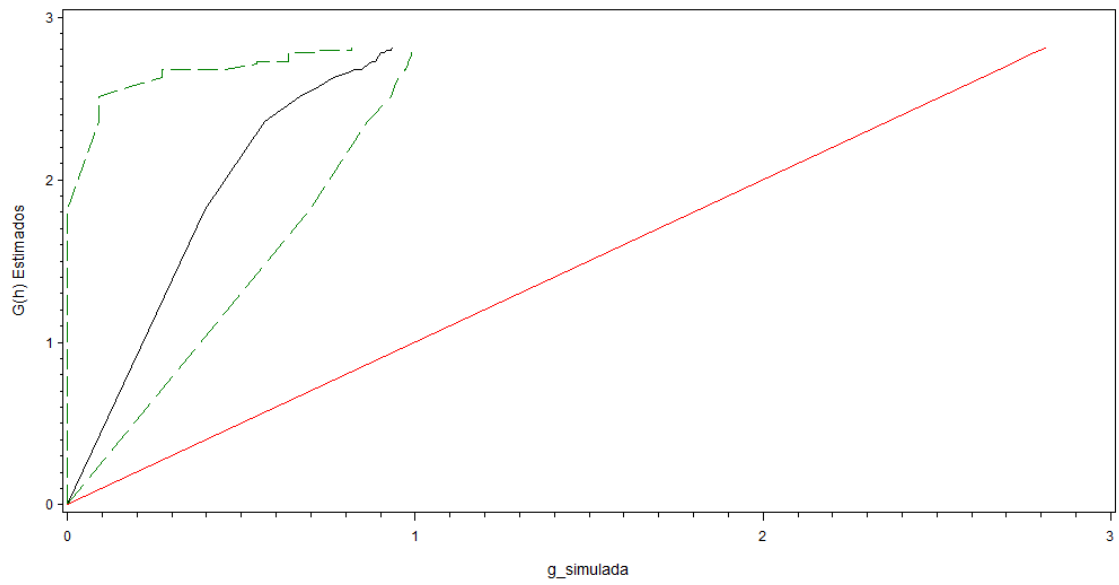


Figura 5.18: Envelope de simulação para Semi-reboque

observado com a produção da superfície de densidade representada na Figura 5.19.

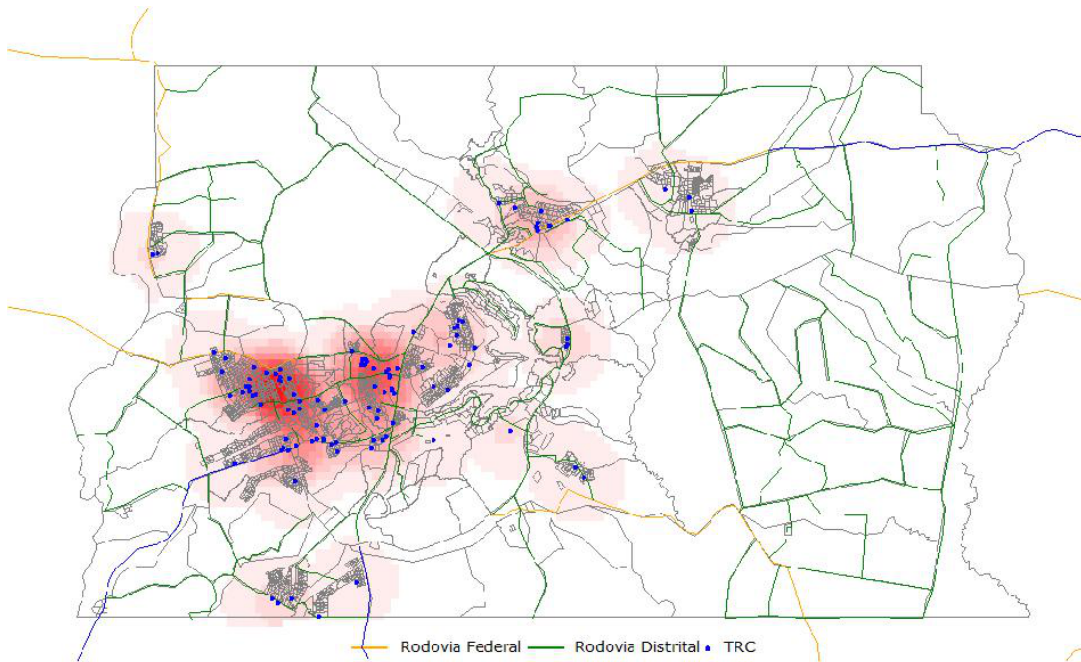


Figura 5.19: Função de intensidade para Semi-reboque com $\tau=5$

Além disso, pelo gráfico da função K para caminhões do tipo Semi-Reboque

(Figura 5.20), assegura-se, assim como no método do vizinho mais próximo, que para valores menores que 10 quilômetros existe tendência a agrupamento.

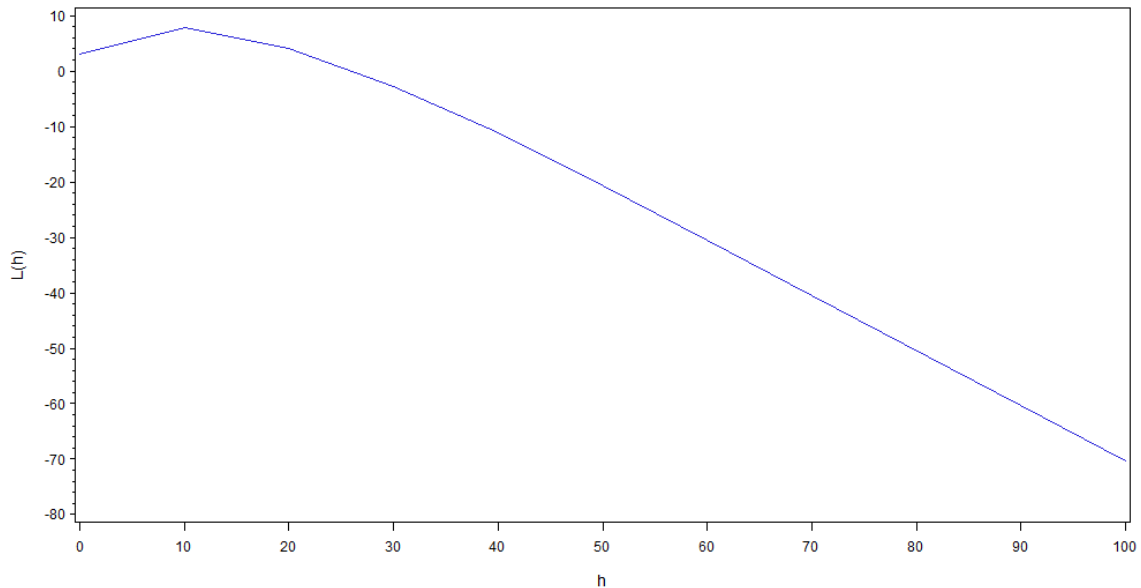


Figura 5.20: Função K para Semi-Reboque

Para a categoria Reboque, observou-se a existência de 60 veículos deste tipo no Distrito Federal. Sua distribuição no território pode ser visualizada na Figura 5.21.

Aparentemente, não observa-se uma aglomeração como nos outros casos. Os pontos estão mais distribuídos pelo território e não está presente a concentração de veículos nas regiões de Taguatinga e Ceilândia, comum nos outros casos. Contudo, para se avaliar de maneira correta, é necessário verificar a existência de interação dos eventos em questão, sendo utilizado o método do vizinho mais próximo, conforme mostra a Figura 5.22.

Diferentemente das outras categorias, pode-se observar que para uma distância h de 10 quilômetros, o gráfico apresenta um crescimento rápido, mas se estabiliza

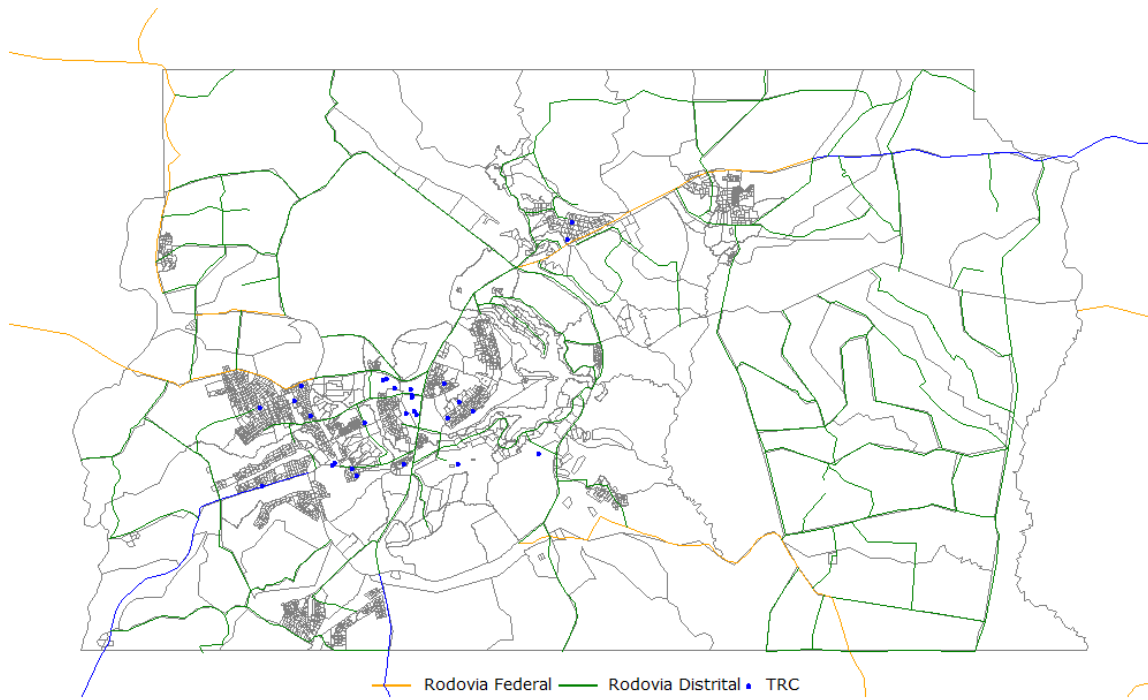


Figura 5.21: Distribuição dos veículos do tipo Reboque

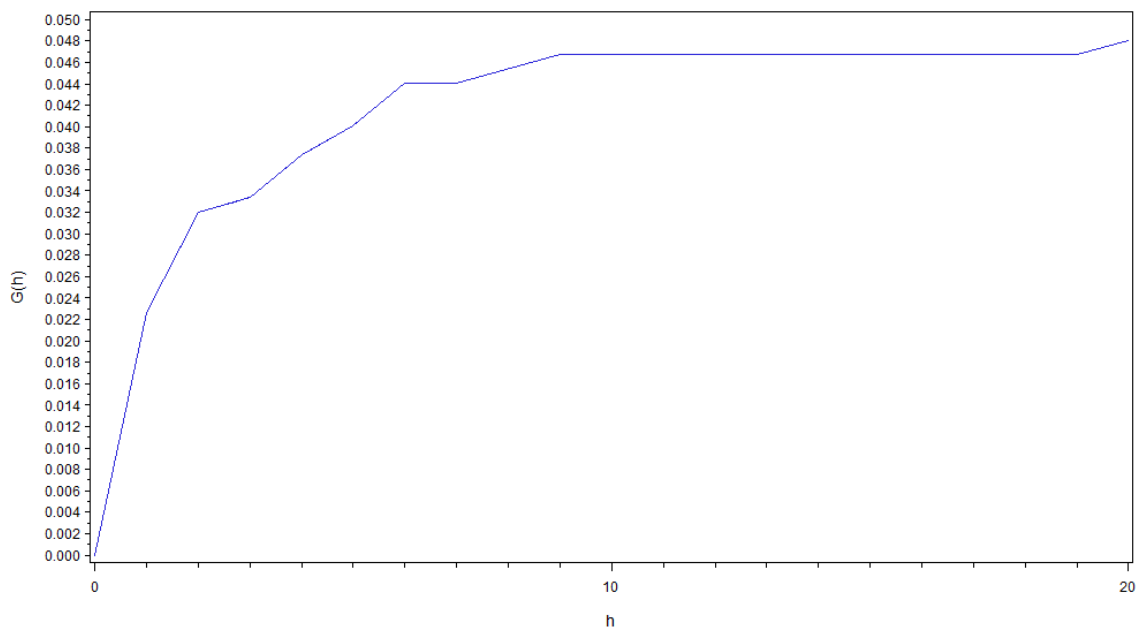


Figura 5.22: Método do vizinho mais próximo para Reboque

definitivamente em torno de 20 km, o que significa dizer que é nessa faixa de distância que podemos afirmar que existem agrupamentos.

Para a verificação da aleatoriedade dessa distribuição, novamente é usada a análise de vizinhança. Dessa vez, em oposição aos outros tipos de veículo, observa-se um padrão de regularidade ao invés de aglomeração, já que os traçados ficaram abaixo da linha de 45 graus (Figura 5.23).

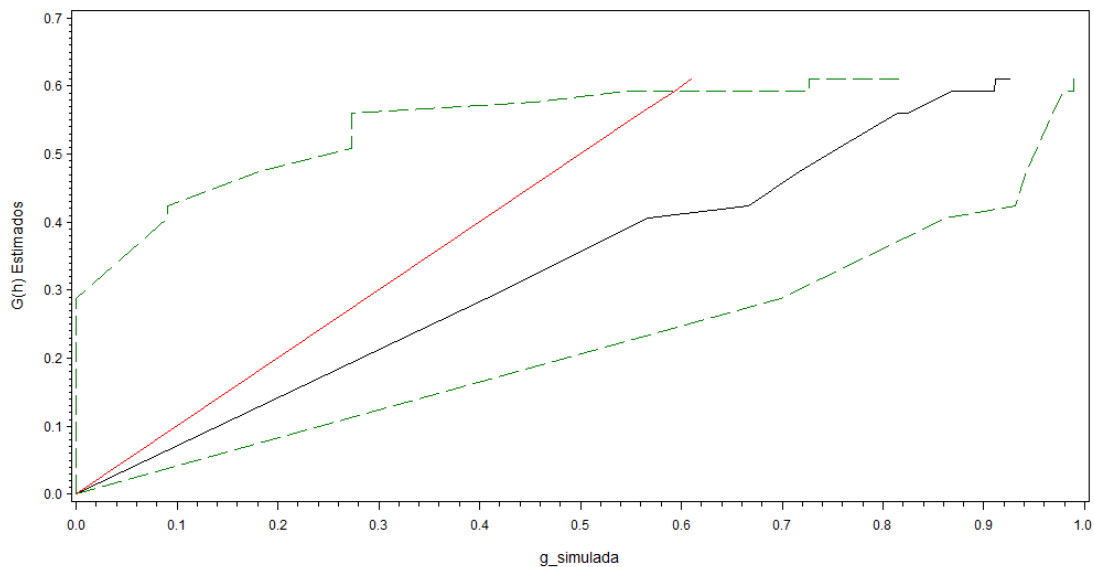


Figura 5.23: Envelope de simulação para Reboque

A superfície de densidade criada utilizando um valor $\tau=10$ mostra que mesmo havendo uma “área quente”, percebe-se a existência de apenas um ponto em seu centro, o que significa que essa área se apresenta dessa forma devido a maior quantidade de pontos próximos a ele. Essa superfície está ilustrada na Figura 5.24. O gráfico da função K (Figura 5.25) mostra que existe tendência a aglomeração para valores abaixo de 10 quilômetros.

Esses mapas e gráficos mostram que existe uma tendência de aglomeração das

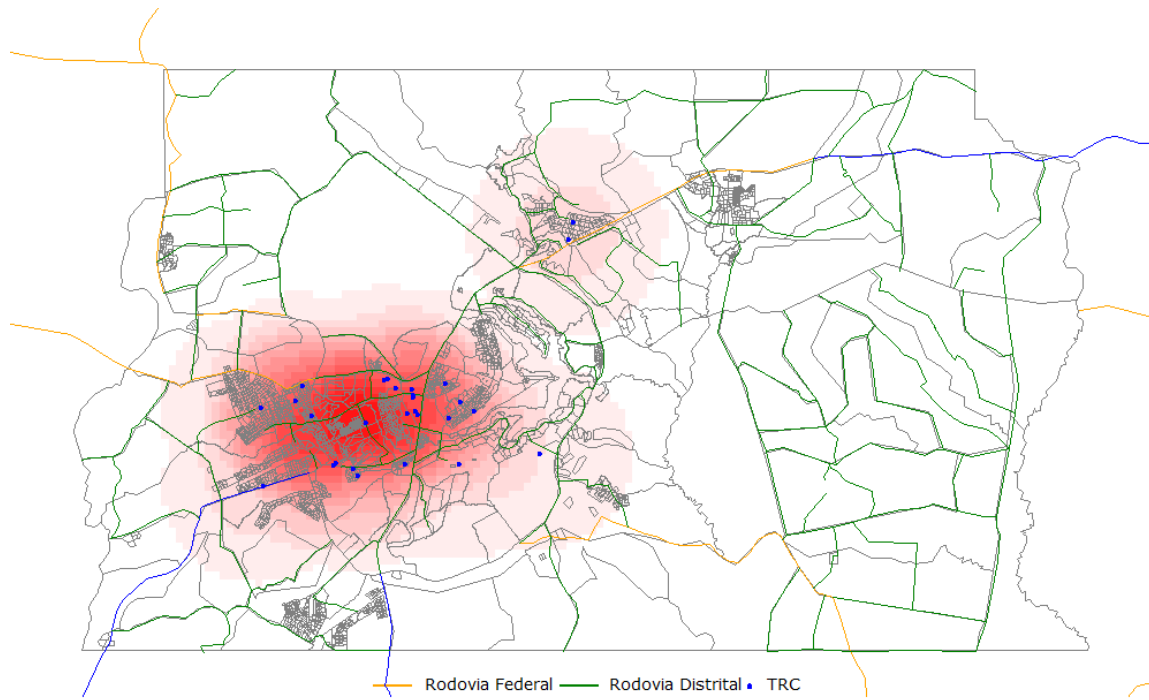


Figura 5.24: Função de intensidade para Reboque com $\tau=10$

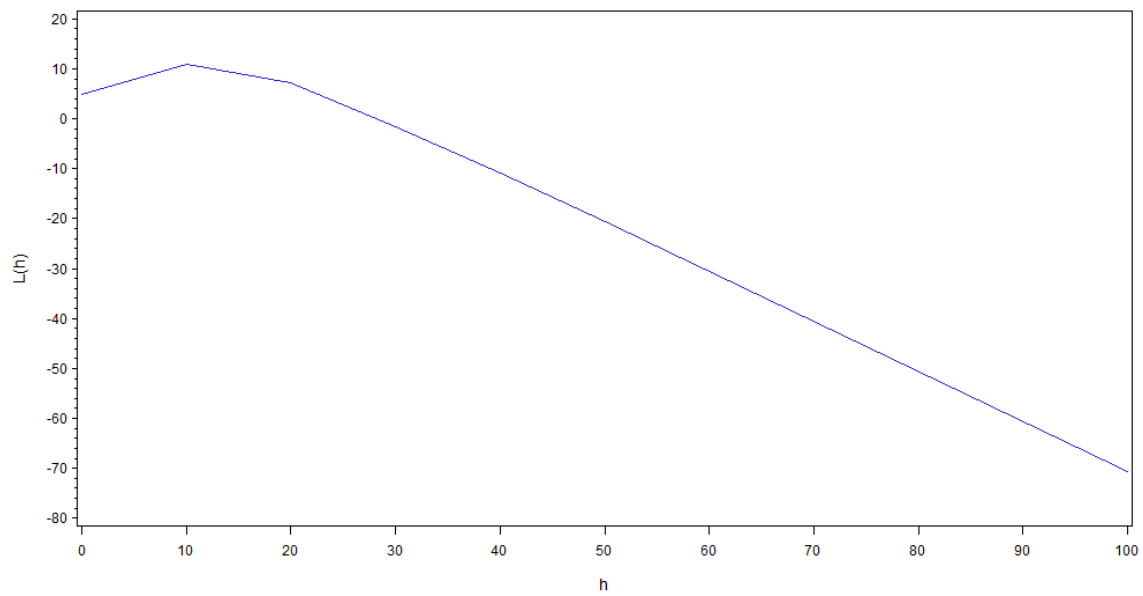


Figura 5.25: Função K para Reboque

empresas nas regiões administrativas de Taguatinga e Ceilândia.

Capítulo 6

Conclusão

A análise da distribuição dos pontos representando as empresas de TRC no Distrito Federal através dos métodos do vizinho mais próximo, análise de vizinhança, função de intensidade e função K de Ripley se mostrou de grande valia.

Pôde-se concluir pelo modelo geral que há uma aglomeração das empresas nas regiões de Taguatinga, Ceilândia e Guará, o que pode ser explicado pela existência de Setores de Indústria nessas Regiões Administrativas, e pela proximidade dessas regiões com a Saída Sul, uma rodovia federal de grande fluxo, facilitando o embarque e desembarque de mercadorias, além de custos mais reduzidos quando comparados a outras localidades mais distantes de rodovias. As empresas localizadas nas outras regiões como Sobradinho, Planaltina e Brazlândia também apresentaram a tendência de se localizar perto das rodovias distritais e federais, uma grande vantagem considerando que desse modo o custo e o deslocamento do veículo da rodovia até a empresa é menor.

Para as categorias de veículo de Caminhão Simples, Caminhão Leve e Semi-Reboque, foram observados resultados semelhantes, que indicaram uma aglomeração nas mesmas regiões (Taguatinga e Ceilândia), e com distâncias próximas de agrupa-

mento em torno de 5 quilômetros. Para os caminhões do tipo Reboque, apesar de existir uma “área quente” na mesma região, não foi possível afirmar a existência de um aglomerado, tendo em vista que tal resultado se deu devido a menor distância entre os pontos daquela região, e não por uma concentração de pontos nesta. Tal fato foi comprovado ao realizar-se a análise pelo método do vizinho mais próximo, que indicou um padrão de regularidade para os veículos na categoria Reboque.

Os resultados dessas análises mostraram o que já era esperado, que a localização das empresas de TRC está realmente associada à localização das rodovias que atravessam o Distrito Federal. Em certas localidades essa proximidade com as rodovias se dá pelo fato da área urbana da Região Administrativa ser pequena, ou seja, não importa onde a empresa esteja, estará próxima a uma rodovia devido ao pequeno tamanho da região, como é o caso de Brazlândia. No geral, verificou-se essa proximidade entre localização de empresas e rodovias, podendo assim afirmar-se que existe uma boa distribuição espacial das empresas de TRC no Distrito Federal.

Referências Bibliográficas

Cressie, N.A.C. Statistics for Spatial Data. USA:Wiley-Interscience Inc, 928p, 1991.

Bailey, T.C. e Gatrell, A.C. Interactive Spatial Data Analysis. England: Prentice Hall, 432p, 1995. .

Druck, S., Carvalho, M. S., Câmara, G., Monteiro, A. M.V. Análise Espacial de Dados Geográficos. Brasília, EMBRAPA, 190p, 2004.

ANTT. Agência Nacional de Transportes Terrestres - ANTT. URL "http://www.antt.gov.br/carga/rodoviario/legislacaobasica", 2001.

IBGE. Instituto Brasileiro de Geografia e Estatística - IBGE. URL "http://www.ibge.gov.br/", 2001.

RNTRC. Registro Nacional de Transporte Rodoviário de Carga - RNTRC. URL "http://www.antt.gov.br/carga/rodoviario/rntrc", 2001.

Santos, M.S.,Souza,W.V. Introdução à Estatística Espacial para a Saúde Pública. Ministério da Saúde: Fundação Oswaldo Cruz, 2007.