UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ENGENHARIA DE SÃO CARLOS
DEPARTAMENTO DE ENGENHARIA DE TRANSPORTES

GEOESTATÍSTICA APLICADA A PROBLEMAS RELATIVOS À DEMANDA POR TRANSPORTES

CIRA SOUZA PITOMBO

Texto apresentado à Escola de Engenharia de São Carlos da
Universidade de São Paulo como parte dos requisitos para
obtenção do título de Livre-Docente em concurso realizado pelo
Departamento de Engenharia de Transportes

SÃO CARLOS
Agosto de 2016

UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ENGENHARIA DE SÃO CARLOS
DEPARTAMENTO DE ENGENHARIA DE TRANSPORTES

**GEOESTATÍSTICA APLICADA A PROBLEMAS
RELATIVOS À DEMANDA POR TRANSPORTES**

CIRA SOUZA PITOMBO

Texto apresentado à Escola de Engenharia de São Carlos da Universidade de São Paulo como parte dos requisitos para obtenção do título de Livre-Docente em concurso realizado pelo Departamento de Engenharia de Transportes

SÃO CARLOS
Agosto de 2016

*“ Uma semente enterrada no solo não aparece, mas água, calor, luz e sombra
agem como estímulos diários. Pouco a pouco, acontece uma transformação
invisível até que um broto aparece aos nossos olhos. Quando o broto alcança a luz
do sol, passa a crescer mais rapidamente”*

Shinichi Suzuki

AGRADECIMENTOS

- Ao Flávio e à pequena Alice, por estarem ao meu lado também durante a realização deste trabalho;
- Aos meus queridos pais, Nilton e Iacina, pelo importante suporte na fase de escrita deste documento;
- Aos amigos e principais colaboradores da linha de pesquisa **Geoestatística aplicada a problemas relativos à demanda por transportes**: Ana Rita Salgueiro, Antonio Jorge Sousa, Cassiano Isler, José A. Quintanilha;
- Ao Antonio Nelson Rodrigues da Silva pela grande ajuda desde a concepção até a finalização do texto. Particularmente, agradeço também ao David pelos “cubos mágicos”;
- Aos amigos e colegas Ana Paula Larocca, José Reynaldo Setti e Adalberto Faxina, por me encorajar a “começar”;
- Aos grandes amigos, e por acaso meus orientados: Aline Costa, Anabele Lindner, Andreza Dornelas, Cintia Isabel de Campos, Claudia Naizer, Daiane Castro Bittencourt, Heymar Arancibia Suárez, Lucas Assirati, Lucas Pierre Clément Vialle, Marina Urano de Carvalho Caldas, Monique Martins Gomes, Murilo Castanho dos Santos, Samille Santos Rocha, Sandra Helena Miranda de Souza e Viviani Antunes Gomes;
- Às agências de fomento à pesquisa CNPQ, FAPESP, CAPES, FAPESB, FCT e Fundação *Calouste Gulbenkian*;
- Aos professores Eiji Kawamoto e Orlando Strambi, principais responsáveis pela minha “paixão” por demanda por transportes e modelagem;
- Aos professores Ilce Marília Dantas de Freitas e Juan Delgado Moreno, pela generosidade profissional e carinho que me receberam na Universidade Federal da Bahia;
- Ao professor e amigo Jefferson Lins pela valiosa ajuda na fase final de preparação de documentos;
- Ao professor André Luiz Cunha pela “parceria” em atividades de docência e pesquisa;
- Aos funcionários e professores do Departamento de Engenharia de Transportes da EESC-USP;
- A todos aqueles que, direta ou indiretamente, colaboraram para conclusão deste trabalho.

RESUMO

O comportamento individual relativo a viagens urbanas é influenciado, dentre outros fatores diversos, por atributos locacionais, sendo fortemente recomendada a análise espacial de demanda por transportes. A partir da hipótese principal de que a localização espacial influencia variáveis relativas à demanda por viagens, a linha de pesquisa **Geoestatística aplica a problemas relativos à demanda por transportes** foi proposta desde 2008. Assim, neste texto são apresentados, de forma sintética, os principais trabalhos subjacentes à linha de pesquisa, realizados nos últimos anos.

É feita uma análise crítica acerca dos resultados obtidos de doze trabalhos, classificados segundo dois grandes temas: geoestatística aplica a geração de viagens e geoestatística aplica a escolha modal. Os trabalhos descritos são dissertações de mestrado já defendidas, artigos científicos publicados em periódicos nacionais e internacionais e artigos apresentados em congressos nacionais e internacionais. São listadas as contribuições e avanços para melhoria das estimativas, além dos desafios científicos e tecnológicos para superação das atuais limitações metodológicas, tais como: agregação de dados associados a áreas irregulares, adequação dos dados a distribuição de probabilidade não gaussiana e incorporação de características da variografia a modelos tradicionais de previsão de demanda por viagens.

Finalmente, são apresentados os temas futuros, com perspectivas de estimativas ainda melhores em cinco anos, através da realização de três doutorados, um mestrado, além de um projeto de pós-doutorado no exterior.

Como conclusão geral dos resultados já obtidos, o uso da geoestatística fornece uma superfície contínua de estimativas de variáveis de demanda por transportes. Dado que não é possível ser obtido através de ferramentas tradicionalmente utilizadas. O uso dos mapas krigados, associados a técnicas de geoprocessamento, pode configurar a obtenção de dados de demanda por transportes exatamente nos domicílios existentes, com pouca informação previamente necessária.

Palavras-chave: Geração de viagens; Escolha modal; Krigagem; Variograma; Modelagem tradicional de demanda por transportes.

ABSTRACT

Individual travel behavior is influenced, among several other factors, by locational attributes, and spatial analysis of travel demand is strongly recommended. From the main hypothesis that spatial location influences travel demand variables, the research line called **Applying geostatistics to travel demand issues** has been proposed since 2008. Thus, in this text, the main studies related to the research line, carried out in recent years, are summarised.

A critical analysis of the results of twelve studies, classified according to two major themes is made: Applying geostatistics to trip generation and to travel mode choice. The studies described are Master's dissertations, scientific papers published in national and international journals and papers presented at national and international conferences. Contributions and advances to improve the estimates are listed in addition to the scientific and technological challenges to overcome current methodological limitations, such as: using aggregated data associated with irregular areas, adequacy of non-Gaussian distribution variables and including variogram features in traditional travel demand forecasting.

Finally, future topics are presented, with prospects of even better estimates in five years, by conducting three doctorate degrees, a Master's degree and a post-doctoral project.

As the general conclusion from the results already obtained, the use of geostatistics provides the continuous surface of travel demand estimates. This output cannot be obtained using traditional tools. By using kriging maps related to geoprocessing techniques, travel demand data can be obtained from existing households, with little previous information.

Keywords: Trip generation; Travel mode choice; Kriging; Variogram; Travel demand modelling.

LISTA DE FIGURAS

Figura 1.1 Construção da linha de pesquisa – analogia com a montagem do cubo mágico.	2
Figura 1.2 Distribuição espacial de uma variável discreta qualquer em uma área fictícia.....	2
Figura 1.3 Principais colaboradores da linha de pesquisa	7
Figura 1.4 Compilação de publicações provenientes da linha de pesquisa	8
Figura 2.1 Representação de uma variável espacial isotrópica (a) e anisotrópica (b). Adaptado de Lindner (2015)	15
Figura 2.2 Análise prévia da estrutura espacial da VR (área fictícia)	16
Figura 2.3 Esquema mostrando a pesquisa de pares de pontos	19
Figura 2.4 Etapa 1 para obtenção do variograma experimental. Fonte: Lindner (2015)	19
Figura 2.5 Etapa 2 para obtenção do variograma experimental. Fonte: Lindner (2015)	19
Figura 2.6 Etapa 3 para obtenção do variograma experimental. Fonte: Lindner (2015)	20
Figura 2.7 Etapa 4 para obtenção do variograma experimental. Fonte: Lindner (2015)	20
Figura 2.8 Etapas para cálculo do variograma experimental	22
Figura 2.9 Exemplos de ajustes para variogramas experimentais e parâmetros variográficos. Fonte: Olea (2006) e Lindner (2015)	23
Figura 2.10 Ajustes para variograma experimental correspondente à área fictícia	24
Figura 2.11 Síntese do processo de estimação geoestatística univariada. Adaptado de Yamamoto e Landim (2013)	25
Figura 2.12 Procedimento ilustrativo para estimativa por Krigagem Ordinária.....	28
Figura 2.13 Variograma teórico correspondente à área fictícia (dados binários)	30
Figura 2.14 Procedimento ilustrativo para estimativa por Krigagem Indicativa da mediana..	31
Figura 3.1 Sequência de informações para descrição das áreas de estudo	40
Figura 3.2 Região Metropolitana de São Paulo (subregiões e municípios). Fonte: METRÔ- SP (1997).....	41
Figura 3.3 RMSP. Densidade populacional. Pesquisa O/D 1997. Fonte: Pitombo (2007)	44
Figura 3.4 RMSP. Densidade de empregos. Pesquisa O/D 1997. Fonte: Pitombo (2007) ...	44

Figura 3.5 RMSP. Zonas de Tráfego. Fonte: METRÔ-SP (2007)	45
Figura 3.6 Representação dos domicílios amostrados na Pesquisa O/D da RMSP (2007). Fonte: Lindner (2015)	46
Figura 3.7 Representação da: (a) preferência por transporte individual motorizado – em vermelho- sobreposta à preferência por transporte público, (b) preferência por transporte público – em verde – sobreposta à preferência por transporte individual motorizado. Fonte: Lindner (2015).....	49
Figura 3.8 Região Metropolitana de Salvador. Fonte: IPEA (2014)	50
Figura 3.9 Subregiões e Macrozonas. Fonte: STEPS (1995)	52
Figura 3.10 Distribuição dos valores padronizados na RMS: Produção de viagens por TCO e Atração de viagens por TCO. Fonte: Pitombo <i>et al.</i> (2014).....	53
Figura 3.11 População por MZ de tráfego. Fonte: IBGE (2010).....	54
Figura 3.12 Renda média do responsável do domicílio, em salários mínimos, por MZ de tráfego. Fonte: IBGE (2010).....	55
Figura 3.13 Base de pontos contento os domicílios selecionados da cidade de São Carlos para a Pesquisa O/D de 2007/2008. Fonte: Rodrigues da Silva <i>et al.</i> (2009).....	56
Figura 3.14 Distribuição espacial das probabilidades de escolha do modo particular motorizado, público e não motorizado. Fonte: Costa <i>et al.</i> (2013).....	58
Figura 3.15 Distribuição da probabilidade dos 3 modos de transporte – região Santa Felícia. Fonte: Costa <i>et al.</i> (2013).....	59
Figura 3.16 Distribuição da probabilidade dos 3 modos (Centro). Fonte: Costa (2013).....	60
Figura 3.17 Ilustração do procedimento metodológico geral	62
 Figura 4.1 Sequência de informações para apresentação das pesquisas	73
Figura 4.2 Variogramas modelados (produção de viagens por automóvel) nas direções 45°, -45° e omnidirecional (0) e 90° . Fonte: Pitombo e Sousa (2009)	75
Figura 4.3 Variogramas modelados (produção de viagens por transporte público) omnidirecional e Componente 1 – 90°. Fonte: Lindner <i>et al.</i> (2016)	83
Figura 4.4 Mapa de superfície de valores estimados de produção de viagens por transporte coletivo. Fonte: Lindner <i>et al.</i> (2016).....	84
Figura 4.5 Valores interpolados de produção de viagens por TCO para 2010. Fonte: Pitombo <i>et al.</i> (2014).....	91

Figura 4.6 Etapas do processo metodológico aplicado ao estudo. Fonte: Rocha <i>et al.</i> (2014)	94
Figura 4.7 Malha gerada para amostragem sistemática artificial. Fonte: Rocha <i>et al.</i> (2014)	94
Figura 4.8 Modelagem dos variogramas. Fonte: Rocha <i>et al.</i> (2014)	98
Figura 4.9 Interpolação espacial para as variáveis Produção e Atração de viagens por TCO: (a) amostra agregada; (b) amostra desagregada; (c) valores observados por área. Fonte: Rocha <i>et al.</i> (2014)	99
Figura 5.1 Sequência de informações para apresentação das pesquisas	108
Figura 5.2 Procedimento metodológico geral proposto nos Trabalhos 1, 2 e 3 (Costa, 2013; Pitombo <i>et al.</i> , 2015a; Pitombo <i>et al.</i> , 2015b).	1110
Figura 5.3 Variogramas modelados na direção 30° (região Santa Felícia). Fonte: Costa <i>et al.</i> (2013).....	113
Figura 5.4 Variogramas modelados nas direções 30° e 25° (região central). Fonte: Costa (2013)	114
Figura 5.5 Mapas interpolados para as duas regiões em análise. Fonte: Costa (2013)	115
Figura 5.6 Representação do método proposto. Fonte Lindner e Pitombo (2015).....	118
Figura 5.7 Variogramas ajustados para a direção principal e ortogonal. Fonte: Lindner e Pitombo (2015)	120
Figura 5.8 Mapa krigado para a probabilidade de preferência por transporte público em relação ao transporte individual motorizado. Fonte: Lindner e Pitombo (2015)	121
Figura 5.9 Mapa krigado para a probabilidade de preferência por automóvel – região Santa Felícia – São Carlos. Fonte: Gomes <i>et al.</i> (2016).....	125
Figura 5.10 Mapa krigado para a variável preferência por modo motorizado. RMSP. Fonte: Lindner (2015)	125
Figura 6.1 Montagem da linha de pesquisa – Avanços metodológicos e etapas futuras	130
Figura 6.2 Limitações e avanços da linha de pesquisa	133
Figura 6.3 Futuras pesquisas.....	134
Figura 6.4 Mapa temático da cidade fictícia relativo à escolha do modo de transporte (0 – ônibus, 1 – carro)	138

Figura 6.5 Representação da cidade após a simulação proposta.....	138
Figura 6.6 Banco de dados a ser utilizado no Doutorado 3.....	139
Figura 6.7 Exemplo de representação do processo de amostragem sistemática: (a) amostra por zona de tráfego; (b) geração da malha e técnicas de simulação. Fonte: Lindner e Pitombo (2014)	141

LISTA DE TABELAS

Tabela 2.1 Modelos teóricos de variogramas. Fonte: Yamamoto e Landim (2013).....	23
Tabela 2.2 Trabalhos que utilizaram a Krigagem Ordinária como estimador	28
Tabela 2.3 Trabalhos que utilizaram a Krigagem Indicativa como estimador	32
Tabela 2.4 Trabalhos que utilizaram a Krigagem com Deriva Externa como estimador.....	32
Tabela 3.1 Síntese dos dados relacionados aos trabalhos abordados	40
Tabela 3.2 Variáveis socioeconômicas agregadas. Fonte: METRÔ (SP) - 1997	42
Tabela 3.3 Variáveis de viagens agregadas. Fonte: METRÔ (SP) - 1997.....	43
Tabela 3.4 Medidas descritivas das principais variáveis agregadas	43
Tabela 3.5 Informações dos municípios da RMSP. Fonte: Lindner (2015).....	46
Tabela 3.6 Informações utilizadas da Pesquisa O/D RMSP 2007. Fonte: Lindner (2015).....	47
Tabela 3.7 Medidas descritivas para as variáveis quantitativas. Fonte: Lindner (2015)	48
Tabela 3.8 Variáveis utilizadas nos trabalhos. Fonte: Rocha (2014).....	51
Tabela 3.9 Variáveis agregadas por MZs - 2010.....	54
Tabela 3.10 Variáveis desagregadas utilizadas nos trabalhos – Costa (2013); Pitombo <i>et al.</i> (2015a); Pitombo <i>et al.</i> (2015b); Gomes <i>et al.</i> (2016a).....	57
Tabela 3.11 Medidas descritivas das variáveis quantitativas – São Carlos 2007/2008.....	58
Tabela 3.12 Identificação das diferenças metodológicas entre trabalhos.....	63
Tabela 4.1 Trabalhos a serem discutidos no Capítulo 4.....	72
Tabela 4.2 Principais resultados da abordagem tradicional. Fonte: Lindner <i>et al.</i> (2016)	84
Tabela 4.3 Medidas de erros das duas abordagens: <i>Factorial Kriging with External Drift</i> e Regressão Linear. Fonte: Lindner <i>et al.</i> (2016).....	85
Tabela 4.4 Resultados do modelo calibrado para estimação de produção de viagens. Fonte: Rocha (2014).....	89
Tabela 4.5 Exemplo de ponderação da área e determinação da variável para amostragem sistemática artificial (célula ou quadrícula 2). Fonte: Rocha <i>et al.</i> (2014)	95
Tabela 4.6 Dados obtidos pela validação cruzada da estimação das variáveis	98

Tabela 5.1 Trabalhos discutidos no Capítulo 5.....	107
Tabela 5.2 Medidas estatísticas avaliadas nas modelagens. Fonte: Lindner e Pitombo (2015)	119
Tabela 5.3 Prós e contras da aplicação dos dois procedimentos utilizados nos seis trabalhos	127

LISTA DE QUADROS

Quadro 4.1 Conceitos básicos da etapa de geração de viagens (OTÚZAR e WILLUMSEM, 2011)	
.....	70
Quadro 4.2 Exemplos de trabalhos relativos ao tema análise espacial em áreas para estimativa de geração de viagens.....	71
Quadro 4.3 Síntese dos seis trabalhos discutidos (geração de viagens x geoestatística)	101
Quadro 5.1 Conceitos básicos da etapa de escolha modal	104
Quadro 5.2 Exemplos de trabalhos relativos ao tema análise espacial em escolhas discretas	106
Quadro 5.3 Síntese dos seis trabalhos discutidos (escolha modal x geoestatística)	128

SUMÁRIO

Capítulo 1: Introdução	1
1.1 Apresentação e contextualização do tema	2
1.2 Objetivos	5
1.3 A linha de pesquisa	5
1.3.1 Os produtos	7
1.3.2 Avanços e desafios	8
1.3.3 Questões de pesquisa	11
1.4 Estrutura do trabalho	11
 Capítulo 2: Geoestatística	 13
2.1 Modelagem geoestatística	13
2.2 Variáveis Regionalizadas (VR)	14
2.3 Variogramas: construção e modelagem	17
2.4 Krigagem	25
2.4.1 Krigagem Ordinária (KO)	26
2.4.2 Krigagem Indicativa (KI)	29
2.4.3 Krigagem com Deriva Externa (KDE)	32
2.5 Validação cruzada	33
2.6 O problema de espacialização de áreas por meio de centroides	34
2.7 O uso da geoestatística para dados de contagens	37
 Capítulo 3: Materiais e método geral	 39
3.1 Sintetizando os estudos de caso	39
3.2 Região Metropolitana de São Paulo (1997)	41
3.3 Região Metropolitana de São Paulo (2007)	44
3.4 Região Metropolitana de Salvador (1995)	49
3.5 Município de Salvador (censo demográfico 2010)	53
3.6 Município de São Carlos (2007/2008)	55
3.7 Aplicativos utilizados	60
3.8 Método geral	61

3.8.1 Tratamento da base de dados	63
3.8.2 Visualização espacial do conjunto de pontos	63
3.8.3 Obtenção de variogramas experimentais	64
3.8.4 Modelagem de variogramas experimentais	64
3.8.5 Validação	65
3.8.6 Aplicação do interpolador ou estimador espacial	65
3.8.7 Obtenção de superfícies (mapas krigados)	65
Capítulo 4: Geração de viagens e geoestatística	67
4.1 Modelagem tradicional de demanda por transportes: um quadro resumo	67
4.1.1 Geração de viagens	68
4.1.2 Distribuição de viagens	68
4.1.3 Divisão modal	68
4.1.4 Alocação de tráfego	69
4.2 Análise espacial de geração de viagens: <i>Background</i>	70
4.3 Sistematizando os trabalhos	71
4.4 Geração de viagens e geoestatística: uma análise agregada	72
4.4.1 Krigagem Ordinária e dados agregados por zonas	73
4.4.2 Análise em Componentes Principais e Krigagem com Deriva Externa: dados agregados	77
4.4.3 Krigagem com Deriva Externa para projeções de viagens futuras: áreas muito agregadas	86
4.5 Geração de viagens e geoestatística: uma análise desagregada	92
4.5.1 Krigagem Ordinária para estimação de geração de viagens: proposta simplificada para desagregação de dados em áreas	92
4.6 Considerações finais	101
Capítulo 5: Escolha modal e geoestatística	103
5.1 Modelos de escolha modal	103
5.2 Análise espacial relativa à escolha modal: <i>Background</i>	105
5.3 Sistematizando os trabalhos	106
5.4 Escolha modal e geoestatística: procedimento metodológico sequencial multivariado ...	107
5.4.1 Método: Árvore de Decisão e Krigagem Ordinária	108
5.4.2 Método: Logit Binomial e Krigagem Ordinária	117

5.5 Escolha modal e geoestatística: etapa única univariada.....	122
5.5.1 Krigagem Indicativa para escolha modal.....	122
5.6 Considerações finais.....	127
Capítulo 6: Conclusões: respostas, avanços e futuras pesquisas	129
6.1 Sintetizando avanços metodológicos	129
6.2 Futuras pesquisas	134
6.2.1 Pós-doutorado: Proposta de modelos autoregressivos logísticos para problemas de escolha modal: uma abordagem através de conceitos geoestatísticos	134
6.2.2 Iniciação Científica 1: Incorporação de localização espacial em modelos de escolha discreta	137
6.2.3 Mestrado 1: Introdução de conceitos de variografia a modelos de escolha discreta ..	138
6.2.4 Doutorado 3: Incorporação de geoestatística a previsões de escolha modal	139
6.2.5 Doutorado 1: Proposta de desagregação de dados para aprimoramento da modelagem geoestatística aplicada à demanda por transportes	139
6.2.6 Doutorado 2: Uso conjunto de técnicas tradicionais e de geoestatística para demanda por transportes: viagens baseadas em atividades	141
6.3 Respondendo às indagações.....	142
Referências bibliográficas	145

Apêndice digital: Trabalhos que fazem parte da análise crítica deste documento

Capítulo

1

INTRODUÇÃO

O cubo mágico é um brinquedo popular e inteligente, que requer paciência e técnica para sua solução. Na matemática, um cubo mágico é uma figura de três dimensões, com números inteiros arranjados da forma $n \times n \times n$. Existem algumas variações desse quebra-cabeça, criadas ao longo dos anos, geralmente por matemáticos, tais como Cubo de Rubik (Ernő Rubik); Cubo Bedlam (Bruce Bedlam); Cubo Soma (Piet Hein). Existem quase 20.000 maneiras de montar o cubo e, ao longo desse texto, será realizada uma série de analogias entre os avanços na montagem do cubo e na linha de pesquisa que é título deste documento.

A primeira etapa, que corresponde ao cubo completamente desmontado, conforme ilustrado no início deste capítulo, é relativa à motivação inicial que seria a incorporação da geoestatística a problemas relativos à demanda por transportes no intuito de melhorar estimativas. A etapa final, que corresponde à peça completamente montada, é relacionada à obtenção de estimativas mais precisas de demanda por transportes; auxílio a planejadores para tomada de decisão relacionada à oferta de transportes – Figura 1.1.

Assim, esta introdução visa essencialmente apresentar a ideia e a motivação para construção da linha de pesquisa. Este texto inicial se propõe a responder indagações importantes: (1) Quando se iniciou a linha de pesquisa e por qual razão? (**Apresentação e contextualização do tema**); (2) O quê fazer? (**Objetivos**); (3) Quais os principais produtos já obtidos e razões para continuar pesquisando acerca do tema? (**A linha de pesquisa**).

<ol style="list-style-type: none"> 1. Construir bases teóricas (Cap 2); 2. Propor método (Cap 3); 3. Testar técnicas para variáveis numéricas (Cap 4); 4. Testar técnicas para variáveis qualitativas (Cap 5); 5. Verificar limitações metodológicas (Cap 6); 6. Observar incrementos (Cap 6); 7. Realizar adequações metodológicas (Cap 6).
<p>Ideia inicial: Incorporar geoestatística a problemas relativos à demanda por transportes.</p>		<p>Resultado final: Estimativas mais precisas de demanda por transporte; auxílio a planejadores para tomada de decisão relacionada à oferta de transportes.</p>

Figura 1.1 Construção da linha de pesquisa – analogia com a montagem do cubo mágico

1.1. APRESENTAÇÃO E CONTEXTUALIZAÇÃO DO TEMA

Este texto apresenta uma síntese e análise crítica das principais conclusões obtidas através das pesquisas realizadas ao longo dos últimos oito anos pela docente. Desta forma, não se trata de uma tese original, mas de um documento que sintetiza, sistematiza e apresenta os resultados dos principais trabalhos desenvolvidos pela autora, desde o seu projeto de pós-doutoramento.

O foco destas investigações é a aplicação de conceitos e técnicas de Geoestatística a problemas relativos à demanda por transportes. Todo o esforço investigativo aqui realizado teve apoio na hipótese sugerida a partir da Figura 1.2.

Figura 1.2 Distribuição espacial de uma variável discreta qualquer em uma área fictícia

A Figura 1.2 ilustra a distribuição espacial de uma variável discreta em uma área fictícia. Observa-se um padrão espacial da variável em questão, considerando uma direção específica. A figura esboça o principal pressuposto da Primeira Lei da Geografia, que afirma que todos os objetos no espaço são relacionados, embora objetos mais próximos sejam mais semelhantes que aqueles mais distantes (TOBLER, 1970).

Imaginando que a variável ilustrada na Figura 1.2 pudesse ser uma variável relativa à demanda por transportes e a posição espacial equivalente à localização dos domicílios de residência dos indivíduos, sugere-se a seguinte hipótese, que norteou o desenvolvimento da linha de pesquisa a ser descrita neste documento:

Hipótese: A localização espacial influencia variáveis relativas à demanda por viagens.

De fato, o comportamento relacionado a viagens é influenciado por atributos eminentemente geográficos, tais como: adensamentos residenciais e de atividades socioeconômicas, proximidade entre zonas, cobertura espacial da rede de transporte, impedância da viagem na malha rodoviária, etc. (LOUREIRO *et al.*, 2006). Assim, é fortemente recomendada a consideração do atributo espacial na análise do comportamento de viajantes urbanos.

A partir dos avanços tecnológicos e da disponibilidade de informações georreferenciadas, a análise espacial de demanda por transportes passa a ser identificada como uma linha de pesquisa emergente (PÁEZ *et al.*, 2013). Nos últimos anos, pesquisadores têm encontrado interdependência entre localização espacial e demanda por transportes a partir de diversas abordagens, como exemplificado na sequência.

Bhat e Zhao (2002) identificaram variáveis espaciais a serem incorporadas à modelagem de demanda por transportes. Os autores propuseram um modelo logit hierárquico espacial para previsão de viagens na Região Metropolitana de Boston, Estados Unidos. Posteriormente, o

uso de estimador de Kernel foi proposto por Ben Akiva *et al.* (2004), associando-o ao modelo logit para estimativa de alocação de tráfego.

Lopes (2005) comparou projeção de demanda por transportes, especificamente na fase de previsão de viagens produzidas de base domiciliar, realizada através de modelos convencionais e de modelos alternativos, nos quais foram introduzidos indicadores para medir a dependência espacial. Concluiu-se que a introdução de variáveis espaciais é importante, já que melhores resultados foram obtidos com modelos alternativos, tanto na etapa de calibração e diagnóstico dos modelos, como na etapa de validação.

Loureiro *et al.* (2006) descreveram uma metodologia de calibração de modelos de Regressão Linear Geograficamente Ponderada, aplicada à geração de viagens no sistema de transporte rodoviário intermunicipal do estado do Ceará. Os autores analisaram resultados obtidos com diferentes tipos de modelos de regressão espacial, comparando-os com um modelo de regressão linear múltipla tradicional. Os melhores resultados obtidos foram atribuídos aos modelos de regressão que incluíam variáveis espaciais.

Peer *et al.* (2013) aplicaram modelos de Regressão Geograficamente Ponderada para estimar tempos de viagens. Em seguida, Páez *et al.* (2013) introduziram um novo indicador espacial, aplicável a modelos de escolha discreta.

Dos métodos e abordagens mencionados (regressão espacial e modelos logit baseados em fatores espaciais), é importante enfatizar que os resultados são provenientes de estimações limitadas a coordenadas geográficas em que os valores (de um atributo de interesse) são previamente conhecidos (domicílios amostrados em pesquisas domiciliares, por exemplo).

Dentre as abordagens de Estatística Espacial, a Geoestatística permite estimações em localizações geográficas onde não se tem conhecimento prévio dos valores das variáveis. (domicílios não amostrados em pesquisas tradicionais, por exemplo), desde que as mesmas possuam estrutura espacial (Primeira Lei da Geografia).

Alguns estudos recentes demonstram o potencial da aplicação da Geoestatística a problemas de Engenharia de Transportes, assim como o uso dos resultados, tais como mapas de krigagem, para auxílio ao Planejamento de Transportes.

Pode-se perceber, no entanto, que a maioria dos trabalhos refere-se a estudos de engenharia de tráfego (CIUFFO *et al.*, 2011; MAZZELLA *et al.*, 2012; ZHANG e WANG, 2013), ou de emissões veiculares (PEARCEA *et al.*, 2009, van de KASSTEELE e VELDER, 2006; van de KASSTEELE e STEIN, 2006). Além disso, a aplicação da geoestatística a problemas relativos à demanda por transportes é extremamente recente (YOON *et al.*, 2014; PITOMBO *et al.*, 2015a; PITOMBO *et al.*, 2015b; CHEN *et al.*, 2015; GOMES *et al.*, 2016a).

1.2. OBJETIVOS

Os objetivos deste documento são:

- Apresentar as principais pesquisas, desenvolvidas nos últimos oito anos pela autora e colaboradores, envolvendo o tema ***Geoestatística aplicada a problemas relativos à demanda por transportes***;
- Sintetizar as principais conclusões relativas aos trabalhos desenvolvidos e em andamento;
- Realizar uma análise crítica acerca dos resultados já obtidos;
- Apontar as contribuições e avanços para melhoria das estimativas;
- Enumerar desafios científicos e tecnológicos e os meios para superá-los;
- Sugerir aplicações futuras relacionadas ao tema.

1.3. A LINHA DE PESQUISA

A linha de pesquisa intitulada ***Geoestatística aplicada a problemas relativos à demanda por transportes*** foi iniciada pela docente em 2008 durante o seu projeto de pós-

doutoramento, realizado na Universidade de Lisboa (Departamento de Engenharia de Minas – Centro de Recursos Naturais e Ambiente - CERENA), em parceria com a *University of Leeds (School of Geography)*. Em seguida, as pesquisas foram realizadas na Universidade Federal da Bahia (atual PPEC – Programa de Pós-Graduação em Engenharia Civil). No momento atual, a linha é desenvolvida no Programa de Pós-Graduação de Engenharia de Transportes da Escola de Engenharia de São Carlos (Universidade de São Paulo), com apoio de pesquisadores de instituições nacionais e internacionais e através de orientações de Mestrado, Doutorado e Iniciação Científica.

Neste contexto, são desenvolvidas pesquisas associadas ao uso de ferramentas da Estatística Espacial, sobretudo para a área de Planejamento de Transportes, mais especificamente, estimação de demanda por transportes.

Além disso, contou e conta com o apoio financeiro de Agências de Fomento (Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq; Coordenação e Aperfeiçoamento de Pessoal de Nível Superior - CAPES; Fundação de Amparo à Pesquisa no Estado de São Paulo - FAPESP; Fundação para Ciência e Tecnologia - FCT; Fundação *Calouste Gulbenkian*), tanto para realizações de trabalhos de graduação e pós-graduação, quanto para desenvolvimento de projeto de pesquisa.

Envolvendo esta temática, foi finalizado um Projeto Universal CNPq (Processo 482754/2010-9) e existem mais dois projetos em andamento, financiados pelo CNPq (Processo 462713/2014-8) e FAPESP (Processo 2013/25035-1). A Figura 1.3 apresenta os principais pesquisadores colaboradores da linha de pesquisa (Universidade de Lisboa, Universidade Federal do Ceará, Universidade de São Paulo – Escola Politécnica, Universidade Federal de Santa Catarina).

Aline Costa – PPEC - UFBA	Anabele Lindner EESC - USP	Ana Rita Salgueiro UFC

Antonio Jorge Sousa Ulisboa	Cassiano Isler UFSC	Cira Souza Pitombo EESC - USP

José A. Quintanilha Poli - USP	Lucas Assirati EESC - USP	Marina Urano EESC - USP

Monique Gomes EESC - USP	Samille Rocha EESC - USP	Viviani Antunes UFTM – EESC - USP

Figura 1.3 Principais colaboradores da linha de pesquisa

1.3.1 Os produtos

Através da referida linha de pesquisa, foram obtidos produtos relacionados à formação de recursos humanos, disseminação e divulgação.

Desta forma, sob a supervisão da docente, quatro dissertações de Mestrado foram defendidas, totalmente relacionadas ao tema (geração de viagens, escolha modal e geoestatística). Atualmente, há um projeto de Mestrado e um de Iniciação Científica, em andamento, envolvendo análise espacial de escolhas discretas de transportes. Além disso, há quatro Doutorados em andamento, também associados ao tema aqui discutido.

Foi publicado um total de cinco artigos em periódicos (três nacionais e dois internacionais). Além disso, ao longo deste período, foram apresentados e publicados vinte e um trabalhos em anais de congressos nacionais e internacionais. Este documento traz uma síntese dos resultados dos principais trabalhos, doze no total, agrupados segundo os temas: (1) Geração de viagens e Geoestatística e (2) Escolha modal e Geoestatística. Uma compilação dos trabalhos é apresentada na Figura 1.4. Destacados em negrito na cor azul estão os trabalhos que fazem parte da análise crítica apresentada neste texto.

Figura 1.4 Compilação de publicações provenientes da linha de pesquisa

1.3.2 Avanços e desafios

No ano seguinte ao início do desenvolvimento da linha de pesquisa (ainda na realização do projeto de pós-doutoramento) foram obtidos os primeiros resultados. Ao longo dos anos, alguns avanços foram realizados, considerando limitações metodológicas e adaptação da técnica para uso em problemas de transportes (em geral variáveis discretas, sem continuidade espacial).

O trabalho de Pitombo e Sousa (2009) apresentou como objetivo estimar taxas de produção e atração de viagens por modo de transporte e motivo de viagem, através de procedimentos geoestatísticos (Krigagem Ordinária). Os resultados demonstraram ser possível

estimar em pontos não amostrados, ou seja, sem conhecimento prévio dos valores das variáveis relativas a viagens. O trabalho, no entanto, possui limitações metodológicas relacionadas ao problema de espacialização de áreas por meio de centroides, à sua característica de estimação univariada e distribuição não-normal da variável de interesse.

No ano seguinte, Pitombo *et al.* (2010) compararam duas diferentes técnicas de análise de dados espaciais para a previsão de produção e atração de viagens urbanas na Região Metropolitana de São Paulo (RMSP). As técnicas utilizadas foram: geoestatística multivariada (Krigagem com Deriva Externa) e a Regressão Geograficamente Ponderada (RGP). Os resultados mostram que os dois modelos são considerados bem ajustados para a previsão de geração de viagens. Detectou-se, como benefício da regressão, a possibilidade de visualizar os resultados (inclusive coeficientes estimados das variáveis explicativas e suas diferenças de acordo com a unidade de área) na região de estudo. Já no caso da Krigagem com Deriva Externa, um benefício é a possibilidade de estimar variáveis em diferentes localizações geográficas, cujos valores são previamente desconhecidos. Apesar de utilizar dados agregados (centroides), há a proposta de desenvolvimento de análises multivariadas. Adicionalmente, é adotada a suposição de normalidade da variável de interesse, sendo esta uma das limitações metodológicas.

Posteriormente, Rocha *et al.* (2014) propuseram um método simplificado para desagregação de dados, aplicado ao caso onde não há pesquisas domiciliares, para estimação de produção e atração de viagens. Neste caso, foram finalmente utilizados dados desagregados sintéticos e geoestatística multivariada (Krigagem com Deriva Externa - KDE). As estimativas obtidas foram significativamente superiores às aquelas anteriormente obtidas. O estudo de caso foi a Região Metropolitana de Salvador, Bahia. Uma das limitações inerentes ao método proposto é referente à suposição de normalidade das variáveis secundária e primária.

Para o caso da previsão da escolha modal, Lindner (2015) propôs um método de estimação espacial univariada para variáveis binárias. Este caso foi aplicado à Região Metropolitana de São Paulo, utilizando a Pesquisa Domiciliar (O/D) realizada em 2007. A autora estimou escolhas por modo de transporte motorizado, através da Krigagem Indicativa. Fazendo uma análise dos resultados, verificou-se que modelos multivariados tradicionais, como o logit binomial, possuem melhor poder preditivo quando comparados à abordagem espacial univariada.

Posteriormente, Pitombo *et al.* (2015a), Pitombo *et al.* (2015b) e Lindner e Pitombo (2015), propuseram um método com etapa multivariada não espacial e etapa univariada espacial. A partir deste método, é possível obter mapas krigados de projeções futuras, havendo valores de variáveis explicativas projetados. Finalmente, Gomes *et al.* (2016a) propuseram uma análise comparativa de estimativas de escolha modal provenientes de técnica espacial univariada (Krigagem Indicativa) e do método sequencial multivariado, anteriormente proposto.

Até o presente momento, foram obtidos avanços metodológicos. No entanto, existem ainda desafios importantes para análises futuras, tais como:

(1) construir modelos onde são consideradas as estruturas espaciais (variogramas) das variáveis explicativas e não somente da variável dependente;

(2) aprimorar métodos para utilização da geoestatística em termos de projeções futuras;

(3) analisar procedimentos para incorporação de conceitos da geoestatística em modelos de escolha discreta, a fim de incrementar previsões de variáveis discretas de demanda por transportes, como escolha modal, período do dia ou atividade, por exemplo;

(4) propor utilização da técnica para problemas de alocação de tráfego e distribuição de viagens;

(5) propor aplicação da técnica em conjunto com a abordagem de viagens baseadas em atividades;

(6) propor técnicas de simulação no processo de obtenção de dados desagregados espacialmente correlacionados.

1.3.3 Questões de pesquisa

Esta linha de pesquisa é uma tentativa de melhor compreender a forma como os indivíduos realizam escolhas relativas a viagens, considerando, dentre outros aspetos, localizações espaciais. O conjunto de trabalhos a ser apresentado neste documento visa responder algumas indagações:

- Que incremento o uso da Geoestatística pode trazer na previsão de demanda por transportes?
- Os conceitos e técnicas de Geoestatística podem ser aplicados para qual tipo de dado de transportes? Qual informação é obrigatória para este fim? Quais as principais restrições à sua aplicação na área de modelagem de demanda por transportes?
- Qual a principal contribuição acadêmica? E técnica?
- Um procedimento conjunto ou sequencial de métodos tradicionais e modelagem geoestatística pode ser um caminho razoável para obtenção de projeções futuras de mapas krigados?
- Já que demanda por transportes é um problema pontual distribuído ao longo de uma área urbana, qual a utilidade de superfícies contínuas de variáveis de demanda?

1.4. ESTRUTURA DO TRABALHO

O presente texto é constituído de cinco capítulos, além dessa introdução, conforme descrito em seguida.

- i. **O Capítulo 2** traz a fundamentação teórica relacionada à técnica utilizada, a Geoestatística. Apresenta conceitos essenciais, além de problemas metodológicos inerentes à sua aplicação relativa à demanda por transportes;
- ii. **O Capítulo 3** descreve o método geral, além dos materiais utilizados nos doze trabalhos que esse texto sintetiza: diferentes bases de dados e aplicativos utilizados;
- iii. **O Capítulo 4** sistematiza os seis trabalhos que abordam o tema geoestatística e geração de viagens;
- iv. **O Capítulo 5** apresenta, sucintamente, os seis trabalhos que fazem parte da temática geoestatística e escolha modal;
- v. **O Capítulo 6**, finalmente, sintetiza as principais conclusões obtidas, objetivando responder às indagações propostas nesta introdução. Também são apresentados trabalhos futuros a serem realizados nos próximos anos.

Capítulo

2

GEOESTATÍSTICA

A estimativa de um determinado valor de um ponto (Z), localizado em um par de coordenadas geográficas (x,y), é influenciada pelos valores nos demais pontos ($Z_1, Z_2, \dots, Z_n$) e inversamente proporcional às distâncias em relação ao ponto de interesse.

Essa é uma definição compartilhada por diferentes procedimentos. A principal diferença concentra-se no método de cálculo dos ponderadores. A geoestatística considera uma ferramenta importante para este cálculo, denominada variograma. O professor Georges Matheron e Daniel Krige, apresentaram, nos anos 1960, uma série de trabalhos que os caracterizaram como “criadores” da geoestatística.

Este capítulo traz definições importantes, relativas à ferramenta comum aos doze trabalhos descritos neste documento. Além disso, descreve o primeiro passo na montagem do “cubo mágico”, o qual corresponde à construção da base teórica.

2.1. MODELAGEM GEOESTATÍSTICA

O processo de modelagem geoestatística é mais bem definido quando esquematizado pelas etapas de: (1) Análise das Variáveis Regionalizadas; (2) Cálculo e modelagem de variogramas experimentais; (3) Validação e (4) Krigagem. Tais etapas são descritas nas seções subsequentes. A autora sugere como fontes mais detalhadas para as definições descritas neste

capítulo: Matheron (1971); Landim (2003); Soares (2006); Wackernagel (2010); Yamamoto e Landim (2013).

2.2. VARIÁVEIS REGIONALIZADAS (VR)

A teoria das Variáveis Regionalizadas, desenvolvida por Matheron, diz que, sendo a Variável Regionalizada (VR) distribuída no espaço, esta pode ser vista como resultado de uma função aleatória e uma componente estrutural não aleatória, associada a um valor médio constante ou uma tendência constante no espaço. Esta teoria consolida o alicerce da geoestatística (CAMARGO *et al.*, 2001).

A VR deve possuir alguma estrutura espacial, podendo apresentar características anisotrópicas ou isotrópicas. A estrutura espacial de uma Variável Regionalizada pode apresentar comportamento diferente dependendo da direção de estudo (anisotropia). Já um fenômeno espacial é considerado isotrópico quando a VR apresenta comportamento semelhante em todas as direções, variando em função apenas da distância.

Em geoestatística, a covariância mede a relação entre valores da mesma variável, obtidos em pontos separados por uma distância h , conforme uma determinada direção. Isso significa que, alterando a direção, a covariância também pode ser alterada e, neste caso, o fenômeno classifica-se como anisotrópico. A Figura 2.1, adaptada de Lindner (2015), ilustra os fenômenos espaciais isotrópicos e anisotrópicos. A variabilidade dos dados pode acontecer em uma dada direção particular (SOARES, 2006). A direção principal refere-se ao ângulo θ de máxima variabilidade espacial dos dados. No exemplo de anisotropia da Figura 2.1 (b), a direção principal é visualizada como $\theta = -45^\circ$.

Figura 2.1 Representação de uma variável espacial isotrópica (a) e anisotrópica (b). Adaptado de Lindner (2015).

Considerando a Variável Regionalizada, ilustrada no capítulo anterior (Figura 1.2), pode-se afirmar que a mesma possui características anisotrópicas (Direção -45°), sendo a sua componente estrutural espacial ilustrada na Figura 2.2.

São selecionados os pontos localizados na direção principal verificada e observa-se, a partir do ponto 1, diminuição dos valores da variável discreta com o aumento da distância observada a partir do ponto 1 (Figura 2.2 – A, B e C). Plotando-se os valores da VR e das distâncias, verifica-se tendência na estrutura espacial dos dados (Figura 2.2 – D). A autora demonstra, através da Figura 2.2, uma tendência espacial linear, com valor alto de Coeficiente de Determinação ($R^2 = 0,9231$).

Figura 2.2 Análise prévia da estrutura espacial da VR (área fictícia).

2.3. VARIOGRAMAS: CONSTRUÇÃO E MODELAGEM

A ferramenta primária da modelagem geoestatística é o variograma. Este representa graficamente a estrutura espacial da VR. A função do variograma é dada segundo a Equação 2.1. “N” representa o conjunto de pares de pontos, $z(x_i)$ e $z(x_i + h)$ são valores de pontos localizados em “i” e “i+h”, respectivamente.

$$\gamma(h) = \frac{1}{2N} \sum_{i=1}^{n(h)} [Z(x_i) - Z(x_i + h)]^2 \quad (2.1)$$

Para construção ou cálculo dos variogramas experimentais, devem-se definir alguns parâmetros, tais como direção, tolerância angular, *Lag*, tolerância do *Lag* e largura máxima, sobretudo para distribuição irregular dos pontos (YAMAMOTO e LANDIM, 2013).

- A largura máxima é o maior valor de distância considerado para cálculo do variograma. Em geral, é equivalente à metade da maior distância entre pontos da amostra;
- *Lag (h)*: distância entre pontos vizinhos;
- Tolerância de *Lag* (Δh): valores dentro dos limites $(h + \Delta h)$ e $(h - \Delta h)$ são considerados como pertencentes a h ;
- Direção angular (θ): direção adotada para elaboração do variograma que melhor representa a distribuição espacial dos dados (anisotropia);
- Tolerância angular ($\Delta\theta$): valores dentro dos limites $(\theta + \Delta\theta)$ e $(\theta - \Delta\theta)$ são considerados como pertencentes a θ ;
- Direção ortogonal ($\theta + 90^\circ$): direção ortogonal à direção adotada para elaboração do variograma para casos de anisotropia.

A Figura 2.3 ilustra tais parâmetros adotados para a VR exemplificada e ilustrada inicialmente na Figura 1.2. Para este caso, foram utilizados os seguintes parâmetros: Lag = 50 m; Tolerância do Lag = 15 m; Direção angular = - 45°; Tolerância angular = $\pm 5^\circ$; Largura Máxima = 600 m.

Ponto	Pares	Total	Ponto	Pares	Total	Ponto	Pares	Total
1	1-2; 1-13	2	8	8-9; 8-7; 8-21; 8-24	4	21	21-18; 21-7; 21-8	3
2	2-1; 2-13; 2-3	3	9	9-24; 9-8; 9-10; 9-25	4	24	24-21; 24-23; 24-8; 24-9; 24-25	4
3	3-2; 3-4; 3-13	3	10	10-9; 10-11; 10-25; 10-31	4	25	25-23; 25-24; 25-10; 25-9	4
4	4-3; 4-15; 4-5	3	11	11-10; 11-12; 11-26; 11-25; 11-35; 11-31	6	26	26-11; 26-12; 26-25	3
5	5-15; 5-4; 5-16; 5-6	4	12	12-11; 12-26; 12-35; 12-36	4	31	31-10; 31-11; 31-30; 31-34; 31-35	4
6	6-5; 6-20; 6-18; 6-7	4	13	13-2; 13-1; 13-3	3	33	33-34; 33-35; 33-36	3
7	7-6; 7-20; 7-18; 7-8	4	18	18-6; 18-7; 18-21	3	35	35-12; 35-11	2
						36	36-12; 36-35	2
E. Pares de pontos para parâmetros adotados. Total = 78								

Figura 2.3 Esquema mostrando a pesquisa de pares de pontos

O processo para construção do variograma experimental segue o procedimento ilustrado por Lindner (2015). Considera-se o Lag de tamanho h com tolerância Δh e a direção angular θ com tolerância angular $\Delta\theta$; nlags o número de Lags, $N = [n_1, n_2, n_3, \dots, n_N]$ o número total de pontos; e d_{corte} a distância de corte ou largura máxima.

1) O processo tem início em um primeiro ponto n_1 da amostra. Para toda observação distante $h \pm \Delta h$ e com ângulo entre $\theta \pm \Delta\theta$ deste ponto, calcula-se a diferença quadrática entre os valores dos pares (Figura 2.4);

Figura 2.4: Etapa 1 para obtenção do variograma experimental. Fonte: Lindner (2015)

2) Refaz-se a etapa 1 para todos os N pontos (Figura 2.5);

Figura 2.5: Etapa 2 para obtenção do variograma experimental. Fonte: Lindner (2015)

3) Obtém-se a variância γ para o lag de tamanho h , através da Equação 2.1. Este valor estará representado por um dos pontos γ no variograma experimental, onde o eixo y é representado pela variância γ e o eixo x pela distância h (Figura 2.6);

Figura 2.6: Etapa 3 para obtenção do variograma experimental. Fonte: Lindner (2015)

4) Repetem-se os passos de 1) até 3) para os lags $2h, 3h, \dots, n_{lags} \cdot h$, a fim de se obter $\gamma(2h), \gamma(3h), \dots, \gamma(n_{lags} \cdot h)$, desde que os valores de distância não excedam a distância de corte (d_{corte}), conforme Figura 2.7.

Figura 2.7 Etapa 4 para obtenção do variograma experimental. Fonte: Lindner (2015)

Considerando as etapas, descritas anteriormente, para cálculo do variograma experimental e os parâmetros adotados, a Figura 2.8 ilustra o cálculo considerando Lag = 50 metros, Lag 2h, Lag 3h até Lag 12h (600m).

Para distância igual a Lag (50 metros), deve ser calculada a média das variâncias (Equação 2.1) dos 78 pares de pontos encontrados considerando os parâmetros para varredura de pontos (Lag = 50 m; Tolerância do Lag = 15 m; Direção angular = - 45°; Tolerância angular = ± 5° ; Largura Máxima = 600). Em seguida, é calculada, segundo Equação 2.1, a média das variâncias para os X pares de pontos considerados a partir dos critérios de varredura e Lag 2h (100 metros). Neste exemplo, o procedimento é realizado repetidamente até a distância equivalente a Lag 12h (Largura máxima).

$\gamma(h) = \frac{1}{2N} \sum_{i=1}^{n(h)} [Z(xi) - Z(xi + h)]^2$ <p>Para Lag h = 50 metros</p>	Par 1 x 2	$\gamma(h) = \frac{1}{2}[10 - 10]^2=0$
	Par 1 x 13	$\gamma(h) = \frac{1}{2}[10 - 10]^2=0$
	Par 2 x 1	$\gamma(h) = \frac{1}{2}[10 - 10]^2=0$
	Par 2 x 3	$\gamma(h) = \frac{1}{2}[10 - 10]^2=0$
	Par 2 x 13	$\gamma(h) = \frac{1}{2}[10 - 10]^2=0$
	Par 3 x 2	$\gamma(h) = \frac{1}{2}[10 - 10]^2=0$
	Par 3 x 4	$\gamma(h) = \frac{1}{2}[10 - 8]^2=2$
	Par 3 x 13	$\gamma(h) = \frac{1}{2}[10 - 10]^2=0$
Considerando-se os demais 70 pares de pontos, tem-se o valor médio da variância equivalente a 1. A figura ao lado representa o primeiro ponto do variograma experimental.

Figura 2.8 Etapas para cálculo do variograma experimental.

Observa-se, a partir da Figura 2.8, que a função variograma mede a variância entre os pares de pontos, separados por uma distância h . Aumentando-se a distância, os valores dos pontos tornam-se mais diferentes e a variância média γ , consequentemente, aumenta. Geralmente, a variância estabiliza-se em um valor máximo, denominado patamar (C). No entanto, em alguns casos, como no exemplo da Figura 2.8, a variância continua aumentando, caracterizando os variogramas sem patamar. Já o alcance ou amplitude (a) é a distância, no eixo y , relacionada ao valor do Patamar (C). Corresponde à distância a partir da qual os pontos não se apresentam correlacionados espacialmente.

À medida que h tende a zero, a variância γ se aproxima de um valor que revela a descontinuidade do variograma para distâncias menores do que a distância média entre os pontos. Este valor é definido como efeito pepita (C_0). O efeito pepita representa a alta variabilidade da variável em distâncias pequenas (SOARES, 2006).

Finalmente, é necessário o ajuste dos pontos do variograma experimental para modelos teóricos. Os modelos teóricos que explicam a maioria dos fenômenos espaciais são o esférico, exponencial e gaussiano, representados na Figura 2.9 (a, b e c). Na mesma figura também são

representados os principais parâmetros do variograma (Figura 2.9 d). A Tabela 2.1 apresenta as equações dos modelos teóricos de variogramas ilustrados na Figura 2.9.

Tabela 2.1 Modelos teóricos de variogramas. Fonte: Yamamoto e Landim (2013).

Modelo	Equação
Esférico	$\gamma(h) = C_o + C \left[1,5 \frac{h}{a} - 0,5 \left(\frac{h}{a} \right)^3 \right] \text{ para } h < a$ $\gamma(h) = C_o + C \text{ para } h \geq a$
Exponencial	$\gamma(h) = C_o + C \left[1 - e^{-\frac{h}{a}} \right]$
Gaussiano	$\gamma(h) = C_o + C \left[1 - e^{-\left(\frac{h}{a} \right)^2} \right]$
Para C_o = Efeito pepita; C_1 = Contribuição, h = Lag; a = Amplitude ou Alcance

Figura 2.9: Exemplos de ajustes para variogramas experimentais e parâmetros variográficos. Fonte: Olea (2006) e Lindner (2015)

Quando o variograma experimental não apresenta patamar, o modelo teórico pode ser representado pelo variograma de potência, tal como na Equação 2.2 (OLEA, 2000).

$$\gamma(h) = \alpha h^\beta, \text{ com } 0 < \beta < 2$$

(2.2)

Considerando o exemplo utilizado ao longo deste texto (área fictícia), são propostos os seguintes ajustes dos variogramas experimentais, ilustrados na Figura 2.10.

Figura 2.10: Ajustes para o variograma experimental correspondente à área fictícia

2.4. KRIGAGEM

A Krigagem é um processo de predição linear, pois suas estimativas são combinações lineares ponderadas por dados existentes adjacentes, quando considerados espacialmente dependentes pela análise do variograma. Fornece, em geral, estimativas não tendenciosas e com variância mínima (LANDIM, 2003).

A Krigagem pode ser univariada, abrangendo uma série de algoritmos, tais como Krigagem Simples, Ordinária, Universal e Indicativa. A geoestatística multivariada abrange outros tipos de estimadores como Krigagem com Deriva Externa e Cokrigagem, por exemplo. Estimativas geoestatísticas são superiores a demais interpoladores pelo fato de considerarem, além da função distância entre os pontos, efeito pepita, amplitude e presença ou não de anisotropia (YAMAMOTO e LANDIM, 2013).

A Variável Regionalizada (VR) pode ser contínua ou discreta e, além disso, apresentar diferentes distribuições possíveis. Por esta razão, transformação de dados pode ser um procedimento necessário para obtenção de melhores estimativas. A Figura 2.11 ilustra um processo de estimativa geoestatística de diferentes tipos de VR, na qual estão destacados os estimadores de modelagem univariada utilizados nos trabalhos descritos neste documento.

Figura 2.11: Síntese do processo de estimativa geoestatística univariada. Adaptado de Yamamoto e Landim (2013).

Neste texto, serão definidas sinteticamente, nas próximas subseções, apenas os estimadores que foram utilizados nos trabalhos que fazem parte desta análise crítica dos principais produtos da linhas de pesquisa **Geoestatística aplicada a problemas relativos à demanda por transportes**.

2.4.1 Krigagem Ordinária (KO)

A Krigagem Ordinária (KO) é um método de estimativa local, que calcula a ponderação a partir das curvas ajustadas de variogramas. Desta forma, valores de efeito pepita (C_0), contribuição (C_1) e amplitude (a) são utilizados para a definição de pesos λ_i (ponderadores). A função dos ponderadores é a de ditar a influência dos dados amostrais na estimação de novos valores. O estimador de KO é dado pela Equação 2.2. Seja x_0 um local não amostrado e n valores obtidos em pontos adjacentes.

$$Z^*(x_0) = \sum_{i=1}^n \lambda_i \times Z(x_i) \quad (2.2)$$

Os valores dos ponderadores são calculados com o objetivo de conduzir estimativas a erros nulos e minimização da variância (JOURNEL e HUIJBREGTS, 1978). O sistema de equações da KO em termos de função variograma é descrito conforme Equação 2.3

$$\left\{ \begin{array}{l} \sum_{j=1}^n \lambda_j \gamma(x_i - x_j) + \mu = \gamma(x_0 - x_i) \\ \sum_{j=1}^n \lambda_j = 1 \end{array} \right. \quad (2.3)$$

A solução para os ponderadores da Krigagem Ordinária é dada pelas representações matriciais:

$$\lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \dots \\ \lambda_n \\ \mu \end{pmatrix}, \quad G = \begin{pmatrix} \gamma(x_1, x_1) & \gamma(x_1, x_2) & \dots & \gamma(x_1, x_n) & 1 \\ \gamma(x_2, x_1) & \gamma(x_2, x_2) & \dots & \gamma(x_2, x_n) & 1 \\ \dots & \dots & \dots & \dots & 1 \\ \dots & \dots & \dots & \dots & 1 \\ \gamma(x_n, x_1) & \gamma(x_n, x_2) & \dots & \gamma(x_n, x_n) & 1 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix}, \quad M = \begin{pmatrix} \gamma(x_1, x_0) \\ \gamma(x_2, x_0) \\ \dots \\ \dots \\ \gamma(x_n, x_0) \\ 1 \end{pmatrix}.$$

A Figura 2.12 ilustra o procedimento de cálculo de um ponto de valor desconhecido (ponto 37) na área fictícia exemplo deste texto. Finalmente, é ilustrado o mapa de krigagem para estimativa de diversos pontos de valores não conhecidos. Em seguida, a Tabela 2.2 apresenta os trabalhos, discutidos ao longo deste texto, nos quais foi utilizada a Krigagem Ordinária como ferramenta principal.

Figura 2.12: Procedimento ilustrativo para estimativa por Krigagem Ordinária

Tabela 2.2: Trabalhos que utilizaram a Krigagem Ordinária como estimador

TRABALHOS QUE UTILIZARAM KRIGAGEM ORDINÁRIA	TEMA PRINCIPAL
PITOMBO e SOUSA (2009)	GERAÇÃO DE VIAGENS
COSTA (2013)	ESCOLHAMODAL
ROCHA <i>et al.</i> (2014)	GERAÇÃO DE VIAGENS
ROCHA (2014)	GERAÇÃO DE VIAGENS
PITOMBO <i>et al.</i> (2014)	GERAÇÃO DE VIAGENS
PITOMBO <i>et al.</i> (2015a)	ESCOLHAMODAL
PITOMBO <i>et al.</i> (2015b)	ESCOLHAMODAL
LINDNER e PITOMBO (2015)	ESCOLHAMODAL

2.4.2. Krigagem Indicativa (KI)

A Krigagem Indicativa, ou da Indicadora, foi inicialmente proposta para modelar distribuições lognormais e é obtida por uma transformação não-linear da Variável Regionalizada. A KI evita possíveis problemas com a presença de *outliers*, ou contaminação pela presença de poucos valores altos na interpolação de regiões predominantemente com baixos valores da VR (JOURNEL, 1983; YAMAMOTO e LANDIM, 2013).

O primeiro passo da KI consiste em transformar os dados iniciais em indicadores, isto é, transformar valores que estão acima de um determinado valor de corte em “1” e os que estão abaixo em “0”.

$$i_j(v_c) = \begin{cases} 0, & \text{se } v_j \leq v_c \\ 1, & \text{se } v_j > v_c \end{cases} \quad (2.4)$$

v_c é o teor de corte e v_j é o valor observado.

A frequência acumulada dos valores observados acima do teor de corte é representada pela Equação 2.5. Analogamente, a proporção de valores acima do teor de corte pode também ser considerada como a média ponderada dos indicadores situados na vizinhança do local avaliado segundo a Equação 2.6.

$$F(v_c) = \frac{1}{n} \sum_{j=1}^n i_j(v_c) \quad (2.5)$$

$$\hat{F}(v_c) = \sum_{j=1}^n \lambda_j i_j(v_c) \quad (2.6)$$

Sendo λ_j os ponderadores, i_j os indicadores e v_c o teor de corte.

A partir dos cálculos das frequências acumuladas, da proporção de valores e dos variogramas experimentais e teóricos, são calculadas as probabilidades para determinados locais de valores desconhecidos (LIMA *et al.*, 2008).

Pode-se definir o teor de corte v_c de tal modo que este esteja no intervalo de amostragem da Variável Regionalizada $Z(x)$. A ideia é discretizar uma distribuição contínua em K teores de

corte. A KI requer um variograma teórico para cada teor de corte v_c . De forma geral, o melhor variograma corresponde ao teor de corte igual à mediana da distribuição. Este valor corresponde à divisão exata de metade dos valores equivalentes a 1 ou a zero (JOURNAL, 1983).

A Figura 2.14 ilustra, a partir da função do variograma teórico da variável indicadora da mediana representada na Figura 2.13, o procedimento de cálculo de um ponto de valor desconhecido (ponto 37) na área fictícia exemplo deste texto, através da Krigagem Indicativa da mediana. Ao final do processo, é ilustrado o mapa de KI para estimativa de superfície de probabilidades.

Figura 2.13: Variograma teórico correspondente à área fictícia (dados binários).

<table><tr><th>ID</th><th>x</th><th>y</th><th>valor</th></tr><tr><td>1</td><td>25</td><td>650</td><td>1</td></tr><tr><td>2</td><td>50</td><td>600</td><td>1</td></tr><tr><td>13</td><td>75</td><td>600</td><td>1</td></tr><tr><td>37</td><td>30</td><td>655</td><td>?</td></tr></table>	ID	x	y	valor	1	25	650	1	2	50	600	1	13	75	600	1	37	30	655	?	<table><tr><td>Ponto 1-1</td><td>Ponto 1-2</td><td>Ponto 1-13</td></tr><tr><td>Ponto 2-1</td><td>Ponto 2-2</td><td>Ponto 2-13</td></tr><tr><td>Ponto 13-1</td><td>Ponto 13-2</td><td>Ponto 13-13</td></tr></table> <table><tr><td>0</td><td>55,9</td><td>70,71</td></tr><tr><td>55,9</td><td>0</td><td>25</td></tr><tr><td>70,71</td><td>25</td><td>0</td></tr></table>	Ponto 1-1	Ponto 1-2	Ponto 1-13	Ponto 2-1	Ponto 2-2	Ponto 2-13	Ponto 13-1	Ponto 13-2	Ponto 13-13	0	55,9	70,71	55,9	0	25	70,71	25	0
ID	x	y	valor																																				
1	25	650	1																																				
2	50	600	1																																				
13	75	600	1																																				
37	30	655	?																																				
Ponto 1-1	Ponto 1-2	Ponto 1-13																																					
Ponto 2-1	Ponto 2-2	Ponto 2-13																																					
Ponto 13-1	Ponto 13-2	Ponto 13-13																																					
0	55,9	70,71																																					
55,9	0	25																																					
70,71	25	0																																					
C. Pontos de dados vizinhos para estimativa da localização (x=30; y =655) por meio de Krigagem Indicativa.	D. Cálculo das distâncias euclidianas entre os pontos de valores conhecidos.																																						
<table><tr><td>Ponto 1 -37</td><td>7,07</td></tr><tr><td>Ponto 2-37</td><td>58,52</td></tr><tr><td>Ponto 13-37</td><td>71,06</td></tr></table>	Ponto 1 -37	7,07	Ponto 2-37	58,52	Ponto 13-37	71,06	$\gamma(h) = 0,11 + 0,49 \left[1 - e^{-\left(\frac{h}{420}\right)^2} \right]$ <table><tr><td>0</td><td>0,12</td><td>0,12</td><td>1</td></tr><tr><td>0,12</td><td>0</td><td>0,11</td><td>1</td></tr><tr><td>0,12</td><td>0,11</td><td>0</td><td>0</td></tr><tr><td>1</td><td>1</td><td>1</td><td>0</td></tr></table>	0	0,12	0,12	1	0,12	0	0,11	1	0,12	0,11	0	0	1	1	1	0																
Ponto 1 -37	7,07																																						
Ponto 2-37	58,52																																						
Ponto 13-37	71,06																																						
0	0,12	0,12	1																																				
0,12	0	0,11	1																																				
0,12	0,11	0	0																																				
1	1	1	0																																				
E. Cálculo das distâncias euclidianas entre cada ponto e o ponto a ser estimado.	F. Conhecida a função variograma, a matriz de variâncias pode ser obtida.																																						
<table><tr><td>0,11</td></tr><tr><td>0,12</td></tr><tr><td>0,12</td></tr><tr><td>1</td></tr></table>	0,11	0,12	0,12	1	<table><tr><td>0</td><td>0,12</td><td>0,12</td><td>1</td><td>λ_1</td><td>0,11</td></tr><tr><td>0,12</td><td>0</td><td>0,11</td><td>1</td><td>λ_2</td><td>0,12</td></tr><tr><td>0,12</td><td>0,11</td><td>0</td><td>0</td><td>λ_3</td><td>0,12</td></tr><tr><td>1</td><td>1</td><td>1</td><td>0</td><td>μ</td><td>1</td></tr></table>	0	0,12	0,12	1	λ_1	0,11	0,12	0	0,11	1	λ_2	0,12	0,12	0,11	0	0	λ_3	0,12	1	1	1	0	μ	1										
0,11																																							
0,12																																							
0,12																																							
1																																							
0	0,12	0,12	1	λ_1	0,11																																		
0,12	0	0,11	1	λ_2	0,12																																		
0,12	0,11	0	0	λ_3	0,12																																		
1	1	1	0	μ	1																																		
G. Também é obtida a matriz de variâncias para o ponto de valor desconhecidos 37.	H. Constrói-se o sistema de equações a partir das etapas D, E, F e G.																																						
<table><tr><td>$0,11 \lambda_1 + 0,12 \lambda_2 + 0,12 \lambda_3 + \mu = 0,11$</td></tr><tr><td>$0,12 \lambda_1 + 0,12 \lambda_2 + 0,11 \lambda_3 + \mu = 0,12$</td></tr><tr><td>$0,12 \lambda_1 + 0,11 \lambda_2 + 0 \lambda_3 = 0,12$</td></tr><tr><td>$\lambda_1 + \lambda_2 + \lambda_3 = 1$</td></tr></table> <table><tr><td>$\lambda_1 = 0.56$</td><td>$\lambda_2 = 0.48$</td></tr><tr><td>$\lambda_3 = -0.04$</td><td>$\mu = 0.0572$</td></tr></table>	$0,11 \lambda_1 + 0,12 \lambda_2 + 0,12 \lambda_3 + \mu = 0,11$	$0,12 \lambda_1 + 0,12 \lambda_2 + 0,11 \lambda_3 + \mu = 0,12$	$0,12 \lambda_1 + 0,11 \lambda_2 + 0 \lambda_3 = 0,12$	$\lambda_1 + \lambda_2 + \lambda_3 = 1$	$\lambda_1 = 0.56$	$\lambda_2 = 0.48$	$\lambda_3 = -0.04$	$\mu = 0.0572$	<div>$Z * (x_o) = 0,56*1 + 0,48*1 - 0,04*1 = 1$ </div>																														
$0,11 \lambda_1 + 0,12 \lambda_2 + 0,12 \lambda_3 + \mu = 0,11$																																							
$0,12 \lambda_1 + 0,12 \lambda_2 + 0,11 \lambda_3 + \mu = 0,12$																																							
$0,12 \lambda_1 + 0,11 \lambda_2 + 0 \lambda_3 = 0,12$																																							
$\lambda_1 + \lambda_2 + \lambda_3 = 1$																																							
$\lambda_1 = 0.56$	$\lambda_2 = 0.48$																																						
$\lambda_3 = -0.04$	$\mu = 0.0572$																																						
I. A partir da solução do sistema de equações, determinam-se os ponderadores.	<div>J. Aplicando-se os ponderadores obtidos na função de Krigagem Indicativa, obtém-se a estimativa de probabilidade do valor ser maior ou igual à mediana no ponto 37.</div> <div>K. Considerando a estimativa em vários outros pontos de valores conhecidos, tem-se o mapa final de probabilidades (KI)</div>																																						

Figura 2.14: Procedimento ilustrativo para estimativa por Krigagem Indicativa da mediana.

A Tabela 2.3 apresenta os trabalhos, que fazem parte desta análise crítica, nos quais foi utilizada a Krigagem Indicativa como ferramenta principal.

Tabela 2.3: Trabalhos que utilizaram a Krigagem Indicativa como estimador

TRABALHOS QUE UTILIZARAM KRIGAGEM INDICATIVA	TEMA PRINCIPAL
LINDNER (2015)	ESCOLHAMODAL
GOMES <i>et al.</i> (2016)	

2.4.3 Krigagem com Deriva Externa (KDE)

Levando-se em conta fenômenos espaciais, por muitas vezes valores de diversas variáveis podem ser mensurados ou amostrados e associados a uma localização espacial. Problemas relativos à demanda por transportes, por exemplo, geralmente utilizam dados provenientes de pesquisas domiciliares, que coletam uma série de variáveis socioeconômicas e de viagens.

Assim, a geoestatística multivariada proporciona um conjunto de ferramentas capazes de estimar a Variável Regionalizada primária, muitas vezes subamostrada, a partir da Variável Regionalizada secundária (Cokrigagem e Krigagem com Deriva Externa – KED). É descrita aqui a Krigagem com Deriva Externa, a qual foi utilizada nos trabalhos apresentados na Tabela 2.4. A autora recomenda alguns autores para maiores detalhes teóricos e de aplicação da KED (Wackernagel, 2010; Yamamoto e Landim, 2013).

Tabela 2.4: Trabalhos que utilizaram a Krigagem com Deriva Externa como estimador

TRABALHOS QUE UTILIZARAM KRIGAGEM COM DERIVA EXTERNA	TEMA PRINCIPAL
PITOMBO <i>et al.</i> (2010)	GERAÇÃO DE VIAGENS
ROCHA (2014)	
PITOMBO <i>et al.</i> (2014)	
LINDNER <i>et al.</i> (2016)	

A KED procura fazer a estimativa de uma variável primária $Z(x)$ baseando-se em uma variável secundária $Y(x)$ altamente correlacionada, sobretudo quando a amostragem da variável primária é insuficiente. Assim, a variável secundária é usada para auxiliar na estimativa da primeira. Tal correlação pode ser descrita por meio de uma relação linear, conforme Equação 2.7. O procedimento poderia ser aplicado, por exemplo, a problemas de demanda por

transportes para estimação de viagens domiciliares a partir de número de membros da família, que é uma informação mais facilmente obtida através de pesquisas de recenseamento.

$$E[Z(x)] = a_0 + b_1 Y(x) \quad (2.7)$$

A KED considera duas restrições importantes, apresentadas na Equação 2.8 e Equação 2.9.

$$Y(x_0) = \sum_{j=1}^n \lambda_j Y(x_j) \quad (2.8)$$

$$\sum_{j=1}^n \lambda_j = 1 \quad (2.9)$$

λ_j = ponderadores; $Y(x)$ = variável secundária; $Y(x_0)$ = valor estimado da variável secundária

A estimativa dos ponderadores é dada de acordo com a matriz, onde C é a função covariância e μ é o multiplicador de Lagrange que minimiza a variância da estimação para as restrições apresentadas na Equação 2.8 e Equação 2.9.

$$\begin{bmatrix} C & 1 & Y \\ 1^T & 0 & 0 \\ Y^T & 0 & 0 \end{bmatrix} \begin{bmatrix} w \\ -\mu \\ -\mu \end{bmatrix} = \begin{bmatrix} C_0 \\ 1 \\ Y_0 \end{bmatrix} \quad (2.10)$$

2.5. VALIDAÇÃO CRUZADA

A estimativa de Krigagem calcula valores da Variável Regionalizada para os pontos não amostrados $Z^*(x_0)$. Sendo que, para pontos não amostrados, não existe conhecimento sobre valores observados $Z(x_0)$, é supostamente impossível calcular o erro associado à estimativa $Z^*(x_0) - Z(x_0)$.

Tal restrição, associada aos métodos de Krigagem, pode ser parcialmente resolvida quando os valores estimados são pontuais e coincidentes com a amostragem. Assim, a *Validação Cruzada* ou *Teste do Ponto Fictício* consiste na eliminação sucessiva do valor de um ponto conhecido $Z(x_1)$, pertencente à amostra. Em seguida, é realizada a estimação deste mesmo ponto $Z^*(x_1)$, considerando os valores dos pontos vizinhos e os ponderadores,

dependentes da função variograma (variogramas teóricos). Repete-se, seguidamente, o procedimento para um segundo ponto pertencente à amostra $Z(x_2)$. A partir da obtenção dos pares de valores observados $Z(x_i)$ e estimados $Z^*(x_i)$ da amostra de tamanho “N” é possível se calcular algumas medidas de erro para mensuração da acurácia das estimativas de Krigagem. Algumas medidas de erros são sugeridas, tais como *Erro Quadrático Médio (EQM)*, *Erro Médio Absoluto (EMA)*, *Erro Médio Relativo (EMR)*.

$$EQM = \frac{\sum (Z(x_i) - Z^*(x_i))^2}{N} \quad (2.11)$$

$$EMA = \frac{\sum (Z(x_i) - Z^*(x_i))}{N} \quad (2.12)$$

$$EMR = \frac{\sum (Z(x_i) - Z^*(x_i))}{N \sum Z(x_i)} \quad (2.13)$$

2.6. O PROBLEMA DE ESPACIALIZAÇÃO DE ÁREAS POR MEIO DE CENTROIDES

Inúmeras vezes, problemas relativos à análise espacial de dados e previsão de demanda por transportes estão associados a dados geograficamente agregados, ou seja, unidades de área. No entanto, uma questão importante conhecida na análise espacial de dados em áreas é o problema da *Unidade de Área Modificada (MAUP - Modifiable Areal Unit Problem)* ou *Falácia Ecológica*. Este problema ocorre quando se realizam análises com resultados derivados de agregação de valores por unidade de área, inferindo que tais valores correspondem ao nível individual.

Esta questão seria facilmente resolvida com obtenção de dados individuais associados às suas coordenadas geográficas. Variáveis relacionadas à demanda por transportes, tais como viagens, indivíduos e domicílios são objetos (pontuais) distribuídos ao longo da cidade. No entanto, muitas vezes há, sobretudo no caso de dados públicos, a necessidade de preservar a

confidencialidade de registros individuais, sendo a agregação geográfica a única forma facilmente disponível para este fim (MARTIN, 2000).

Uma das questões, apontada por Páez e Scott (2004), é que diferentes sistemas de zoneamento, aplicados na mesma cidade, podem gerar um leque de resultados, dependentes da unidade de agregação. Outro aspecto relevante é o nível de agregação (o efeito da escala). Resultados gerados através da mesma ferramenta de estatística espacial podem variar segundo diferentes níveis de agregação espacial para a mesma região.

Em seu trabalho, Páez e Scott (2004) revisaram diferentes estudos que objetivavam mitigar tais problemas, como Método dos polígonos de *Thiessen* (FOTHERINGHAM e WONG, 1991; HORNER e MURRAY, 2002) e divisão em células através de diferentes *grids* (WONG, 1996).

Desde a aplicação, mais recentemente, da geoestatística nas ciências sociais e geografia médica, por exemplo, ocorrem alguns problemas ou desafios metodológicos relacionados à natureza e suporte dos dados. Problemas relativos à demanda por transportes podem ser incluídos nessa categoria de trabalhos, que utiliza dados agregados por unidades de áreas de diferentes escalas, tais como setores censitários, zonas de tráfego, distritos, cidades, estados, etc. Muitas vezes tais variáveis são medidas em forma de taxas associadas a áreas de tamanhos e formas irregulares. Essa é uma mudança notável em relação à análise tradicional pontual geoestatística, desenvolvida inicialmente para estimação espacial de medidas físicas, tais como metais em solos, por exemplo.

Grande parte dos trabalhos subjacentes à área de saúde associam as coordenadas geográficas dos centroides das áreas irregulares aos valores da Variável Regionalizada (CHRISTAKOS e SERRE, 2000, GOOVAERTS, 2005). Na área de demanda por transportes, alguns trabalhos, inclusive objetos de estudo deste documento, são tratados através do mesmo procedimento (PITOMBO e SOUSA, 2009; PITOMBO *et al.*, 2010, ROCHA, 2014, PITOMBO *et*

al., 2014; LINDNER *et al.*, 2016a). No entanto, através deste procedimento, assume-se que todas as unidades – suporte – têm o mesmo tamanho e forma. Em geoestatística, define-se suporte como o domínio geométrico de onde se obtém o valor da amostra, dotado de um volume, forma e orientação. Os variogramas geralmente são calculados a partir de dados que possuam suportes iguais (JOURNEL e HUIJBREGTS, 1978).

Analogamente, em estudos agregados de demanda por transportes, o suporte é equivalente às unidades de áreas. No entanto, tais suportes seriam irregulares e considerados, erroneamente, como iguais pela ferramenta geoestatística. Trabalhos que consideram o impacto de suportes diferentes equivalentes às áreas irregulares ainda são raros na literatura. Há algumas abordagens na última década, propostas de forma a mitigar tal problema.

A abordagem de krigagem área – para – ponto (*Aera-to-point –ATP- kriging*) prevê valores pontuais a partir de unidades de área. Tal análise permite mapear a variabilidade dentro das unidades de áreas considerando a restrição de que a soma ou média dos valores previstos desagregados é equivalente ao dado agregado original (KYRIAKIDIS, 2004; GOTWAY e YOUNG, 2004). No entanto, esta abordagem necessita conhecimento da função variograma/covariância pontual da variável regionalizada. Esta é uma restrição metodológica haja vista que apenas dados agregados estariam disponíveis. A obtenção da função variograma pontual a partir do variograma experimental por áreas é chamada, em geoestatística, de deconvolução. Para o caso de suportes irregulares, algumas abordagens foram propostas recentemente (GOOVAERTS, 2008; GOOVAERTS, 2010).

Para trabalhos futuros, relativos à demanda por transportes e dados agregados por áreas, sugere-se a utilização e testes desses procedimentos propostos por Pierre Goovaerts para melhor adequação dos dados e suporte espacial às técnicas utilizadas.

2.7. O USO DA GEOESTATÍSTICA PARA DADOS DE CONTAGENS

A Figura 2.11 ilustra aplicação de diferentes estimadores para Variáveis Regionalizadas que seguem uma distribuição normal (simétrica ou assimétrica). A mesma figura aponta um estimador potencial para problemas de distribuição não-normal ou dados de contagem – Krigagem com Distribuição de Poisson.

Procedimentos alternativos para uma VR que segue distribuição populacional não-normal são fundamentais para aprimoramento de estimativas de variáveis de demanda por transportes através, ou com auxílio, da ferramenta geoestatística. Variáveis relativas à demanda por transportes geralmente são variáveis de contagens, tais como Número de viagens domiciliares ou zonais, por exemplo.

Avanços metodológicos ocorrem na área de saúde ou medicina geográfica. Na década de 90, foram observados raros trabalhos com aplicação de um modelo binomial de cokrigagem para mapeamento de riscos de câncer infantil (OLIVER *et al.*, 1993; WEBSTER *et al.*, 1994). Mais recentemente, a abordagem de Krigagem com Distribuição de Poisson foi aplicada para o caso de análise de taxas de mortalidade provenientes de câncer (GOOVAERTS, 2005).

O objetivo geral do estimador de Krigagem com Distribuição de Poisson seria estimativa de uma variável $Z(x)$ inteira e não negativa ao longo de uma área D a partir de valores conhecidos e amostrados de $Z(x_i)$ e de covariáveis colocalizadas. Uma revisão detalhada de estimadores de Krigagem com Distribuição de Poisson é descrita por Oliveira (2014). Vale ressaltar que o uso de Krigagem com Distribuição de Poisson está previsto para futuras pesquisas, com orientação da docente.

Capítulo

3 MATERIAIS E MÉTODO GERAL

Este capítulo apresenta os materiais utilizados para desenvolvimento dos doze trabalhos analisados neste documento. Inicialmente, são descritos os estudos de caso utilizados. Ressalta-se, ainda, que todos os dados utilizados foram secundários e provenientes de Pesquisas Origem/Destino ou do último censo demográfico, realizado em 2010. Posteriormente, os aplicativos utilizados para auxílio à modelagem geoestatística, análise e visualização de resultados, são sumariados. Finalmente, é sistematizado o método geral comum aos trabalhos aqui tratados. A proposta metodológica corresponde à segunda etapa da “montagem do cubo mágico”.

3.1. SINTETIZANDO OS ESTUDOS DE CASO

Para os doze trabalhos, discutidos neste documento, foram utilizados cinco bancos de dados, provenientes de pesquisas domiciliares oriundas de Pesquisas Origem/Destino ou do último censo demográfico (IBGE, 2010).

A Pesquisa O/D fornece dados para o conhecimento da natureza dos deslocamentos da população, identificando as principais viagens diárias, conforme motivo, modo de transporte, etc. Além disso, permite estabelecer, posteriormente, relações quantitativas entre as viagens realizadas e outras variáveis, como características socioeconômicas da população e aspectos de

ocupação urbana, de forma a realizar projeções futuras das necessidades de viagens das pessoas.

Foram utilizados dados agregados, espacializando áreas por meio de centroides, e também dados desagregados, associando coordenadas geográficas a localizações dos domicílios de residência. Os dados utilizados, ao longo dos últimos oito anos, foram: (1) Região Metropolitana de São Paulo (Pesquisa O/D 1997); (2) Região Metropolitana de São Paulo (Pesquisa O/D 2007); (3) Região Metropolitana de Salvador (Pesquisa O/D 1995); (4) Salvador (censo 2010 – IBGE); (5) São Carlos (Pesquisa O/D 2007/2008). A Tabela 3.1 relaciona os dados aos doze trabalhos abordados.

Tabela 3.1 Síntese dos dados relacionados aos trabalhos abordados

DADOS	TRABALHOS	TEMA PRINCIPAL	ANÁLISE
REGIÃO METROPOLITANA DE SÃO PAULO (1997)	PITOMBO e SOUSA (2009)	GERAÇÃO DE VIAGENS X GEOESTATÍSTICA	Agregada por Zonas de Tráfego
	PITOMBO <i>et al.</i> (2010)		Agregada por Zonas de Tráfego
REGIÃO METROPOLITANA DE SÃO PAULO (2007)	LINDNER (2015)	ESCOLHA MODAL X GEOESTATÍSTICA	Desagregada por domicílios
	LINDNER e PITOMBO (2015)	ESCOLHA MODAL X GEOESTATÍSTICA	Desagregada por domicílios
	LINDNER <i>et al.</i> (2016a)	GERAÇÃO DE VIAGENS X GEOESTATÍSTICA	Agregada por Zonas de Tráfego
MUNICÍPIO DE SÃO CARLOS (2007/2008)	COSTA (2013)	ESCOLHA MODAL X GEOESTATÍSTICA	Desagregada por indivíduos
	PITOMBO <i>et al.</i> (2015a)		
	PITOMBO <i>et al.</i> (2015b)		
	GOMES <i>et al.</i> (2016a)		
REGIÃO METROPOLITANA DE SALVADOR (1995)	ROCHA (2014)	GERAÇÃO DE VIAGENS X GEOESTATÍSTICA	Agregada por Macrozona de Tráfego
	ROCHA <i>et al.</i> (2014)		Desagregada Artificialmente
	PITOMBO <i>et al.</i> (2014)		Agregada por Macrozona de Tráfego
MUNICÍPIO DE SALVADOR (2010)	ROCHA (2014)	GERAÇÃO DE VIAGENS X GEOESTATÍSTICA	Agregada por Macrozona de Tráfego
	PITOMBO <i>et al.</i> (2014)		Agregada por Macrozona de Tráfego

Para a apresentação das regiões de estudo nas próximas seções, a autora sugere uma sequência de informações para melhor caracterização dos dados, conforme Figura 3.1.

Figura 3.1 Sequência de informações para descrição das áreas de estudo

3.2. REGIÃO METROPOLITANA DE SÃO PAULO (1997)

Os estudos iniciais (PITOMBO e SOUSA, 2009; PITOMBO *et al.*, 2010) foram baseados nos dados da Pesquisa Origem-Destino (O-D) da Região Metropolitana de São Paulo (RMSP), realizada em 1997, por meio de entrevista domiciliar, pela Companhia do Metropolitano de São Paulo (METRÔ-SP, 1997). Na época, a RMSP contava com uma população de aproximadamente 17 milhões de habitantes, distribuídos em 39 municípios.

A região foi subdividida em sete subregiões (centro, leste, nordeste, norte, oeste, sudeste e sudoeste) e 389 zonas de tráfego. A pesquisa domiciliar, aplicada a todos os residentes nos domicílios amostrados, levantou dados socioeconômicos, características das viagens realizadas, localização do emprego e da escola de cada indivíduo residente e dados gerais dos domicílios, tais como condição de ocupação e número de veículos.

O banco de dados da RMSP é composto originalmente de 98.780 indivíduos e 26.278 domicílios, representando um total de 0,61% do total de domicílios existentes na RMSP na época. A Figura 3.2 representa as subregiões e municípios que compõem a região de estudo.

Figura 3.2 Região Metropolitana de São Paulo (subregiões e municípios) – Fonte: METRÔ-SP (1997)

Nos trabalhos mencionados anteriormente, foram utilizados dados agregados por Zonas de Tráfego (ZTs), ou seja, características por zonas de tráfego, associadas aos centroides das respectivas áreas. Foram disponibilizados, pelo METRÔ-SP, dados gerais como: número de domicílios, famílias, população, matrículas, empregos, automóveis, viagens produzidas, viagens atraídas, densidade populacional, densidade de emprego e taxa de motorização.

Também foram disponibilizados dados mais específicos, como população por faixa etária (em cada uma das zonas de tráfego), população por gênero, população por faixa de renda familiar, população por ocupação principal, tipo de emprego (agrícola, construção civil, indústria, etc.) e número de matrículas por grau de instrução (creche/pré-escola, 1º grau, 2º grau e superior). A Tabela 3.2 e a Tabela 3.3 descrevem as principais variáveis agregadas disponibilizadas. Desta forma, a base de dados final era composta por 389 observações (Zonas de Tráfego), associadas às coordenadas geográficas dos centroides, além de variáveis socioeconômicas e de viagens (produção e atração de viagens por modo e motivo).

Tabela 3.2 Variáveis socioeconômicas agregadas. Fonte: METRÔ - SP (1997).

Zona	Dados Gerais Socioeconômicos	Dados Específicos Socioeconômicos			
		População por faixa etária	População por Gênero	População por Condição de Atividade	Empregos por setor de Atividade
		Até 3 anos	Masculino	Ocupado	Agrícola
	Latitude do centróide	4 a 6 anos	Feminino	Ocupado Eventualmente	Construção Civil
	Longitude do centróide	7 a 10 anos		Em Licença	Indústria
	Número de Domicílios	11 a 14 anos		Não Ocupado	Comércio
	Número de Famílias	15 a 17 anos		Aposentado	Serviços de transp de carga
	População	18 a 22 anos		Nunca Trabalhou	Serviços de transp de passageiros
	Número de Matrículas Escolares	23 a 29 anos		Dona de casa	Alimentação
	Número de Empregos	30 a 39 anos			Saúde
	Número de Automóveis	40 a 49 anos			Educação
	Área	50 a 59 anos			Adm Pública
	Taxa de Motorização	mais de 60 anos			
		População por Ocupação Principal	Renda por Zona	Número de Matrículas Escolares	População por Faixa de Renda (10/97)
		Assalariado com carteira	Renda Total	Creche e Pré-Escola	250 reais
		Assalariado sem carteira	Renda Média Familiar	1º 2º 3º Graus	251 a 500 reais
		Funcionário Público		Outros	501 a 1000 reais
		Autônomo			1001 - 1800 reais
		Empregador			1801 - 3600 reais
		Profissional Liberal			acima de 3600 reais
		Trab Doméstico			
		Dono de Negócio Familiar			
		Trabalhador Familiar			

Tabela 3.3 Variáveis de viagens agregadas. Fonte: METRÔ - SP (1997).

Zona	Produção/Atração de viagens	
	Motivo de viagem	Modo de viagem
	Trabalho Indústria	Ônibus
	Trabalho Comércio	Ônibus Fretado
	Trabalho Serviços	Escolar
	Educação	Dirigindo Automóvel
	Compras	Passageiro Automóvel
	Saúde	Táxi
	Lazer	Lotação
	Residência	Metrô
	Outros	Trem
	Total	Motocicleta
		Bicicleta
		Apé

A Tabela 3.4 apresenta as medidas descritivas associadas aos dados gerais, além da atração e produção de viagens totais por zona de tráfego.

Tabela 3.4 Medidas descritivas das principais variáveis agregadas

		Média	Desvio Padrão	Máximo	Mínimo
Dados Gerais	Domicílios	11804,98	11091,44	74033	29
	Famílias	12627,09	11841,54	79369	31
	População	46516,4	45928,18	311131	111
	Matrículas	13616,05	14217,83	92740	14
	Empregos	17890,48	16371,31	108059	0
	Automóveis	8565,756	8234,19	62823	15
	Viagens Produzidas	84043,33	73084,36	454169	66
	Viagens Atraídas	83022,81	71090,12	440333	64

Em relação à dependência espacial, observa-se um padrão variável do centro até a periferia da RMSP, de variáveis fortemente relacionadas à produção e atração de viagens, respectivamente, tais como densidade populacional e densidade de empregos, conforme ilustrado na Figura 3.3 e Figura 3.4, respectivamente. Observa-se que os valores das densidades decrescem do centro em direção à periferia da RMSP.

Figura 3.3 RMSP. Densidade populacional. Pesquisa O/D 1997. Fonte: Pitombo (2007)

Figura 3.4 RMSP. Densidade de empregos. Pesquisa O/D 1997. Fonte: Pitombo (2007)

3.3. REGIÃO METROPOLITANA DE SÃO PAULO (2007)

Os dados utilizados nos trabalhos – (LINDNER, 2015; LINDNER e PITOMBO, 2015; LINDNER *et al.*, 2016) - são referentes à Pesquisa Origem/Destino (O/D), realizada na Região

Metropolitana de São Paulo (RMSP), em 2007, pela Companhia do Metropolitano de São Paulo. Na RMSP estão contidas 460 Zonas de Tráfego em 39 municípios (Figura 3.5).

Foram levantadas informações de 30 mil domicílios (Figura 3.6), escolhidos aleatoriamente e distribuídos nas 460 Zonas de Tráfego na RMSP (Figura 3.5). A distância entre o par de domicílios mais afastados entre si é de aproximadamente 142 quilômetros, enquanto a média entre os pares de domicílios mais próximos é de 146 metros.

A RMSP possuía, na época, uma população de 19,8 milhões de habitantes (IBGE, 2010) em uma área de 7.946 km². A Tabela 3.5 apresenta as informações relacionadas aos municípios da RMSP e aos domicílios da Pesquisa O/D. Observa-se que a maior quantidade de domicílios entrevistados (Pesquisa O/D) está representada pelos municípios de São Paulo, Osasco, Santo André, Mogi das Cruzes e Guarulhos. Estes municípios são os municípios mais populosos, no entanto, não são os que apresentam as maiores áreas ou densidades populacionais (LINDNER, 2015).

Figura 3.5 RMSP. Zonas de Tráfego. Fonte: METRÔ - SP (2007)

Figura 3.6 Representação dos domicílios amostrados na Pesquisa O/D da RMSP (2007). Fonte: Lindner (2015)

Tabela 3.5 Informações dos municípios da RMSP. Fonte: Lindner (2015)

Cód	Município	Domicílios*	Habitantes**	Área** km ²	Densidade hab/km ²
36	São Paulo	77,25%	11.253.503	1521,11	7.398
24	Osasco	2,52%	666.740	64,954	10.265
32	Santo André	2,08%	676.407	175,782	3.848
23	Mogi das Cruzes	2,04%	387.779	712,541	544
15	Guarulhos	2,03%	1.221.979	318,675	3.835
33	São Bernardo do Campo	1,54%	765.463	409,508	1.869
34	São Caetano do Sul	1,01%	149.263	15,331	9.736
2	Barueri	0,97%	240.749	65,701	3.664
22	Mauá	0,86%	417.064	61,909	6.737
6	Carapicuíba	0,79%	369.584	34,546	10.698
27	Ribeirão Pires	0,71%	113.068	99,075	1.141
7	Cotia	0,66%	201.150	323,994	621
37	Suzano	0,65%	262.480	206,236	1.273
8	Diadema	0,61%	386.089	30,756	12.553
9	Embu das Artes	0,55%	240.230	70,398	3.412
18	Itaquaquecetuba	0,48%	321.770	82,622	3.894
19	Jandira	0,42%	108.344	17,449	6.209
11	Ferraz de Vasconcelos	0,39%	168.306	29,564	5.693
26	Poá	0,39%	106.013	17,264	6.141
28	Rio Grande da Serra	0,38%	43.974	36,341	1.210
38	Taboão da Serra	0,38%	244.528	20,388	11.994
10	Embu-Guaçu	0,33%	62.769	155,641	403
5	Cajamar	0,30%	64.114	131,386	488
17	Itapevi	0,29%	200.769	82,658	2.429
31	Santana de Parnaíba	0,26%	108.813	179,949	605
12	Francisco Morato	0,23%	154.472	49,001	3.152
13	Franco da Rocha	0,23%	131.604	132,775	991
1	Arujá	0,21%	74.905	96,167	779

39	Vargem Grande Paulista	0,21%	42.997	42,489	1.012
16	Itapeverica da Serra	0,17%	152.614	150,742	1.012
21	Mairiporã	0,16%	80.956	320,697	252
35	São Lourenço da Serra	0,16%	13.973	186,456	75
30	Santa Isabel	0,15%	50.453	363,332	139
3	Biritiba-Mirim	0,14%	28.575	317,406	90
25	Pirapora do Bom Jesus	0,14%	15.733	108,489	145
20	Juquitiba	0,13%	28.737	522,169	55
29	Salesópolis	0,09%	15.635	424,997	37
14	Guararema	0,07%	25.844	270,816	95
4	Caieiras	0,02%	86.529	97,642	886

*Informação obtida pela proporção de domicílios entrevistados pela Pesquisa O/D (2007).

**Informação obtida pelo censo 2010 (IBGE).

Obs: Em verde estão representados os cinco menores valores para cada atributo e em vermelho os cinco maiores valores.

Ressalta-se que, nos trabalhos nos quais esta base de dados foi utilizada, fez-se utilização de dados desagregados, relativos a coordenadas geográficas de localização dos domicílios, além de dados agregados por zonas de tráfego, associados aos centroides. Por se tratar de um banco de dados referente à Pesquisa O/D, observou-se que havia casos em que coordenadas geográficas estavam associadas a mais de um domicílio. Como uma das suposições da geoestatística sugere que cada elemento do espaço tenha representação única, houve necessidade da adoção de um valor singular para cada elemento de espaço. Dessa forma, para cada caso de coordenada geográfica repetida, sorteou-se (através de pseudoaleatoriedade baseada em semente temporal) um domicílio para sua representação espacial. Desta forma, a amostra final continha 18.733 domicílios. A Tabela 3.6 apresenta as variáveis desagregadas utilizadas nos trabalhos. A Tabela 3.7, em seguida, apresenta as medidas descritivas das variáveis quantitativas.

Tabela 3.6 Informações utilizadas da Pesquisa O/D RMSP 2007. Fonte: Lindner (2015)

Informações por domicílio	Natureza	
Coordenada geográfica x	UTM	
Coordenada geográfica y	UTM	
Zona	Qualitativa	
Município	Qualitativa	
Número de pessoas	Quantitativa	Discreta
Critério de renda	Qualitativa	
Quantidade de automóveis	Quantitativa	Discreta

Quantidade de motocicletas	Quantitativa	Discreta
Quantidade de bicicletas	Quantitativa	Discreta
Número total de viagens	Quantitativa	Discreta
% de viagens por transporte individual motorizado	Quantitativa	Contínua
% de viagens por transporte não motorizado	Quantitativa	Contínua
% de viagens por transporte público	Quantitativa	Contínua
Distância total de viagens	Quantitativa	Contínua
Distância por transporte individual motorizado	Quantitativa	Contínua
Distância por transporte não motorizado	Quantitativa	Contínua
Distância por transporte público	Quantitativa	Contínua
Preferência por modo motorizado	Categórica Dicotômica	

Tabela 3.7 Medidas descritivas para as variáveis quantitativas. Fonte: Lindner (2015)

Informações por domicílio	média	desvio padrão	moda	mínimo	máximo
Número de pessoas	3,3	1,6	3	1	24
Quantidade de automóveis	0,9	0,9	1	0	8
Quantidade de motocicletas	0,1	0,3	0	0	9
Quantidade de bicicletas	0,5	0,9	0	0	12
Número total de viagens	6,8	4,6	2	1	68
% de viagens por transporte individual motorizado	27%	32%	0%	0%	100%
% de viagens por transporte não motorizado	17%	24%	0%	0%	100%
% de viagens por transporte público	33%	32%	0%	0%	100%
Distância total de viagens (km)	38,4	34,5	7,2	0%	427,1
Distância por transporte individual motorizado (km)	12,1	19,9	0%	0%	273,0
Distância por transporte não motorizado (km)	2,0	7,5	0%	0%	124,5
Distância por transporte público (km)	19,8	27,5	0%	0%	423,9

Quanto ao padrão espacial, a Figura 3.7 (a, b) demonstra a preferência por transporte motorizado na região. É notável que a representação pontual dos valores observados não produz conclusões sem que haja outro tipo de análise. Dessa forma, a geoestatística, através da análise estrutural dos dados (variografia), se mostra mais eficaz para tal fim.

Figura 3.7 Representação da: (a) preferência por transporte individual motorizado - em vermelho - sobreposta à preferência por transporte público, (b) preferência por transporte público - em verde - sobreposta à preferência por transporte individual motorizado. Fonte: Lindner (2015).

3.4. REGIÃO METROPOLITANA DE SALVADOR (1995)

A área de estudo utilizada nos trabalhos – (ROCHA, 2014; ROCHA *et al.*, 2014; PITOMBO *et al.*, 2014) – foi a Região Metropolitana de Salvador (RMS). Atualmente, a RMS é composta por 13 municípios (Figura 3.8). A RMS tem 4.375 km² de extensão e possui uma população de cerca de 3.600.000 habitantes. Deste total, cerca de 2.700.000 residem na capital (IBGE, 2010). Em 2010, Salvador ainda detinha 75% da população da RMS. Além de ser centro financeiro, polo educacional e de saúde, Salvador tem a sua economia voltada ao setor de serviços, a exemplo do comércio, transportes, telecomunicações, turismo e atividades culturais (RIBEIRO, 2012).

Figura 3.8 Região Metropolitana de Salvador. Fonte: IPEA (2014)

Para os três trabalhos abordados, foram utilizados dados agregados por Macrozonas de Tráfego (MZs), além de uma amostragem desagregada sintética proposta pelos autores. Os dados da pesquisa OD da RMS foram agrupados em 63 MZs. Estas macrozonas são resultado do agrupamento de 173 zonas de tráfego. Este processo foi orientado pela equipe técnica do Sindicato das Empresas de Transporte de Passageiros de Salvador (SETPS), que dividiu a RMS de acordo com suas características gerais de uso e ocupação do solo, renda, além dos fatores geográficos e viários. Estas macrozonas se agrupam somando 13 subregiões maiores. A subregião Miolo Norte é que possui maior quantidade de MZs, um total de 10. Já a subregião Subúrbio Pirajá é constituída por apenas uma macrozona. As demais subregiões são formadas, em média, por cinco MZs, conforme ilustrado na Figura 3.9.

A Tabela 3.8 descreve as variáveis agregadas utilizadas nos trabalhos. Para a análise geoestatística, as coordenadas geográficas dos centroides das macrozonas foram associadas às variáveis socioeconômicas mencionadas na Tabela 3.8. A Figura 3.9 mostra a subdivisão das MZs de tráfego da RMS em 1995.

Tabela 3.8: Variáveis utilizadas nos trabalhos. Fonte: Rocha (2014)

Variáveis	Média	Desvio Padrão	Mínimo	Máximo
População Residente por MZ	37533,21	23761,47	0	99855
Empregos por MZ	13381,35	11671,92	874	55867
Produção de viagens por Transporte coletivo por MZ*	6573,00	4333,44	69	18872
Atração de viagens por Transporte coletivo por MZ*	6574,44	7871,85	52	47415
* Para o pico da manhã (6 às 8h)				

As variáveis descritas acima foram padronizadas, conforme a Equação 3.1. Optou-se por este procedimento devido à alta variabilidade dos valores absolutos. Ressalta-se ainda que os menores valores das variáveis padronizadas foram transformados em valores nulos de forma que não existissem valores negativos. Os demais valores foram transformados segundo os novos valores nulos.

$$z = \frac{x - \mu}{\sigma} \quad (3.1)$$

z é o valor padronizado da variável; **x** é o valor absoluto da variável; μ é a média aritmética dos dados observados; σ é o desvio padrão dos dados observados.

Desta forma, o banco de dados utilizado para estimação geoestatística foi formado por 63 macrozonas, as coordenadas geográficas (UTM) correspondentes aos centroides das respectivas MZs e pelos dados padronizados referentes às variáveis: produção de viagens por TCO (Transporte Coletivo) e população, atração de viagens por TCO (Transporte Coletivo) e empregos.

Observando o padrão espacial das variáveis objeto de estudo, observa-se que valores de produção de viagens por TCO têm uma tendência de crescimento conforme proximidade do subúrbio ferroviário (subregião com menor renda per capita e mais populosa) enquanto que valores de atração de viagens por TCO crescem em direção à subregião central (Área Urbana Consolidada – AUC). A Figura 3.10 ilustra a distribuição espacial, através dos centroides, das variáveis Produção e Atração de Viagens por Transporte Coletivo no pico da manhã.

Figura 3.9: Subregiões e Macrozonas. Fonte: STEPS (1995)

Figura 3.10 Distribuição dos valores padronizados na RMS: Produção de viagens por TCO e Atração de viagens por TCO – Fonte: Pitombo *et al.* (2014)

3.5. MUNICÍPIO DE SALVADOR (CENSO DEMOGRÁFICO 2010)

Utilizaram-se dados de população e renda do responsável da família, divulgadas pelo IBGE (2010), nos trabalhos de Rocha (2014) e Pitombo *et al.* (2014). A razão de utilização de tais dados foi a aplicação da geoestatística em termos de projeções futuras de produção de viagens por transporte público.

Inicialmente, foram compatibilizados os setores censitários, referentes ao último censo, com as MZs de tráfego da RMS da Pesquisa OD de 1995. Após a compatibilização, foi extraído o total da população e a renda média do responsável por domicílio por MZ em 2010. Os dados de renda foram corrigidos a partir do percentual da diferença do Salário Mínimo (SM) de 2010 em relação a 1995. Foi possível verificar os valores dos SM e obter a taxa de crescimento desta variável de 1995 a 2010. Encontrou-se um valor de aproximadamente 170% de aumento do SM entre os dois períodos, levando-se em conta a inflação. A Tabela 3.9 apresenta as medidas descritivas das duas variáveis agregadas por MZs para o ano de 2010.

Tabela 3.9 Variáveis agregadas por MZs - 2010

Variável	Média	DesvioPadrão	Mínimo	Máximo
População2010	44418,48	28382,08	563,00	141221,00
Renda2010	1863,42	1716,80	455,08	8436,23
Renda2010_Ajustada	980,75	903,58	239,52	4440,12

Em relação à distribuição espacial das duas variáveis, a população (Figura 3.11) apresenta maiores valores no Subúrbio Ferroviário e em algumas MZs da subregião do Miolo.

Figura 3.11 População por MZ de tráfego (valores padronizados). Fonte: IBGE (2010)

Quanto à renda (Figura 3.12), maiores valores são observados em algumas MZs da Área Urbana Consolidada (AUC) e também na Orla. Os valores da variável renda diminuem com a aproximação ao Subúrbio Ferroviário.

Figura 3.12 Renda média do responsável do domicílio, em salários mínimos, por MZ de tráfego. Fonte: IBGE (2010).

3.6. MUNICÍPIO DE SÃO CARLOS (2007/2008)

A área de estudo utilizada nos trabalhos de Costa (2013), Pitombo *et al.* (2015a), Pitombo *et al.* (2015b) e Gomes *et al.* (2016a) foi a cidade de São Carlos (SP). São Carlos é uma cidade de porte médio, que em 2010 possuía 221.936 habitantes, com área total de 1.137,30 km² (IBGE, 2010). Observa-se que 96% da população reside na zona urbana, em uma área de aproximadamente 105 km². Os valores de PIB per capita de R\$ 20.519,06 (IBGE, 2010) e de IDH igual a 0,841 (ONU, 2003) são superiores à média do país.

Os dados utilizados nos quatro trabalhos foram obtidos através de entrevista domiciliar e pesquisa de opinião a respeito da avaliação dos transportes urbanos, aplicadas em conjunto na Pesquisa Origem-Destino de 2007/2008 na cidade de São Carlos (SP) (RODRIGUES DA SILVA, 2008).

Para a efetivação da entrevista domiciliar e da pesquisa de opinião foram selecionados aleatoriamente 6% dos domicílios contidos na base de informações cadastral cedida pelo Serviço Autônomo de Água e Esgoto (SAAE) da cidade. A referida base é formada de pontos, com coordenadas geográficas referentes ao cadastro dos hidrômetros dos domicílios (Figura 3.13) (RODRIGUES DA SILVA *et al.*, 2009).

Figura 3.13 - Base de pontos contendo os domicílios selecionados da cidade de São Carlos para a pesquisa O-D de 2007/2008. Fonte: Rodrigues da Silva *et al.* (2009)

Em todos os trabalhos que utilizaram essa base de dados, foram realizadas análises desagregadas por domicílios. Assim, após tratamento de dados, a amostra final continha 1.216 casos, 26 variáveis, sendo 22 qualitativas e 4 quantitativas e coordenadas geográficas (latitude e longitude em metros) referentes aos domicílios de residência. Essa amostra representava 0,5% da população total da cidade. As principais variáveis que compunham a amostra, bem como sua descrição, encontram-se na Tabela 3.10. As medidas descritivas das quatro variáveis quantitativas são apresentadas na Tabela 3.11.

Tabela 3.10 Variáveis desagregadas utilizadas nos trabalhos – Costa (2013); Pitombo *et al.* (2015a); Pitombo *et al.* (2015b); Gomes *et al.* (2016a).

VARIÁVEL	NOME NO BANCO DE DADOS	DESCRIÇÃO
VARIÁVEIS QUALITATIVAS		
Principal Problema - modo a Pé	(PrincProbPÉ)	1 - Risco de Atropelamento, 2 - Roubos e Assaltos, 3 - Péssimas Condições de Calçadas, 4 - Falta de Arborização
Usa ônibus	(ônibus)	1-Sim, 2-Não
Lotação dos ônibus	(F2resposta)	1- Vazios, 2-Com lotação adequada, 3-Cheios, 4-Superlotados
Frota (N. de ônibus)	(F3resposta)	1-Muito pequeno, 2-Pequeno, 3-Suficiente, 4-Ótimo
Usa Automóvel	(AUTOMOVEL)	1-Sim, 2-Não
Mudança de automóvel para ônibus	(F7resposta)	1-Sim, 2-Não
Principal Problema - modo Ônibus	(PrincProbOnib)	1 - Rapidez, 2 - Segurança, 3 - Conforto, 4 - Itinerário, 5 – Horário
Principal Problema - modo Automóvel	(PrincProbAuto)	1 - Congestionamento, 2 - Falta de estacionamento, 3 - Custo elevado
Usa bicicleta	(BICICLETA)	1-Sim, 2-Não
Principal Problema - modo Bicicleta	(PrincProbBic)	1 - Insegurança no trânsito, 2 - Roubos e Assaltos, 3 - Estacionamento, 4 - Clima, 5 – Topografia
Possuir Carteira de Habilitação	(morCNH)	1-Sim, 2-Não
Sexo do chefe da família	(morSexo)	1-Masculino, 2-Feminino
Estuda regularmente	(morEsstuda)	1-Não estuda, 2- Da educação infantil até a 8ª série do ensino fundamental; 3-Ensino médio, 4-Superior / universitário, 5-Outros.
Grau de instrução	(morGrauIns)	1- Segundo grau completo ou superior, 2 - Segundo grau incompleto ou nível inferior, 3-Não alfabetizado
Condição de Atividade	(morCondia)	1-Sem atividade diária, 2-Estudante, 3-Ocupado (tem trabalho), 4-Ocupado eventualmente (faz bico)
Condição de Renda	(morCondi_1)	1- 0,0 a 2 SM; 2- 2,1 a 8 SM 3- 8,1 a 20 SM; 4- não respondeu
Ocupação	(traboCupação)	1 - Assalariado, 2- Profissional liberal, 3- Não se aplica
Motivo principal das viagens	(MotiVo_Pri)	1 - Trabalho, 2 - Educação, 3 - Lazer/Saúde, 4 - Outros
Modo mais utilizado nas viagens	(Modo_Princ)	1-Público, 2-Particular motorizado, 3-Não motorizado
Se o morador utiliza o modo Particular Motorizado	(Part_Mot)	0 - Não utiliza o modo, 1 - Utiliza o modo
Se o morador utiliza o modo Público	(Publico)	0 - Não utiliza o modo, 1 - Utiliza o modo
Se o morador utiliza o modo Não Motorizado	(Nao_Mot)	0 - Não utiliza o modo, 1 - Utiliza o modo
VARIÁVEIS QUANTITATIVAS		
Quantidade de Motocicletas	(grfmotocic)	Quantidade de motocicletas por domicílio
Quantidade de Automóveis	(grfAutomo)	Quantidade de automóveis por domicílio
Idade	(morIdade)	Idade do chefe da família
Total de Viagens Realizadas	(Max_ViaNum)	Número total viagem realizada pelo chefe de família

Tabela 3.11 Medidas descritivas das variáveis quantitativas – São Carlos 2007/2008

VARIÁVEL	MÉDIA	DESVIOPADRÃO	MÍNIMO	MÁXIMO
Quantidade de motocicletas	0,22	0,46	0	3
Quantidade de automóveis	0,81	0,71	0	4
Idade	47,18	14,58	0	89
Total de viagens realizadas	2,56	1,37	1	17

A variável objeto de estudo nos trabalhos foi a escolha do modo principal (Particular Motorizado; Público e Não Motorizado). Observando-se a distribuição espacial tanto da escolha modal ou da probabilidade de escolha modal, não é verificado nenhum padrão espacial aparente, conforme ilustrado na Figura 3.14.

Figura 3.14 Distribuição espacial das probabilidades de escolha do modo particular motorizado, público e não motorizado – Fonte: Costa *et al.* (2013)

De forma a observar padrões espaciais em regiões menores da cidade, optou-se por realizar a modelagem geoestatística em duas regiões distintas: (1) Região de Santa Felícia, onde se observou distribuição espacial aparente para modos de transporte motorizados e (2) Região Central, onde havia maior número de pontos amostrados. A Figura 3.15 e a Figura 3.16 ilustram

a distribuição espacial da probabilidade de escolha modal para as duas regiões, respectivamente.

A região de Santa Felícia, a qual possui 110 pontos ou domicílios amostrados e georreferenciados, é constituída tanto por bairros considerados periféricos, como Loteamento Habitacional São Carlos I (Núcleo Residencial Ivo Morganti), Dário Rodrigues (São Carlos II), São Carlos III e IV (Lourival Maricondi), Bairros Arnon de Melo (São Carlos V) e Conj. Habitacional Sta. Angelina, quanto por bairros de população de maior renda, como Parque Santa Felícia, Jardim Ipanema, Montreal e Quebec. A área engloba ainda a Área 2 da Universidade de São Paulo na cidade.

Figura 3.15 Distribuição da probabilidade dos 3 modos de transporte – região Santa Felícia.
Fonte: Costa *et al.* (2013)

Posteriormente, foi selecionada a região central, a qual aparentemente menos apresentou tendência espacial dos dados de probabilidade de escolha modal, como se pode ver na Figura 3.16. Além disso, é geograficamente maior e tem características socioeconômicas e de ocupação urbana diferentes da Região de Santa Felícia, o que poderia representar um comportamento diferenciado no resultado do modelo proposto. Adicionalmente, apresenta maior número de pontos amostrados (474 domicílios).

Figura 3.16 - Distribuição da probabilidade dos 3 modos (Centro). Fonte: Costa (2013).

A região central é caracterizada pela presença de estabelecimentos comerciais, residências e serviços, com população de 12.466 habitantes (IBGE 2010), basicamente de renda média. Contém a Área 1 do campus da Universidade de São Paulo, Estação Rodoviária, Estação Norte de Integração do Transporte Coletivo, centro comercial e de serviços e engloba, também, importantes corredores de tráfego da cidade.

3.7. APLICATIVOS UTILIZADOS

Os principais materiais utilizados nos doze trabalhos abordados neste documento foram essencialmente os dados e os aplicativos. Desta forma, um conjunto de ferramentas computacionais auxiliaram em diversas etapas metodológicas, tais como: análises estatísticas tradicionais, análise espacial e modelagem geoestatística aplicadas à demanda por transportes. Vale ressaltar que os aplicativos utilizados estão associados à sua disponibilidade (institucional ou *software* livre). As etapas metodológicas não estão restritas a usos de *software* específico.

No âmbito das estatísticas tradicionais, análises univariadas e multivariadas não espaciais foram realizadas através do *software* IBM SPSS Statistics 22. Com relação à análise espacial, os sistemas de informação geográfica (SIGs) provêm ferramental necessário para visualização de

dados por meio de mapas. A geração de superfícies espaciais foi realizada com auxílio do *software* ESRI ArcGIS Platform, versão 10.1 e TRANSCAD 5.0.

No campo da geoestatística, as etapas de visualização estatística dos dados, geração de variogramas experimentais, ajuste por curvas teóricas, interpolação por Krigagem e validação cruzada (teste do ponto fictício) foram obtidas através do *software* geoMS (*Geostatistical Modelling Software*), desenvolvido em 1999, pelo Instituto Superior Técnico de Lisboa. O geoMS é um *software* direcionado para analisar dados de geoestatística que apresentam dependência espacial. É composto por vários módulos que, no seu conjunto, são necessários para modelagem geoestatística.

Análises espaciais através de Regressão Geograficamente Ponderada foram realizadas com o auxílio do *software* GWR3, desenvolvido no *GeoDa Center for Geospatial Analysis and Computation, School of Geographical Sciences and Urban Planning, University of Illinois*.

3.8. MÉTODO GERAL

O método comum aos doze trabalhos é referente à etapa de modelagem geoestatística. A modelagem é composta pelas etapas: (1) Tratamento da base de dados; (2) Visualização espacial dos dados; (3) Obtenção de variogramas experimentais; (4) Modelagem de variogramas teóricos; (5) Validação; (6) Aplicação de interpolador ou estimador espacial; (7) Obtenção de superfícies (mapas Krigados).

O interpolador ou estimador espacial variou entre os trabalhos. Além disso, alguns estudos apresentaram uma etapa anterior ou posterior à modelagem geoestatística, caracterizada como aplicação de modelagem tradicional não espacial. A Figura 3.17 ilustra o procedimento metodológico geral aplicado, já a Tabela 3.12 aponta as diferenças peculiares a cada trabalho. As subseções seguintes descrevem sucintamente as etapas do método comum.

Figura 3.17 – Ilustração do procedimento metodológico geral

Tabela 3.12 – Identificação das diferenças metodológicas entre trabalhos

Trabalhos	Etapa anterior	Etapa posterior	Interpolador	Tema
PITOMBO e SOUSA (2009)	Não	Não	Krigagem Ordinária	Geração de viagens
PITOMBO <i>et al.</i> (2010)	Análise em Componentes Principais	Regressão Geograficamente Ponderada	Krigagem com Deriva Externa	
ROCHA (2014)	Calibração de modelos lineares	Não	Krigagem com Deriva Externa	
ROCHA <i>et al.</i> (2014)	Calibração de modelos lineares	Não	Krigagem com Deriva Externa	
PITOMBO <i>et al.</i> (2014)	Não	Não	Krigagem com Deriva Externa	
LINDNER (2015)	Não	Regressão Logística Binomial	Krigagem Indicativa	Escolha modal
LINDNER e PITOMBO (2015)	Regressão Logística Binomial	Não	Krigagem Ordinária	
COSTA (2013)	Árvore de Decisão	Não	Krigagem Ordinária	
PITOMBO <i>et al.</i> (2015a)	Árvore de Decisão	Não	Krigagem Ordinária	
PITOMBO <i>et al.</i> (2015b)	Regressão Logística Multinomial	Não	Krigagem Ordinária	

3.8.1 Tratamento da base de dados

A etapa de tratamento de dados corresponde, primeiramente, à análise dos dados georreferenciados disponíveis e preparação das amostras considerando os objetivos dos trabalhos. É importante ressaltar que devem ser eliminados dados associados a coordenadas geográficas repetidas, além de inconsistências. De forma geral, o banco de dados é formado por identificador do ponto, localização (latitude e longitude – em metros) dos pontos, além de características sociodemográficas e de viagens associadas à localização.

O tratamento de dados é feito com auxílio de ferramentas usuais de gerenciamento de banco de dados (*Microsoft Office Access, Microsoft Office Excel, IBM SPP 22.0*), além do módulo *geoData* do aplicativo *geoMS*.

3.8.2 Visualização espacial do conjunto de pontos

O objetivo desta etapa é possibilitar a avaliação qualitativa a respeito da real necessidade da consideração do atributo espacial. Neste estágio, deve-se considerar a hipótese de que, em fenômenos eminentemente geográficos, os modelos de estimação através de geoestatística, por exemplo, podem auxiliar na explicação de variáveis relativas à demanda por transportes.

O entendimento inicial do comportamento das variáveis, assim como possíveis padrões espaciais, pode ser auxiliado por mapas temáticos e estatística descritiva, como histogramas, mapas de dispersão, etc.

Através da visualização da dispersão de cada uma das variáveis no espaço pode-se ter uma ideia de como estas se comportam, quais as direções que provavelmente devem ser observadas e, enfim, adotar parâmetros para a próxima etapa dos procedimentos geoestatísticos, que é o cálculo de variogramas experimentais.

Ainda no módulo *geoData* do *geoMS* são visualizados os mapas de cores das variáveis e, com auxílio de um Sistema de Informação Geográfica, são vistos mapas temáticos a fim de definir previamente o comportamento espacial de cada uma das variáveis (variável dependente e variáveis explicativas) – Figura 3.17 (a).

3.8.3 Obtenção de variogramas experimentais

Inicialmente, são gerados alguns variogramas experimentais exploratórios para se identificar quais as direções de maior variação, bem como valores adequados para o número e tamanho de incrementos (*Lags*), direção principal, largura máxima. Esta etapa pode ser realizada através do módulo *geoVAR* do *geoMS* – Figura 3.17 (b).

3.8.4 Modelagem dos variogramas experimentais

Depois de escolhidos os parâmetros e direções a estudar, na etapa de obtenção de variogramas experimentais, é necessário transpor toda essa informação para uma função geral representativa. Sendo assim, deve ser realizado o ajustamento dos variogramas experimentais por uma função geral, sendo esta etapa extremamente importante no processo de estimação – Figura 3.17 (c). O módulo *geoMOD* do *geoMS* permite um ajuste visual da curva, no entanto, o pesquisador pode indicar algum procedimento alternativo, a fim de minimizar os erros, por exemplo. Além disso, existem pacotes computacionais de ajuste automático.

3.8.5 Validação

A etapa de validação visa testar a qualidade da modelagem dos variogramas, ou seja, o ajuste dos variogramas teóricos. O objetivo é obtenção de valores estimados e observados da Variável Regionalizada, nas coordenadas geográficas em que os valores da VR são previamente conhecidos. A partir desses dois vetores de valores (observados x estimados) é possível fazer uma análise da magnitude dos erros, bem como realizar testes estatísticos de hipótese a fim de testar similaridades de distribuições dos dois vetores, bem como características centrais (Kolmogorov-smirnov, Mann-whitney, teste t, etc.). O módulo *geoKrig*, do *geoMS*, permite realizar a validação cruzada das estimações geoestatísticas.

3.8.6 Aplicação do Interpolador ou estimador espacial

Através de um interpolador, estimam-se valores de um atributo em locais não amostrados a partir de dados amostrados na mesma área. A Krigagem (Universal, Indicativa, Ordinária, etc.) permite a estimativa da Variável Regionalizada considerando a distribuição dos seus valores no espaço, além dos parâmetros definidos na etapa de ajuste dos variogramas teóricos. Além disso, devem-se considerar os parâmetros da malha de pontos ou blocos a estimar. A exemplo da etapa anterior, esta etapa também pode ser realizada através do módulo *geoKrig* do *geoMS*.

3.8.7 Obtenção de superfícies (mapas krigados)

Com auxílio de algum Sistema de Informação Geográfica, é possível visualizar as superfícies obtidas a partir dos valores estimados em diversas coordenadas geográficas. Os mapas krigados, bem como os testes de validação, são os principais produtos da modelagem geoestatística - Figura 3.17 (f).

Capítulo

4

GERAÇÃO DE VIAGENS E GEOESTATÍSTICA

Este capítulo trata da terceira grande etapa da linha de pesquisa, que é o uso da geoestatística em dados numéricos agregados por unidades de áreas. Sistematiza os seis trabalhos que fazem parte do tema Geração de viagens e Geoestatística. Inicialmente, são sumariados os conceitos básicos relativos ao tema em forma de quadro resumo. Em seguida, são descritos alguns trabalhos nos quais foram utilizados basicamente técnicas de estatística espacial em áreas para auxílio na modelagem de geração de viagens. Finalmente, são descritos os trabalhos desenvolvidos pela autora e colaboradores, acompanhados de uma análise crítica dos resultados obtidos e procedimentos metodológicos aplicados.

4.1. MODELAGEM TRADICIONAL DE DEMANDA POR TRANSPORTES: UM QUADRO RESUMO

Estimativas de demanda por transportes auxiliam a tomada de decisão relativa ao planejamento urbano. A previsão da demanda por modelos tradicionais considera a utilização do sistema de transportes em função de aspectos sociodemográficos e do meio. O modelo mais utilizado ainda é o modelo quatro etapas.

Apesar de ser conhecido como um modelo, se configura mais como um método ou um procedimento composto por uma sequência de vários “submodelos”. Sua finalidade é, principalmente, a de previsão de viagens entre diversas zonas de tráfego. O método se baseia

na suposição de que a realização de uma viagem tem estreita relação com um processo desenvolvido em quatro fases: i) Geração de viagens, ii) Distribuição de viagens, iii) Escolha Modal e iv) Alocação de Tráfego (NOVAES, 1986; KAWAMOTO, 1994; ORTÚZAR e WILLUMSEM, 2011). As quatro etapas são descritas, de maneira sintetizada, em seguida. Posteriormente, o Quadro 4.1 traz um resumo das principais definições relativas à etapa de geração de viagens.

4.1.1 Geração de viagens

Origina o processo com a finalidade de prever o número de viagens de e para uma determinada região em estudo (viagens produzidas e atraídas por unidade de área considerada). Os resultados dessa etapa servem de ponto de partida para as demais etapas do processo.

4.1.2 Distribuição de viagens

Definido o volume de tráfego produzido e atraído nas zonas de tráfego, por exemplo, é necessário conhecer a distribuição das viagens, identificando o volume de tráfego entre a zona de origem e as demais zonas (NOVAES, 1986). Desta forma, pode ser gerada uma matriz de viagens, conhecida como matriz origem-destino (O-D). Existem alguns modelos tradicionalmente conhecidos para estimação de matrizes, tais como: (a) Modelos baseados em Fatores de Crescimento; (b) Modelos Gravitacionais e entrópicos; (c) Modelos econométricos e (d) Métodos sintéticos.

4.1.3 Divisão modal

Conhecendo-se a distribuição das viagens em relação às zonas de tráfego, em seguida é feita a divisão modal, ou seja, a análise da escolha do modo de transporte do total de viagens realizadas. Diversos fatores podem influenciar na escolha modal, como por exemplo, velocidade, conforto e custo (BRUTON, 1979). Em geral, escolhas modais são previstas através de uma

abordagem desagregada e modelos probabilísticos. Os modelos probabilísticos mais utilizados são os econométricos (logit e probit).

4.1.4 Alocação de tráfego

A última etapa do processo consiste na alocação dos fluxos de cada modo para cada par O/D da rede de transportes. Os modelos de alocação são utilizados para fornecer estimativas de fluxos de tráfego em vias estratégicas. Um conceito que deve ser considerado em quase todos os modelos de alocação de tráfego é o custo de deslocamento. Diante das diversas alternativas que o usuário tenha para efetuar uma viagem, naturalmente é ponderada uma série de fatores para decidir qual percurso seguir.

A técnica mais simples empregada para analisar a alocação do tráfego é o modelo “Tudo ou Nada”. Esta técnica admite que, definida a melhor rota entre um par OD, todo fluxo passaria por esta via, sem considerar sua capacidade. Outros modelos surgiram, posteriormente, com o intuito de aprimorar esta técnica, tais como alocação incremental e métodos estocásticos baseados em simulações. Estes últimos possibilitam uma análise microscópica, no qual se analisam inclusive os problemas de congestionamento nas interseções (PAPACOSTAS e PREVEDOUROS, 1993).

Quadro 4.1 Conceitos básicos da etapa de geração de viagens (ORTÚZAR e WILLUMSEM, 2011).

Viagem	É o movimento de único sentido a partir de um ponto de origem para um ponto de destino
Geração de viagens	Total de viagens atraídas e/ou produzidas por unidade de área ou domicílio
Atração de viagens	Quantidade de viagens atraídas por unidade de área
Produção de viagens	São as viagens originadas por unidade de área
Abordagens	Agregadas (zonas de tráfego, macrozona, cidades, etc.); Desagregadas (Domiciliar)
Caracterização das viagens	Motivo; Modo de Transporte; Período do dia; etc.
Fatores que influenciam a produção de viagens	Renda; Posse de automóveis; Tamanho da família; Estrutura domiciliar; Uso do solo; Densidade Residencial; Acessibilidade.
Fatores que influenciam a atração de viagens	Disponibilidade de serviços, indústrias, comércio, empregos, acessibilidade
Modelos ou técnicas	Modelo do fator de crescimento; Análise de Regressão Linear e Classificação Cruzada
Fator de Crescimento	A previsão do número de viagens é determinada através da aplicação de um fator de crescimento aos dados atuais
	$T_i = F_i t_i$; T_i = número de viagens futuras produzidas na zona i ; F_i = Fator de crescimento; t_i = Número de viagens no ano base
Análise de Regressão Linear	Representa uma relação linear entre o número de viagens existentes (variável dependente) e os vários fatores que influenciam as viagens (variáveis independentes).
	$y_i = bX_i + cZ_i + a$; y_i = número de viagens futuras produzidas ou atraídas na zona i ; X_i = Variável Socioeconômica; Z_i = Variáveis de uso do solo, a,b, c (coeficientes estimados)
Classificação Cruzada	Utiliza dados desagregados por categorias de residência
	As viagens são agrupadas e relacionadas à estrutura familiar e às condições econômicas de cada família

4.2. ANÁLISE ESPACIAL DE GERAÇÃO DE VIAGENS: *BACKGROUND*

A partir da popularização do uso de SIG e disponibilidade de fatores locacionais nas bases de dados relativas ao Planejamento de Transportes, o uso da análise espacial de dados urbanos e de transportes tornou-se uma linha de pesquisa promissora, trazendo importantes contribuições para área (THILL, 2000; MILLER E SHAW, 2001; KWAN, 2000; BHAT e ZHAO, 2002; PÁEZ e SUZUKI, 2001; PÁEZ e SCOTT, 2004; PÁEZ *et al.*, 2011)

A análise espacial de dados, voltada à etapa de geração de viagens, em geral é realizada através de dados relativos a áreas com contagens e taxas agregadas. Tais dados geralmente são provenientes de levantamentos censitários ou pesquisas Origem-Destino. O Quadro 4.2 sintetiza alguns trabalhos encontrados na literatura que realizaram análise espacial em áreas para geração de viagens. Existem algumas limitações metodológicas relacionadas a tal abordagem, tais como o problema de MAUP, associação espacial e heterogeneidade espacial (PÁEZ e SCOTT, 2004).

A aplicação de modelagem geoestatística a estimativas de geração de viagens também possui algumas restrições, descritas sucintamente ao final do Capítulo 2. No entanto, a principal motivação para uso de tal ferramenta, ainda incomum para a área de demanda por transportes, seria a sua capacidade de estimar valores de Variável Regionalizada em locais não amostrados, aumentando a capacidade de aquisição de dados relativos a transportes. A geoestatística mostra-se mais poderosa do que outros interpoladores, pois considera, além das distâncias entre os pontos, parâmetros do variograma teórico.

Quadro 4.2 Exemplos de trabalhos relativos ao tema análise espacial em áreas para estimativa de geração de viagens

Trabalho	Objetivos	Técnicas
Bhat e Zhao (2002)	Identificar fatores espaciais que devem ser incorporados a modelos de geração de viagens baseadas em atividades	Modelos espaciais de múltiplos níveis
Paez <i>et al.</i> (2005)	Utilizar procedimentos de análise espacial em áreas para previsão de geração de viagens no contexto de pessoas idosas	Regressão Linear Tradicional Regressão Geograficamente Ponderada
Soltani e Ivaki (2011)	Investigar influência de variáveis geográficas na estimativa de viagens intra e interzonais	Regressão Linear Tradicional Análise em Componentes Principais
Amavi <i>et al.</i> (2014)	Investigar incrementos na estimativa de produção e atração de viagens zonais a partir de modelos espaciais	Regressão Linear Tradicional Regressão Geograficamente Ponderada Global Regressão Geograficamente Ponderada Local
Lopes <i>et al.</i> (2014)	Analisar resultados provenientes de diferentes abordagens de estatística espacial em áreas para geração de viagens	Regressão Linear Tradicional Regressão tradicional com covariáveis espaciais Regressão Geograficamente Ponderada
Sanchez-Diaz <i>et al.</i> (2014)	Realizar uma análise exploratória de atração de viagens por transporte de carga a partir de variáveis locais	Modelos econométricos espaciais

4.3. SISTEMATIZANDO OS TRABALHOS

A apresentação dos trabalhos a serem tratados neste documento, classificados na temática geração de viagens e geoestatística, foi sistematizada de acordo com o tipo de abordagem. Inicialmente, serão apresentados trabalhos que utilizaram dados agregados por Zonas de Tráfego ou Macrozonas de Tráfego. Posteriormente, são descritos os trabalhos que utilizaram dados desagregados pontuais.

A Tabela 4.1 descreve os trabalhos a serem discutidos neste capítulo e suas respectivas abordagens: Análise Agregada (análises baseadas na modelagem geoestatística e localizações

geográficas relativas às coordenadas geográficas de centroides de áreas) e Análise Desagregada (análises baseadas em localizações pontuais a partir de amostragem sintética).

Tabela 4.1 Trabalhos a serem discutidos no Capítulo 4

Trabalho	Ano	Título	Autores	AC AR DM*	Análise
1	2009	APLICAÇÃO DE CONCEITOS GEOESTATÍSTICOS PARA ANÁLISE DE GERAÇÃO DE VIAGENS URBANAS	Cira Souza Pitombo Antonio Jorge de Sousa	XXIV ANPET (AC)	Agregada
2	2010	COMPARING DIFFERENT SPATIAL DATA ANALYSIS TO FORECAST TRIP GENERATION	Cira Souza Pitombo Antonio Jorge de Sousa Mark Birkin José Alberto Quintanilha	12th WCTR (AC)	Agregada
3	2014	ANÁLISE DE GERAÇÃO DE VIAGENS URBANAS POR TRANSPORTE COLETIVO ATRAVÉS DE TÉCNICAS DE GEOESTATÍSTICA	Samille Santos Rocha	Dissertação (Mestrado) UFBA(DM)	Agregada e desagregada
4	2014	USO DE KRIGAGEM PARA ESTIMAÇÃO DE GERAÇÃO DE VIAGENS	Cira Souza Pitombo Samille Santos Rocha Ana Rita Salgueiro Daiane Bittencourt	VI PLURIS (AC)	Agregada
5	2014	MODELAGEM GEOESTATÍSTICA APLICADA À AMOSTRA SISTEMÁTICA DESAGREGADA: ESTIMAÇÃO DE GERAÇÃO DE VIAGENS	Samille Santos Rocha Anabele Lindner Cira Souza Pitombo	XXIX ANPET (AC)	Desagregada
6	2016	ESTIMATION OF TRANSIT TRIP PRODUCTION USING FACTORIAL KRIGING WITH EXTERNAL DRIFT: AN AGGREGATED DATA CASE STUDY	Anabele Lindner Cira Souza Pitombo Samille Santos Rocha José Alberto Quintanilha	Geo-spatial Information Science (AR)	Agregada

*AC: Artigo publicado em anais de congresso; AR : Artigo publicado em revista; DM: Dissertação de mestrado defendida

4.4. GERAÇÃO DE VIAGENS E GEOESTATÍSTICA: UMA ANÁLISE AGREGADA

A partir do tema “Aplicação de geoestatística numa análise agregada em áreas para geração de viagens” foram apresentados e publicados quatro trabalhos em anais de congressos e um trabalho em periódico, além de defendida uma dissertação de mestrado. Para a apresentação dos trabalhos, nas próximas seções, a autora sugere uma sequência de informações, ilustrada na Figura 4.1, para melhor organização do texto.

Figura 4.1 Sequência de informações para apresentação das pesquisas

4.4.1 Krigagem Ordinária e dados agregados por zonas

✓ Trabalho 1 - APLICAÇÃO DE CONCEITOS GEOESTATÍSTICOS PARA ANÁLISE DE GERAÇÃO DE VIAGENS URBANAS

Primeiro produto divulgado da linha de pesquisa, este trabalho foi apresentado na XXIV ANPET (Congresso da Associação Nacional de Pesquisa e Ensino em Transportes) e publicado nos anais do evento em 2009 (Vitória, Espírito Santo). Foi escrito em coautoria com o Professor Catedrático Antonio Jorge Sousa, da Universidade de Lisboa, Portugal.

- Objetivos

O trabalho teve como principal objetivo estimar taxas de produção e atração de viagens por modo de transporte e motivo de viagem, através de modelagem geoestatística. Além disso, o trabalho se propunha a: (1) Fazer uma análise exploratória das variáveis relativas a viagens, detectando padrões espaciais na região de estudo; (2) Estimar taxas de produção de viagens por modo de transporte e motivo de viagem; e (3) Estimar taxas de atração de viagens por modo de transporte e motivo de viagem.

- Região

A região de estudo foi a RMSP e os dados provenientes da Pesquisa Origem-Destino, realizada em 1997, conforme descrito no capítulo anterior. Foram consideradas 14 variáveis relativas a viagens (produção ou atração) por centroide de zona de tráfego (389 no total).

O banco de dados apresentou um total de 389 zonas de tráfego, associados às taxas médias (razão entre a produção/atração de viagens por modo de transporte ou motivo de viagem e a população total da zona de tráfego).

- Materiais e Método

O procedimento metodológico, seguido neste trabalho, foi equivalente àquele apresentado no Capítulo 3, referente ao processo de modelagem geoestatística. O estimador espacial ou interpolador utilizado, neste caso, foi a Krigagem Ordinária (KO).

Assim, o método é composto pelas etapas de: (1) Tratamento do banco de dados; (2) Visualização espacial dos dados; (3) Obtenção de variogramas experimentais; (4) Modelagem de variogramas experimentais; (5) Validação; (6) Aplicação de interpolador ou estimador espacial; (7) Obtenção de superfícies (mapas Krigados). O aplicativo aqui utilizado foi o GeoMs (*Geostatistical Modelling Software*).

- Resultados

Observou-se, neste trabalho, que existe aparentemente comportamento espacial similar para todas as direções (isotropia). A partir de qualquer direção analisada as variáveis relativas à produção e atração de viagens aumentam de valor da periferia para o centro. Foi calculado um total de 70 variogramas experimentais.

A Figura 4.2 apresenta os variogramas experimentais e teóricos para as direções 45° e - 45° da variável taxa de produção de viagens por automóvel, por exemplo. Em ambos os casos observa-se uma forma aproximada de parábola, ou seja, a dependência espacial cresce até

certo ponto atingindo o máximo (periferia para o centro) e depois de certa distância diminui até que seja nula. Também são apresentados os variogramas omnidirecional e na direção 90°, onde se pode verificar ajuste para o modelo esférico.

Para a variável Produção de viagens por automóvel, por exemplo, observam-se os seguintes valores dos parâmetros: • Variograma esférico; • Efeito pepita igual a zero $C_0 = 0$; • Patamar ($C_1 = 0,072$); • Amplitude ($a = 159253,46$) para o variograma omnidirecional.

Figura 4.2 Variogramas modelados (produção de viagens por automóvel) nas direções 45°, -45° omnidirecional (0) e 90°. Fonte: Pitombo e Sousa (2009)

Neste trabalho foi realizada a validação cruzada, ou teste do ponto fictício, estimando sequencialmente o valor da variável regionalizada em coordenadas geográficas de valores conhecidos. Assim, foi gerada uma tabela de resultados de pontos previamente amostrados (374 zonas de tráfego no total, pois foram excluídas aquelas cujos valores eram ausentes) com valores observados, estimados, coordenadas nas 374 zonas de tráfego e variância da estimação.

A fim de averiguar a qualidade dos modelos dos variogramas, foram calculadas as seguintes medidas para a variável Produção de viagens por automóvel, considerando-se os vetores de valores estimados e observados: coeficiente de correlação de *Pearson* (0.47); média dos erros (0.0005); média do quadrado dos erros (0.06).

Considerados razoáveis os valores das medidas de ajuste na validação cruzada, foi realizada a estimação por Krigagem Ordinária. Foi encontrando um total de aproximadamente 40.000 valores de pontos conforme número de blocos adotados na malha (200x200).

- Análise crítica

Tratou-se de um trabalho em fase incipiente da linha de pesquisa. Foi possível concluir, a partir dos primeiros resultados obtidos, que a geoestatística é uma ferramenta viável para aplicação em estudos relativos à demanda por transportes, havendo ainda diversas possibilidades de utilização, a depender dos objetivos a serem atingidos.

Através da estimação geoestatística, usando o estimador de Krigagem Ordinária, foi possível estimar os valores das variáveis relativas a viagens em aproximadamente 40.000 pontos de valores previamente não amostrados. Esta é a principal vantagem em relação a ferramentas tradicionais de modelagem de geração de viagens.

Em relação a restrições ou limitações metodológicas, pode-se citar o fato de a análise ser univariada, ou seja, neste artigo foram considerados os valores das Variáveis Regionalizadas e as distâncias entre pontos. Neste caso, não é possível realizar um entendimento adequado do fenômeno ou mesmo realizar projeções futuras a partir de valores futuros de covariáveis.

O trabalho ainda merece uma análise complementar relativa à validação metodológica. É necessária a estimativa, através de métodos tradicionais, como ajuste de modelos de Regressão Linear Múltipla, por exemplo, para comparação de qualidade de estimativas. Tal validação seria interessante para tornar mensurável a possível contribuição da modelagem geoestatística a

modelos de geração de viagens. Sabidamente, é bastante improvável que a atração ou a produção de viagens seja melhor explicada apenas pelo atributo espacial em detrimento de população e empregos, respectivamente.

Além disso, a abordagem aqui proposta também possui as limitações relativas ao ajuste de variogramas experimentais. Uma alternativa aos modelos aqui ajustados seriam variogramas esféricos sem patamar. Adicionalmente, verifica-se o problema subjacente à ferramenta geoestatística quando aplicada à áreas heterogêneas (suportes geométricos diferentes). A aplicação realizada supõe que todas as áreas são de igual tamanho e forma.

A partir das restrições metodológicas deste primeiro trabalho, pode-se propor aplicação de geostatística multivariada, como Krigagem com Deriva Externa - KDE. Além disso, é interessante analisar a possibilidade de aplicação de procedimentos alternativos para homogeneização de suportes, além de uso futuro de estimadores que considerem distribuições populacionais diferentes da gaussiana.

4.4.2 Análise em Componentes Principais e Krigagem com Deriva Externa: dados agregados

- ✓ Trabalho 2: *COMPARING DIFFERENT SPATIAL DATA ANALYSIS TO FORECAST TRIP GENERATION*
- ✓ Trabalho 6: *ESTIMATION OF TRANSIT TRIP PRODUCTION USING FACTORIAL KRIGING WITH EXTERNAL DRIFT: AN AGGREGATED DATA CASE STUDY*

Ambos os trabalhos adotaram um procedimento metodológico que consistia no uso sequencial de Análise em Componentes Principal e Krigagem com Deriva Externa (*Factorial Kriging with External Drift*) para estimativas de geração de viagens agregadas por zonas de tráfego. A diferença entre os dois trabalhos, além da base de dados utilizada, consiste na comparação dos resultados com outra técnica de estatística espacial – Regressão Geograficamente Ponderada Global e Local.

O Trabalho 2 foi apresentado em 2010 e publicado nos anais da 12ª *World Conference on Transport Research* (Lisboa, Portugal), sendo realizado em coautoria com mais três pesquisadores de universidades europeias e brasileira: Antonio Jorge Sousa (Ulisboa, IST); José Alberto Quintanilha (USP, Escola Politécnica) e Mark Birkin (*University of Leeds, School of Geography*).

O Trabalho 6 foi aceito recentemente para publicação na revista *Geo-Spatial Information Science*. A autoria principal é da aluna de doutorado Anabele Lindner, contando com a colaboração, além da docente, da também aluna de doutorado Samille Santos Rocha e do professor José Alberto Quintanilha (USP, Escola Politécnica).

- Objetivos

O objetivo do Trabalho 2 é apresentar técnicas de análise espacial e multivariadas de dados para estimação de geração de viagens por motivo e modo de viagem. O objetivo secundário está associado à comparação das técnicas de *Factorial Kriging with External Drift* (Análise em Componentes Principais e Krigagem com Deriva Externa) e Regressão Geograficamente Ponderada Global e Local. Já o Trabalho 6 está associado ao uso da primeira abordagem para estimação da produção de viagens por transporte público.

- Região

O Trabalho 2 utilizou como área de estudo a Região Metropolitana de São Paulo e os dados provenientes da Pesquisa Origem-Destino, realizada em 1997. Foram adotadas 12 variáveis relativas a viagens (produção ou atração) por centroide de zona de tráfego (389 no total e 7 subregiões) e 24 variáveis explicativas socioeconômicas, também agregadas por unidades de área.

O banco de dados apresentou um total de 389 zonas de tráfego associadas às taxas médias de viagens (razão entre a produção/atração de viagens por modo de transporte ou

motivo de viagem e a população total da zona de tráfego), taxas médias socioeconômicas e latitude e longitude dos centroides das zonas de tráfego.

Já o Trabalho 6 utilizou a cidade de São Paulo e os dados provenientes da Pesquisa Origem-Destino, realizada na RMSP em 2007. Foi utilizada apenas uma variável relativa a viagens (produção de viagens por transporte público) por centroides de zonas de tráfego (320 no total) e 23 variáveis explicativas socioeconômicas, também agregadas por unidades de área.

- Materiais e Método

O procedimento metodológico, seguido no Trabalho 2, foi equivalente ao método geral mencionado anteriormente (Capítulo 3), referente ao processo de modelagem geoestatística. O estimador ou interpolador utilizado, neste caso, é a Krigagem com Deriva Externa (KDE). A modelagem geoestatística é precedida pela aplicação de Análise em Componentes Principais. Já etapa posterior ao uso da geoestatística é a aplicação de modelos de Regressão Geograficamente Ponderada. Na etapa de análise de resultados há um procedimento comparativo das técnicas de análise multivariada espacial de dados.

O método proposto do Trabalho 6 foi equivalente àquele utilizado anteriormente no Trabalho 2. No entanto, a etapa posterior à modelagem geoestatística foi a calibração de modelos tradicionais de Regressão Linear, com a utilização de variáveis explicativas equivalentes às Componentes extraídas e também aos valores originais das mesmas variáveis.

Ambos os trabalhos utilizaram o *software* estatístico IBM SPSS. Além disso, foram utilizados os aplicativos GeoMs, GWR 3.0 e Transcad.

- Resultados – Trabalho 2

Anteriormente às etapas relativas à modelagem geoestatística, foi realizada aplicação da Análise em Componentes Principal (ACP). Através do critério da raiz latente, foram extraídas 4 Componentes que explicam, aproximadamente, 89% da variância total dos dados.

Observando-se os valores dos escores fatoriais, foi possível propor uma nomenclatura para as quatro componentes a partir da contribuição das variáveis originais em cada uma das componentes.

Componente 1: Características socioeconômicas domiciliares;

Componente 2: Empregos;

Componente 3: Alta renda;

Componente 4: População.

Após extração das Componentes, para comparação das duas técnicas de análise espacial de dados (Regressão Geograficamente Ponderada - RGP e Krigagem com Deriva Externa - KDE), foi necessária a escolha de apenas uma das quatro componentes como variáveis explicativas para calibração de modelos univariados. A opção da escolha de apenas uma das componentes por modelo justifica-se pelo fato da KDE trabalhar apenas com *uma* variável secundária. Neste trabalho, tanto a variável explicativa dos modelos de Regressão Geograficamente Ponderada quanto a variável secundária utilizada na KDE é uma Componente, que seria uma composição das 24 variáveis originais.

Assim, as Componentes foram selecionadas como variáveis explicativas para RGP, considerando aquela de mais alta correlação com a variável dependente relativa a taxas de viagens. As Componentes escolhidas nos doze modelos variavam entre a Componente 1 (Características socioeconômicas domiciliares) e Componente 2 (Empregos). Componente 3 (Alta renda) e Componente 4 (População) não entraram nos modelos. Os modelos de RGP globais foram considerados razoáveis, com valores de Coeficientes de Determinação variando entre 0,5 e 0,8.

Uma das grandes vantagens da aplicação da Regressão Geograficamente Ponderada é a de observar o comportamento das variáveis dependentes e parâmetros estimados nas diferentes

unidades de áreas. Assim, em dada zona de tráfego, pode-se ter parâmetros estimados não significativos. Os mesmos parâmetros podem ser importantes para explicação do fenômeno em outras regiões. Observações importantes de diferenças regionais são possíveis a partir da mencionada técnica.

Para aplicação da KDE, foi necessário o ajuste de 14 variogramas teóricos para as seguintes configurações de variáveis primárias e variáveis secundárias. Todos os variogramas foram isotrópicos e ajustados pelo modelo esférico.

- ✓ **Modelo 1: Variável Primária** – Produção de viagens por transporte coletivo (variograma 1); **Variável Secundária** – Componente 2 (variograma 2);
- ✓ **Modelo 2: Variável Primária** – Produção de viagens por automóvel (variograma 3); **Variável Secundária** – Componente 1 (variograma 4);
- ✓ **Modelo 3: Variável Primária** – Produção de viagens por modo não motorizado (variograma 5); **Variável Secundária** – Componente 1 (variograma 4);
- ✓ **Modelo 4 : Variável Primária** – Produção de viagens por motivo trabalho na indústria (variograma 6) ; **Variável Secundária** – Componente 1 (variograma 4);
- ✓ **Modelo 5 : Variável Primária** – Produção de viagens por motivo trabalho no setor de comércio (variograma 7) ; **Variável Secundária** – Componente 1 (variograma 4);
- ✓ **Modelo 6 : Variável Primária** – Produção de viagens por motivo trabalho no setor de serviços (variograma 8) ; **Variável Secundária** – Componente 1 (variograma 4);
- ✓ **Modelo 7 : Variável Primária** – Atração de viagens por transporte coletivo (variograma 9); **Variável Secundária** – Componente 2 (variograma 2);
- ✓ **Modelo 8: Variável Primária** – Atração de viagens por automóvel (variograma 10); **Variável Secundária** – Componente 1 (variograma 4);

- ✓ **Model 9: Variável Primária** – Atração de viagens por modo não motorizado (variograma 11); **Variável Secundária** – Componente 1 (variograma 4);
- ✓ **Modelo 10 : Variável Primária** – Atração de viagens por motivo trabalho na indústria (variograma 12); **Variável Secundária** – Componente 1 (variograma 4);
- ✓ **Modelo 11 : Variável Primária** – Atração de viagens por motivo trabalho no setor de comércio (variograma 13); **Variável Secundária** – Componente 2 (variograma 2);
- ✓ **Model 12 : Variável Primária** – Atração de viagens por motivo trabalho no setor de serviços (variograma 14) ; **Variável Secundária** – Componente 2 (variograma 2).

Na etapa de validação, foram calculadas algumas métricas, tais como coeficiente de correlação de *Pearson*, Erro Médio Absoluto e Variância dos erros, considerando os valores observados e estimados pelas duas técnicas de análise espacial de dados em 374 pontos de coordenadas geográficas. Foram excluídas 15 zonas de tráfego, com valores nulos de produção ou atração de viagens.

Foram obtidos valores mais altos de coeficientes de *Pearson* entre valores observados e estimados através da RGP. No entanto, a KDE gerou menores valores de erros e variâncias de erros.

• Resultados – Trabalho 6

Inicialmente, foi realizada aplicação da Análise em Componentes Principal (ACP). Através do critério da raiz latente foram extraídas 4 componentes que explicavam aproximadamente 88% da variância total dos dados. Observando-se os valores dos escores fatoriais, foi possível propor uma nomenclatura para as quatro componentes a partir da contribuição das variáveis originais em cada uma das componentes.

Componente 1: População de baixa renda;

Componente 2: População de alta renda;

Componente 3: Empregos;

Componente 4: Matrículas escolares.

Para a aplicação da Krigagem com Deriva Externa, foi escolhida a Componente 1 por ser essa, dentre as quatro componentes, aquela que possui maior coeficiente de correlação com a variável primária (Produção de viagens por transporte coletivo). Além disso, a Componente 1 explica 48% da variância total dos dados.

Desta forma, foram ajustados variogramas experimentais, tanto para a variável primária (Produção de viagens por transporte coletivo) quanto para variável secundária (Componente 1) – Figura 4.3. Para a variável primária, foi ajustado um variograma omnidirecional (isotrópico), enquanto que para a variável secundária foi adotada, como direção principal, 90° (anisotrópico).

Figura 4.3 Variogramas modelados (produção de viagens por transporte público) omnidirecional e Componente 1 - 90°. Fonte: Lindner *et al.* (2016)

A partir dos parâmetros dos variogramas teóricos das variáveis primária e secundária, foram calculados os ponderadores e estimados os valores da variável primária em locais de valores previamente desconhecidos. O mapa, apresentado na Figura 4.4, associa valores mais altos de produção de viagens por transporte coletivo em zonas de maiores densidades de população de baixa renda e também de maiores densidades populacionais. Tais áreas concentram-se predominantemente no centro, oeste e leste da cidade.

Figura 4.4 Mapa de superfície de valores estimados de produção de viagens por transporte coletivo. Fonte: Lindner *et al.* (2016)

Com objetivo de validar o procedimento metodológico proposto – *Factorial Kriging with External Drift* para estimação de produção de viagens por transporte coletivo – foi calibrado um modelo de Regressão Linear tradicional com variável explicativa equivalente à Componente 1. A Tabela 4.2 apresenta os principais resultados do modelo de Regressão Linear tradicional.

Tabela 4.2 Principais resultados da abordagem tradicional. Fonte: Lindner *et al.* (2016)

Modelo		Variáveis explicativas			
R ²	0.66	coeficientes	t	sig.	
Sig	0.00				
F	632.40				
Soma dos Quadrados					
Regressão	7.49	Constante	0.03	3.035	0.003
Resíduos	3.82	Componente 1	0.69	24.97	0.000

Conforme esperado, as variáveis Componente 1 (População de baixa renda) e produção de viagens por transporte coletivo são diretamente relacionadas e o valor de

Coeficiente de Determinação é considerado razoável para fins de previsão de viagens agregadas. Contudo, a abordagem tradicional possui algumas restrições importantes, já que este é um procedimento que, além de não considerar a influência de fatores locacionais, é incapaz de determinar superfícies de valores de viagens estimadas.

Finalmente, foram calculadas medidas de erros, considerando valores estimados através da abordagem proposta e também da abordagem tradicional não espacial. Neste trabalho, foram calculados valores de Erro médio quadrático, Raiz do erro médio quadrático, Erro médio absoluto e Coeficiente de Correlação de *Pearson*. Tais medidas são apresentadas na Tabela 4.3.

Tabela 4.3 Medidas de erros das duas abordagens: *Factorial Kriging with External Drift* e Regressão Linear. Fonte: Lindner *et al.* (2016)

Método	Erro médio quadrático	Raiz do erro médio quadrático	Erro médio absoluto	Coeficiente de <i>Pearson</i>
FKriging	0.013	0.113	0.080	0.807
Regressão	0.012	0.109	0.075	0.812

Pode-se observar que ambos os procedimentos apresentam medidas similares de erros e Coeficiente de Correlação. Podem ser considerados razoáveis para fins de estimação de produção de viagens. No entanto, o método proposto pelo trabalho mencionado viabiliza a análise espacial dos resultados.

- Análise crítica

Ambos os trabalhos propuseram a aplicação de *Factorial Kriging with External Drift* (FKED) – Análise em Componentes Principais seguida de Krigagem com Deriva Externa – para problemas de geração de viagens agregadas.

Os resultados provenientes da proposta metodológica mostraram que a abordagem pode ser promissora para estudos de previsão de demanda por transportes. O procedimento adotado

permite a estimação de viagens em coordenadas geográficas, cujos valores da variável podem ser previamente conhecidos ou não.

É importante ressaltar que o método é adequado para projeções futuras baseando-se nas covariáveis e também fornece uma superfície contínua de valores estimados de produção e atração de viagens. Estas podem ser apontadas como as principais contribuições do Trabalho 2 e do Trabalho 6.

No entanto, é importante mencionar que, para ambos os estudos, uma suposição inicial foi considerada: todas as unidades de áreas são consideradas homogêneas. Outra suposição importante é a de que os valores das variáveis são uniformemente distribuídos em cada área.

Tais suposições são razoáveis desde que as unidades de áreas sejam de tamanho pequeno, o que não é o caso de nenhum dos dois trabalhos. Procedimentos metodológicos capazes de mitigar tais problemas são propostos no trabalho de Goovaerts (2006), conforme descrito no Capítulo 2. Além disso, em ambos os trabalhos considera-se que as variáveis possuem, predominantemente, distribuição populacional normal.

O Trabalho 2 possui, especificamente, um outro problema relacionado a escolhas das variáveis explicativas ou variáveis secundárias na explicação de produção ou atração de viagens urbanas. Além da alta correlação entre variáveis, seria necessário considerar, em termos críticos, se as variáveis escolhidas realmente são capazes de explicar o fenômeno. Assim, neste trabalho, a Componente 2 (Empregos), por exemplo, foi escolhida erroneamente como variável secundária para interpolação de produção de viagens.

4.4.3 Krigagem com Deriva Externa para projeções de viagens futuras: áreas muito agregadas

- ✓ Trabalho 3: ANÁLISE DE GERAÇÃO DE VIAGENS URBANAS POR TRANSPORTE COLETIVO ATRAVÉS DE TÉCNICAS DE GEOESTATÍSTICA.

✓ Trabalho 4: USO DE KRIGAGEM PARA ESTIMAÇÃO DE GERAÇÃO DE VIAGENS

Ambos os trabalhos adotaram um método que consistia no uso de Krigagem com Deriva Externa para projeções futuras de produção de viagens por transporte público, agregadas por Macrozonas de tráfego (MZs).

O Trabalho 3 é uma dissertação de mestrado, defendida no ano de 2014 no antigo Mestrado em Engenharia Urbana e Ambiental (MEAU - UFBA), atual PPEC – Programa de Pós-Graduação em Engenharia Civil (UFBA), por Samille Santos Rocha, hoje aluna de doutorado no Programa de Pós Graduação em Engenharia de Transportes (EESC – USP).

O Trabalho 4 foi apresentado e publicado nos anais do VI PLURIS (Congresso Luso Brasileiro para o Planejamento Urbano, Regional, Integrado e Sustentável), Lisboa – Portugal em 2014. Este artigo foi elaborado em coautoria das alunas Samille Santos Rocha e Daiane Bittencourt, além da participação da pesquisadora Ana Rita Salgueiro, atualmente pós-doutoranda do programa Ciências sem Fronteiras da Universidade Federal do Ceará. O artigo utilizou resultados parciais da mencionada dissertação. A dissertação relativa ao Trabalho 3 gerou dois artigos científicos, descritos deste documento como Trabalho 4 e Trabalho 5. O Trabalho 5 será tratado mais adiante.

- Objetivos

O principal objetivo dos trabalhos descritos (Trabalho 3 e Trabalho 4) foi realizar projeções futuras de produção de viagens por Transporte Coletivo, na hora pico, em coordenadas de valores conhecidos e desconhecidos, através de Krigagem com Deriva Externa em dados agregados por grandes unidades de áreas (Macrozonas de Tráfego).

- Região

Em ambos os trabalhos a área de estudo foi a Região Metropolitana de Salvador. Inicialmente, foram utilizados dados referentes à Pesquisa Origem-Destino (OD) de 1995 e consideradas 63 Macrozonas de Tráfego. Para projeções futuras, foram utilizados dados

populacionais do ano de 2010, provenientes do último censo demográfico e compatibilizados com as unidades geográficas (Macrozonas - MZs) da Pesquisa Origem-Destino.

O banco de dados utilizado para estimação geoestatística foi formado pelas coordenadas geográficas (UTM) correspondentes aos centroides das respectivas MZs e pelos dados padronizados positivos referentes às variáveis Produção de viagens por Transporte Coletivo e População, Atração de viagens por Transporte Coletivo e Empregos, pares de variáveis altamente correlacionadas entre si.

- Materiais e Método

O estimador utilizado foi a Krigagem com Deriva Externa (KDE). A modelagem geoestatística foi precedida pela calibração de modelos lineares de produção de viagens para o ano de 1995, compatibilização de áreas (setores censitários e Macrozonas de Tráfego) e estimação de produção de viagens para o ano de 2010 a partir dos modelos previamente calibrados.

Finalmente, com os valores da variável população de 2010 (variável secundária) e valores estimados de produção de viagens para o ano de 2010 (variável primária), houve a aplicação da KDE e pôde-se obter um mapa de superfície contínua de valores estimados de produção de viagens por transporte coletivo.

- Resultados

A técnica de Regressão Linear Múltipla foi utilizada para prever a Produção de viagens por Transporte Coletivo (TCO) para 2010, através das variáveis população e renda média do responsável por domicílio. O modelo foi inicialmente calibrado com a amostra de dados correspondente à Pesquisa OD de 1995 e as mesmas variáveis. A equação calibrada com os dados de 1995 corresponde à Equação 4.1. Com a utilização deste modelo, foi possível estimar dados de produção de viagens por TCO para 2010.

$$Prod_{TCO} = 0.17 \times Pop - 0.218 \times Ren \quad (4.1)$$

$Prod_{TCO}$ = Produção de viagens por TCO;

Pop = População da macrozona equivalente

Ren = Renda média do responsável do domicílio.

A Tabela 4.4 traz os principais resultados do modelo linear obtido com os dados de 1995. Esta tabela mostra que o modelo de regressão linear múltipla obteve um elevado Coeficiente de Determinação, além disso, os parâmetros estimados fazem sentido na explicação do fenômeno produção de viagens. É possível verificar que a variável população foi a mais significativa para estimar a variável analisada.

Tabela 4.4 - Resultados do modelo calibrado para estimação de produção de viagens. Fonte Rocha (2014)

Modelo		Variáveis independentes		
R ²	0.964	Coeficientes		t
Soma dos quadrados				
Regressão	3.75E+09	Constante	0	
Resíduos	1.38E+08	Renda	-2.18E-01	-3.706
Total	3.88E+09	População	0.177	30.39

Vale ressaltar que, para aplicação da Krigagem com Deriva Externa, são necessários valores de variáveis primárias e secundárias. Por esta razão, primeiramente foram obtidos dados de população (variável secundária) em 2010 pelo último censo demográfico, para posteriormente, com base nos valores populacionais e de renda, se conhecer valores de produção de viagens para 2010 (variável primária) em diversas coordenadas.

Posteriormente, foram compatibilizados os setores censitários, referentes ao último censo, com as MZs de tráfego da Pesquisa OD de 1995 (dados originais). A análise das duas áreas e sua compatibilização foi feita através do *software Arcgis 9.3*. Neste SIG foi feita a sobreposição dos setores censitários do censo 2010 com MZs da pesquisa OD 1995, para extrair dados atuais de população para as MZs. Selecionados os setores pertencentes a cada MZ, foi utilizada a ferramenta *PROC-V* do *Excel* para localizar os setores dentro de uma zona, através do seu

código. Após a compatibilização, foi extraído o total da população e a renda média do responsável por domicílio por MZ equivalente em 2010.

Os dados de renda foram corrigidos em relação ao percentual da diferença do salário mínimo quando comparado ao valor referente ao ano de 1995. Para a correção da renda em relação ao valor do Salário Mínimo (SM) de 1995, foi encontrado o percentual de reajuste do valor do SM por ano. Foi possível verificar os valores dos SM, e obter a taxa de crescimento de 1995 a 2010. Encontrou-se um valor de aproximadamente 170% de aumento do SM entre os dois períodos. Finalmente, substituídos os valores das variáveis explicativas para 2010, foram estimados os valores de Produção de viagens por Transporte Coletivo para 2010.

Para construção e modelagem dos variogramas, foram considerados modelos esféricos e VR anisotrópicas (Produção 2010 e População 2010), com direção principal 50°. Vale ressaltar que foi possível a construção de mapas de krigagem para estimação de produção de viagens por TCO em aproximadamente 10.900 células desconhecidas. Observou-se ainda uma tendência de aumento da produção de viagens por TCO crescente, do centro para o subúrbio ferroviário na região. A Figura 4.5 ilustra o mapa de KDE para produção de viagens por TCO.

Figura 4.5 Valores interpolados de Produção de viagens por TCO para 2010. Fonte Pitombo *et al.* (2014)

A validação cruzada mostrou bons resultados, considerando erros médios de estimação e variância dos erros. No entanto, foram observados valores baixos de coeficiente de correlação entre viagens estimadas (krigagem) e observadas (produção 2010).

- Análise crítica

A partir de dados populacionais do censo demográfico brasileiro de 2010 e de compatibilização de unidades de área, foi possível fazer uma interpolação espacial considerando estimativas de produção de viagens por TCO para o ano de 2010 (variável primária) a partir da variável população de 2010 (variável secundária).

Através da interpolação espacial proposta nestes trabalhos, foi possível obtenção de estimativas de produção de viagens em qualquer ponto que compõe a superfície interpolada. Esta é uma contribuição significativa, sobretudo considerando a carência de dados relativos a viagens urbanas da região.

Uma questão importante a ser levada em consideração é que as variáveis krigadas não são naturais, mas sim foram levantadas a partir da Pesquisa O/D. Além disso, salienta-se a alta agregação dos dados e a falta de continuidade espacial das variáveis. Deve-se considerar, no entanto, o caráter inédito e incipiente dos estudos, no que diz respeito à utilização de técnicas pouco comuns à análise de demanda por transportes.

Além disso, nesses trabalhos são apontados como pontos negativos a suposição de distribuição normal das Variáveis Regionalizadas e também o ajuste de variogramas teóricos que, possivelmente, não teriam patamar.

4.5. GERAÇÃO DE VIAGENS E GEOESTATÍSTICA: UMA ANÁLISE DESAGREGADA

4.5.1 Krigagem Ordinária para estimação de geração de viagens: proposta simplificada para desagregação de dados em áreas.

- ✓ *Trabalho 3: ANÁLISE DE GERAÇÃO DE VIAGENS URBANAS POR TRANSPORTE COLETIVO ATRAVÉS DE TÉCNICAS DE GEOESTATÍSTICA.*
- ✓ *Trabalho 5: MODELAGEM GEOESTATÍSTICA APLICADA À AMOSTRA SISTEMÁTICA DESAGREGADA: ESTIMAÇÃO DE GERAÇÃO DE VIAGENS*

Ambos os trabalhos propunham a desagregação de dados provenientes de áreas muito agregadas, através da extração de centroides de uma malha regular. Em seguida, foi aplicada a Krigagem Ordinária para estimação de produção e atração de viagens por transporte público.

Conforme mencionado anteriormente, o Trabalho 3 é uma dissertação de mestrado, defendida pela discente Samille Santos Rocha, no ano de 2014. Já o Trabalho 5 foi apresentado e publicado nos anais do XXIX ANPET (Congresso Anual da Associação Nacional de Pesquisa e Ensino em Transportes - ANPET), realizado em Curitiba, Paraná, em 2014. Este artigo foi elaborado em parceria com as alunas Samille Santos Rocha e Anabele Linder. O artigo utilizou resultados parciais da mencionada dissertação.

- Objetivos

O objetivo dos trabalhos foi avaliar melhorias nas estimativas de modelagem geoestatística para geração de viagens por transporte coletivo, a partir de uma proposta simplificada de desagregação dos dados associados a Macrozonas de tráfego.

- Região

A área de estudo foi a Região Metropolitana de Salvador (RMS) e o banco de dados utilizado foi referente à Pesquisa Origem-Destino (OD) de 1995 e compreende o município de Salvador, além de 10 municípios. A unidade de área da pesquisa OD foi Macrozona de tráfego (MZ). Estas macrozonas (63 no total) foram classificadas em subregiões e em quatro grandes regiões.

Para a análise geoestatística, as coordenadas geográficas dos centroides das macrozonas foram associadas às variáveis de produção de viagens por TCO e atração de viagens por TCO. Foram utilizados valores padronizados positivos das duas variáveis em questão.

- Materiais e Método

Os trabalhos descreviam a estimação de geração de viagens por Transporte Coletivo (TCO) em coordenadas cujos valores da variável de interesse são desconhecidos, através da técnica de interpolação espacial - Krigagem Ordinária (KO). Os trabalhos utilizaram amostras de pontos (dados da Pesquisa Origem/Destino da RMS de 1995). Além dos dados originais da Pesquisa O/D, foi proposta uma amostra sistemática artificial para geração de dados desagregados. Resultados provenientes de modelagem geoestatística aplicada às duas amostras foram apresentados. A Figura 4.6 esquematiza o método proposto.

Figura 4.6: Etapas do processo metodológico aplicado ao estudo. Fonte: Rocha *et al.* (2014)

A fim de desagregar o banco de dados, adotou-se a amostragem sistemática artificial, em que foram seguidos os seguintes passos: (1) Determinação da malha; (2) Geração de centroides por cada célula que compõe a malha; (3) Determinação de valores das variáveis padronizadas Produção e Atração de viagens por TCO associados às coordenadas dos centroides das células.

- Resultados

Após alguns testes, a determinação da malha se deu pela geração de células de 2000 x 2000 m. Em seguida, foram gerados centroides para cada célula obtida. As células quadradas obtidas e os centroides associados estão ilustrados na Figura 4.7.

Figura 4.7 Malha gerada para amostragem sistemática artificial. Fonte: Rocha *et al.* (2014)

Notou-se que cada célula era composta por diferentes macrozonas. Tornou-se necessário, então, mensurar a ocupação, em porcentagem, das macrozonas em cada célula.

A célula 2 (ou quadrícula 2), por exemplo, é composta pela MZ 1, MZ 6, MZ 7, MZ 8 e MZ 11, conforme ilustrado na Figura 4.7. A MZ 7, por exemplo, ocupa 32,8% da célula 2. A fim de determinar o valor das variáveis de geração de viagens para cada quadrícula, a partir dos valores conhecidos das macrozonas, atribuiu-se um peso para cada MZ ocupada nas quadrículas. O exemplo da MZ 7 na quadrícula 2, teve um peso de 0,328, a MZ 11 teve um peso de 0,335, e assim por diante (vide Tabela 4.5). Multiplicou-se o peso de cada MZ pelo seu respectivo valor de geração de viagem. O somatório dessas multiplicações resultou no valor da quadrícula e foi associado à respectiva coordenada geográfica do centroide.

Tabela 4.5 Exemplo de ponderação da área e determinação da variável para amostragem sistemática artificial (célula ou quadrícula 2). Fonte: Rocha *et al.* (2014)

Área da quadrícula (km ²)	ID (MZ), Equivalente a 1 célula da malha	Área ocupada pela MZ em cada célula (km ²)	Peso (%)	Produção de viagens por MZ (padronizada)	Valor da variável após compatibilização	Média ponderada por célula
2,975	1	0,424	0,142	0,18	0,03	0,18
	6	0,366	0,123	0,18	0,02	
	7	0,974	0,328	0,15	0,05	
	8	0,191	0,064	0,22	0,01	
	11	0,997	0,335	0,20	0,07	

Após a determinação dos valores das variáveis de interesse nos centroides da malha regular, foi realizada a modelagem geoestatística por Krigagem Ordinária para determinar os valores estimados da geração de viagens.

A partir da análise do mapa de pontos visualizados no *GeoMS*, verificou-se a direção que apresenta maior variabilidade dos dados para a Variável Regionalizada de produção de viagens por TCO. Perceberam-se, por meio da distribuição espacial dos dados, as áreas que apresentam maiores valores da Variável Regionalizada, sendo estas localizadas na região do Subúrbio Ferroviário de Salvador. Esta área é composta pelos bairros mais populosos e com renda mais baixa da cidade. Já a tendência contrária é observada na região central.

Comparando os resultados da análise das variáveis de produção de viagens da amostra agregada com a desagregada, percebeu-se que a distribuição espacial das variáveis

continuou similar aos pontos analisados por centroides das macrozonas, apresentando, porém, uma quantidade maior de dados e diminuição da distância entre estes.

Após visualização espacial e realização de alguns testes, optou-se por modelar os variogramas experimentais para a direção 50° NE e o seu respectivo ângulo ortogonal 320° SE para os dados de produção de viagens por TCO por centroides das macrozonas. Este ângulo representou melhor a continuidade espacial das amostras.

A variável de atração de viagens por TCO possui comportamento isotrópico, e mesmo adaptada à malha regular de 2000 por 2000 m, não demonstrou nenhuma continuidade ou variação aparente dos dados em alguma direção. O variograma foi modelado com o ângulo máximo de abertura (180° - omnidirecional).

Os dados desagregados não possuem uma direção principal de variabilidade. As variáveis foram também modeladas com o ângulo máximo de abertura (variograma omnidirecional). Verificou-se que a amostragem artificial gerou uma melhora significativa no ajuste dos variogramas experimentais, em relação à amostragem pelos centroides das macrozonas devido à maior quantidade de pontos localizados na área de estudo.

Os parâmetros de azimuth θ , tolerância angular (R), *Lag distance* (h), largura máxima foram utilizados para a modelagem dos variogramas experimentais das duas variáveis analisadas. No total, foram gerados três variogramas para a amostragem original e dois variogramas para a amostragem obtida pela malha regular (amostra sistemática).

Na etapa de modelagem dos variogramas, foram determinados os seguintes parâmetros: efeito pepita, alcance ou amplitude, patamar e o modelo teórico de ajuste do variograma. Os variogramas ajustados estão ilustrados na Figura 4.8.

Observa-se claramente a melhoria na estruturação espacial da Variável Regionalizada Atração de viagens por TCO, através da proposta da amostragem sistemática artificial. O

variograma correspondente à Atração de viagens para a amostra original não apresenta nenhuma estrutura espacial (efeito pepita puro - Figura 4.8 (a)). Enquanto que o variograma correspondente à variável Atração de viagens por TCO para amostra desagregada apresenta uma estrutura que se assemelha ao modelo esférico (Figura 4.8 (b)), com valor de efeito pepita reduzido (0,068).

A análise dos resultados da validação cruzada para as duas variáveis, em ambas as amostras de pontos, apresentou, no geral, bons resultados (Tabela 4.6). A média de erros e a variância dos resíduos apresentaram valores baixos.

No entanto, o coeficiente de correlação entre valores observados e estimados das variáveis para as amostras agregadas apresentou baixos valores, devido ao modelo teórico do variograma ter baixa variância espacial e apresentar alto efeito pepita. Considerando a amostra desagregada, o coeficiente de correlação apresentou significativo aumento tanto para produção de viagens, quanto para atração de viagens por TCO. Os demais resultados são satisfatórios, com valores baixos de média de erros e variância de resíduos.

Figura 4.8: Modelagem dos variogramas. Fonte: Rocha *et al.* (2014)

Tabela 4.6: Dados obtidos pela validação cruzada da estimação das variáveis

Variável	Coefficiente de correlação	Média dos erros	Variância dos erros
Produção de viagens por TCO	0,20	0,0005	0,476
Atração de viagens por TCO	0,40	-0,0030	0,747
Produção de viagens por TCO (desagregada)	0,84	-0,0041	0,386
Atração de viagens por TCO (desagregada)	0,77	-0,0208	0,409

Os mapas de interpolação, obtidos através da Krigagem Ordinária, para o caso da amostra agregada e amostra sistemática artificial são mostrados na Figura 4.9 (a,b). A mesma figura ainda apresenta mapas temáticos de valores observados das variáveis produção e atração de viagens por TCO para facilitar comparação visual de valores obtidos através da interpolação e valores reais (Figura 4.9c).

Fazendo-se apenas uma análise visual entre as interpolações (Figura 4.9 - a,b) e os valores observados por área (Figura 4.9 – c), observa-se que, para produção de viagens por TCO, a interpolação proveniente da amostra desagregada assemelha-se mais aos dados

observados, enquanto que, para Atração de viagens por TCO, a interpolação proveniente da amostra agregada aproxima-se dos valores observados.

Figura 4.9 Interpolação espacial para as variáveis Produção e Atração de viagens por TCO: (a) amostra agregada; (b) amostra desagregada; (c) valores observados por área. Fonte: Rocha et al. (2014).

- Análise crítica

A proposta metodológica destes trabalhos alcançou resultados satisfatórios para a aplicação da geoestatística na análise de demanda por transportes. Contudo, ressalvas devem ser apontadas, como o ajuste de variogramas com patamar, por exemplo.

Na obtenção de dados para a amostra sistemática artificial, atentou-se para algumas considerações:

- A desagregação de dados desconsidera mudanças abruptas nos valores de variáveis socioeconômicas e de viagens, que em geral não apresentam continuidade espacial equivalente a amostras de solos ou de água, por exemplo, comumente utilizadas em modelagem geoestatística;

- Adotou-se a premissa da continuidade espacial dos valores das variáveis através da utilização da técnica de geração de amostras desagregadas. No entanto, sabe-se que as amostras agregadas possuem descontinuidades mais evidentes;

- Os dados devem ser preferencialmente gerados através de técnicas mais sofisticadas, tais como o procedimento proposto por Goovaerts (2006) e não média ponderada de valores das macrozonas.

O método empregado neste estudo é uma tentativa inicial de melhorar a modelagem geoestatística a partir de dados muito agregados, sabendo que muitas vezes, apenas estes estão disponibilizados. Portanto, o método proposto para desagregação amostral tem o intuito de suavizar os efeitos da falta de continuidade espacial das variáveis em estudo e suportes geométricos heterogêneos. Outro aspecto favorável da desagregação dos dados foi verificado pelo aprimoramento da estrutura espacial das Variáveis Regionalizadas. Considerando os resultados de validação cruzada, observou-se significativa melhora das estimativas e interpolações por Krigagem Ordinária com a amostragem sistemática artificial.

4.6. CONSIDERAÇÕES FINAIS

Este capítulo descreveu os seis trabalhos relativos ao tema Geração de viagens e Geoestatística. Mostrou diferentes aplicações voltadas a bases de dados geograficamente agregadas, além de ter apresentado uma proposta preliminar para desagregação de dados em áreas para aplicação da geoestatística. O Quadro 4.3 sintetiza os seis trabalhos e aponta os pontos positivos e negativos de cada aplicação aqui descrita.

Quadro 4.3 Síntese dos seis trabalhos discutidos (geração de viagens e geoestatística)

Trabalho	Objetivos e Método	Pontos Positivos	Pontos Negativos
1	Estimar taxas de produção e atração de viagens por modo de transporte e motivo de viagem, através de modelagem geoestatística.	Resultados Razoáveis	Univariada
		Apresentação da Geoestatística	Centroides/suportes heterogêneos
		Pouca informação necessária	Ausência de validação com ferramenta tradicional
	Uso de KO		Ajustes de variogramas Suposição de normalidade Método não viável para projeções futuras
2	Estimar taxas de produção e atração de viagens por modo de transporte e motivo de viagem, através de Fkriging e Regressão Geograficamente Ponderada	Resultados Razoáveis	Centroides/suportes heterogêneos
		Análise espacial multivariada	Suposição de normalidade
		Possível realizar projeções futuras	Uso indevido de todas as variáveis explicativas nos modelos de produção e atração de viagens
		Proposta metodológica	
3	3.1 Realizar projeções futuras de produção de viagens por Transporte Coletivo na hora pico em coordenadas de valores conhecidos e desconhecidos através de Krigagem com Deriva Externa	3.1 Resultados Razoáveis	3.1 Centroides/suportes heterogêneos
		Análise espacial bivariada	Suposição de normalidade
		Possível realizar projeções futuras	Ajustes de variogramas
		Proposta metodológica	
	3.2 Avaliar melhorias nas estimativas de modelagem Geoestatística para geração de viagens por transporte coletivo, a partir de uma proposta simplificada de desagregação dos dados associados a Macrozonas de tráfego	3.2 Bons resultados	3.2 Ajustes de variogramas
		Proposta metodológica incipiente	Proposta metodológica simplificada
4	Realizar projeções futuras de produção de viagens por Transporte Coletivo na hora pico em coordenadas de valores conhecidos e desconhecidos através de Krigagem com Deriva Externa	Dados desagregados	Suposição de normalidade
		Melhor estrutura espacial das VR	Ajustes de variogramas
		Resultados Razoáveis	Centroides/suportes heterogêneos
		Análise espacial bivariada	Suposição de normalidade
5	Avaliar melhorias nas estimativas de modelagem Geoestatística para geração de viagens por transporte coletivo, a partir de uma proposta simplificada de desagregação dos dados associados a Macrozonas de tráfego	Possível realizar projeções futuras	Ajustes de variogramas
		Proposta metodológica	
		Bons resultados	Ajustes de variogramas
		Proposta metodológica incipiente	Proposta metodológica simplificada
6	Estimar taxas de produção de viagens por transporte público, através de Fkriging	Dados desagregados	Suposição de normalidade
		Melhor estrutura espacial das VR	
		Resultados Razoáveis	Centroides/suportes heterogêneos
		Análise espacial multivariada	Suposição de normalidade

Capítulo

5

ESCOLHA MODAL E GEOESTATÍSTICA

Este capítulo trata da quarta etapa da linha de pesquisa, que é o uso da geoestatística em dados qualitativos desagregados (domicílios ou indivíduos). Sistematiza os seis trabalhos que fazem parte do tema Escolha Modal e Geoestatística. Conceitos básicos da etapa de escolha modal são apresentados em forma de quadro resumo. Seguidamente, são descritos trabalhos que envolvem a incorporação de fatores espaciais aos modelos de escolha discreta. Finalmente, são descritos os seis trabalhos desenvolvidos pela autora e colaboradores, acompanhados de uma análise crítica dos resultados obtidos e procedimentos metodológicos aplicados.

5.1. MODELOS DE ESCOLHA MODAL

As razões que levam um usuário a escolher o modo de transporte são bastante complexas e diversas, tais como: uso do solo; infraestrutura de transportes; características ligadas ao indivíduo, às viagens, às alternativas, além do local em que este reside, trabalha, estudo, etc.

A etapa de escolha modal é, possivelmente, aquela mais importante em termos de planejamento urbano e de transportes. A escolha de modos de transporte coletivo em detrimento dos particulares motorizados, por exemplo, pode acarretar benefícios, tais como redução de acidentes e emissões veiculares (ORTÚZAR e WILLUMSEM, 2011). É importante o desenvolvimento de modelos sensíveis para detectar a influência de mudanças de cenários de diferentes modos de transportes (tarifa, tempo de viagem, conforto, etc.) em relação às escolhas individuais do modo de transporte.

Modelos de escolha modal podem ser agregados, baseados em zonas de tráfego, ou desagregados, baseados em dados domiciliares ou individuais. Os modelos desagregados de escolha modal geralmente são probabilísticos, são os chamados modelos econométricos ou modelos de escolha discreta, explorados inicialmente por Ben-Akiva (1973), McFadden (1974) e Domencich e McFadden (1975). O Quadro 5.1 traz as principais definições de modelos de escolha discreta e a etapa de escolha modal. Os principais conceitos foram extraídos de Ortúzar e Willumsem (2011); Ben-Akiva (2010); Ben-Akiva e Lerman (1985).

Quadro 5.1 Conceitos básicos da etapa de escolha modal

Variáveis que influenciam a escolha modal	Característica do viajante, da viagem e facilidades de transporte
Suposições básicas de modelos de escolha discreta	A probabilidade de indivíduos escolherem uma dada opção é função de suas características socioeconômicas e da atratividade relativa da opção.
Principais modelos	Logit, Probit, Logit Aninhado, Logit misto
Pressupostos	Ser humano é capaz de perceber a utilidade de um objeto
	É capaz de comparar duas ou mais utilidades
	Escolhe o objeto de maior utilidade
	A utilidade de um modo de viagem pode ser quantificada em função de atributos do modo e de atributos socioeconômicos do usuário
	O erro na percepção de utilidade segue a distribuição acumulada de Gumbel
A função utilidade	$U_{in} = V_{in} + \varepsilon_{in}$
Forma linear da componente sistemática da utilidade	$V_{in} = \beta_0 + \beta_1 x_{in1} + \beta_2 x_{in2} + \beta_3 x_{in3} + \dots + \beta_k x_{ink}$
Probabilidade da alternativa 1 (duas alternativas)	$P_1 = \frac{\exp(U_1)}{\exp(U_1) + \exp(U_2)}$
Dados utilizados	Preferência Revelada e Preferência Declarada
Preferência Revelada	As escolhas efetivamente realizadas pelos indivíduos a partir de uma situação existente
Preferência Declarada	Referem-se às escolhas diante de um conjunto de opções em que são apresentados cenários hipotéticos

U_{in} a utilidade global de uma alternativa i para um indivíduo n , V_{in} é a componente sistemática da utilidade de uma alternativa i para um indivíduo n e ε_{in} , a componente aleatória (que pode ser uma função), representa uma parcela desconhecida da função utilidade que captura a dispersão das escolhas e os fatores não controláveis ou desconhecidos do pesquisador, tais como atributos relevantes não considerados, erros de medição dos atributos, erros de percepção do entrevistado e diferenças não observáveis na avaliação. β_0 : constante específica da alternativa; β_k : peso relativo do atributo x_{ink} na composição da função utilidade.

5.2. ANÁLISE ESPACIAL RELATIVA À ESCOLHA MODAL: **BACKGROUND**

O uso convencional de modelos de escolha discreta a problemas de escolha modal, em geral, ocorre a partir de variáveis explicativas, relacionadas a atributos socioeconômicos dos indivíduos, a impedâncias dos modos de transporte e nível de serviço. Contudo, na abordagem tradicional, a localização espacial das variáveis não é considerada.

Ao longo dos anos, muitos estudos corroboraram a afirmação de que comportamentos subjacentes a viagens urbanas, sobretudo associados à escolha do modo de transporte, são fortemente relacionados à distribuição espacial das atividades no meio urbano, presença de zonas de tráfego de uso do solo misto, etc. (CERVERO e RADISCH, 1996; KITAMURA *et al.*, 1997).

Avanços tecnológicos e maior disponibilidade de informações georreferenciadas tornaram a inclusão de efeitos espaciais em modelos matemáticos cada vez mais popular na área de modelagem de demanda por transportes (PÁEZ *et al.*, 2013; PÁEZ e FISHER, 2007; PÁEZ e SCOTT, 2005). O Quadro 5.2 sintetiza diferentes abordagens de análise espacial de dados, aplicadas especificamente a problemas de escolha modal. Todas as abordagens mencionadas referem-se a estimativas relativas a pontos ou localizações amostradas.

Quadro 5.2 Exemplos de trabalhos relativos ao tema Análise espacial em escolhas discretas.

Trabalho	Objetivos	Técnicas
Bhat e Guo (2004)	Propor um modelo logit espacialmente correlacionado	Modelos logísticos mistos e multinomiais e variáveis relativas a adjacências espaciais de cada alternativa
Miyamoto <i>et al.</i> (2004)	Combinar conceitos de autocorrelação espacial em modelos logísticos hierárquicos	Modelos logísticos hierárquicos e variáveis relativas a autocorrelação espacial
Bhat e Sener (2009)	Acomodar dependência espacial a modelos logit binomiais	Modelos logit com erros correlacionados (dependência espacial)
Dugundji e Walker (2005)	Investigar decisões individuais relativas à escolha modal considerando a influência de indivíduos que possuem o mesmo status social e residem em locais próximos (rede espacial de relacionamentos)	Kemel Logit Aninhado
Paez <i>et al.</i> (2013)	Introduzir um indicador espacial em modelos de escolha discreta	Logit Multinomial Associação espacial Simulação Monte Carlo

5.3. SISTEMATIZANDO OS TRABALHOS

A apresentação dos trabalhos foi classificada de acordo com o tipo de procedimento proposto. Inicialmente, são apresentados trabalhos que utilizaram um procedimento metodológico sequencial, composto de etapa multivariada não espacial seguida de etapa espacial univariada. Posteriormente, são descritos os trabalhos que utilizaram o método composto apenas pela modelagem geoestatística univariada – Krigagem Indicativa para dados discretos.

A Tabela 5.1 descreve os trabalhos a serem discutidos neste capítulo e seus respectivos procedimentos metodológicos. Vale ressaltar que todos os trabalhos utilizaram dados desagregados por domicílios ou indivíduos.

Tabela 5.1 Trabalhos a serem discutidos no Capítulo 5

Trabalho	Ano	Título	Autores	AC AR DM*	Procedimento
1	2013	PROPOSTA DE UM MÉTODO PARA ESTIMAÇÃO DE ESCOLHA MODAL ATRAVÉS DA GEOESTATÍSTICA	Aline Schindler Gomes da Costa	Dissertação (Mestrado) UFBA (DM)	Sequencial multivariado
2	2015a	A TWO-STEP METHOD FOR MODE CHOICE ESTIMATION WITH SOCIOECONOMIC AND SPATIAL INFORMATION	Cira Souza Pitombo Ana Rita Salgueiro Aline S. Gomes da Costa Cassiano A. Isler	Spatial Statistics (AR)	Sequencial multivariado
3	2015b	PROPOSAL OF A SEQUENTIAL METHOD FOR SPATIAL INTERPOLATION OF MODE CHOICE	Cira Souza Pitombo Aline S. Gomes da Costa Ana Rita Salgueiro	BCG - Boletim de Ciências Geodésicas (AR)	Sequencial multivariado
4	2015	INTERPOLAÇÃO ESPACIAL DE PROBABILIDADES DE ESCOLHA DE MODO DE TRANSPORTE MOTORIZADO	Anabele Lindner Cira Souza Pitombo	XXIX Anpet (2015) (AC)	Sequencial multivariado
5	2015	ANÁLISE DESAGREGADA DE DADOS DE DEMANDA POR TRANSPORTES ATRAVÉS DE MODELAGEM GEOESTATÍSTICA E TRADICIONAL	Anabele Lindner	Dissertação (Mestrado) USP (DM)	Etapa única univariada
6	2016	KRIGING GEOSTATISTICAL METHODS FOR TRAVEL MODE CHOICE: A SPATIAL DATA ANALYSIS TO TRAVEL DEMAND FORECASTING	Viviane Antunes Gomes Cira Souza Pitombo Samille Santos Rocha Ana Rita Salgueiro	Open Journal of Statistics Spatial Statistics Research (Special Issue) (AR)	Etapa única univariada
*AC: Artigo publicado em anais de congresso; AR: Artigo publicado em revista; DM: Dissertação de mestrado defendida					

5.4. ESCOLHA MODAL E GEOESTATÍSTICA: PROCEDIMENTO METODOLÓGICO SEQUENCIAL MULTIVARIADO

Esta seção apresenta os trabalhos nos quais houve contribuição metodológica importante. O método sequencial consiste em um procedimento de duas etapas, onde a primeira era determinada pela aplicação de uma técnica não espacial de análise multivariada, seguida da modelagem geoestatística. A finalidade do procedimento é a de gerar uma superfície contínua de probabilidades de escolhas modais. Importante ressaltar que o método proposto pode também ser utilizado para projeções futuras, baseadas nos valores futuros das covariáveis.

Foram publicados dois trabalhos em periódicos nacionais e internacionais de grande impacto (Boletim das Ciências Geodésicas e *Spatial Statistics*), apresentado um trabalho no congresso da ANPET 2015 – ASSOCIAÇÃO NACIONAL DE PESQUISA E ENSINO EM TRANSPORTES (Ouro Preto, MG), além de defendida uma dissertação de mestrado.

Analogamente ao Capítulo 4, para a apresentação dos trabalhos, nas próximas subseções, a autora sugere uma sequência de informações, ilustrada na Figura 5.1, para melhor organização do texto.

Figura 5.1 Sequência de informações para apresentação das pesquisas

5.4.1 Método: Árvore de Decisão e Krigagem Ordinária

- ✓ Trabalho 1 - PROPOSTA DE UM MÉTODO PARA ESTIMAÇÃO DE ESCOLHA MODAL ATRAVÉS DA GEOESTATÍSTICA;
- ✓ Trabalho 2: A TWO-STEP METHOD FOR MODE CHOICE ESTIMATION WITH SOCIOECONOMIC AND SPATIAL INFORMATION;
- ✓ Trabalho 3: PROPOSAL OF A SEQUENTIAL METHOD FOR SPATIAL INTERPOLATION OF MODE CHOICE

O Trabalho 1 é uma dissertação de mestrado, defendida no ano de 2013 no antigo Mestrado em Engenharia Urbana e Ambiental (MEAU - UFBA), atual PPEC – Programa de Pós-Graduação em Engenharia Civil (UFBA), por Aline S. Gomes da Costa. Esta dissertação gerou dois artigos científicos, relativos ao Trabalho 2 e ao Trabalho 3.

O Trabalho 2 é o primeiro produto da linha de pesquisa, publicado em periódico internacional *Spatial Statistics* (<http://www.journals.elsevier.com/spatial-statistics>) em fevereiro de

2015. Este artigo foi elaborado em coautoria com a ex-aluna Aline S. Gomes da Costa, além da participação dos pesquisadores Ana Rita Salgueiro (Universidade Federal do Ceará) e Cassiano A. Isler (Universidade Federal de Santa Catarina). O artigo utilizou resultados parciais da mencionada dissertação, que foram aprimorados para publicação.

O Trabalho 3 foi publicado no periódico nacional BCG – Boletim das Ciências Geodésicas (<http://revistas.ufpr.br/bcg>), também em 2015. Este artigo foi elaborado em coautoria com a ex-aluna Aline S. Gomes da Costa, além da participação da pesquisadora Ana Rita Salgueiro (Universidade Federal do Ceará).

Há variações metodológicas entre o Trabalho 2 e o Trabalho 3, além do uso de bases de dados diferentes. As diferenças são apontadas nos itens seguintes.

- Objetivos

O objetivo geral destes trabalhos (Trabalho 1, Trabalho 2 e Trabalho 3) é propor um método em duas etapas (sequencial) para estimar escolhas modais dos viajantes em coordenadas geográficas cujos valores da variável são previamente conhecidos ou não.

- Região

A área de estudo relativa aos trabalhos foi a cidade de São Carlos (São Paulo) e os dados provenientes da Pesquisa Origem/Destino, realizada em 2007/2008 (RODRIGUES DA SILVA, 2008). Após o tratamento da amostra, foram utilizados dados desagregados por indivíduo (chefe da família), associados a características socioeconômicas, de viagens e latitude e longitude do domicílio (total de 1216 registros).

A amostra de 1216 indivíduos foi utilizada na etapa não espacial e multivariada dos trabalhos – aplicação da Árvore de Decisão. Para a etapa espacial univariada – Krigagem Ordinária – foram utilizados dados provenientes de duas regiões menores da cidade, onde se observou um padrão espacial da variável *Probabilidades de Escolha Modal*.

As regiões analisadas no Trabalho 1 e 2 foram: (1) Região de Santa Felícia (110 pontos amostrados), onde se observou distribuição espacial aparente para modos de transporte motorizados e (2) Região Central (474 pontos amostrados), onde havia maior número de pontos amostrados. Já no Trabalho 3 foi utilizada apenas a Região de Santa Felícia.

A variável dependente ou objeto de estudo é a escolha modal, que possui três categorias: (1) Modo de transporte individual motorizado (carro ou automóvel); (2) Modo de transporte público (ônibus) e (3) Modo de transporte não motorizado (a pé).

- Materiais e Método

O procedimento metodológico seguido pelos três trabalhos possui uma etapa anterior à modelagem geoestatística, que consiste na etapa não espacial com aplicação de algoritmo de Árvore de Decisão. O estimador espacial utilizado nos três trabalhos foi a Krigagem Ordinária (KO).

O método é ilustrado na Figura 5.2. A figura representa o método completo proposto no Trabalho 2. O diferencial metodológico entre o Trabalho 2 e os Trabalhos 1 e 3 consiste na etapa de calibração do modelo Logit Multinomial para validação do algoritmo de Árvore de Decisão.

Além disso, os mesmos trabalhos utilizaram os aplicativos *GeoMS* e *IBM SPSS 22.0*.

Figura 5.2 Procedimento metodológico geral proposto nos Trabalhos 1, 2 e 3 (Costa, 2013; Pitombo *et al.*, 2015a; Pitombo *et al.*, 2015b).

- Resultados

Resultados da etapa não espacial: Árvore de Decisão

Primeiramente, com a amostra de 1216 indivíduos, foi gerada a AD utilizando o algoritmo CART (*Classification and Regression Tree*), adotando-se o parâmetro de mínimo de 25 observações por nó terminal. A variável dependente foi “Modo de transporte principal”, composta por três categorias. Já as variáveis independentes foram as características socioeconômicas dos usuários, as características da viagem e as qualitativas sobre sistema de transporte. Uma das finalidades da aplicação do CART foi investigar as variáveis que contribuem para o entendimento da escolha modal. Outro objetivo essencial para o uso da AD foi a determinação da variável contínua, necessária para a aplicação da KO - valor estimado pelo modelo de AD das probabilidades de utilização dos modos nos locais amostrados.

Na etapa não espacial, percebe-se, por exemplo, a influência positiva das seguintes características para o uso do modo de transporte particular motorizado: possuir CNH, possuir automóveis ou motocicletas no domicílio, ter idade inferior a 59 anos, ser do sexo masculino e considerar como principais problemas de utilização desse modo a falta de estacionamento e o custo elevado. Verificou-se também uma boa porcentagem de acertos de estimação para os três modos analisados: 78% para o modo particular motorizado, 83% para o público e 80% para o modo não motorizado.

Com o intuito de validar a aplicação do algoritmo CART foi feita, apenas no Trabalho 2, a calibração de um modelo Logit Multinomial (MNL), considerando as seguintes funções utilidade.

$$U_{\text{carro}} = \beta_{\text{MOTO}} \cdot \text{MOTO} + \beta_{\text{CARRO}} \cdot \text{CARRO} + \beta_{\text{idade}} \cdot \text{IDADE} + \beta_{\text{GNERO}} \cdot \text{GENERO} + \beta_{\text{INSTRUÇÃO}} \cdot \text{INSTRUÇÃO} + \beta_{\text{nVIAGENS}} \cdot \text{nVIAGENS} \quad (5.1)$$

$$U_{\text{Ônibus}} = \beta_{\text{Ônibus}} \cdot \text{Ônibus} \quad (5.2)$$

$$U_{APF} = ASC_{APF}$$

(5.3)

Observa-se que os resultados, tanto da AD quanto do MNL, foram similares. As variáveis explicativas consideradas significativas no modelo foram também àquelas selecionadas pelo algoritmo CART. O percentual de acertos também foi similar. O MNL apresentou taxas que variavam entre 70% e 80%. Assim, ambas as técnicas são adequadas para serem utilizadas na primeira etapa do procedimento proposto, sendo a AD não paramétrica.

Resultados da etapa espacial: Modelagem geoestatística

A etapa seguinte à aplicação da AD foi elaborar mapas exploratórios das probabilidades geradas pela AD (variáveis a serem utilizadas na KO) com o objetivo de observar padrões espaciais. A partir da análise dos mapas exploratórios, percebeu-se que as variáveis a serem krigadas não apresentaram padrão espacial aparente. Desta forma, optou-se por segregar a cidade de São Carlos em seis regiões considerando o critério da renda, na tentativa de encontrar padrões espaciais das variáveis em análise em regiões menores.

Ao observar os mapas de probabilidades para cada uma das seis regiões separadamente, percebeu-se que a região de Santa Felícia (110 pontos) foi aquela que apresentou um padrão espacial, variando do centro para a periferia. O Trabalho 3 considerou apenas esta região para etapa de KO, enquanto que os Trabalho 1 e 2 levaram em conta também o centro (474 pontos).

Assim, nas etapas correspondentes à modelagem geoestatística, para a região Santa Felícia, foram estabelecidos os parâmetros para construção dos variogramas experimentais, referentes aos valores de probabilidades de escolha modal. Após diversos testes, foi determinada a direção principal (30° - maior dependência espacial dos dados) e a ortogonal (-60°). A distância entre amostras (h – *Lag distance*) foi considerada na faixa de 130 metros (próxima a uma quadra – distância média entre as coordenadas geográficas) e a largura máxima a ser considerada para o cálculo do variograma foi de 100.000 metros. Em seguida, foi feita a modelagem dos variogramas,

determinando outros parâmetros, tais como efeito pepita (C_0), Amplitude (a) e modelo teórico ajustado. A Figura 5.3 ilustra os variogramas modelados na direção 30° para a região de Santa Felícia. Destaca-se o fato de todos os variogramas terem sido ajustados por modelos teóricos esféricos, bem como a ocorrência de todos os modos de transporte apresentarem maiores alcances na direção 30° , com exceção do modo não motorizado, que apresenta maior correlação espacial entre as amostras na direção ortogonal -60° .

Figura 5.3 Variogramas modelados na direção 30° (região Santa Felícia). Fonte: Costa *et al.* (2013)

Já a Figura 5.4 ilustra os variogramas modelados para os três modos de transporte para a região central (Trabalho 2). A direção principal foi 25° para o transporte público e modo não motorizado e 30° para o modo de transporte particular motorizado.

Figura 5.4 Variogramas modelados nas direção 30° e 25° (região central). Fonte: Costa (2013)

Com o objetivo de aferir a qualidade dos modelos determinados pelos variogramas teóricos, foram calculados parâmetros como Coeficiente de correlação, Erro médio Absoluto, Erro médio Quadrático e Variância dos erros. Para todos os variogramas modelados, o teste do ponto fictício mostrou resultados razoáveis para as métricas de desempenho, sendo os melhores resultados relativos à estimativa de modo de transporte público e não motorizado.

Finalmente, houve a realização da interpolação através da Krigagem Ordinária, ao estimar em locais não amostrados os valores das variáveis relativas à utilização dos modos de viagens estudados. Foram estimadas 588 células. A Figura 5.5 ilustra os mapas krigados para os três modos de transporte para a região de Santa Felícia e central de São Carlos. Assim, foram observadas tendências espaciais de escolhas modais em ambas as regiões e foram estimados os valores de probabilidade de escolha em qualquer ponto da superfície contínua.

Figura 5.5 Mapas interpolados para as duas regiões em análise. Fonte: Costa (2013)

- Análise crítica

Os trabalhos descritos nessa seção objetivaram apresentar um método sequencial multivariado para estimar probabilidades de escolha modal em coordenadas geográficas de valores conhecidos e desconhecidos. Os resultados provenientes da etapa não espacial (algoritmo de Árvore de Decisão) foram considerados interessantes e validados pela calibração de um modelo tradicional Logit Multinomial (Trabalho 2).

A etapa não espacial foi essencial para melhor investigação do fenômeno, determinando a influência das variáveis explicativas nas escolhas individuais. Além disso, tal etapa foi fundamental para transformar a variável objeto de estudo em contínua, viabilizando o uso do estimador espacial Krigagem Ordinária.

Os mapas obtidos através da KO mostraram uma tendência do crescimento da escolha pelo automóvel que cresce do centro para a periferia na região de Santa Felícia. Tal tendência acompanha a configuração espacial de renda da mesma região. Já na região central, observa-se uma tendência que segue a direção sudeste – noroeste.

Em termos de contribuições metodológicas, tais trabalhos apresentam uma forma de, não somente investigar o fenômeno de interesse com ajuda de variáveis explicativas, como também analisar estimativas no espaço de forma contínua. Além disso, o procedimento proposto possibilita projeções futuras da variável de interesse (inclusive na forma de mapas krigados), através de projeções das covariáveis.

Em relação a restrições ou limitações metodológicas, pode-se citar a não-normalidade da Variável Regionalizada, além de limitações referentes ao ajuste de variogramas experimentais. Uma alternativa aos modelos aqui ajustados, especialmente aqueles relativos ao transporte público, seriam variogramas esféricos sem patamar.

Além disso, o método sequencial pode acarretar maior propagação de erros, já que a Variável Regionalizada é proveniente de um modelo anterior não espacial. Uma solução estratégica para isso seria o uso de modelagem geoestatística multivariada, com o uso de variáveis explicativas na cokrigagem, ou Krigagem com Deriva Externa, por exemplo.

5.4.2 Método: Logit Binomial e Krigagem Ordinária

✓ Trabalho 4: INTERPOLAÇÃO ESPACIAL DE PROBABILIDADES DE ESCOLHA DE MODO DE TRANSPORTE MOTORIZADO

O Trabalho 4 foi apresentado em 2015 e publicado nos anais do XXIX Congresso da ANPET, realizado em Ouro Preto (MG). Tem como autora principal a aluna de doutorado Anabele Lindner.

O Trabalho 4 aplicou o procedimento sequencial, proposto anteriormente, para estimativa espacial da escolha modal. No entanto, na etapa não espacial, foi calibrado um modelo logit binomial para estimar probabilidades de escolhas domiciliares do modo de transporte motorizado.

- Objetivos

O objetivo principal do Trabalho 4 foi aplicar o método proposto anteriormente para realizar interpolação espacial de probabilidades de escolha do modo de transporte motorizado (individual ou público). A ideia foi estimar dados em diversas coordenadas geográficas de valores não pesquisados. Através da superfície de divisão modal, foi possível se obter “manchas” de preferência modal, associadas a valores estimados e erros relativos a estimativas.

- Região

A região de estudo referente ao Trabalho 4 foi o município de São Paulo e os dados provenientes da Pesquisa Origem-Destino, realizada em 2007. Foram utilizados dados desagregados relativos aos domicílios, variáveis socioeconômicas e de viagens.

As variáveis relativas ao modo de transporte público e individual motorizado por domicílio, obtidas pela Pesquisa O/D, foram transformadas em uma variável binária. A variável resultante desta discretização, utilizada neste estudo, foi a preferência domiciliar por modo motorizado, em que o valor “0” representa o uso predominante de modo individual motorizado e o valor “1” o uso de transporte público por domicílio. Consequentemente, os domicílios em que não havia viagens por transporte motorizado foram eliminados. Ao final, foram analisados 14.443 domicílios, para a cidade de São Paulo.

- Materiais e Método

O Trabalho 4 seguiu os passos propostos pela Figura 5.6

Figura 5.6: Representação do método proposto. Fonte: Lindner e Pitombo (2015)

Nesta aplicação metodológica, foram separados dois conjuntos de amostras. A amostra de 70% dos dados foi reservada para a calibração de dois procedimentos: Modelo logit e modelagem geoestatística. A amostra de 30% foi delimitada com o propósito de prover a validação dos procedimentos. Por conseguinte, pode-se resumir este trabalho em duas etapas: (1) abordagem clássica: calibração do logit e (2) abordagem espacial: aplicação de Geoestatística.

A validação se deu pela análise de medidas de desempenho considerando valores observados (probabilidades obtidas na abordagem clássica) e valores estimados pela krigagem em coordenadas geográficas de valores conhecidos. As medidas adotadas foram aquelas representadas na Tabela 5.2.

Tabela 5.2: Medidas estatísticas avaliadas nas modelagens. Fonte: Lindner e Pitombo (2015)

Média do erro absoluto	Erro quadrático médio	Desvio do erro quadrático médio
$\frac{1}{N} \sum_{i=1}^N x_i - y_i $ (5.4)	$\frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2$ (5.5)	$\sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2}$ (5.6)
Desvio padrão	Erro padrão	Correlação de Pearson
$\sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N - 1}}$ (5.7)	$\frac{\text{Desvio padrão}}{\sqrt{N}}$ (5.8)	$\frac{\sum_{i=1}^N (x_i - \bar{x}) \times (y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \times \sum_{i=1}^N (y_i - \bar{y})^2}}$ (5.9)

Em que N: número de valores da amostra, x: valor calculado, y: valor observado, $\bar{x}$: média dos valores calculados e $\bar{y}$: média dos valores observados.

• Resultados

Resultados da etapa não espacial

A etapa não espacial do Trabalho 4 consistiu na calibração de um modelo logit para estimativa de probabilidade do modo de transporte motorizado. A equação de calibração da variável de probabilidade de escolha do transporte público em relação ao transporte individual motorizado é apresentada pela Equação 5.10.

$$\pi = \frac{1}{1 + e^{-g(x)}}$$

$$g(x) = -0,547 + 0,271\beta_1 + 0,318\beta_2 - 0,878\beta_3 - 1,180\beta_4 - 0,172\beta_5 + 0,000193\beta_6 - 0,000177\beta_7 + 0,00003\beta_8 \quad 5.10$$

1-número de pessoas, 2-critério de renda, 3-quantidade de motocicletas, 4-quantidade de automóveis, 5-total de viagens, 6-distância de viagens transporte público, 7-distância de viagens por transporte individual motorizado, 8-distância de viagens por transporte não motorizado.

Houve um ótimo ajuste pelo modelo logístico. Os valores de erro padrão e significância foram extremamente baixos. Além disso, foi observado valor alto das medidas *Cox & Snell* e *Nagelkerke*. Adicionalmente, a taxa de acertos foi superior a 95%. A Equação 5.10 foi, então, utilizada ao conjunto de 30% para ser realizada a validação do método. Esta validação resultou em uma taxa de acertos alta, sendo esta, também, igual a 96%. Um excelente ajuste na etapa não espacial do procedimento sequencial é essencial para mitigar os problemas da propagação de erros. A Equação 5.10, aplicada ao conjunto de 70%, gerou resultados de probabilidade a serem utilizados na etapa de Geoestatística.

Resultados da etapa espacial

A etapa de modelagem geoestatística utilizou, como valores de Variável Regionalizada, os valores estimados do conjunto de 70% pela Equação 5.10. A partir dos variogramas experimentais calculados, pode-se concluir que a estrutura espacial da variável é anisotrópica. A direção principal se deu para o ângulo de 60° e a consequente direção ortogonal foi de -30° .

A curva teórica exponencial apresentou melhor ajuste para os pontos do variograma experimental das direções. O patamar foi equivalente à média de variância e o alcance foi ajustado para os dados. Os variogramas ajustados da direção principal e da direção ortogonal estão demonstrados na Figura 5.7.

Figura 5.7: Variogramas ajustados para a direção principal e ortogonal. Fonte: Linder e Pitombo (2015)

Os valores de direção principal e ortogonal, bem como o efeito pepita (Co), contribuição (C1) e amplitude (a) foram utilizados para a definição dos ponderadores de krigagem. O resultado da estimação é apresentado pelo mapa da Figura 5.8.

Figura 5.8 Mapa krigado para a probabilidade de preferência por transporte público em relação ao transporte individual motorizado. Fonte: Lindner e Pitombo (2015)

O mapa apresentado representa a superfície de probabilidades de escolha do modo de transporte motorizado. Interessante observar uma tendência central ao uso do automóvel e motocicleta e uma tendência periférica ao uso de transporte público. Pode-se pressupor escolhas por transporte coletivo no entorno do município, onde concentram-se as maiores distâncias de viagens.

Verifica-se ainda que, através da superfície obtida, é possível estimar probabilidades de escolha modal em diversos pontos, bem como associar erros de estimativa, sendo uma forma espacial e confirmatória de observar tendências ou “manchas” urbanas de escolha modal.

Os resultados da validação da modelagem geoestatística demonstram que os erros foram baixos, o que é atestado pelo valor aceitável de desvio do erro quadrático médio. O coeficiente de correlação pode ser considerado razoável (0,5), demonstrando correlação moderada entre valores observados e estimados pela interpolação espacial. Pelos resultados de validação, pode-se afirmar que a técnica é adequada com um poder preditivo razoável.

- Análise crítica

A partir da aplicação do método sequencial, proposto anteriormente, foi possível a obtenção de valores de escolha modal interpolados espacialmente, a partir de uma outra base de dados desagregada.

A diferença principal entre o Trabalho 4 e os anteriores (Trabalho 1, 2 e 3) se trata da separação aleatória das amostras de calibração e validação para as duas grandes etapas metodológicas. As limitações metodológicas são similares aos trabalhos descritos anteriormente com exceção da propagação de erros, já que o ajuste do modelo logit binomial na etapa não espacial foi preciso.

5.5. ESCOLHA MODAL E GEOESTATÍSTICA: ETAPA ÚNICA UNIVARIADA

5.5.1 Krigagem Indicativa para escolha modal

- ✓ Trabalho 5: *ANÁLISE DESAGREGADA DE DADOS DE DEMANDA POR TRANSPORTES ATRAVÉS DE MODELAGEM GEOESTATÍSTICA E TRADICIONAL*
- ✓ Trabalho 6: *KRIGING GEOSTATISTICAL METHODS FOR TRAVEL MODE CHOICE: A SPATIAL DATA ANALYSIS TO TRAVEL DEMAND FORECASTING*

Ambos os trabalhos adotaram como procedimento metodológico comum: Etapa espacial única e univariada. Os trabalhos utilizaram a Krigagem Indicativa para obtenção da superfície de probabilidades de escolha modal.

O Trabalho 5 é uma dissertação de mestrado, defendida pela discente Anabele Lindner, no ano de 2015, pelo Programa de Pós-Graduação em Engenharia de Transportes (EESC-USP). Já o Trabalho 6 foi publicado recentemente na edição especial (*Spatial Statistics Research*) do periódico internacional *Open Journal of Statistics* (<http://www.scirp.org/journal/ojs/>). Este artigo tem autoria principal da aluna de doutorado Viviane Antunes Gomes e conta com a colaboração,

além da docente, da aluna Samille Santos Rocha e da pesquisadora Ana Rita Salgueiro (Universidade Federal do Ceará).

O primeiro trabalho (Trabalho 5) calibra modelo logit binomial a partir de variáveis domiciliares socioeconômicas e, finalmente, compara resultados da abordagem não espacial tradicional e da abordagem espacial univariada (Krigagem Indicativa). Já o último trabalho (Trabalho 6) compara os resultados dos mapas obtidos com a Krigagem Indicativa aos mapas obtidos anteriormente pelo procedimento sequencial multivariado, proposto no Trabalho 3.

- Objetivos

O objetivo de ambos os trabalhos foi obter uma superfície contínua de escolha modal a partir de procedimento espacial univariado (Krigagem Indicativa).

Como objetivo secundário relativo ao Trabalho 5 pode-se citar a comparação das estimativas à abordagem tradicional para investigar contribuição da modelagem geoestatística para previsão de escolhas modais.

Já o Trabalho 6 possuía, como objetivo secundário, a comparação aos resultados obtidos previamente, a partir de procedimento sequencial espacial e multivariado.

- Região

O Trabalho 5 utilizou dados domiciliares desagregados provenientes da Pesquisa Origem-Destino da Região Metropolitana de São Paulo de 2007. Já o Trabalho 6 utilizou dados desagregados relativos a Pesquisa Origem-Destino, realizada em 2007/2008 na cidade de São Carlos.

A variável de interesse do Trabalho 5 é escolha domiciliar predominante pelo modo de transporte motorizado (público ou particular motorizado), enquanto que a variável objeto de estudo do Trabalho 6 é Uso do transporte particular motorizado (uso ou não uso do automóvel ou motocicleta).

- Materiais e Método

O método proposto no Trabalho 5 consistiu na calibração do modelo logit binomial (amostra de calibração) e verificação de medidas de desempenho a partir de amostra de validação. Em seguida, com a mesma amostra de calibração do modelo logit, foram modelados variogramas experimentais. Finalmente, foram estimados valores de probabilidades de escolha modal na amostra de validação a partir dos parâmetros dos variogramas teóricos determinados previamente. Em seguida, foi obtida superfície contínua de probabilidade de escolha do modo de transporte motorizado a partir da Krigagem Indicativa.

Já o método proposto pelo Trabalho 6 consistiu na comparação dos resultados obtidos pela Krigagem Indicativa àqueles resultados obtidos previamente pelo método sequencial composto de Árvore de Decisão e Krigagem Ordinária, desenvolvidos em trabalhos anteriores para o mesmo banco de dados (Trabalho 3).

- Resultados

Para ambos os trabalhos, foram modelados variogramas experimentais. Em ambos os casos a variável objeto de estudo (dicotômica) apresentava comportamento anisotrópico. Determinados parâmetros de variogramas teóricos, foram calculados os ponderadores da Krigagem Indicativa e estimados valores na superfície de escolha modal. A Figura 5.9 ilustra a superfície krigada para a cidade de São Carlos (Trabalho 6), enquanto a Figura 5.10 ilustra o mapa obtido para a Região Metropolitana de São Paulo (Trabalho 5). Através da validação cruzada e das métricas de erros considerando valores estimados e observados, o procedimento foi considerado viável para ambos os trabalhos, com taxas de acertos em torno de 60% nas coordenadas geográficas cujos valores eram previamente conhecidos.

Figura 5.9 Mapa krigado para a probabilidade de preferência por automóvel – região Santa Felícia – São Carlos. Fonte: Gomes *et al.* (2016a)

Figura 5.10. Mapa krigado para a variável *preferência por modo motorizado*. RMSP. Fonte: Lindner (2015)

O modelo logit binomial calibrado para fins comparativos no Trabalho 5 é descrito conforme Equação 5.11.

$$\pi = \frac{1}{1 + e^{-g(x)}}$$

$$g(x) = -0,547 + 0,271\beta_1 + 0,318\beta_2 - 0,878\beta_3 - 1,180\beta_4 - 0,172\beta_5 + 0,000193\beta_6 - 0,000177\beta_7 + 0,00003\beta_8 \quad (5.11)$$

Onde:

1 - número de pessoas, 2 - critério de renda, 3 - quantidade de motocicletas, 4 - quantidade de automóveis, 5 - total de viagens, 6 - distância de viagens transporte público, 7 - distância de viagens por transporte individual motorizado, 8 - distância de viagens por transporte não motorizado.

O modelo calibrado teve um ajuste quase perfeito, com taxa de acertos equivalente a aproximadamente 95%, muito superior ao método espacial univariado (KI).

- Análise crítica

Os trabalhos que utilizaram o procedimento espacial de etapa única univariada (Krigagem Indicativa) comprovaram que o presente método é adequado para obtenção de superfície contínua de probabilidades de escolha modal, especialmente para análises espaciais no presente e na ausência de dados (como covariáveis).

Tal método não seria adequado para o caso de projeções futuras, com base em variáveis explicativas ou, até mesmo, investigação do fenômeno, já que a escolha modal depende de diversos fatores, inclusive comportamentais, além da localização das escolhas.

O ajuste do modelo logit binomial, a partir das variáveis explicativas no Trabalho 5, mostra a superioridade da abordagem tradicional, quando comparada à técnica de interpolação espacial univariada. No entanto, considera-se adequada a taxa de acertos de 60%, levando-se em conta apenas os valores da Variável Regionalizada, associados às suas localizações geográficas. Na ausência de outros dados relativos aos domicílios e indivíduos, tal método é promissor. A Tabela 5.2 faz uma comparação entre o método sequencial proposto nos Trabalhos 1, 2, 3 e 4 e o método de KI, utilizado nos Trabalhos 5 e 6.

Tabela 5.3 Prós e contras da aplicação dos dois procedimentos utilizados nos seis trabalhos

Abordagem	Vantagens	Desvantagens
Sequencial multivariado	<ul style="list-style-type: none"> - Considera variáveis socioeconômicas e dependência espacial - Adequado a futuras projeções 	<ul style="list-style-type: none"> - VR não são originais e provenientes de etapa não espacial - Pode ocorrer propagação de erros ao longo do procedimento
Krigagem Indicativa	<ul style="list-style-type: none"> - Requer pouca informação - A VR é original e não proveniente de um modelo prévio 	<ul style="list-style-type: none"> - Útil para análises no presente - Apenas a dependência especial é considerada

Além disso, para o caso de Variáveis Regionalizadas qualitativas com mais de duas categorias, como no caso de escolha modal, recomenda-se a utilização de equações multiquadráticas (YAMAMOTO *et al.*, 2012). No caso da KI, há necessidade de um variograma por categoria da variável.

5.6. CONSIDERAÇÕES FINAIS

Este capítulo descreveu os seis trabalhos relativos ao tema Escolha modal e Geoestatística. Mostrou diferentes aplicações considerando base de dados desagregados por domicílios e indivíduos. Além disso, os seis trabalhos apresentaram dois diferentes procedimentos metodológicos espaciais – Sequencial multivariado e Etapa única univariada. O Quadro 5.3 sintetiza os trabalhos e aponta os pontos positivos e negativos de cada aplicação aqui descrita.

Quadro 5.3 Síntese dos seis trabalhos discutidos (escolha modal e geoestatística)

Trabalho	Objetivos e Método	Pontos Positivos	Pontos Negativos
2	Propor método sequencial e multivariado para estimativa de escolha modal em uma superfície contínua	2.1 Bons resultados de métricas de desempenho	2.1 Pode haver maior propagação de erros
		2.2 Permite uma análise do comportamento espacial da escolha em forma de superfície contínua	2.2 Ajustes de variogramas (transporte público)
		2.3 Método adequado também para projeções futuras	2.3 Suposição de normalidade
		2.4 Validação da etapa não espacial	
3	Propor método sequencial e multivariado para estimativa de escolha modal em uma superfície contínua	3.1 Bons resultados de métricas de desempenho	3.1 Pode haver maior propagação de erros
		3.2 Permite uma análise do comportamento espacial da escolha em forma de superfície contínua	3.2 Ajustes de variogramas (transporte público)
		3.3 Método adequado também para projeções futuras	3.3 Suposição de normalidade
4	Aplicar o método anterior para realizar interpolação espacial de probabilidades de escolha do modo de transporte motorizado (individual ou público).	4.1 Separação de amostras para validação e calibração nas duas etapas do método	Suposição de normalidade
		4.2 Método proposto anteriormente funciona adequadamente para outras bases de dados	
		Menor propagação de erros devido ao bom ajuste do modelo logit	
5	5.1 Obter uma superfície contínua de escolha modal a partir de procedimento espacial univariado (Krigagem Indicativa).	5.1 VR original e não proveniente de modelo prévio	5.1 Krigagem Indicativa funciona melhor para variáveis numéricas codificadas a partir de um valor de corte
	5.2 Comparar as estimativas da KI com uma abordagem tradicional não espacial	5.2 Pouca informação necessária	5.2 Equações multiquadráticas mais indicadas para variáveis qualitativas
			5.3 Procedimento útil apenas para análises no presente
			5.4 Procedimento menos eficaz do que a modelagem tradicional multivariada
6	6.1 Obter uma superfície contínua de escolha modal a partir de procedimento espacial univariado (Krigagem Indicativa).	6.1 VR original e não proveniente de modelo prévio	6.1 Krigagem Indicativa funciona melhor para variáveis numéricas codificadas a partir de um valor de corte
	6.2 Comparar resultados da KI com o procedimento sequencial proposto previamente	6.2 Pouca informação necessária	6.2 Equações multiquadráticas mais indicadas para Variáveis qualitativas
			6.3 Procedimento útil apenas para análises no presente

Capítulo

6 CONCLUSÕES: RESPOSTAS, AVANÇOS E FUTURAS PESQUISAS

Quais as etapas restantes para a montagem completa do “cubo mágico”? Quais seriam os próximos passos para obtenção de estimativas cada vez mais precisas de demanda por transportes, e consequente auxílio a planejadores para tomada de decisão relacionada à oferta de transportes?

Este capítulo final sistematiza os avanços metodológicos realizados ao longo dos anos, além das próximas etapas de pesquisas que serão desenvolvidas nos próximos cinco anos, através de projeto de pós-doutoramento e orientações de Iniciação Científica, Mestrado e Doutorado. Além disso, se propõe a responder às indagações de pesquisas, principais motivações para continuidade do desenvolvimento da linha, que foram apresentados no capítulo inicial deste texto.

6.1. SINTETIZANDO AVANÇOS METODOLÓGICOS

Os avanços metodológicos, ocorridos desde 2008, que foram sumariados no primeiro capítulo deste documento e apresentados ao longo do texto (Capítulo 4 e Capítulo 5), são sistematizados a partir da montagem completa do “cubo mágico” (Figura 6.1).

As etapas de montagem correspondem aos avanços ocorridos ao longo do desenvolvimento da pesquisa (*Passo 1; Passo 2; Passo 3 e Passo 4*). Pode-se situar a linha de pesquisa, atualmente, no *Passo 5* da montagem. O *Passo 5* representa etapas futuras a serem realizadas nos próximos anos.

<p>Ideia inicial: Incorporar geoestatística a problemas relativos à demanda por transportes (Capítulo 1).</p>	<p>Passo 1: Construção da base teórica (Capítulo 2)</p>	<p>Passo 2: Preparação de materiais e elaboração metodológica (Capítulo 3)</p>

<p>Passo 3: Estimções de dados numéricos agregados em áreas – Geração de viagens (Capítulo 4)</p>	<p>Passo 4: Estimções de dados qualitativos desagregados – Escolha modal (Capítulo 5)</p>	<p>Passo 5: Trabalhos futuros</p>

<p>Resultado final: Estimativas mais precisas de demanda por transporte; auxílio a planejadores para tomada de decisão relacionada à oferta de transportes.</p>		

Figura 6.1 Montagem da linha de pesquisa - Avanços metodológicos e etapas futuras

O Passo 1 corresponde à construção da base teórica e entendimento da principal ferramenta de análise espacial utilizada. Já o Passo 2 é relativo a propostas metodológicas com o uso da Geoestatística. Nesta etapa um método geral comum foi elaborado para desenvolvimento dos trabalhos.

O Passo 3 corresponde aos trabalhos relacionados a dados de geração de viagens. Os primeiros trabalhos publicados na linha de pesquisa (Pitombo e Sousa, 2009) são relativos à estimação por Krigagem Ordinária (KO) de taxas de produção e atração de viagens, em centroides associados a unidades de áreas. As principais limitações das primeiras pesquisas podem ser sumariadas como: abordagem univariada, suposições de distribuição normal da Variável Regionalizada e associação de centroides a áreas desiguais.

Com o objetivo de resolver a limitação da abordagem univariada, já que estimação de viagens é um problema conhecidamente multivariado e não apenas espacial, foram desenvolvidos trabalhos imediatamente posteriores com o foco na Geoestatística Multivariada (Pitombo *et al.*, 2010). No entanto, essa etapa apresenta ainda importantes restrições metodológicas relacionadas à agregação dos dados e suposição de distribuição normal da VR.

Em 2014, Rocha *et al.* (2014) propuseram um método simplificado para desagregação de dados, aplicado ao caso onde não há pesquisas domiciliares, para estimação de geração de viagens desagregada. Desta forma, a limitação relacionada à agregação dos dados é parcialmente resolvida. No entanto, ainda há o problema de suposição de distribuições populacionais normais e ferramentas mais adequadas para desagregação de dados e simulação.

Considerando as restrições ditas acima, há duas pesquisas de doutorado, a serem realizadas futuramente (Passo 5). A primeira partirá da proposta de empregar técnicas de simulação de dados espacialmente correlacionados para desagregações de dados em áreas. A

segunda irá empregar técnicas de Krigagem voltadas para dados de contagem (Krigagem com distribuição de Poisson, por exemplo).

Já o *Passo 4* da montagem do “cubo mágico”, relaciona-se a dados discretos de demanda por transportes. No âmbito da escolha modal, Lindner (2015) propôs um método de estimação espacial univariada para variáveis binárias. Posteriormente, Pitombo *et al.* (2015a), Pitombo *et al.* (2015b) e Lindner e Pitombo (2015), propuseram um método com etapa multivariada não espacial e etapa univariada espacial. A ideia era associar ferramentas tradicionais à geoestatística. Partindo-se ainda da mencionada proposta metodológica (Método conjunto para uso de ferramentas tradicionais e geoestatística a dados desagregados discretos), há algumas pesquisas em andamento, que serão descritas na subseção subsequente. A Figura 6.2 resume os trabalhos já realizados, limitações e avanços, além das pesquisas futuras (caixas em tons de cinza).

Figura 6.2 Limitações e avanços da linha de pesquisa

6.2. FUTURAS PESQUISAS

O tema apresentado neste documento ainda tem muitas pesquisas em desenvolvimento, além de ideias para continuidade de aprimoramento do método até a montagem completa do “cubo mágico”.

Ficou constatado, ao longo de oito anos, o potencial da ferramenta, além de adaptações necessárias e amadurecimento para que ocorram, de fato, incrementos significativos das estimativas de demanda por transportes, através do uso da geoestatística.

Esta seção descreve sucintamente as pesquisas ilustradas na Figura 6.3. Trata-se de três Doutorados, um Mestrado, um trabalho de Iniciação Científica, além do Pós-Doutorado, previsto para ser realizado pela docente durante o período de Dezembro de 2016 a Março de 2017.

Figura 6.3 Futuras pesquisas

6.2.1 Pós-doutorado: Proposta de modelos autoregressivos logísticos para problemas de escolha modal: uma abordagem através de conceitos geoestatísticos

Este projeto, com duração prevista de três meses, a ser realizado pela docente no *Transport and Mobility Laboratory, School of Architecture, Civil and Environmental Engineering,*

EPFL (Lausanne, Suíça), visa propor um modelo logit autoregressivo para problemas de escolha modal, considerando dados espacialmente correlacionados. Inicialmente, será realizada a análise da estrutura espacial da variável dependente, através da obtenção de variogramas experimentais e ajuste de variogramas teóricos. Em seguida, serão calculadas covariáveis espaciais por domicílio, considerando as distâncias entre observações, valores das variáveis e parâmetros dos variogramas teóricos. Posteriormente, tais covariáveis serão incorporadas às funções utilidade dos modos de transportes. Finalmente, a eficácia dos modelos propostos será mensurada considerando medidas de erros e testes não paramétricos. Para fins de validação metodológica, será realizada a comparação com modelos logit tradicionais. Os dados utilizados serão da Pesquisa Origem-Destino realizada em 2007/2008 na cidade de São Carlos (SP). Espera-se trazer benefícios ao meio acadêmico, constituindo referências para futuras aplicações, sobretudo em termos de projeções futuras da escolha do modo de transporte.

Autocorrelação espacial muitas vezes não é modelada satisfatoriamente a partir das variáveis explicativas (ou covariáveis) usuais. Assim, Augustin *et al.* (1996) propuseram um modelo logístico autoregressivo, com intuito de representar a autocorrelação espacial para previsão de distribuição de cervos em uma região da Escócia.

Neste caso, a probabilidade de presença ou não de cervos, em determinadas áreas, era calculada baseada em um modelo logístico que incorporava uma “autocovariável”. Esta “autocovariável” é derivada das respostas das suas áreas vizinhas. Este modelo é conhecido como autoregressivo logístico espacial.

Desta forma, o modelo logístico tradicional seguia a formulação proposta da Equação 6.1, enquanto que o modelo autoregressivo logístico segue as formulações explicitadas na Equação 6.2 e Equação 6.3, respectivamente.

$$\log\left(\frac{p_i}{1-p_i}\right) = \alpha + \beta_1 \text{altitude}_i^2 + \beta_2 \text{northing}_i + \beta_3 \text{mires}_i + \beta_4 \text{easting}_i + \beta_5 \text{pine}_i. \quad (6.1)$$

$$\log\left(\frac{p_i}{1-p_i}\right) = \alpha + \beta_1 \text{altitude}_i^2 + \beta_2 \text{northing}_i + \beta_3 \text{mires}_i + \beta_4 \text{easting}_i + \beta_5 \text{pine}_i + \beta_6 \text{autocov}_i \quad (6.2)$$

Altitude: altitude; Northing: coordenada geográfica Norte-Sul; Easting: coordenada geográfica Leste-oeste; pine e mire: área coberta por lama e pinhos; pi :probabilidade de presença de cervo.

$$\text{autocov}_i = \frac{\sum_{j=1}^{k_i} w_{ij} y_j}{\sum_{j=1}^{k_i} w_{ij}} \quad (6.3)$$

A “autocovariável” é ponderada pela média de presenças de cervos em um conjunto de k_i vizinhos de uma observação i . O peso dado a uma observação j é $w_{ij} = 1/h_{ij}$, onde h_{ij} é a distância euclidiana entre as observações i e j .

Outros autores propuseram modelos similares, com o uso voltado a variáveis relativas a escolhas, por exemplo, como no trabalho de Adjemian *et al.* (2010), onde foi proposto um modelo baseado em função utilidade (V_{ij}), com a inclusão de covariável espacial para escolha do tipo/marca de automóvel ou veículo automotor.

$$V_{ij} = \beta' x_i + \rho' Wf(V_{ij}) \quad (6.4)$$

W é a matriz de ponderação espacial que define a vizinhança de cada domicílio i .

O projeto de pós-doutorado baseia-se no modelo proposto por Augustin *et al.* (1996) e propõe o cálculo da “autocovariável” espacial, através de parâmetros dos variogramas teóricos (matriz de variâncias).

6.2.2 Iniciação Científica 1: Incorporação de localização espacial a modelos de escolha discreta

O projeto de Iniciação Científica (IC), atualmente em desenvolvimento pela aluna de graduação do curso de Engenharia Civil, tem como objetivo principal incorporar fatores locais (como coordenadas geográficas, por exemplo) a modelos de escolha discreta. O intuito é investigar a influência de tais variáveis na estimação de probabilidades de escolha modal. Além disso, também é investigada a significância de variáveis socioeconômicas e de viagens na escolha do modo de transporte.

O método desenvolvido compreende as etapas de (1) Revisão da literatura; (2) Proposta de cidade fictícia; (3) Calibração de modelos com variáveis socioeconômicas e de viagens; (4) Validação dos modelos calibrados em (3); (5) Calibração de modelos com a inclusão de variáveis locais; (6) Validação de modelos calibrados em (5); (7) Comparação de modelos calibrados em (3) e (5). Após as sete etapas, todo o procedimento é repetido para uma nova região fictícia, obtida através de técnica de interpolação bivariada, para simulação de dados espacialmente dependentes.

O banco de dados artificial inicial possui 60 indivíduos, com valores da posição na direção x, posição na direção y, renda, distância para o polo atrativo, custo da viagem para o caso do automóvel e para o caso do ônibus, tempo de viagem para o caso do automóvel e para o caso do ônibus e, por fim, a escolha do modo de transporte. A Figura 6.4 ilustra a cidade fictícia inicial, antes da simulação por interpolação bivariada e escolhas modais. A Figura 6.5 ilustra a cidade após a simulação de dados.

Figura 6.4 Mapa temático da cidade fictícia relativo à escolha do modo de transporte (0 – ônibus, 1- carro).

Figura 6.5 Representação da cidade após a simulação proposta.

6.2.3 Mestrado 1: Introdução de conceitos de variografia a modelos de escolha discreta

O projeto de Mestrado, em desenvolvimento desde março de 2016, visa incorporar “autocovariáveis” espaciais a modelos de escolha discreta. O intuito é investigar a melhoria das estimativas considerando modelos omnidirecionais, direção principal e ortogonal da variável dependente (no caso, escolha modal) a partir de variogramas teóricos ajustados. Numa etapa seguinte, serão repetidas as análises para variáveis explicativas.

O banco de dados será proveniente do trabalho de IC, descrito anteriormente, obtido através da simulação por interpolação bivariada (Figura 6.5). A base final corresponde a uma cidade fictícia, com 580 observações finais, com dados espacialmente correlacionados (variáveis independentes e variável dependente).

6.2.4 Doutorado 3: Incorporação da geoestatística a previsões de escolha modal

O Doutorado 3, em desenvolvimento desde 2014, está estritamente relacionado aos três trabalhos descritos anteriormente. A ideia principal é incorporar “autocovariáveis” que consideram variogramas teóricos das variável dependente e também das variáveis explicativas.

Neste caso, o banco de dados é relativo à Pesquisa de Mobilidade Urbana de 2012 da Região Metropolitana de São Paulo. Na Pesquisa de Mobilidade de 2012 foi selecionada uma amostra com aproximadamente 8100 domicílios (Figura 6.6b), e a RMSP teve um zoneamento mais agregado, 31 zonas de tráfego (Figura 6.6a).

Figura 6.6 Banco de dados a ser utilizada no Doutorado 3

6.2.5 Doutorado 1: Proposta de desagregação de dados para aprimoramento da modelagem geoestatística aplicada à demanda por transportes

Quando se faz um planejamento de transportes em uma cidade, é praticamente indispensável realizar pesquisas domiciliares. Contudo, países em desenvolvimento, como o Brasil,

não costumam ter tais pesquisas disponíveis devido ao alto custo. Consequentemente, a eficiência do planejamento de sistemas de transportes é prejudicada. Por outro lado, outros dados têm maior disponibilidade e maior frequência. Dados censitários, por exemplo, estão disponíveis a cada 10 anos.

A diferença entre estas informações censitárias e as pesquisas domiciliares é, dentre outros fatores, referente à agregação espacial. Dados domiciliares têm coordenadas geográficas relacionadas pontualmente, enquanto dados censitários são relacionados a áreas administrativas. Da mesma forma, os dados censitários não englobam pesquisas relacionadas aos indivíduos e suas preferências de transporte.

O Doutorado 1, em desenvolvimento desde 2015, visa a transformação de dados censitários em dados convenientes para o planejamento de transportes através de simulação, levando-se em conta dados espacialmente correlacionados. O trabalho utiliza dados discretos de transporte e a geoestatística, como ferramenta de análise estatística espacial.

Os dados utilizados neste trabalho são referentes à Pesquisa Origem-Destino (O/D), realizada na Região Metropolitana de São Paulo (RMSP) em 2007. Para a obtenção da amostragem sistemática elaborada, pretende-se promover a desagregação dos dados originais, de forma que estes possam garantir boa representação da realidade, sobretudo quando comparados aos dados da amostra desagregada real. Cada Zona de Tráfego será representada por um valor, por exemplo, de viagem urbana (razão entre variável de demanda e número de domicílios) associado ao seu respectivo centroide (Figura 6.7 (a)).

O processo de desagregação é detalhado como:

- Geração de uma malha quadricular (Figura 6.7 (b));
- Geração de centroides (novas coordenadas geográficas desagregadas) de cada quadricula correspondente à malha quadricular;
- Determinação de valores de viagens para cada quadricula a partir de técnicas de simulação de dados (Monte Carlo, *bootstrapping*).

Figura 6.7: Exemplo de representação do processo da amostragem sistemática: (a) amostra por Zona de Tráfego; (b) geração da malha e técnicas de simulação. Fonte: Lindner e Pitombo (2014)

6.2.6 Doutorado 2: Uso conjunto de técnicas tradicionais e de geostatística para demanda por transportes: viagens baseadas em atividades

Este projeto de Doutorado 2, em desenvolvimento desde 2014, pretende utilizar sequencialmente técnicas tradicionais e de análise espacial de dados para previsão de demanda por viagens baseadas em atividades.

São utilizados dados desagregados domiciliares relativos a viagens da Região Metropolitana de São Paulo (Pesquisa Origem/Destino 2012). Serão obtidos padrões de viagens encadeadas considerando a sequência de atividades e também modos de transportes utilizados pelos indivíduos.

Na etapa não espacial serão utilizados modelos de escolha discreta para prever a escolha do padrão de viagem, baseado em características socioeconômicas e domiciliares, gerando probabilidades de escolhas nos padrões de viagens e atividades, tal como realizado por Pitombo (2007).

A etapa espacial consiste no uso de Krigagem com distribuição de Poisson. Os resultados em domicílios observados serão comparados a resultados obtidos através do método sequencial (logit – Krigagem Poisson).

Neste trabalho será essencialmente analisado o incremento na estimativa e entendimento do fenômeno a partir do uso de técnicas de análise espacial de dados. A principal contribuição desta tese de doutorado será o aprimoramento de modelagem de viagens baseadas em atividades, considerando fatores locais, além das variáveis usualmente utilizadas na abordagem tradicional.

6.3. RESPONDENDO ÀS INDAGAÇÕES

Levando-se em conta as indagações motivadoras, apresentadas no capítulo introdutório, pode-se afirmar que algumas questões foram completamente respondidas, enquanto outras parcialmente. Para finalização deste documento, a autora faz uma reflexão considerando os passos já desenvolvidos e a desenvolver.

- **Que incremento o uso da Geoestatística pode trazer na previsão de demanda por transportes?**
- **Já que demanda por transportes é um problema pontual distribuído ao longo de uma área urbana, qual a utilidade de superfícies contínuas de variáveis de demanda?**

O principal diferencial do uso da geoestatística se trata do produto da superfície contínua de estimativas de variáveis de demanda por transportes, dado que não é possível ser obtido através de ferramentas tradicionalmente utilizadas.

No entanto, a superfície contínua de estimativas ainda não seria o produto ideal para utilização como ferramenta de apoio à tomada de decisão, já que a menor unidade de informação para transporte é o domicílio. As localizações domiciliares são discretas no espaço.

Dessa forma, o uso dos mapas krigados, associados ao geoprocessamento, por exemplo, pode configurar a obtenção de dados de demanda por transportes exatamente nos domicílios existentes, com pouca informação previamente necessária.

- **Os conceitos e técnicas de Geoestatística podem ser aplicados para qual tipo de dado de transportes? Qual informação é obrigatória para este fim? Quais as principais restrições à sua aplicação na área de modelagem de demanda por transportes?**

Para uso da geoestatística, é necessária a associação entre valores de variáveis de interesse e coordenadas geográficas. Além disso, a variável objeto de estudo deve ter uma estrutura espacial para que seja possível o bom ajuste dos variogramas teóricos – principal ferramenta de análise. A ferramenta pode ser utilizada tanto para dados discretos como contínuos. No entanto, há algumas restrições importantes para aplicação na modelagem de demanda por transportes, tais como: (1) fenômeno geralmente espacialmente discreto; (2) necessidade de adequação de distribuições de probabilidade da variável e técnica de krigagem utilizada; (3) necessidade de consideração da estrutura espacial das variáveis explicativas.

- **Qual a principal contribuição acadêmica? E técnica? Um procedimento conjunto ou sequencial de métodos tradicionais e geoestatística pode ser um caminho razoável para obtenção de projeções futuras de mapas krigados?**

Até o presente momento, a principal contribuição acadêmica, relativa a estimativas de variáveis de demanda por transportes, através da geoestatística, está relacionada ao uso sequencial de técnicas tradicionais com ferramentas de análise espacial. Esse método de duas etapas (não espacial e espacial) é mais eficaz quando comparado a métodos apenas não espaciais ou a uso apenas de krigagem (caso univariado).

REFERÊNCIAS

ADJEMIAN, M. K.; LIN, C. Y. C.; WILLIAMS, J. Estimating spatial interdependence in automobile type choice with survey data. **Transportation Research Part A**, [S.l.], v.44, p. 661–675, 2010. Doi:10.1016/j.tra.2010.06.001.

AMAVI, A. *et al.* Advanced trip generation/attraction models. **Procedia - Social and Behavioral Sciences**, [S.l.], v.160, p.430–439 2014. Doi:10.1016/j.sbspro.2014.12.155.

ANDRADE, G. R.; PITOMBO, C. S. Efeito da urbanização na velocidade de fluxo livre de rodovias de pista dupla paulistas: uma abordagem através de interpolação espacial. In: ANPET - CONGRESSO DE PESQUISA E ENSINO EM TRANSPORTES, 29., 2015, Ouro Preto. **Anais...**, Ouro Preto, 2015.

AUGUSTIN, N. H.; MUGGLESTONE, M. A.; BUCKLAND, S. T. An autologistic model for the spatial distribution of wildlife. **Journal of Applied Ecology**, Londres, v.33, p.339–347. 1996. Doi: 10.2307/2404755.

BEN-AKIVA, M. E. **Structure of Passenger Travel Demand Models**. 1973. (Ph.D. Dissertation) - Massachusetts Institute of Technology, Cambridge. 1973.

_____.; SCOTT, M. R.; BEKHOR, S. Route choice models. **Human Behaviour and Traffic Networks**. Cambridge, p.23-45, 2004. Doi:10.1016/S1755-5345(13)70005-8.

_____. Planning and Action in a Model of Choice. In.: HESS, S.; DAL, A. Choice Modelling: the State-of-the-Art and the State-of-Practice, Proceedings from the Inaugural International Choice Modelling Conference. **Emerald**, [S.l.], p.19-34. 2010.

_____.; LERMAN, S. **Discrete choice analysis: theory and application to travel demand**. The MIT Press, Cambridge, v.390, p.1985. Doi: 10.1007/978-1-4615-5203-1_2.

BHAT, C.; ZHAO, H. The spatial analysis of activity stop generation. **Transportation Research Part B Methodical**, Texas, v.36, p.557–575, 2002. Doi: 10.1016/S0191-2615(01)00019-4.

_____.; SENER, I. A copula-based closed-form binary logit choice model for accommodating spatial correlation across observational units. **J Geogr Syst.**, Texas, v.11, p.243–272, 2009. Doi: 10.1007/s10109-009-0077-9.

_____.; GUO, J. A mixed spatially correlated logit model: formulation and application to residential choice modeling. **Transportation Research Part B**, Texas, v.38, n.2, p.147–168, 2004. Doi: 10.1016/S0191-2615(03)00005-5

BITTENCOURT, D. C.; DELGADO, J. P. M.; PITOMBO, C. S. Análise espacial de segurança em transporte coletivo. In: ANPET - PANORAMA NACIONAL DA PESQUISA EM TRANSPORTES, 2014, Curitiba. **Anais...** Curitiba, 2014.

_____. **Avaliação espacial de ocorrências de roubo em transporte coletivo urbano por ônibus**. Dissertação (Mestrado em Engenharia Ambiental Urbana) - Universidade Federal da Bahia, 2016.

BRUTON, M. J. **Introdução ao planejamento dos transportes**. Rio de Janeiro: Interciência 1979.

CAMARGO, E. C. G.; FELGUEIRAS, C. A.; MONTEIRO, A. M. V. A importância da modelagem da anisotropia na distribuição espacial de variáveis ambientais utilizando procedimentos geoestatísticos. In.: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, FOZ DO IGUAÇU, 10., 2001, Foz do Iguaçu. **Anais...** Foz do Iguaçu: INPE, 2001.

CERVERO, R.; RADISCH, C. Pedestrian versus automobile oriented neighborhoods. **Transport Policy**, [S.l.], v.3, p.127–141, 1996. Doi:10.1016/0967-070X(96)00016-9.

CHEN, X. *et al.* A Bayesian Stochastic Kriging Metamodel for Simultaneous Optimization of Travel Behavioral Responses and Traffic Management. In.: TRB 94th Annual Meeting Compendium of Papers. Washington, 2015. Doi:10.1016/j.trpro.2015.06.056.

CHRISTAKOS G.; SERRE M. L. Spatiotemporal analysis of environmental exposure-health effect associations. **J. Expo Anal Environ Epidemiol**, Berlin, v.10, n.2, p.168-187. 2000. Doi: 10.1038/sj.jea.7500077

CIUFFO, B. F.; PUNZO, V.; QUAGLIETTA, E. Kriging Meta-Modelling to Verify Traffic Micro-Simulation Calibration Methods. In.: TRB 90th Annual Meeting Compendium of Papers. Washington, 2011.

COSTA, A. S.; PITOMBO, C. S. Análise da tendência de mudança modal com foco na mobilidade sustentável. In: ANPET, 25., 2011, Belo Horizonte. **Anais...** Belo Horizonte, 2011.

_____.; _____.; SALGUEIRO, A. R. Estimção de escolha modal através da Geoestatística. In: XXVII ANPET, 2013, Belém. **Anais...** Belém, 2013.

_____. **Proposta de um método para estimção de escolha modal através da Geoestatística**. 2013. Mestrado em Engenharia Ambiental Urbana - Universidade Federal da Bahia, Salvador, 2013.

DOMENCICH, T.; MCFADDEN, D. L. **Urban Travel Demand: A Behavioral Analysis**. Amsterdam: North-Holland Press, 1975.

DUGUNDJI, E.; WALKER, J. Discrete choice with social and spatial network interdependencies: an empirical example using mixed generalized extreme value models with field and panel effects. **Journal of the Transportation Research Board**, Washington, v.1921, p.70-78, 2005. Doi:10.3141/1921-09.

FOTHERINGHAM, A. S.; WONG, D.W.S. The modifiable areal unit problem in multivariate statistical analysis. **Environment and PlanningA**, New York, v.23, p.1025–1044, 1991. Doi: 10.1068/a231025.

GOMES, M. M.; PITOMBO, C. S. Estimação de acidentes de trânsito com óbito através de técnicas de análise espacial dedados. In.: ANPET 2014, Curitiba. **Anais...** Curitiba, 2014a.

GOMES, M. M. *et al.* ESTIMATING THE PROBABILITY OF ROAD CRASH FATALITY USING GEOSTATISTICS: A BRAZILIAN CASE STUDY. Geostats 2016. Valencia, Espanha. 2016b. [Aceito para publicação]

GOMES, V. A. *et al.* Kriging Geostatistical Methods for Travel Mode Choice: A Spatial Data Analysis to Travel Demand Forecasting. **Open Journal of Statistics**, [S.l.], v.6, p.514-527, 2016a. Doi:10.4236/ojs.2016.63044.

_____; PITOMBO, C.S. Incorporação da Geoestatística a previsões de escolha modal. In: ANPET, 2014, Curitiba. **Artigo...** Curitiba, 2014b.

GOOVAERTS, P. Geostatistical analysis of disease data: estimation of cancer mortality risk from empirical frequencies using Poisson Kriging. **International Journal of Health Geography**, Bethesda, v.4, n.31, 2005. Doi:10.1186/1476-072X-4-31.

_____. Kriging and semivariogram deconvolution in presence of irregular geographical units. **Mathematical Geosciences**, Bethesda, v.40, p.101-128, 2008. Doi: 10.1007/s11004-007-9129-1

_____. Combining areal and point data in geostatistical interpolation: Applications to soil science and medical geography. **Mathematical Geosciences**, Bethesda, v.42, p.535-554, 2010. Doi:10.1007/s11004-010-9286-5.

GOTWAY C. A.; YOUNG L. J. **A geoestatistical approach to linking geographically-aggregated data from diferente sources. Technical report #2004-012.** Florida: Departament of stastics, University of Florida, 2004.

HORNER, M. W.; MURRAY, A. T. Excess commuting and the modifiable areal unit problem. **Urban Studies**, [S.l.], v.39, p.131–139, 2002. Doi: 10.1080/00420980220099113.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATISTICA. Censo Demográfico Brasileiro 2010. IBGE, 2010. Disponível em <<http://www.ibge.gov.br>>. Acesso em: 20 ago. 2012.

INSTITUTO DE PESQUISA ECONÔMICA APLICADA Malha digital dos municípios do Estado da Bahia. IPEA, 2014. Disponível em:<<http://www.ipea.gov.br/ipeageo/malhas.html>> Acesso em: 29 mar. 2014.

JOURNAL, G. D; HUIJBREGTS, J. C.H. **Mininig geoestatistics.** Nova York. Academic Press, 1978.

JOURNAL, A. G. Non-parametric estimation of spatial distribution. **Journal of the International Association for Mathematical Geology**, [S.l.], v.15, p.445-468. 1983. Doi:10.1007/BF01031292.

KASSTEELE, J. van de; VELDERS, G.J.M. Uncertainty Assessment of Local NO2 Concentrations Derived from Error-in-Variable External Drift Kriging and its Relationship to

the 2010 Air Quality Standard. **Atmos Environ**, [S.l.], v.40, n.14, p.2583-2595, 2006. Doi: 10.1016 / j.atmosenv.2005.12.023.

_____.; STEIN, A. A model for external drift kriging with uncertain covariates applied to air quality measurements and dispersion model output. **Environmetrics**, [S.l.], v.17, n.4, p.309-322, 2006. Doi: 10.1002 / env.771.

KAWAMOTO, E. **Análise de sistemas de transporte**. 2 ed. São Carlos: Departamento de Engenharia de Transportes. Escola de Engenharia de São Carlos, 1994.

KITAMURA, R.; MOKHTARIAN, P. L.; LAIDET, L. A micro-analysis of land use and travel in five neighborhoods in the San Francisco Bay Area. **Transportation**, Netherlands, v.24, p.125–158, 1997. Doi: 10.1023/A:1017959825565.

KYRIAKIDIS, P. A geostatistical framework for área-to-point spatial interpolation. **Geography Anal**, Ohio, v.36, n.2, p.259-289, 2004. Doi: 10.1111/j.1538-4632.2004.tb01135.x

KWAN, M. P. Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: a methodological exploration with a large data set. **Transportation Research C**, [S.l.], v.8, p.185–203, 2000. Doi:10.1016/S0968-090X(00)00017-6.

LANDIM, P. M. B. **Análise Estatística de Dados Geológicos**. 2.ed. São Paulo: Universidade Estadual Paulista, 2003.

LIMA, M. L. A. *et al.* Krigagem Indicativa aplicada à pesquisa de nitrato nas águas subterrâneas da área urbana de Porto Velho – RO. In.: CONGRESSO BRASILEIRO DE ÁGUAS SUBTERRÂNEAS. 15., 2008, Natal. **Anais...** Natal, 2008.

LINDNER, A.; PITOMBO, C. S. Interpolação espacial de probabilidade de escolha de modo de transporte motorizado. In: ANPET - CONGRESSO DE PESQUISA E ENSINO EM TRANSPORTES, 29. 2015, Ouro Preto. **Anais...** Ouro Preto, 2015.

_____.; _____. Modelo logit binomial com componentes principais para estimação de preferência por modo de transporte motorizado. **Journal of Transport Literature**, Manaus, v. 10, p. 5-9, 2016. Doi: 10.1590/2238-1031.jtl.v10n3a1.

_____.; _____. Análises de diferentes amostragens no aprimoramento de estimativas de modelagem geoestatística na demanda por transportes. In: ANPET 2014, Curitiba. **Anais...** Curitiba, 2014.

_____. **Análise desagregada de dados de demanda por transporte através de modelagem geoestatística e tradicional**. 2015. Dissertação de Mestrado em Programa de Pós-graduação em Engenharia de Transportes - Escola de Engenharia de São Carlos/USP, 2015.

LINDNER, A. *et al.* Estimation of transit trip production using Factorial Kriging with External Drift: An aggregated data case study. **Geo-spatial Information Science**. [Sl.], 2016a.[Aceito para publicação].

_____.; *et al.* Evaluating the potential of spatial information when estimating motorized travel mode choices - A geostatistical approach. **Geostatistics**, Valência, 2016b. [Aceito para publicação].

LOPES, S. B. **Efeitos da dependência espacial em modelos de previsão de demanda por transporte**. 153f., 2005. Trabalho de conclusão de curso (Dissertação de Mestrado) - Escola de Engenharia de São Carlos, USP. São Carlos, 2005.

_____.; BRONDINO, N. C. M.; RODRIGUES DA SILVA, A. N.. GIS-Based Analytical Tools for Transport Planning: Spatial Regression Models for Transportation Demand Forecast. **ISPRS International Journal of Geo-Information**, Basel, v.3, p.565-583, 2014. Doi:10.3390/ijgi3020565.

LOUREIRO, C.F.G.; SILVA, H.N.; CARVALHO, L.E.X. Metodologia de análise de regressão geograficamente ponderada aplicada ao fenômeno das viagens intermunicipais. Panorama Nacional de Pesquisa em Transportes. In.: ANPET, 20, Curitiba. **Anais...** Curitiba, v.1, p.479-491. 2006.

MARTIN, D. Census 2001: making the best of zonal geographies. In.: THE CENSUS OF POPULATION: 2000 AND BEYOND, 2000, Manchester. **Anais...** Manchester: University of Manchester 2000.

MATHERON, G. **The theory of regionalized variables and its applications**. Paris: École Nationale Supérieure des Mines de Paris, 1971.

MAZZELLA, A.; PIRAS, C.; PINNA, F. Use of Kriging Technique to Study Roundabout Performance. **Transportation Research Record**. [S.l.], v.2241, 2011. Doi: 10.3141/2241-09.

MCFADDEN, D. The measurement of urban travel demand. **Journal of Public Economics**. Berkeley, v.3, n.4, p.303-328, 1974. Doi:10.1016/0047-2727(74)90003-6.

METRÔ-SP (1997) Companhia do Metropolitano de São Paulo. Pesquisa Origem e Destino - 1997. Disponível em < <http://www.metro.sp.gov.br/> > Acesso em: 5 de Março de 2007.

METRÔ-SP (2007) Companhia do Metropolitano de São Paulo. Pesquisa Origem e Destino - 2007. Disponível em < <http://www.metro.sp.gov.br/> > Acesso em: 8 de Janeiro de 2014.

MILLER, H. J.; SHAW, S. L. **Geographic Information System for Transportation: Principles and Applications**. Oxford: Oxford University Press, 2001. Doi: 10.1111/gec3.12204.

MIYAMOTO, K. *et al.* Discrete choice model with structuralized spatial effects for location analysis. **Transportation Research Record**. Washington, v.1898, p.183-190, 2004. Doi: 10.3141/1898-22

NOVAES, A. G. **Sistemas de Transportes: Análise da Demanda**. São Paulo: Edgard Blucher, 1986.

OLEA, R. A. A six step practical approach to semivariogram modeling **Stochastic Environmental Research and Risk Assessment**, [S.l.], v.20, p.307–318, 2006. Doi: 10.1007/s00477-005-0026-1

_____. Geostatistics for Engineers and Earth Scientists, **Technometrics**, [S.l.], v.42-4 p.444-445, 2000. Doi: 10.1080/00401706.2000.10485748.

OLIVER, M. A. *et al.* **Geostatistics for engineers and Earth scientists**. Boston: Kluwer Academic Publishers. 1999. Doi: 10.1007 / 978-1-4615-5001-3

_____. *et al.* Estimating the risk of childhood cancer. **Geostatistics Troia** 92, 1993. v.2

OLIVEIRA, V. D. Poisson kriging: A closer investigation. **Spatial Statistics**, [S.l.], v.7, p.1–20. 2014. Doi: 10.1016 / j.spasta.2013.11.001

ORGANIZAÇÃO DAS NAÇÕES UNIDAS. Programa das Nações Unidas para o Desenvolvimento. Atlas de Desenvolvimento Humano no Brasil 2003. [S.l.], 2003. Disponível em <http://www.pnud.org.br/atlas/ranking/IDH_Municipios_Brasil_2000.aspx?indiceAccordion=1&li=li_Ranking2003> Acesso em: 12 ago. 2012.

ORTÚZAR, J. D.; WILLUMSEN, L.G. **Modelling Transport**. Wiley, 4.ed. [S.l.], 2011.

PAPACOSTAS C. S.; PROVEDOUROS, P. D. **Transportation Engineering and Planning**. 2.ed. New Jersey: Prentice Hall, 1993.

PÁEZ, A. *et al.* Development of an indicator to assess the spatial fit of discrete choice models. **Transportation Research Part B**, [S.l.], v.56, p.217-233, 2013. Doi:10.1016/j.trb.2013.08.009.

_____.; SCOTT, D. M. Spatial statistics for Urban analysis: a review of techniques with examples. **Geo Journal**, Holanda, v.61, n.1, p. 53–67, 2004. Doi:10.1007/s10708-005-08775.

_____.; _____. Spatial statistics for urban analysis: a review of techniques with examples. **Geo Journal**. v.61, n.1, p.53-67, 2005. Doi:10.1007/s10708-005-0877-5.

_____. SUZUKI, J. Transportation impacts on land use change: an assessment considering neighborhood effects. **Journal of the Eastern Asia Society for Transportation Studies**, Tokio, v.4, p.47–59, 2001.

_____.; *et al.* The Spatial Analysis of Trip Generation: A Case Study of the Hamilton CMA, In.: CONFERENCE ON COMPUTERS IN URBAN PLANNING AND URBAN MANAGEMENT, 10., 2005, London. **Anais...** Londres, 2005.

_____.; FARBER, S.; WHEELER, D. A simulation-based study of geographically weighted regression as a method for investigating spatially varying relationships. **Environment and Planning A**, [S.l.], v.43 n.12, p.2011. Doi: 10.1068/a44111.

_____. FISCHER, M. M. Spatial perspectives in urban systems: developments and directions. **Journal of Geographical Systems**, [S.l.], v.9, n.1, p.1–6. 2007. Doi:10.1007/s10109-007-0041-5.

PEARCEA, J. L. *et al.* Characterizing the spatiotemporal variability of PM2.5 in Cusco, Peru using kriging with external drift. **Atmospheric Environment**. [S.l.], v.43, n.12, p.2060–2069, 2009. Doi: 10.1016/j.atmosenv.2008.10.060

PEER, S. *et al.* Door-to-door travel times in RP departure time choice models: An approximation method using GPS data. **Trans-*portation Research Part B***, [S.l.], v.58, p. 134-150, 2013. Doi: 10.1016/j.trb.2013.10.006.

PITOMBO, C. S. **Estudos de relações entre variáveis socioeconômicas, de uso do solo, participação em atividades e padrões de viagens encadeadas urbanas**. 2007, 268f. Tese de Doutorado em Engenharia de Transportes- Escola de Engenharia de São Carlos da Universidade de São Paulo, São Paulo, 2007.

_____.; SOUSA, A. J. . Aplicação de conceitos Geoestatísticos para análise de geração de viagens urbanas. In: ANPET 2009, Vitória. **Anais...** Vitória, 2009.

_____. *et al.* Comparing different spatial data analysis to forecast trip generation. In: WORLD CONFERENCE ON TRANSPORT RESEARCH SOCIETY, 2010, Lisboa. **Proceedings of the 12th WCTR**. Lisboa, 2010.

_____. Combined application of principal component analysis and geographically weighed regression in a trip generation analysis. In: CONGRESSO PAN-AMERICANO DE ENGENHARIA DE TRÁFEGO E TRANSPORTES E LOGÍSTICA, 16., 2010, Lisboa. **Anais...** Lisboa, 2010.

_____. *et al.* Uso de krigagem para estimação de geração de viagens. In.: CONGRESSO LUSO-BRASILEIRO PARA PLANEAMENTO URBANO, REGIONAL, INTEGRADO E SUSTENTÁVEL, 60., 2014, Lisboa. **Anais...** Lisboa: Pluris, 2014.

_____.; COSTA, A. S. G. Decision Tree application for modal choice. In: PANAM, 2014, Santander. **Anais...** Santander: Panam, 2014.

_____.; _____. Aplicação conjunta de modelos não paramétricos e paramétricos para previsão de escolha modal. **Journal of Transport Literature**, São José dos Campos – SP, v. 9, p. 30-34, 2015. Doi.org/10.1590/2238-1031.jtl.v9n1a6.

_____. *et al.* A Two-step method for mode choice estimation with socioeconomic and spatial information. **Spatial Statistics**, [S.l.], v. 11, p. 45-64, 2015. 2015a. Doi: 10.1016/j.spasta.2014.12.002.

_____. *et al.* Proposal of a sequential method for spatial interpolation of mode choice. **Boletim de Ciências Geodésicas JCR**, Curitiba, v. 21, p. 3, 2015b. Doi.org/10.1590/S1982-21702015000200016.

RIBEIRO, L. C. Os Estados e as Regiões Metropolitanas constitutivas do Observatório das Metrôpoles no Censo 2010. Instituto Nacional de Ciência e Tecnologia CNPq/FAPERJ/CAPES. Observatório das Metrôpoles: Rio de Janeiro. 2012.

ROCHA, S. S.; PITOMBO, C. S. Geração de viagens através de dados censitários com uso de técnica de Geoestatística. In: ANPET 2012, Joinvile. **Anais...** Joinvile, 2012.

_____. **Análise de geração de viagens urbanas por transporte coletivo através de técnicas de Geoestatística**. 2014. Dissertação - Mestrado em Engenharia Ambiental Urbana - Universidade Federal da Bahia. Salvador, 2014.

_____.; LINDNER, A.; PITOMBO, C. S. Modelagem geoestatística aplicada à amostra sistemática desagregada: estimação de geração de viagens. In: ANPET 2014, Curitiba. **Anais...** Curitiba, 2014.

RODRIGUES DA SILVA, A. N. Elaboração de um Banco de Dados de Viagem para Auxílio ao Desenvolvimento de Pesquisas na Área de Planejamento dos Transportes. Relatório FAPESP, Processo nº 04/15843-4. Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2008.

_____. *et al.* Avaliação dos Resultados da Pesquisa de Opinião Realizada no Município de São Carlos – SP. Relatório Final. São Carlos, 2009.

SANCHEZ-DIAZ I., HOLGUIN VERAS J., WANG, X. An exploratory analysis of spatial effects on freight trip attraction. **Transportation**, New York, 2014. Doi: 10.1007/s11116-014-9570-1.

SANTOS, M. C.; PITOMBO, C. S. Modelos espaciais de acidentes de trânsito com óbitos. In.: ANPET, 2014, Curitiba. **Anais...** Curitiba, 2014.

SINDICATO DAS EMPRESAS DE TRANSPORTE DE PASSAGEIROS DE SALVADOR Pesquisa Domiciliar Origem/Destino. Salvador, 1995.

SOARES, A. **Geoestatística para as Ciências da Terra e do Ambiente**. 2.ed. Lisboa: IST Press, 2006.

SOLTANI, A.; IVAKI, Y. E. The influence of urban physical form on trip generation, evidence from metropolitan Shiraz, Iran. **Indian Journal of Science and Technology**, Bangalore, v.4, n.9, p. 1168-1174, 2001. Doi:10.17485/ijst/2011/v4i9/30250

THILL, J.C. Geographic Information System for Transportation in perspective. **Transportation Research C**, [S.l.], v.8, p.3-12, 2000. Doi: 10.1016/S0968-090X(00)00029-2

TOBLER, W.R. A computer movie simulating urban growth in the Detroit region, **Economic Geography**, [S.l.], v.46, p.234-240, 1970. Doi:10.2307/143141.

YOON, S. Y; RAVULAPARTHY, S.; GOULIAS, K. Dynamic diurnal social taxonomy of urban environments using data from a geocoded time use activity-travel diary and point-based business establishment inventory, **Transportation Research Part A: Policy and Practice**, [S.l.], v.68, p.3-17, 2014. Doi.org/10.1016/j.tra.2014.01.004.

YAMAMOTO, J. K.; LANDIM, P. M, B. **Geoestatística: Conceitos e Aplicações**. São Paulo: Oficina de Textos. 2013.

WACKERNAGEL, H. **Multivariate Geostatistics: An Introduction with applications**. 3.ed. [S.l.], 2010.

WEBSTER, R. *et al.* Kriging the local risk of a rare disease from a register of diagnoses. **Geographical Analysis**, [S.l.], v.26, p.168-18 1994. Doi:10.1111/j.1538-4632.1994.tb00318.x.

WONG, D.W.S. Aggregation effects in geo-referenced data. In: ARLINGHAUS, S. L. *et al.*, **Practical Handbook of Spatial Statistics**. Boca Raton: CRC Press, 1996.

YAMAMOTO, J.K. *et al.* Mapping an uncertainty zone between interpolated types of a categorical variable. **Computers and Geosciences**, [S.l.], v.40, p.146-152, 2012. Doi:10.1016/j.cageo.2011.09.005.

ZHANG, D.; WANG, X. Traffic Volume Estimation using Network Interpolation Techniques: An Application on Transit Ridership in NYC Subway System. Final Report. New York, 2013.