



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

INPE-10286-TDI/905

**ANÁLISE ESPACIAL DE ESTRUTURAS INTRA-URBANAS: O
CASO DE SÃO PAULO**

Frederico Roman Ramos

Dissertação de Mestrado do Curso de Pós-Graduação em Sensoriamento Remoto,
orientada pelos Drs. Antônio Miguel Monteiro e Gilberto Câmara Neto, aprovada em 12
de agosto de 2002.

INPE
São José dos Campos
2004

528.711.7

RAMOS, F. R.

Análise espacial de estruturas intra-urbanas: o caso de São Paulo / F. R. Ramos. – São José dos Campos: INPE, 2002.

139p. – (INPE-10286-TDI/905).

1.Planejamento urbano. 2.Sistemas de Informação Geográfica (SIG). 3.Análise espacial. 4.Dependências espaciais. 5.Krigeagem. I.Título.


Aprovado pela Banca Examinadora em cumprimento a requisito exigido para a obtenção do Título de **Mestre** em **Sensoriamento Remoto**.

Dra. Corina da Costa Freitas



Presidente/INPE/SJCampos-SP

Dr. Antônio Miguel Vieira Monteiro




Orientador/INPE/SJCampos-SP

Dr. Gilberto Câmara



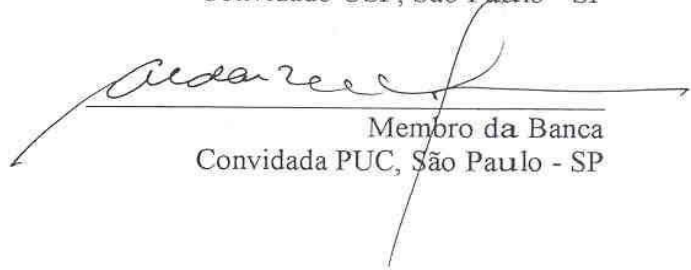
Orientador/INPE/SJCampos - SP

Dr. Ricardo Toledo Silva



Membro da Banca
Convidado USP, São Paulo - SP

Dra. Aldaíza de Oliveira Sposati



Membro da Banca
Convidada PUC, São Paulo - SP

Candidato: Frederico Roman Ramos

São José dos Campos, 12 de agosto de 2002.

“Essa sinfonia de raciocínios que contemplam a ponta do conhecimento, contemplam também usos, costumes e configuram as cidades, deste modo é a essência da questão da arquitetura que, na dureza da técnica, da ciência, da mecânica indispensável para que tudo fique em pé, faz com que também um projeto flua pela poesia, pela história e, num certo sentido, por todos os campos de conhecimento”.

Paulo Mendes da Rocha

A meus pais, Femke e Xavier.

AGRADECIMENTOS

Aos orientadores deste trabalho, Dr. Antônio Miguel Vieira Monterio e Dr. Gilberto Câmara pela confiança, apoio, disposição e amizade.

À equipe do Mapa da Exclusão/Inclusão Social, Dirce, Kazuo, Profa. Aldaíza e Jorge, pelos instigantes diálogos.

À João Carlos Scatena da Companhia do Metropolitano de São Paulo pela cessão dos dados da Pesquisa OD 97 e 87.

Ao pessoal do INPE, em especial ao pessoal da DPI, sempre dispostos a ajudar e sempre ajudando muito.

RESUMO

O uso de Sistemas de Informação Geográfica (SIG) vem crescendo dentro dos mais diversos campos do conhecimento onde a informação de localização é relevante. Os estudos intra-urbanos não são exceção a esta regra. Entretanto, nestes estudos, a complexidade do objeto de análise exige uma abordagem multidimensional. Dificilmente os processos que atuam na definição destes espaços podem ser estudados isoladamente. Esta complexidade desafia os estudiosos a formularem modelos conceituais capazes de capturar a informação necessária à análise que se queira fazer. Em geoprocessamento, um modelo é uma representação simplificada de determinado fenômeno que se fundamenta na construção de abstrações em diferentes níveis. Para que se possa utilizar o geoprocessamento em estudos intra-urbanos é necessário que se conheça os alcances e limitações da tecnologia, reconhecendo as especificidades da linguagem matemática subjacente a ela. O processo de modelagem de dados parte do reconhecimento dos conceitos relevantes à análise e se consolida nas possibilidades de manipulação matemático-computacional destes dados. Aqui, o conceito que fundamenta a análise é o de Estrutura Intra-Urbana como definido por Villaça em seu trabalho “O Espaço Intra-Urbano no Brasil”. A partir dele, apresenta-se um estudo de caso sobre a Cidade de São Paulo onde são discutidas as questões relativas à modelagem de dados intra-urbanos e às possibilidades de inferências quantitativas sobre eles. Dentre estas, são abordadas as questões relativas ao Problema das Unidades de Áreas Modificáveis (MAUP), os Indicadores Locais de Autocorrelação Espacial (LISA) e a utilização de técnicas geoestatísticas na construção de campos numéricos para dados intra-urbanos.

SPATIAL ANALYSIS OF INTRA-URBAN STRUCTURES: THE CASE OF SÃO PAULO CITY

ABSTRACT

The use of Geographical Information Systems (GIS) is increasing in a wide range of research fields where information related to location is relevant. Intra-urban studies are not an exception to the rule. However, in this case, the complexity of the object of study demands a multidimensional approach. The processes involved in the definition of these dense and complex spaces can not be studied separately. This complexity challenges the analysts to formulate conceptual models able to capture the necessary information for the analysis. In the geoprocessing domain a model is a simplified representation of certain phenomena based on the construction of different levels of abstraction. The application of geoprocessing technology in intra-urban studies depends on the knowledge of its intrinsic mathematical language. The data modelling process starts at the identification of the relevant entities for the analysis and leads to the consolidation of mathematical digital data manipulation. The concept of Intra-Urban Structure introduced by Villaça in "*O Espaço Intra-Urbano no Brasil*" is the theoretical background in this work. This concept is used as a base for a concrete analysis applied to the City of São Paulo. This example discusses the appliance of intra-urban data modelling and some quantitative inference in intra-urban studies. Within this inferences, are presented the Modifiable Areal Units Problem (MAUP), the Local Indicator for Spatial Autocorrelation (LISA) and the use of some geostatistical techniques to construct numerical fields for intra-urban data.

SUMÁRIO

LISTA DE FIGURAS	
LISTA DE TABELAS	
CAPÍTULO 1 INTRODUÇÃO	19
1.1 Contextualização	19
1.2 Objetivos	22
1.3 Organização da Dissertação	23
CAPÍTULO 2 O CONCEITO DE ESTRUTURA INTRA-URBANA	25
2.1 Definição de Estrutura Intra-Urbana	25
2.1.1. A escala intra-urbana	26
2.1.2. Os elementos da estrutura intra-urbana	28
2.1.3. A topologia da estrutura intra-urbana	32
CAPÍTULO 3 REPRESENTAÇÕES COMPUTACIONAIS DE ESTRUTURAS INTRA-URBANAS	37
3.1. Da Natureza Representacional dos Computadores	37
3.1.1. Os universos de abstração	37
3.1.2. O universo ontológico para modelos de estruturas intra-urbanas	39
3.2. Territórios Intra-Urbanos Digitais	41
3.2.1. Os geodados fisicoterritoriais e socioeconômicos	41
3.2.2. Os territórios intra-urbanos como objetos e campos	44
3.2.3. Imagens de sensores remotos	48
3.3. Modelo de Banco de Dados Geográficos	50
3.3.1. Vetores e matrizes	50
3.3.2. O modelo OMT-G para dados intra-urbanos	52
3.4. Conclusões	53
CAPÍTULO 4 ESTUDO DE CASO: SÃO PAULO EM GEODADOS	55
4.1. Dados: Fontes e Formatos	55
4.1.1. Pesquisa origem e destino do Metrô	56
4.1.2. Mapa da exclusão/inclusão social da cidade São Paulo	57
4.1.3. Imagens digitais de sensores remotos orbitais	59
4.2. O Projeto do BDG	60
CAPÍTULO 5 ANÁLISE ESPACIAL I: O PROBLEMA DAS UNIDADES DE ÁREA MODIFICÁVEIS	63
5.1. Contextualização	63
5.2. Abordagens ao MAUP	64
5.3. O Projeto do Zoneamento Territorial como Análise Espacial	65
5.3.1. Fundamentação teórica	65
5.3.2. Experimento	68
5.3.2.1. Quantificação do efeito de escala	69
5.3.2.2. Quantificação do efeito de zoneamento	73
5.4. Estimativa Empírica de Bayes	75
5.4.1. Fundamentação teórica	76
5.4.2. Experimento	78
5.4.3. Resultados	79
5.5. Conclusão	81

CAPÍTULO 6 ANÁLISE ESPACIAL II: INDICADORES LOCAIS DE ASSOCIAÇÃO ESPACIAL	83
6.1 Fundamentação Teórica	83
6.1.1. A estatística espacial	83
6.1.2. Estatística espacial local	85
6.2. As Estatísticas Locais nas Análises de Estruturas Intra-Urbanas	88
6.3. Exemplos de Aplicações dos LISAs em Análises Intra-Urbanas	90
6.3.1. A autocorrelação espacial para a densidade de empregos	92
6.3.2. A autocorrelação espacial para a variável baixa escolaridade	95
6.4. Conclusão	99
CAPÍTULO 7 ANÁLISE ESPACIAL III: DE UNIDADES DE ÁREA A SUPERFÍCIES CONTÍNUAS	101
7.1. Fundamentação Teórica	101
7.1.1. Krigeagem ordinária	103
7.1.2. Krigeagem ordinária por indicação	106
7.2. A Representação da Estrutura Intra-Urbana como Geo-campos	107
7.2.1. A continuidade espacial intra-urbana	108
7.2.2. A interpolação a partir de dados zonais	110
7.2.3. A krigeagem na interpolação de geodados zonais socioeconômicos	112
7.3. Experimentos	113
7.3.1. Krigeagem ordinária	117
7.3.2. Krigeagem ordinária por indicação	120
7.4. Conclusões	123
CAPÍTULO 8 CONCLUSÕES	123
8.1. Síntese das Contribuições do Trabalho	123
8.2. O Estado Atual da Tecnologia e Seu Horizonte	125
8.3. Trabalhos Futuros	126
8.4. Considerações Finais	127
REFERÊNCIAS BIBLIOGRÁFICAS	129
APÊNDICE A	135
APÊNDICE B	139

LISTA DE FIGURAS

3.1	Os quatro universos de abstração proposto em Câmara (2001)	39
3.2	Categorias representacionais usuais em SIG	41
3.3	Esquema de representação dos dados sobre os territórios intra-urbanos	43
3.4	Mapa coroplético e mapa de superfície dos homicídios em São Paulo	47
3.5	Imagem TM/Landsat 5 bandas espectrais 3,4 e 5 da RMSP de 1997	49
3.6	Modelo conceitual de banco de dados geográficos	52
4.1	Bases territoriais dos dados acessados	58
4.2	Primeiro nível hierárquico do modelo OMT-G	60
4.3	Segundo nível hierárquico do modelo OMT-G, categoria imagem	61
4.4	Segundo nível hierárquico do modelo OMT-G, categoria cadastral	61
5.1.	Esquema ilustrativo dos efeitos das diferentes possibilidades de agrupamento de um mesmo grupo de indivíduos nos coeficientes de correlação intra-áreas e inter-áreas	66
5.2.	Gráfico de espalhamento das 389 zonas OD97 cruzando as populações totais e o tamanho das zonas	68
5.3	As 389 zonas OD97 classificadas de acordo com a sua população total e a mancha urbana extraída da imagem TM de 1997	69
5.4.	As 270 zonas OD97 e os 96 distritos do MSP agrupados de acordo com as variáveis selecionadas.	72
5.5.	As 96 novas zonas construídas interativamente pelo critério de homogeneidade da variável renda individual per capita	54
5.6.	Gráfico de espalhamento da taxa de mortalidade infantil e a população até um ano de idade para os 96 distritos do MSP e dos 38 municípios restantes da RMSP	58
5.7.	Gráfico de espalhamento da taxa estimada de mortalidade infantil e a população até um ano de idade para os 96 distritos do MSP e dos 38 municípios restantes da RMSP	59
5.8.	Mapas coropléticos com a taxa de mortalidade infantil observada e estimada	60
6.1.	Exemplo de divisão zonal com matriz de vizinhança associada definida por propriedade de contiguidade	63
6.2.	Esquema de gráfico de espalhamento de Moran	88
6.3.	a) densidade de empregos classificadas por quintís; b) porcentagem da população de baixa escolaridade classificada por quintís	91
6.4.	Mapa de significância dos resultados obtidos para o índice local de Moran para variável emp_area após 999 permutações	93
6.5.	Os índices G_i^* normalizados calculados para variável emp_area classificados por desvios em relação à média	94
6.6.	Mapa de espalhamento de Moran obtido para variável densidade de empregos	95
6.7.	Mapa de significância dos resultados obtidos para o índice local de Moran para variável low_edu após 999 permutações	96

6.8.	Os índices G_i^* normalizados calculados para variável low_edu classificados por desvios em relação à média	97
6.9.	Mapa de espalhamento de Moran obtido para variável low_edu	98
7.1.	Semivariograma experimental idealizado e seus parâmetros utilizados na definição de um modelo teórico de ajuste	105
7.2.	Métodos de representação cartográfica para dados populacionais	109
7.3.	Esquema ilustrativo das etapas do processo de construção de superfícies contínuas a partir de dados zonais por krigeagem	111
7.4.	Conjunto amostral de 254 pontos gerado a partir dos centroídes dos polígonos OD87 e 389 pontos da OD97	112
7.5.	Análise das distribuições dos conjuntos amostrais de auto_pop OD87 e OD97	114
7.6.	Superfícies de semivariograma referentes aos campos de amostras para auto_pop OD 87 e OD97	114
7.7.	Semivariogramas experimentais calculados para as amostras auto_pop OD87 e OD97 com respectivas funções exponenciais ajustadas	116
7.8.	Histograma de erros de estimação para as funções de interpolação por krigeagem ordinária para auto_pop OD87 e OD97	116
7.9.	Superfícies contínuas geradas por krigeagem ordinária para variável auto_pop OD87 e OD97 representadas por variações em níveis de cinza	117
7.10	Superfícies contínuas geradas por krigeagem probabilística para variável auto_pop OD87 e OD97 representadas por variações em níveis de cinza	119
7.11.	Grade de incerteza associada as grades estimadas por krigeagem probabilística definida por distancias interdecís; OD87 e OD97	120
7.12.	Representação temática por faixas de valores de auto_pop das superfícies contínuas geradas para OD87 e OD97	122
8.1.	Descrição esquemática das etapas envolvidas no processo de aplicação de análises urbanísticas em ambiente SIG	124

LISTA DE TABELAS

4.1.	Especificações das imagens incorporadas ao BDG	59
5.1.	Matrizes de correlação para as variáveis selecionadas para zonas OD97 e distritos do MSP	71
5.2.	Matriz de correlação entre as variáveis do sistema de zonas gerado por critério de homogeneidade intra-zona	74
7.1.	Parâmetros dos modelos de semivariogramas teóricos ajustados aos semivariogramas experimentais para os conjuntos amostrais binários definidos pelos decis para OD87 e OD97	119

CAPÍTULO 1

INTRODUÇÃO

1.1. Contextualização

Estudar as cidades requer um olhar amplo, capaz de identificar e relacionar os inúmeros processos sociais, culturais, econômicos e ambientais que atuam na definição dos espaços urbanos. Hoje são bilhões os que vivem em cidades, muitas delas, imensas aglomerações urbanas que colocam os teóricos frente ao desafio de elaborar novos modelos teórico-conceituais capazes de traduzir a totalidade das inter-relações presentes nestas estruturas multidimensionais. A utilização de modelos teórico-conceituais é necessária para que se elaborem análises e propostas de intervenção. Entretanto, modelos são sempre reducionistas, a realidade é selecionada e representada sob determinados códigos que estabelecem a relação entre o objeto real observado e sua representação num modelo.

Os modelos espaciais desenvolvidos para análise de estruturas urbanas inauguram-se a partir do esforço teórico da ecologia humana da Escola de Chicago. Dentre estes, o modelo de “zonas concêntricas” proposto por Burgess em 1925 iniciou esta forma de abordagem e influenciou uma série de outros modelos como o modelo de “setores” proposto por Hoyt em 1939 e o modelo de “núcleos múltiplos” de Harris e Ullman em 1945 (Villaça,1998). Os modelos matemáticos desenvolvidos a partir da década de 60 surgiram paralelamente ao desenvolvimento das tecnologias de computação, e apostavam na capacidade de manipulação e armazenamento de grande quantidade de dados para a descrição de estruturas urbanas e predição de padrões de fluxos e ocupações (Echenique,1975). Apesar da importância que a elaboração destes modelos representaram para o desenvolvimento do planejamento urbano enquanto disciplina, a extrema simplificação da realidade e as distâncias entre os diferentes contextos urbanos para os quais estes foram desenvolvidos, limitaram sua aplicação em experimentos concretos.

Hoje, assiste-se ao crescimento do uso de Sistemas de Informação Geográfica (SIG) e do geoprocessamento em empresas e instituições públicas e privadas que atuam sobre as questões da cidade. Nestes sistemas, informações com conteúdo geográfico de natureza diversa podem ser armazenadas, manipuladas, visualizadas e principalmente transformadas através de processamentos matemáticos. Define-se como Análise Espacial a coleção de técnicas matemático-computacionais que operam sobre conjuntos de dados com informação geográfica, ou georreferenciados. Estas técnicas englobam desde operações de sobreposição de camadas de informação (*overlays*), às álgebras definidas para operar sobre representações digitais de mapas e estatísticas espaciais, que derivam da estatística tradicional, porém procuram considerar o “lugar” geográfico da ocorrência da variável observada como informação significativa (Bailey e Gattrel,1995) (Câmara e Monteiro,2001b). Os métodos e técnicas da análise geográfica quantitativa aplicados em ambiente digital, que encontram sua mais evidente expressão em SIG, objetivam a formulação de modelos de distribuição espacial para variáveis geográficas, análise de padrões e problemas relativos à localização de variáveis, e à investigação e predição de dinâmicas espaço-temporais (Longley,1996).

Em estudos urbanos, estas técnicas têm sido exploradas por grupos de pesquisa na Inglaterra, com importantes trabalhos em visualização e modelagem de fluxos e dinâmicas urbanas (Batty,1996) e em análises sintáticas (*space syntax*) dos ambientes intra-urbanos, análises dos elementos físicos e a representação espacial da morfologia construída, objetivando entender de que maneira o espaço físico está relacionado aos diversos fenômenos sociais e econômicos nas cidades (Hillier,1998); na Itália, trabalhos em visualização de fluxos intra-urbanos a partir da construção de campos de atração consegue interessantes representações dos deslocamentos internos cotidianos (Guez,2000), e na França, a técnica de regressão bidimensional vem sendo utilizada em estudos sobre os impactos dos novos sistemas de transporte no espaço urbano e regional (Cauvin,2000). No Brasil, alguns grupos de pesquisa têm se dedicado a aplicação do geoprocessamento a estudos urbanos. Dentre estes, o grupo da Universidade Federal do Rio Grande do Sul explora a utilização de SIG associado a modelos matemáticos como ferramenta para o planejamento urbano (Saboya,2000). Mais especificamente com o

conceito de sintaxe espacial de Hillier existem pesquisas na UFRN - Universidade Federal do Rio Grande do Norte - (Rufino et al.,2001). Em São Paulo, alguns importantes estudos intra-urbanos (Rolnik,2000) (Sposati,2000) e centros de pesquisa já utilizam o geoprocessamento em suas análises e têm apontado o potencial desta ferramenta nestes estudos.

Porém, como regra geral, a utilização do geoprocessamento em estudos urbanos tem se limitado às aplicações de cadastro urbano e a visualização de mapas digitais, dando-se pouca atenção à capacidade de processamento de dados espaciais através da aplicação de técnicas matemático-computacionais. Um dos obstáculos para efetivar a Análise Espacial de dados georreferenciados como instrumental para o planejamento urbano reside na dificuldade de se encontrar representações computacionais capazes de modelar adequadamente os dados disponíveis sobre o ambiente urbano. Em estudos urbanos, onde a diversidade de elementos é tão grande quanto a complexidade de relações existente entre eles, este ponto é particularmente crítico. Se por um lado, uma excessiva simplificação de conceitos é aplicada, o risco de ignorar-se aspectos importantes da realidade pode afetar significativamente o alcance de suas aplicações; por outro, representações altamente complexas podem tornar a análise sobrecarregada e de difícil interpretação. O equilíbrio é alcançado quando se estabelecem critérios claros sobre quais conceitos são relevantes e como eles podem estar representados coerentemente dentro do sistema.

A natureza multidimensional do fenômeno urbano exige múltiplas abordagens e um suporte de análise flexível, que possibilite um compartilhamento das múltiplas visões possíveis para um mesmo dado no espaço geográfico. Assim, os dados disponíveis, mapas digitais e levantamentos cadastrais e censitários, uma vez incorporados em ambiente digital, devem, com os cuidados necessários, admitir a representação computacional mais adequada para o estudo que se pretende fazer, embasado no arcabouço teórico do urbanismo e sob o olhar do especialista, responsável sempre pela reflexão crítica sobre sua prática tecnológica, sendo esta a receita para o êxito de estudos urbanos apoiados em geotecnologias.

1.2. Objetivos

Esta dissertação pretende examinar a possibilidade de aplicação de técnicas matemático-computacionais para tratamento de dados espaciais como suporte a estudos urbanos. Para isso, o enfoque deste trabalho será a aplicação destas técnicas na modelagem e análise da estrutura socioeconômica do ambiente intra-urbano. A dimensão socioeconômica do espaço intra-urbano está diretamente relacionada à população, às suas condições de vida, à maneira como ela se distribui sobre o espaço, seu hábitat e às relações que ela estabelece com o ambiente nos seus diferentes níveis. O desafio que se coloca então, é estabelecer métodos de análise da dimensão socioespacial dos ambientes urbanos, utilizando-se as capacidades de análises quantitativas do geoprocessamento como suporte.

Para isso, são realizadas análises para a Cidade de São Paulo sobre uma coleção de dados obtidas em sua maioria da Pesquisa Origem/Destino da Companhia Metropolitana de São Paulo de 1987 e 1997. Este conjunto de dados, além de ser extremamente rico, permite estudar o caso de São Paulo, sem dúvida um espaço laboratório privilegiado dentro da realidade urbana brasileira, para o estudo de estruturas intra-urbanas.

Espera-se com esta dissertação, ampliar o entendimento de como a utilização de SIG e Análise Espacial pode contribuir aos estudos urbanísticos, possibilitando novas formas de tratamentos de dados e ampliando a capacidade de visualização e percepção do fenômeno urbano. Com a certeza de que a oferta de dados digitais georreferenciados sobre as cidades será cada vez maior, e o avanço das tecnologias e técnicas de tratamento destes dados não sofrerá interrupções, espera-se que esta dissertação seja também uma pequena contribuição de um arquiteto e urbanista aos seus pares, pelo simples fato de demonstrar que é possível assimilar estas técnicas no estudo das estruturas intra-urbanas, e contribuir assim para uma mudança qualitativa no ofício dos planejadores.

1.3. Organização da Dissertação

Esta dissertação está organizada em oito capítulos, onde nos capítulos 2 e 3 são apresentados os conceitos necessários à compreensão dos métodos e análises contidas nos capítulos posteriores. Ao longo de todo o texto tem-se sempre como condutor o estudo de caso para o espaço intra-urbano em São Paulo. Abaixo, a estrutura de cada um dos capítulos.

No próximo capítulo, *Capítulo 2 - O Conceito de Estruturas Intra-Urbanas*, são apresentados os conceitos urbanísticos que sustentam a definição de estrutura intra-urbana tendo como base o trabalho teórico de Villaça (Villaça, 1998). Destes conceitos serão estabelecidas as pontes entre os universos conceituais do urbanismo e do geoprocessamento.

No *Capítulo 3 - Representações Computacionais de Estruturas Intra-Urbanas*, as questões relativas às representações computacionais para o estudo destas estruturas são apresentadas. A definição de modelos de dados geográficos para estudos urbanos é o primeiro passo para a implementação de estruturas territoriais em ambiente SIG, e caracterizam as abstrações utilizadas para a representação da realidade geográfica. As abstrações de conceitos e entidades do mundo real é parte fundamental na criação de um sistema de informação voltado para esta aplicação. Elas servem como ferramentas auxiliares à compreensão do sistema, dividindo-o em componentes menos complexos e isolados. Cada um destes componentes pode ser visualizado em diferentes níveis de complexidade e detalhe, de acordo com a necessidade de compreensão e representação das diversas entidades e de suas interações.

No *Capítulo 4 - Estudo de caso: São Paulo Digital* é apresentada a modelagem do banco de dados geográficos utilizado nesta dissertação. Os *Capítulos 5, 6 e 7* tratam do uso da Análise Espacial, entendida como um conjunto de técnicas matemático-computacionais aplicadas ao estudo de fenômenos geográficos, em estudos intra-

urbanos e apresentam exemplos práticos realizados sobre o conjunto de dados disponível.

O **Capítulo 5 - Análise Espacial I – O Problema das Unidades de Áreas Modificáveis**, discute e apresenta métodos para lidar com a agregação de dados coletados ao nível do indivíduo e agrupados em unidades de área, caso comum para dados intra-urbanos. Assim, o dado acessível diz respeito somente às características estatísticas de um grupo de indivíduos, e não mais ao indivíduo que muitas vezes é o interesse da análise. Apesar da indiscutível validade deste procedimento, ele é sensível a alguns fatores que muitas vezes são desconsiderados nas análises que se apoiam nestes dados. Fatores como escala, homogeneidade interna do grupo de indivíduos e definição das fronteiras das zonas de coleta interferem diretamente sobre os resultados. Esta é uma característica inerente a este tipo de dado conhecida na literatura como *Problema das Unidades de Área Modificáveis* (Openshaw,1984) (Wrigley et al.,1996), que pode ser avaliado e trabalhado com auxílio do geoprocessamento.

O **Capítulo 6 – Análise Espacial II - As Análises de Autocorrelação Espacial Local**, discute e analisa o conceito de *autocorrelação espacial*, suas possibilidades de identificação e mensuração, sua interpretação e aplicação em ambientes urbanos. A interdependência espacial é a noção por trás deste conceito. Buscam-se evidências em padrões localizados de valores semelhantes indicando tendências de que algo que ocorra em uma determinada localização estar relacionado à outras localizações vizinhas (Gattrel,1979). As análises de autocorrelação espacial têm por finalidade captar, caso existam, comportamentos globais de uma determinada variável distribuída sobre um território, e também inferir padrões locais para esta dependência espacial. Estes métodos são úteis para a detecção de locais de agregações espaciais significativas dentro do conjunto de localizações. Em estudos urbanos, o cálculo dos índices de autocorrelação local permite identificar e quantificar padrões de segregação intra-urbana sob determinadas variáveis. A importância deste tipo de análise cresce quando se reconhece a segregação como um processo fundamental para a compreensão da estrutura intra-urbana (Villaça,1998).

O **Capítulo 7 – Análise Espacial III - Das Unidades de Área às Superfícies Contínuas**, discute a aplicação de técnicas geoestatísticas na construção de *superfícies* a partir de dados censitários. A representação de dados socioeconômicos através de superfícies vêm sendo proposta como alternativa a representação zonal típica dos geodados populacionais (Martin,1996a) (Martin,1995). Nesta dissertação, a construção de superfícies está relacionada a idéia de dissolução das fronteiras rígidas das representações zonais, na procura por tendências na distribuição espacial dos atributos. A idéia é gerar campos numéricos para variáveis sociodemográficas do espaço interno da cidade. Apesar das superfícies estarem estreitamente relacionadas ao conjunto de zonas utilizadas no processo de interpolação, este outro olhar possível sobre os dados pode ser muito útil na análise das tendências das variáveis e no cruzamento de diferentes planos de informação em formatos matriciais.

No **Capítulo 8 - Conclusões e Trabalhos Futuros**, apresenta-se uma discussão geral do trabalho sob a ótica de um arquiteto e urbanista de formação que ao longo deste estudo incluiu em seu ferramental algumas habilidades para o trato quantitativo de dados espaciais e com isso passou a colaborar no seu entendimento da realidade urbana.

CAPÍTULO 2

O CONCEITO DE ESTRUTURA INTRA-URBANA

A definição de métodos de análise e de intervenção para os espaços urbanos requer um amplo arcabouço teórico capaz de situar o fenômeno frente as suas várias dimensões. A aplicação da tecnologia do geoprocessamento em estudos urbanos deve estar apoiada em conceitos urbanísticos, reconhecíveis dentro da especialidade, capazes de garantir a validade dos processos de modelagem e inferências computacionais. As dificuldades de aplicação desta tecnologia no campo do urbanismo advêm da constatação de que nenhuma metodologia predomina sozinha na análise urbana. Há sempre a necessidade de se adotar ampla variedade de abordagens para se analisar as complexidades da realidade urbana.

Em geoprocessamento, um modelo é uma representação simplificada e abstrata de um fenômeno ou situação geográfica concreta e que serve de referência para a observação, estudo e análise deste fenômeno. Para que se possa introduzir o geoprocessamento dentro do ferramental do urbanista, é necessário que se estabeleçam primeiramente os conceitos em que o trabalho se sustenta para daí avaliar quais são as expressões computacionais possíveis destes conceitos. Nesta dissertação, o conceito fundamental é o de Estrutura Intra-Urbana como colocado por Villaça (1998) na medida em que sua análise parte das inter-relações socioespaciais entre os diferentes elementos que compõem esta estrutura.

2.1. Definição de Estrutura Intra-Urbana

O conceito de *estrutura* é aplicado em diversas situações para definir a parte fundamental, essencial de um todo. Também, para definir a forma como as partes ou elementos e seus relacionamentos estão organizadas dentro de um todo, determinando a natureza, as características e o funcionamento deste todo. O estudo da estrutura intra-urbana deve dar conta do que é fundamental neste ambiente, analisar a organização das

partes e elementos que constituem as cidades em termos de seu arranjo espacial e das inter-relações que estes elementos apresentam.

Villaça (1998) aponta que “a expressão *estrutura urbana* – e sua correlata *reestruturação urbana* – tem sido vítima de muitos abusos. É freqüentemente utilizada como sinônimo de cidade enquanto elemento físico, de cidade como um todo material, sem considerar a inter-relação entre seus elementos, aliás sem considerar sequer que elementos são esses.” Ele segue: “o estudo da estrutura intra-urbana não será satisfatório se não der conta das *localizações* dos elementos da estrutura nem das correlações entre eles e outros elementos e/ou partes da metrópole.”

Sob esta perspectiva e dentro do âmbito desta dissertação, a modelagem computacional do conceito de estrutura intra-urbana baseia-se em três outros conceitos aqui considerados chaves: *a escala intra-urbana* definidora do nível de detalhamento ou resolução de apreensão do fenômeno, *os territórios intra-urbanos* entendidos como os elementos componentes da estrutura e *a topologia do espaço intra-urbano* entendida como o arranjo socioespacial dos territórios intra-urbanos caracterizando os inter-relacionamentos geográficos e funcionais dos elementos componentes da estrutura intra-urbana. Estes conceitos não estão isolados, juntos formam o arcabouço teórico do conceito de estrutura intra-urbana, porém serão analisados pontualmente nas próximas seções para direcionar a análise no estabelecimento de uma conexão entre os universos conceituais do urbanismo e do geoprocessamento.

2.1.1. A escala intra-urbana

Em qualquer estudo sobre o espaço geográfico, é preciso definir a unidade mínima de análise, a partir da qual os fenômenos observados são agregados. Os levantamentos cartográficos utilizam o termo *escala* para determinar a relação entre unidades mensuradas no campo e sua representação no mapa. Por exemplo, uma escala de 1:100.000 implica que cada centímetro no mapa corresponde a uma distância aproximada de 1Km no território mensurado. Os levantamentos cartográficos apresentam regras detalhadas sobre quais fenômenos e entidades da realidade devem ser

representados em cada uma das escalas cartográficas. O termo *escala*, acabou por analogia a este significado explícito do contexto cartográfico, sendo utilizado num sentido metafórico mais amplo como indicador da estratégia de desagregação da informação na representação espacial. Assim, escala é entendida aqui não como uma relação entre medidas no mundo e sua representação cartográfica, mas como uma definição imprecisa do nível de desagregação espacial utilizado na análise.

A definição da *escala intra-urbana* parte do entendimento do significado do adjetivo composto que qualifica sua aplicação a determinados casos. O termo intra-urbano surgiu da necessidade de se delimitar o espaço interno da cidade em contraponto ao componente urbano do espaço regional. A utilização do termo *estrutura urbana* em análises interessadas nas redes urbanas regionais e inter-regionais gerou uma confusão semântica que obrigou a utilização do prefixo quando o objeto de estudo é especificamente a compreensão da estrutura interna da cidade. A necessária diferenciação entre espaço intra-urbano e espaço regional define uma escala de percepção do fenômeno.

A análise em escala intra-urbana é onde se observa o alto nível de diferenciação interna das cidades. Os conjuntos de zonas, comunidades ou bairros que freqüentemente são identificáveis em termos de aparência física, composição da população e aspectos relacionados com as características e problemas sociais, que se repetem de uma cidade para outra. Nesta perspectiva, a escala intra-urbana está relacionada à distribuição da população, suas características, suas atividades e seus deslocamentos cotidianos sobre um território urbano contínuo. A necessidade de um olhar abrangente capaz de identificar comportamentos gerais e simultaneamente observar padrões de usos e fluxos ao nível mais local, requer uma flexibilidade nos suportes de análise onde o geoprocessamento certamente apresenta grande potencialidade.

Villaça (1998) assinala que o espaço intra-urbano é estruturado pelo deslocamento de seres humanos e desta maneira a escala intra-urbana se define. O deslocamento espacial cotidiano do ser humano especifica o espaço intra-urbano. Isto

pode ser mais bem compreendido quando ele diz: “são elas (as áreas metropolitanas) assentamentos, ou compartimentos territoriais estruturados pelos deslocamentos de seres humanos enquanto consumidores ou portadores da matéria força de trabalho; são, por isso, cidades – por maior e mais importantes e globais que sejam, e por mais que incluam vários municípios”. A leitura que se faz aqui é que a cidade não está restringida a fronteiras geopolíticas ou recortes administrativos, mas que a sua escala se revela a partir da articulação das diferentes zonas funcionais internas as quais são atratoras ou produtoras de fluxos.

A escala intra-urbana se define então como o grau de desagregação da informação espacial necessária para a compreensão da estrutura intra-urbana, seja de uma metrópole de 19 milhões de habitantes como São Paulo, como de uma cidade pequena com 20 mil habitantes. Evidente que quanto maior for a cidade, maior será a complexidade de sua estrutura intra-urbana e conseqüentemente diferentes padrões de assentamentos e deslocamentos populacionais poderão ser observados sob diferentes escalas de análise. Em muitos casos, esta escala corresponde ao nível de detalhe capturado pelo levantamento de dados utilizado. Por exemplo, se temos dados sobre cada um dos lotes urbanos, pode-se dizer em *escala de lotes*; analogamente pode-se dizer em *escala das quadras*, *escala dos bairros* ou qualquer outra unidade espacial na qual os dados estejam associados. Deste modo, na maior parte dos casos práticos a escala utilizada não é resultante de uma escolha a priori, mas decorrente da desagregação espacial inerente aos dados utilizados.

2.1.2. Os elementos da estrutura intra-urbana

O segundo conceito chave para o processo de modelagem computacional de estruturas intra-urbanas está estreitamente ligado ao conceito de escala intra-urbana. Quando observadas sob a luz do conceito de *território intra-urbano*, expressões como escala do lote, escala da quadra ou do distrito remetem a noções como *escala privada ou íntima*, *escala da vizinhança* ou *escala da comunidade*. O território traz a dimensão mais subjetiva do vínculo que se estabelece entre o conjunto de objetos arranjados no

espaço geográfico¹ e a construção de uma identidade coletiva da população que o ocupa, imprimindo-lhe características socioeconômicas específicas que o torna distinguível enquanto *território*. Assim, a noção de território se constrói a partir da associação indissolúvel entre o *espaço geográfico* e os *usos* que seus habitantes fazem dele (Santos,2000 citado por Koga,2001).

Villaça (1998) se refere a análise da estrutura intra-urbana como a investigação sobre o conjunto da cidade e sobre a articulação entre suas várias *áreas funcionais*. Ele diz: “são considerados elementos dessas estruturas o centro principal da metrópole (a maior aglomeração diversificada de empregos, ou a maior aglomeração de comércio e serviços), os subcentros de comércio e serviços (aglomerações de comércio e serviços, réplicas menores do centro principal), os bairros residenciais, ou melhor, o conjunto de bairros residenciais segundo as classes sociais e as áreas industriais. Esta estrutura está imbricada a outras estruturas territoriais, como sistemas de transporte e de saneamento. Entretanto, consideramos ser a primeira mais importante, pois inclui, incorpora e subjuga as demais, mais do que o contrário, embora não possa existir sem elas”. Pode-se concluir então que para Villaça, os elementos da estrutura intra-urbana devem possuir clara função urbana.

Nesta dissertação, os elementos componentes das estruturas intra-urbanas são identificados como territórios intra-urbanos. Assim, torna-se explícito que a sua caracterização se dá a partir dos aspectos intrinsecamente relacionais que se estabelecem em decorrência das relações sociais estabelecidas no lugar cotidiano (Koga,2001). Faz, dessa forma, a associação da idéia de *uso* inerente ao termo território com a de *função urbana* como utilizada por Villaça, evitando que o termo *função* seja entendido unicamente dentro do contexto da cidade setorizada, da linearidade funcional do racionalismo modernista da Carta de Atenas (Corbusier,1989) e direciona a análise para a complexidade socioambiental ou ecológica² mais ampla da estrutura intra-urbana.

¹ como colocado por Santos (1997) na definição de configuração territorial e paisagem

² ver Argan (1995) capítulo sete: a crise da arte como “ciência européia”, onde o autor aponta para o redirecionamento do *problema urbanístico* na segunda metade do século passado.

A análise da estrutura intra-urbana parte, portanto, da identificação destes territórios. Há uma hierarquia de territórios associados a diferentes esferas de interação social de sua população. Entretanto, nas análises da estrutura intra-urbana a noção de *território* adquire um significado muito próximo ao de *comunidade*. Milton Santos (1993) refere-se a estes lugares como *sítios sociais*, espaços seletivamente transformados pelas exigências funcionais da sociedade urbana. Estes *lugares* são reconhecíveis a determinadas escalas de observação. As fronteiras, por vezes são mais abruptas ou mais diluídas, mas geralmente é possível se observar a divisão da cidade nestes territórios.

A identificação dos territórios intra-urbanos passa necessariamente pela idéia de *segregação espacial*. A segregação intra-urbana é o processo segundo o qual diferentes camadas ou classes sociais tendem a se *concentrar* cada vez mais em diferentes *regiões gerais* ou *conjunto de bairros* da metrópole (Villaça,1998). Em Marcuse (1997) é apresentada uma taxonomia preliminar de algumas categorias territoriais intra-urbanas nas cidades norte-americanas como o Gueto, o Enclave e a Cidadela. No contexto norte-americano, suas características estão mais relacionadas a raça e religião da população. No Brasil, apesar desta segregação também ser estruturadora do espaço intra-urbano¹, é a segregação por classes sociais a que predomina (Villaça,1998). A disputa por vantagens de localização dentro da estrutura intra-urbana foi, e continua sendo, o fator preponderante na formação dos territórios intra-urbanos. Na marcante segmentação socioterritorial que as caracteriza, cada território intra-urbano tem seu *valor de uso*. O preço da terra é definido pelo acesso aos serviços e infra-estruturas ali instalados, ou ainda pela maior facilidade de acesso a estes. Aqui, os enclaves e cidadelas (Caldeira,2000) (Koga,2001) são também categorias territoriais intra-urbanas importantes no padrão de segregação espacial, ou da *estratificação urbana* como define Castells (2000).

¹ ver em Rolnik (1999) os Mapas dos territórios étnicos de São Paulo na década de 30.

A forte polarização centro-periferia observada por muitos estudiosos da questão urbana brasileira (Campos Filho,1989) (Maricato,1996) (Santos,1993) (Villaça,1998), é resultado direto da associação das legislações urbanísticas reguladoras aos interesses do capital imobiliário. O trabalho de Rolnik (1999) é minucioso na demonstração desta associação na definição dos territórios intra-urbanos que se formaram na cidade de São Paulo como consequência desta. Para ela, o fundamento da geografia social da cidade de São Paulo, tem sido desenhado desde os primórdios da expansão urbana. Os territórios segregados surgiram sob uma legislação urbanística que, ao mesmo tempo em que regulava meticulosamente algumas áreas da cidade, sempre foi permissiva quanto à existência de enormes porções com usos e ocupações ilegais. As grandes periferias, no plural porque são muitas e muito desiguais (Kowarick,2000), verdadeiros territórios da subcidadania, surgem em oposição ao centro provido de toda a espécie de serviço e infra-estrutura.

Esta geografia social, ou *topografia social* (Sposati,2000), que se desenha por estes processos é a base estrutural do espaço intra-urbano. Esta configuração socioespacial definirá todos os inter-relacionamentos funcionais presentes na estrutura.

A representação computacional desta base estrutural está diretamente relacionada ao alcance da informação dos dados, ou com o nível de desagregação da informação. Há uma distância que separa os territórios intra-urbanos dos zoneamentos territoriais. Não necessariamente a divisão geopolítica do espaço urbano em zonas, distritos ou sub-regiões está diretamente relacionada aos territórios intra-urbanos. Muitas vezes, o zoneamento da cidade é realizado em função de condicionantes operacionais, como coleta de dados ou de interesses políticos e não da captação das características comuns que possam vir a definir um território intra-urbano. Isto traz consequências às análises urbanísticas, uma vez que os dados coletados individualmente estão agrupados por zonas onde não há homogeneidade interna.

A homogeneidade interna dos territórios intra-urbanos ou *unidades urbanas* como Castells (2000) coloca, está diretamente relacionada à escala de análise. Villaça

(1998) diz que “o que se pretende explorar com esta questão é o tamanho da *área homogênea* ou a forma de seu traçado. Se a favela invade o interior de um bairro tendente à classe média ou média alta, duas situações são possíveis, ambas ligadas à questão da delimitação da área segregada, ou melhor, à área tendente à segregação. Na primeira situação, traçar-se-ia um perímetro que englobasse ambas as áreas, delimitando-se assim uma única área, a qual evidentemente não tenderia à segregação, mas sim à heterogeneidade. Na segunda situação, traçar-se-iam duas áreas, uma dentro da outra. Neste caso, haveriam duas áreas, ambas tendentes à segregação”. Assim, a homogeneidade interna é sempre relativa e depende do fenômeno que está sendo observado.

2.1.3. A topologia da estrutura intra-urbana

Os inter-relacionamentos entre os diversos territórios configuram a estrutura intra-urbana. Neste trabalho, os inter-relacionamentos espaciais se dão fundamentalmente a partir da topologia do espaço intra-urbano, ou seja, das localizações relativas dos elementos espaciais estruturadores deste espaço. O conceito de *topologia intra-urbana* refere-se ao todo, à forma como se dá a interação dos inúmeros territórios configurados através dos processos de urbanização e as conseqüências destas articulações funcionais no espaço. O entendimento das localizações intra-urbanas se constitui como objetivo principal das análises das estruturas territoriais intra-urbanas.

Para Villaça (1998), “estrutura, quando se refere a espaço urbano, diz respeito à localização relativa dos elementos espaciais e suas relações, ou seja, dos centros de negócios (não só o principal, mas também os demais) das áreas residenciais segregadas e, finalmente das áreas industriais. Assim, ao estudo da estrutura urbana interessa saber porque estes bairros e centros exibem certo arranjo territorial, e não outro qualquer, e qual a inter-relação espacial entre estes bairros e centros, ou quais são seus papéis espaciais. Ao estudo da estrutura intra-urbana é irrelevante como surge a classe média e qual a origem nacional de seu peso, isto é, poder político. Interessa saber porque ela se localiza onde se localiza e quais são as implicações disso.”

A noção de localização é utilizada em diversas situações em análises intra-urbanas. As diversas categorias semânticas do termo “lugar” nascem da multidimensionalidade do fenômeno observado. Em Augé (1994) são apresentadas três categorias que parecem abranger os principais contextos: (a) o lugar aristotélico; (b) o lugar antropológico e; (c) o lugar transcontextual ou o não-lugar. A cada uma destas categorias estão associadas diferentes categorias de relacionamentos ou localizações.

O lugar aristotélico relaciona-se com idéia direta de volume, de espaço ocupado por um corpo. Pode-se associar esta idéia à noção de proximidade, contato, apreensão, alcance, ou seja, termos inseridos na lógica métrica do espaço. O lugar antropológico definido como identitário, relacional e histórico por Augé (1994) são os territórios intra-urbanos. Nestes, as articulações se dão por processos de segregação, proteção, dominação, exploração, ou seja, dos aspectos sociológicos do espaço. A terceira categoria é definida como os não-lugares. Definição nascida na última década do século passado da necessidade de se classificar os espaços intersticiais, virtuais, os espaços efêmeros de passagens e transições onde muitas das relações da sociedade globalizada se efetuam. Neste sentido, o não-lugar surge em oposição ao lugar antropológico. Aqui enquadram-se as cidades genéricas de Koolhaas e Mau (1995) ou os terrenos vagos de Brissac e Dercon (2001), elementos de uma configuração urbanística mais ampla, que dá às situações seu caráter essencialmente genérico e indistinto.

Sendo a dimensão urbanística do lugar relativa, ou seja, dificilmente objetivada de uma forma única dentro da complexidade territorial destes ambientes, há que se discutir o alcance do geoprocessamento em representar tais dimensões. Nos modelos topológicos computacionais, fala-se de contiguidade, de conectividade, de proximidade, de continuidade, ou seja, formulações conceituais nascidas no universo matemático da geocomputação. A introdução desta ferramenta lógica na análise das estruturas territoriais intra-urbanas requer a construção de modelos representacionais inevitavelmente reducionistas. Entretanto, as técnicas de Análise Espacial permitem formalizações da dimensão espacial a partir de linguagens interativas que vão além da

cartografia estática dos mapas em papel, e se apresentam como uma importante ferramenta de análise para o urbanista.

O próximo capítulo tratará das possibilidades de formalização e representação dos conceitos apresentados neste capítulo, apontando as etapas do processo de modelagem de geodados intra-urbanos.

CAPÍTULO 3

REPRESENTAÇÕES COMPUTACIONAIS DE ESTRUTURAS INTRA-URBANAS

A aplicação das tecnologias da engenharia de informação espacial em estudos urbanos parte necessariamente da formulação de um *modelo de dados*. Neste, são estabelecidas as abstrações formais que possibilitam o salto dimensional entre a realidade concreta e a sua expressão computacional. O ponto focal deste capítulo é validar a aplicação do geoprocessamento neste tipo de análise sem imprimir uma visão excessivamente reducionista sobre o fenômeno, sabendo-se infactível capturar todas as dimensões do fenômeno urbano num modelo de dados. A resposta pode estar na compreensão da natureza representacional dos sistemas computacionais e num posicionamento crítico do analista, reconhecendo as especificidades desta tecnologia como uma ferramenta de análise complementar aos seus estudos. Para se alcançar tal reconhecimento é necessário situar dentro do universo matemático da computação as possibilidades de formalização de conceitos e entidades do mundo real, além de conhecer em profundidade o significado e alcance de cada uma das operações aplicadas sobre essas formulações.

3.1. Da Natureza Representacional dos Computadores

3.1.1. Os universos de abstração

A evolução dos sistemas computacionais tem levado a uma grande penetração desta tecnologia em quase todas as atividades da vida humana. A ubiqüidade desta tecnologia tem gerado discussões sobre o complicado relacionamento institucional entre o mundo do computador e o resto do mundo. Surgem questionamentos como se os computadores não estariam se constituindo como uma espécie de novo imperialismo, reinventando a virtualidade dos diversos aspectos da vida em que estão inseridos à sua própria imagem (Agre,2000). Essa discussão é importante quando se estabelece o fundamento básico da Ciência da Geoinformação como a *construção de representações computacionais do espaço*. Segundo Câmara et al.(2000): “a geração atual de SIG oferece ferramentas que permitem a expressão de procedimentos lógicos e matemáticos

sobre as variáveis georreferenciadas com uma economia de expressão e repetitividade impossíveis de alcançar em análises tradicionais. No entanto, a tecnologia de SIG resolveu apenas os problemas simples de representação computacional do espaço. Os atuais sistemas são fortemente baseados numa lógica “cartográfica” do espaço, exigindo sempre a construção de “mapas computacionais”, tarefa sempre custosa e nem sempre adequada ao entendimento do problema em estudo.”

Freqüentemente, o processo de modelagem de dados é associado à construção de abstrações em diferentes níveis, partindo-se da identificação dos elementos da realidade relevantes ao modelo e transitando até o nível puramente sintático de implementação em ambiente computacional. A associação entre o processo de modelagem e níveis de abstração pode ser exemplificada em Burrough e McDonnell (1998), Câmara e Monteiro (2001c), Gomes e Velho (1995) e Peuquet (1984). Os níveis de abstração estabelecem as necessárias transições entre representações no nível *infológico* e *datalógico*. Segundo Schuurman (1999), o nível infológico está relacionado à representação semântica tradicional e o nível datológico à representação digital. Ela diz: “Os dados apoiam-se sobre estruturas computacionais de *lower-level*. Estas constróem a arquitetura básica dos computadores, os lugares onde os dados binários são arquivados. Embora muitas das idéias expressas em termos infológicos sejam complexas, a familiaridade com este método de expressão torna-o transparente. O pensamento infológico não é natural, porém, um método de articulação com o qual fomos acostumados durante nosso aprendizado. A tradução para o universo datológico aparentemente pesado e de difícil manuseio, tende a reforçar o senso de que o universo digital está muito mais distante dos princípios organizacionais humanos do que os analógicos.” Ela argumenta entretanto, que “ambos são comunicações construídas com diferentes áreas de rigidez e flexibilidade. Simplesmente, os seres humanos têm muito mais experiência em manipulações de estruturas fenomenológicas do que com estruturas de dados.”

Em Câmara e Monteiro (2001c), a transição do nível infológico ao datológico é relacionada a partir de quatro universos de abstração: *o ontológico, o formal, o*

estrutural e o de implementação (Fig.3.1). Os dois primeiros associados ao nível infológico e os dois últimos ao datológico. A cada um desses níveis estão associados diferentes momentos dentro do mesmo processo de formulação de abstrações. De maneira simplificada, pode-se dizer que no nível ontológico encontram-se os aspectos relevantes da realidade selecionados para a análise; o segundo nível refere-se às possibilidades de formalização conceituais destas entidades; no terceiro nível estão associadas às possibilidades de representação geométrico-matemáticas; no último, estão associadas às questões de implementação computacional, códigos e arquiteturas de sistemas de dados. Apesar dos diferentes níveis de abstração estarem estreitamente inter-relacionados, idealmente espera-se que o processo de análises geográficas em ambiente SIG se estabeleça sobretudo no nível infológico, baseando-se em ontologias e formalizações, ou seja, trabalhando em níveis de abstração mais altos, tratando sobre conceitos e não sobre especificidades de sistemas computacionais e operacionais (Câmara et al. 1996).

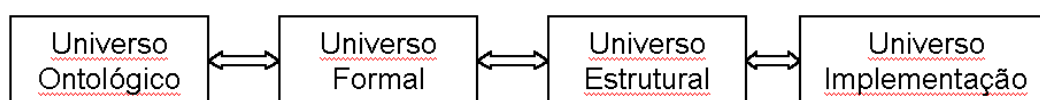


FIGURA.3.1. Os quatro universos de abstração proposto em Câmara e Monteiro (2001c).

3.1.2. O universo ontológico para modelos de estruturas intra-urbanas

O termo *ontologia* traz diferenças entre suas definições no sentido filosófico e na maneira como este é aplicado na comunidade da Ciência da Computação (Fonseca et al., 2000). Em computação, ontologia é entendido como um artefato da engenharia que descreve uma determinada realidade dentro de um vocabulário específico, estabelecendo um conjunto de definições relacionadas ao significado desejado de cada vocábulo. Neste sentido, a formulação de um *domínio ontológico* é aspecto crucial para questões como interoperabilidade e compartilhamento de dados entre múltiplos usuários. Em SIG, as ontologias estabelecem correspondências entre diferentes domínios de entidades espaciais e relações. Segundo Fonseca et al.(2000), no caso de SIG para aplicações urbanas as ontologias devem incluir: *objetos* (e.g. zonas político-administrativas, ruas, quadras, lotes, equipamentos); *relações* (e.g. uma escola gerencia um distrito educacional; um lote pertence àquela quadra); *eventos* (e.g. acidentes de

trânsito, manutenção de infra-estruturas); e *processos* (e.g. poluição sonora, fluxo de tráfego).

Dentro do escopo deste trabalho, nas análises de estruturas intra-urbanas em SIG, o domínio ontológico deve conter formulações para os elementos componentes desta estrutura, ou seja, o centro principal, os subcentros de comércio e serviços, o conjunto de bairros residenciais, as áreas industriais e seus inter-relacionamentos. O compromisso ontológico necessário à modelagem destas entidades em ambiente SIG deve permitir diferentes maneiras de olhar e identificar estas entidades. Por exemplo, quando se define o centro principal da metrópole enquanto a maior aglomeração diversificada de empregos, ou a maior aglomeração de comércio e serviços (Villaça,1998). Em SIG, a concentração dessas funções centrais deverá ser quantificada, medida, muitas vezes a partir de dados cadastrais e censitários, atrelados a determinados recortes administrativos ou a um conjunto de localizações. Neste caso, em análises de estruturas intra-urbanas o domínio ontológico deve incluir olhares possíveis sobre os territórios intra-urbanos de maneira a romper com a imobilidade dos suportes originais de dados. Estes olhares englobam operações sobre dados como agregações, sobreposições e análises estatísticas. Estas operações devem ser consideradas ontologias, expressões computacionais possíveis dos territórios intra-urbanos e seus inter-relacionamentos.

Quer-se demonstrar aqui que existe uma distância conceitual grande na construção de ontologias para entidades e processos do ambiente intra-urbano como ruas, lotes e escolas, e a construção de ontologias para análises da estrutura intra-urbana como definida no capítulo anterior. No primeiro caso pode-se abstrair com relativa naturalidade uma rua como uma linha ou um lote como um polígono. Em análises de estruturas intra-urbanas são conceitos mais subjetivos que devem ser representados. Os territórios e seus inter-relacionamentos dificilmente poderão ser representados simplesmente como imagens estáticas, polígonos coloridos, linhas e pontos. Eles deverão ser revelados num processo interativo de manipulação e transformação dos dados. Neste sentido, o geoprocessamento enquanto ferramenta analítica, permite a

necessária flexibilidade aos estudos intra-urbanos a partir da sua capacidade de manipular e processar geodados, transformando-os entre diferentes categorias representacionais a partir de técnicas de Análise Espacial, gerando assim novos dados, ampliando (ou realçando) a capacidade de observar um conjunto de dados (Fig.3.2).

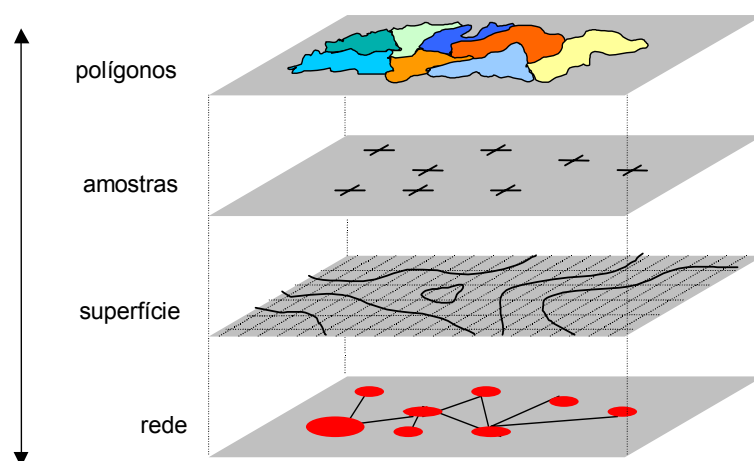


FIGURA.3.2. Categorias representacionais usuais em SIG.

3.2. Territórios Intra-Urbanos Digitais

Quando se observam os territórios enquanto elementos componentes da estrutura intra-urbana, pressupõem-se a possibilidade de identificação destes territórios dentro do espaço intra-urbano como um todo. Para tanto, deve-se ter a clareza do conjunto de características necessárias para que se qualifique uma determinada porção do espaço enquanto território. Em termos gerais, este conjunto de características pode ser subdividido em duas grandes subcategorias de informação: as *fisicoterritoriais* e as *socioeconômicas*. Nos territórios intra-urbanos, é grande a sobreposição destas categorias, tanto pelo impacto da densa ocupação humana no meio físico como pela intensa teia de relações sociais que estes ambientes suportam.

3.2.1. Os geodados fisicoterritoriais e socioeconômicos.

Grande parte do desenvolvimento das geotecnologias, em termos de metodologias de análise e tecnologia de aquisição de dados, deu-se dentro das

geociências e, portanto, focalizado em coleta e tratamento de dados fisicoterritoriais. Pode-se dizer que há um consenso em como adquirir e representar dados como a altitude, a temperatura, a concentração de poluentes, as estruturas geológicas, os rios ou a superfície do oceano. A objetividade destes dados torna-os palpáveis, mensuráveis e formalizáveis. É certo que os dados fisicoterritoriais, sob certos enfoques, podem trazer informação socioeconômica, por exemplo, a concentração de poluentes ou altos níveis de ruídos, ou ainda áreas de risco ou alagáveis trazem informações sobre as condições da população que habita estas áreas. Entretanto, usualmente nos dados de natureza socioeconômica, os processos de aquisição e tratamento são mais complexos. Pergunta-se: Como obter dados confiáveis sobre uma determinada população? Como representar um território a partir de suas dimensões humanas como qualidade de vida e desenvolvimento? Quais as dimensões da existência humana capazes de transmitir tal informação? Estas questões trazem a tona a subjetividade da dimensão humana presente nestes ambientes. Pode-se dizer então que a representação computacional de territórios intra-urbanos em suas dimensões fisicoterritoriais e socioeconômicas dependerá de duas condicionantes principais: primeiro, dos dados disponíveis, do alcance e representatividade das pesquisas socioeconômicas e da qualidade e precisão dos levantamentos de dados fisicoterritoriais; segundo, dos processamentos aplicados a este conjunto de variáveis no sentido de gerar o mosaico necessário para a análise que se deseja fazer.

Da mesma forma em que se deu o desenvolvimento do geoprocessamento focalizado nas geociências, hoje há uma demanda muito grande para que se disponibilize esta tecnologia em análises socioeconômicas. Um grupo crescente de pesquisadores das ciências humanas vem rompendo com o preconceito inicial da utilização de técnicas computacionais, muitas vezes tidas como positivista e de alcance relativo, fomentando assim um rico debate entre essa comunidade e pesquisadores da área da engenharia da geoinformação.

Grande parte das características dos territórios intra-urbanos é de natureza socioeconômica, pois se tratam fundamentalmente de territórios apropriados,

construídos e habitados por determinados grupos populacionais com características culturais, econômicas e sociais que os definem enquanto grupo e que estão intimamente associadas à maneira como se dá o estabelecimento destas populações sobre o território.

Tipicamente, as fontes de dados socioeconômicos são as pesquisas censitárias realizadas sistematicamente através de questionários entre intervalos constantes de tempo (no caso do censo brasileiro a cada 10 anos). Estas pesquisas contemplam um amplo conjunto de informações socioeconômicas relacionando aspectos de uma determinada população como renda, educação, habitação, entre outras variáveis. Na maioria dos casos, as informações censitárias são as fontes mais completas e confiáveis de dados dessa natureza, mesmo sabendo-se das limitações inerentes ao processo de levantamento, tais como veracidade das respostas coletadas ou representatividade do grupo ao qual foi aplicado o questionário no caso de pesquisas amostrais domiciliares. Porém, há uma outra dimensão socioeconômica que não pode ser revelada pelas pesquisas censitárias uma vez que estas contemplam apenas a população residente no determinado território. Esta outra dimensão é observada sobre os dados aqui chamados de territoriais cadastrais.

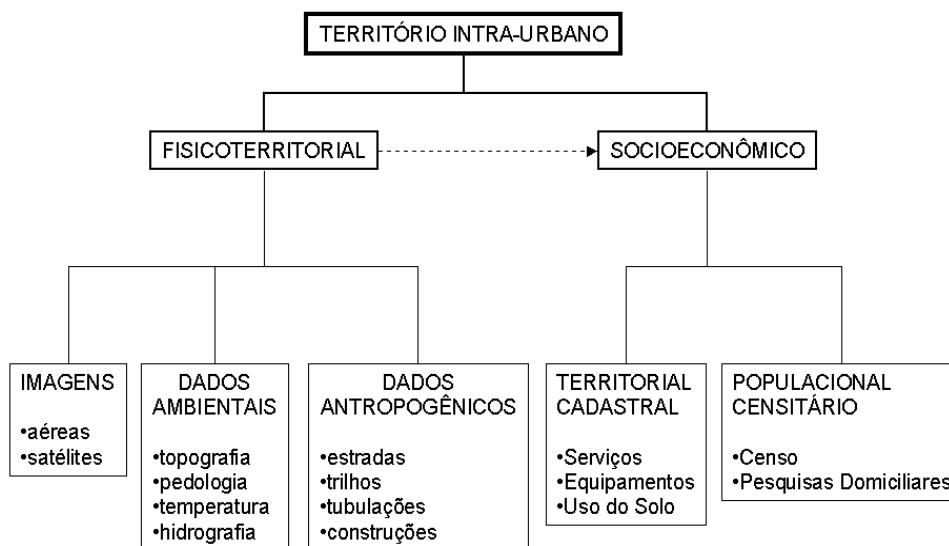


FIGURA.3.3. Esquema da representação dos dados sobre os territórios intra-urbanos.

Os dados cadastrais referem-se às características dos ambientes construídos que não estão necessariamente relacionadas à população ali residente. Exemplificando, imagine uma parcela da cidade em que quase não há residentes mas que corresponda a um importante núcleo de serviços e empregos para a população desta cidade. Neste caso a dimensão socioeconômica destes espaços dificilmente seria captada por um levantamento censitário domiciliar. Esta somente poderia ser revelada a partir de fontes cadastrais como número de estabelecimentos empregadores ou capacidade de atendimento dos serviços e equipamentos ali instalados. Portanto, pode-se dizer que as características socioeconômicas de um determinado território se revelam a partir da composição dos dados *populacionais censitários* e dos dados *territoriais cadastrais*¹ (Fig.3.3).

3.2.2. Os territórios intra-urbanos como objetos e campos

O modelo conceitual de referência mais utilizado na teoria de sistemas de informação espacial representa a realidade geográfica a partir da dicotomia de entidades definíveis (geo-objetos ou feições) e entidades com variação espacial contínua (geo-campos ou superfícies). Embora esta dicotomia simples tenha sido objeto de análise crítica (Couclelis,1992), mostrou-se uma base de referência consistente e está sendo utilizada, tanto no projeto da nova geração de sistemas de informação geográfica, como no desenho teórico de modelos de dados para geoprocessamento (Davis,1999). Neste trabalho, esta dicotomia é aplicada na representação dos territórios intra-urbanos e de seus inter-relacionamentos.

Os conceitos de campos e objetos em representações de entidades geográficas em ambiente SIG estão associados a distintas formalizações da realidade geográfica. Pode-se dizer que os geo-campos estão associados a uma percepção *contínua* do fenômeno, com cada variável associada a uma camada numérica, estas sobrepostas umas às outras e subdivididas em células elementares. Já o modelo de geo-objetos está associado a construção de conjuntos de entidades *discretas*, identificáveis e

¹ Os termos *territorial cadastral* e *populacional censitário* foram definidos em conjunto com o

individualizáveis com coleções de atributos associados. Mais formalmente, Câmara et al.(1999) define: “um geo-campo como a representação da distribuição espacial de uma variável com valores em todos os pontos pertencentes a uma região geográfica R, num determinado tempo t. Um geo-objeto é um elemento único que pode possuir atributos não-espaciais e está associado a uma localização geográfica exata dentro de uma região.”

Quando se estabelecem estas duas possibilidades de formalização aos geodados intra-urbanos, surgem as limitações inerentes a cada um destes modelos, associadas a condicionamentos lógicos que atuam tanto no nível infológico como datalógico. Nos dados de natureza fisicoterritorial, sua concretude torna a conceitualização em campos e objetos mais direta, e para estes a dicotomia conceitual se apresenta como forte e amplo arcabouço representacional. Nos dados socioeconômicos o alcance deste arcabouço é mais limitado. Geralmente, a aplicação de SIG em modelagem socioeconômica sustenta-se em modelos nascidos de aplicações fisicoterritoriais.

Como regra geral, dados socioeconômicos são disponibilizados como características estatísticas de um determinado grupo de indivíduos habitantes de uma parcela ou zona do território e representados como objetos (bairro, distrito, setor) no nível infológico, ou conjunto de polígonos coloridos (mapas coropléticos) no nível datalógico (Fig.3.4a). O modelo de geo-objetos quando aplicado a dados populacionais intra-urbanos implica numa percepção fragmentada do ambiente. Se aplicadas apenas como suporte para apresentação de dados há uma grande possibilidade de mascaramento de informação significativa. Porém, quando aplicadas técnicas que se valem das características topológicas dos geo-objetos, este modelo permite a aplicação de técnicas matemáticas capazes de revelar padrões espaciais encobertos por essa fragmentação.

As propriedades topológicas dos geo-objetos resumem-se em dois conceitos básicos: contiguidade e conectividade. O conceito de contiguidade ou vizinhança em ambientes intra-urbanos pode ser relacionado à representação de padrões de segregação intra-urbana. A idéia de agregação, de convizinhança, é capaz de trazer expressões representacionais dos territórios intra-urbanos rompendo com a fragmentação inicial do suporte, revelando padrões de associações espaciais inicialmente ocultos nas tabelas. O conceito de conectividade por sua vez, está relacionado basicamente ao fluxo, à noção de rede. O fluxo geralmente é associado à interconexão, pontos conectados dois a dois, um gerador e outro atrator de movimento. Em estruturas intra-urbanas o estudo dos fluxos está relacionado ao estudo dos padrões de movimento que, em última análise estabelecem os inter-relacionamentos presentes no ambiente intra-urbano.

Os mapas coropléticos certamente são os mais difundidos para a representação dos geo-objetos com informações populacionais. Mesmo sendo tão difundida esta forma de representação, Tufte (1994) aponta três inconveniências básicas na representação por mapas coropléticos: “(a) o tamanho das áreas não são uniformes; (b) as áreas preenchidas com cores geralmente são proporcionais à área territorial e não às características representadas, geralmente com áreas despovoadas recebendo grande ênfase visual; (c) as mudanças históricas nas fronteiras políticas impedem a construção de séries estatísticas”. Portanto, para que se possam estabelecer leituras adequadas da complexidade intra-urbana a partir de uma escala de análise coerente a este tipo de representação, o parcelamento do espaço intra-urbano deveria respeitar algumas condições básicas como homogeneidade interna do grupo. Sabe-se, porém, que muitas vezes o projeto de zoneamento está menos atento a estas questões e mais relacionado a questões administrativas ou operacionais de levantamentos. As conseqüências disso serão mais bem discutidas adiante, por ora, o importante é reafirmar a distância que separa os territórios intra-urbanos dos diferentes zoneamentos territoriais aos quais os dados estão associados.

Apesar da larga utilização dos mapas coropléticos na comunicação de dados socioeconômicos, devido às dificuldades expostas anteriormente, tem-se procurado

trabalhar em formas alternativas de representação baseadas em modelos de geo-campos. O trabalho de Martin (1995) é referência nesta linha de pesquisa. Para ele: “a excessiva fragmentação do território no modelo de polígonos impõem limitações na percepção da totalidade do fenômeno. Assim, a compreensão do território não pode estar restrita a este tipo de representação, deve-se complementar esta visão com representações do espaço intra-urbano através de *imagens* e *superfícies*”.

O modelo de superfície, quando aplicado aos dados socioeconômicos intra-urbanos, representa a população de forma contínua. Isto implica numa leitura da realidade onde é mais difícil estabelecer as fronteiras entre os diferentes territórios, as transições são graduais e contém graus relativos de incerteza (Fig.3.4b). De certo modo é apropriado pensar que a população de uma cidade se distribua continuamente, mesmo sabendo-se que aspectos como vias de circulação, zoneamento urbano, ou acidentes geográficos muitas vezes definem fronteiras abruptas entre um território e outro. Ainda assim, a representação por campos numéricos é capaz de transmitir mais adequadamente de que maneira se dá a ocupação do território, possibilitando uma leitura da tendência global do comportamento das variáveis.

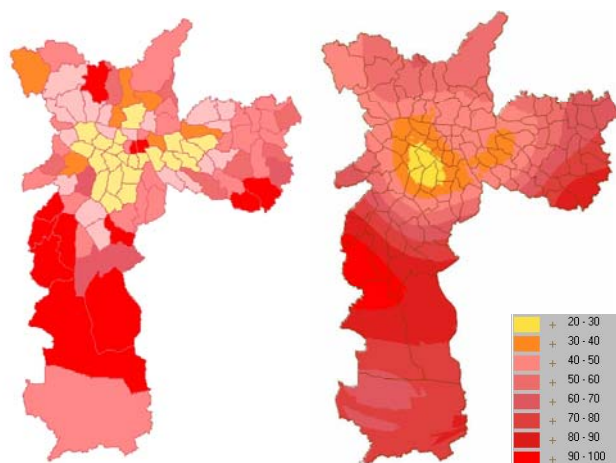


FIGURA 3.4 - (a) À esquerda: Mapa coroplético representando os distritos de Município de São Paulo classificados de acordo com o total de homicídios no ano de 1999. (b) À direita: Superfície interpolada por métodos geoestatísticos para o mesmo conjunto de dados. FONTE: Seade (2000).

Excetuando-se as imagens digitais que são um tipo particular de dado representado como campo numérico, o grande desafio na utilização deste modelo de representação é construir tais superfícies, uma vez que a aquisição de dados em todos os pontos do território nem sempre é possível. Dessa forma, são necessários processamentos para se inferir valores nos pontos onde não se tenha valor conhecido associado. Estes processamentos geralmente estão relacionados a processos determinísticos ou estocásticos de interpolação a partir de pontos amostrais. Alguns exemplos destes processamentos serão apresentados e discutidos no capítulo 7.

3.2.3. *Imagens de sensores remotos*

A particularidade dos dados do tipo imagem está nos processos de aquisição dos dados e nas metodologias de processamento digital de imagens. Segundo Richards (1995): “a principal característica de uma imagem de sensor remoto é a faixa espectral que ela representa. Algumas imagens são medidas da disposição espacial da radiação solar refletida nos comprimentos de onda do ultravioleta, do espectro visível (entre o azul e o vermelho), ao infravermelho próximo ou médio. Outras são mensurações da distribuição espacial da energia emitida pela superfície da Terra (a principal na faixa do *infravermelho termal*); outras ainda, nos comprimentos de onda na faixa das microondas, medidas do retorno relativo da energia emitida pelo veículo.” Esta faixa define a resolução espectral da imagem.

Além da resolução espectral, os sensores remotos podem também ser classificados pela sua resolução temporal e espacial. A resolução temporal pode ser definida a partir do intervalo de revisita do sensor sobre um determinado ponto da superfície terrestre. Já a resolução espacial está associada à geometria de aquisição da imagem, resultado da abertura ou ângulo de visão do sensor e de sua altitude. Até anos recentes, a principal limitação na aplicação de dados de sensores orbitais em estudos intra-urbanos era a baixa resolução espacial, incompatível com a granulosidade desses ambientes. As imagens orbitais comerciais eram capazes de transmitir apenas informações sobre a extensão da ocupação e de alguns aspectos físicos da região (Fig.3.5). Até então, apenas levantamentos aerofotogramétricos e as imagens de

sensores aerotransportados apresentavam a resolução espacial necessária às análises desta natureza. Porém, esses levantamentos dificilmente eram realizados sistematicamente o que acarretava uma baixa resolução temporal. Hoje, já existem satélites comerciais capazes de gerar imagens com resolução espacial muito próxima às dos levantamentos aéreos. Com isso, abre-se a perspectiva de uma oferta crescente de dados tipo imagem associada a possibilidade de coleta sistemática.

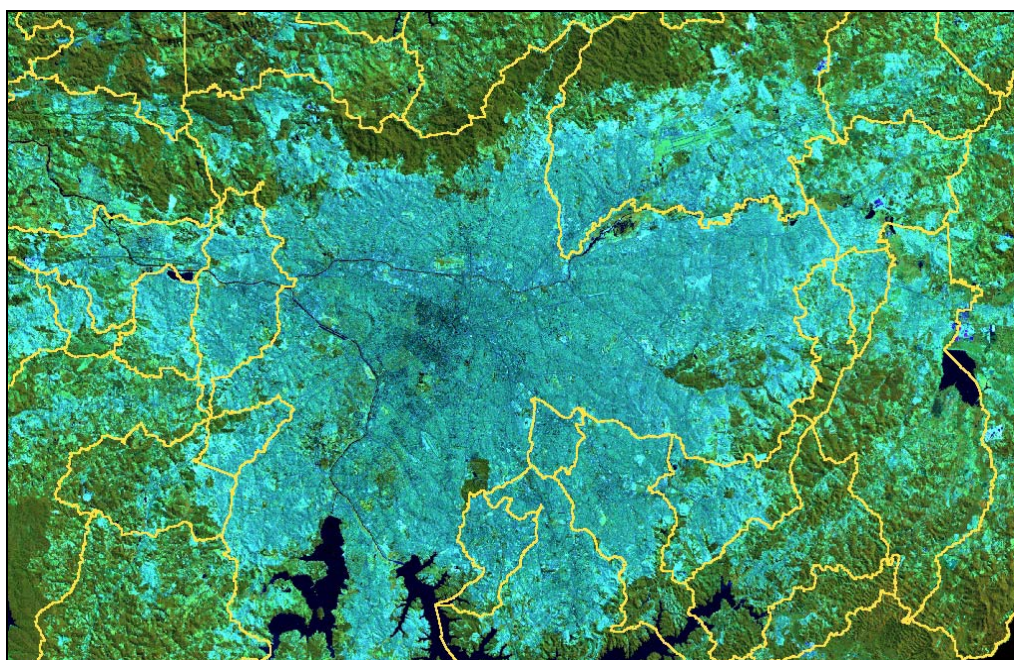


FIGURA.3.5. Divisão Municipal da Região Metropolitana de São Paulo sobrepostos a uma composição colorida TM_Landsat5 bandas espectrais 3, 4 e 5 de 1997.

Está fora do escopo deste trabalho uma discussão mais ampla sobre técnicas de processamento de imagens digitais de sensores remotos. Para isto, há uma vasta literatura que pode ser consultada (Crosta,1992) (Richards,1995) (Schowengert,1997). O interesse específico aqui é no tipo de informação que as imagens podem transmitir sobre os ambientes intra-urbanos. Neste sentido, esta dissertação aposta numa tendência crescente das imagens se tornarem a fonte principal de dados fisicoterritoriais neste tipo de análise. Na medida em que as limitações de resolução temporal e espacial são superadas juntamente com o desenvolvimento de novos equipamentos e sistemas em termos de capacidade de processamento e armazenamento de dados, as informações que se podem extrair das imagens serão cada vez mais ricas e precisas. Uma imagem é

capaz de transmitir com alto grau de naturalidade visual informações sobre o tecido urbano, incluindo os sistemas hídricos e viários, as construções, as áreas de lazer, as áreas de uso especial como aeroportos e aterros sanitários, a vegetação intra-urbana entre várias outras, ou seja, um conjunto muito amplo de temas pode ser visualizado. Quando associada ao conhecimento da realidade *in loco* e a bases auxiliares como topografia e zoneamento urbano, a imagem se torna a fonte básica de informação sobre a morfologia construída do ambiente intra-urbano, constituindo-se como o pano de fundo capaz de referenciar diretamente as análises que porventura se façam ao lugar urbano.

3.3. Modelo de Banco de Dados Geográficos

O domínio ontológico das estruturas intra-urbanas conceitualizado em campos e objetos é, em última instância, formalizado e armazenado em sistemas de arquivos digitais condicionados a regras de lógica computacional de consistência e consulta. Esta outra dimensão presente no nível datalógico interfere diretamente nos níveis superiores de abstração. Agora, pretende-se discutir a formalização e implementação das entidades relacionadas nas seções anteriores e apresentar o modelo de banco de dados geográficos construído para a elaboração deste trabalho.

3.3.1. Vetores e matrizes

Pode-se dizer que um Banco de Dados Geográficos (BDG) é o repositório das geometrias e dos atributos das entidades e fenômenos geográficos. As geometrias são formalizações geométrico-matemáticas dos conceitos de geo-campos e geo-objetos, e por este motivo encontram-se no nível datalógico. Há uma certa confusão de conceitos entre os níveis infológicos e datalógicos. Esta confusão se dá principalmente entre as noções de campos e objetos e representações vetoriais e matriciais. Apesar de haver uma estreita relação entre o modelo de geo-objetos e as representações vetoriais, assim como entre conceitos de geo-campos e as representações matriciais, isto não necessariamente exclui a possibilidade de representação de campos como linhas ou pontos e objetos como matrizes. Citando Schuurman (1999), “ambos campos e objetos

podem ser implementados usando representações matrizes (grade regular) ou vetorial (pontos e linhas). Campos e objetos são conceitualizações (infológico), enquanto matriz e vetor são meios de implementação (datalógico)”.

Sendo o BDG repositório das geometrias, sejam estas matriciais, vetoriais ou ambas, eles devem possuir capacidades de armazenamento, manipulação e visualização desses dados. Estas capacidades estão diretamente vinculadas às arquiteturas dos SIGs. Em Paiva et al.(1998) é apresentado um painel dos atuais projetos de pesquisa que utilizam tecnologias emergentes de BDG e apresenta uma arquitetura de referência baseada em camadas para gerência de objetos geográficos. Em Davis (1999) são apresentados os vários modelos de dados já criados desde o surgimento dos primeiros Sistemas Gerenciadores de Banco de Dados (SGBD). A partir deste panorama, os autores apontam os modelos orientados-a-objetos como uma tendência nos modelos para representação de aplicações geográficas.

É importante ressaltar a distância conceitual entre a noção de objeto nos níveis *infológico* e *datalógico*. No nível infológico, encontram-se as entidades do mundo geográfico vistas como geo-objetos como apresentado anteriormente. No nível datalógico, um objeto é uma representação abstrata dos elementos do universo da aplicação. Uma *classe de objetos* descreve um conjunto de objetos com atributos comuns, mesmo comportamento (operações) e a mesma semântica (Davis,1999). Aqui, entende-se por orientação-a-objeto a associação a classes e subclasses estabelecendo correspondências hierárquicas através de relacionamentos entre essas classes por *generalização* e *agregação* (Davis,1998). Uma generalização é um relacionamento hierárquico entre classes de objetos, assim classes de níveis inferiores generalizam-se em classes superiores, estabelecendo um relacionamento do tipo “*is_a*” (é um). Na especialização, também chamada de “*part_of*” (parte de) um objeto agregado é feito de objetos componentes. A figura 3.6 apresenta um esquema conceitual de modelo de dados geográficos segundo as regras de agregação e especialização (Câmara e Monteiro,2001c).

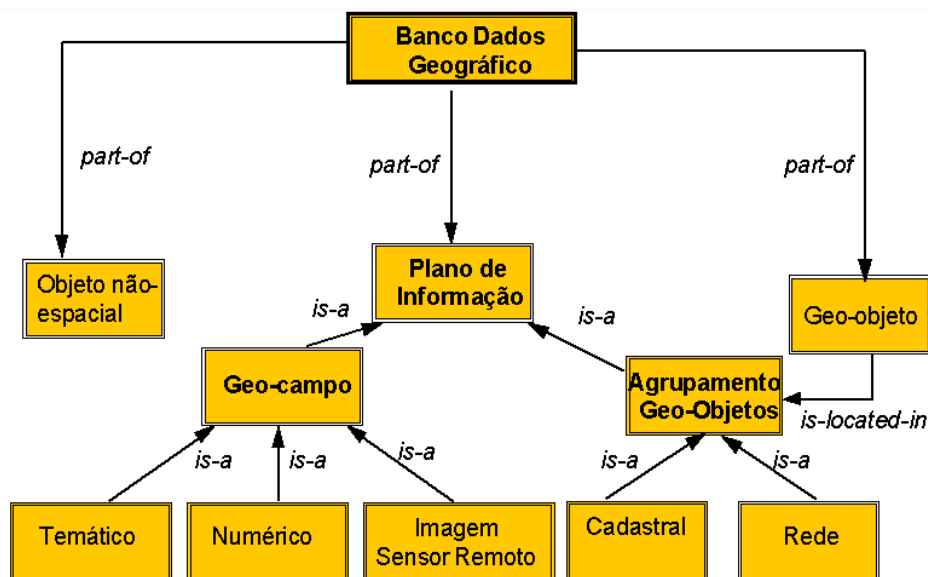


FIGURA.3.6. - Modelo conceitual de banco de dados geográficos conforme Câmara e Monteiro (2001c).

3.3.2. O modelo OMT-G para dados intra-urbanos

O modelo OMT-G foi proposto por Karla Borges (Borges,1997 citado por Davis,1998) como uma extensão do OMT (*Object Modeling Technic*) proposto por Rumbaugh em 1991 (Rumbaugh,1996 citado por Davis,1999). A partir das primitivas do Modelo OMT convencional, foram introduzidas primitivas geográficas, aumentando a capacidade semântica destes. O OMT-G divide as entidades modeladas em duas classes: as georreferenciadas e convencionais. A classes georreferenciadas incluem dois grandes grupos de classes: os geo-campos e os geo-objetos, representando de maneira integrada os três grandes grupos de fenômenos modelados em geoprocessamento: os de variação contínua ou geo-campos, os de variação discreta ou geo-objetos e os não espaciais.

O modelo OMT-G aplica uma notação gráfica específica para a representação das classes e dos objetos geográficos. Segundo Davis (1999), “no modelo OMT-G, uma *Classe Georreferenciada* é representada graficamente por um retângulo, subdividido em quatro partes. A parte superior contém à direita o nome da classe e à esquerda o símbolo representando a forma gráfica da *Classe Georreferenciada*. Na segunda parte, aparece a lista dos atributos gráficos. Na terceira parte a lista dos atributos alfanuméricos (quando

existirem) e, na última parte, a lista das operações que são aplicadas à classe.” A coleção de pictogramas e símbolos utilizadas por este modelo pode ser consultada no Anexo I.

Além de seguir o paradigma de orientação-a-objetos suportando conceitos de classe e herança, em modelagem de dados intra-urbanos o modelo OMT-G apresenta algumas vantagens. A principal delas é não utilizar o conceito de camadas e sim o de nível de informação (temas), não limitando o aparecimento de uma classe geográfica em apenas um nível de informação. Assim, um mesmo objeto pode ser representado como pontos, matrizes, ou polígonos. Além disso, permite a representação direta de hierarquias espaciais de zoneamento do tipo um estado é composto por municípios, estes são formados pelo agrupamento de bairros, situação comum em dados intra-urbanos. De fato, o modelo OMT-G nasceu das necessidades da PRODABEL/Belo Horizonte, equipe da prefeitura responsável pelo desenvolvimento de sistemas georreferenciados para os dados intra-urbanos.

3.4. Conclusões

Neste capítulo, foram apresentadas as bases conceituais do processo de modelagem de dados para geodados de natureza intra-urbana. Foram expostas as etapas de construção de abstrações necessárias a este processo a partir da transição dos níveis infológicos ao datalógicos e as condicionantes que cada uma desta etapa impõem às análises do conjunto de dados disponível. A partir do próximo capítulo será apresentado um estudo de caso aplicado à Região Metropolitana de São Paulo. Serão exemplificados os processos de modelagem e análises dos geodados intra-urbanos desde sua coleta, compatibilização até a aplicação de técnicas de análise espacial.

CAPÍTULO 4

ESTUDO DE CASO: SÃO PAULO EM GEODADOS

Neste capítulo, descrevem-se as etapas de pesquisa e coleta de dados digitais e o projeto do banco de dados geográficos para o estudo de caso apresentado neste trabalho. Todas as análises realizadas foram aplicadas para a Região Metropolitana de São Paulo (RMSP) e sobre um conjunto de dados provenientes do banco de dados do Mapa da Exclusão/Inclusão (E/I) da Cidade de São Paulo de 96 e 2000 (Sposati,1996)(Sposati,2000) e das Pesquisas Origem/Destino (OD) de 1987 e de 1997 da Empresa Metropolitana de São Paulo - METRÔ. Algumas imagens digitais de sensores remotos orbitais Landsat5-TM, Landsat7-TM e SPOT-HRV fornecidas pelo INPE também foram incorporadas ao banco de dados.

4.1. Dados: Fontes e Formatos

Como regra geral, há uma escassez de dados intra-urbanos digitais no Brasil. Em muitos casos, o Instituto Brasileiro de Geografia e Estatística (IBGE) é a única fonte de dados disponível. Ainda que muito distante de uma situação ideal, São Paulo foge a regra graças à pesquisa OD realizada a cada dez anos pelo METRÔ. Porém, mesmo quando da existência de fontes alternativas e complementares ao IBGE, a tarefa de pesquisar e acessar os dados necessários às análises urbanísticas em SIG é geralmente custosa. Há ainda um longo caminho institucional por percorrer para que se estabeleça uma sistematização e padronização nos procedimentos de coleta e difusão dos dados por parte de seus produtores. A penetração do geoprocessamento e a evolução de suas técnicas dentro dos diversos campos de conhecimento deve estimular o debate sobre a propriedade dos dados. É importante que haja uma ampla democratização dos dados para que se possa discutir e planejar as intervenções públicas ou privadas em seu ambiente da forma mais democrática e transparente possível.

4.1.1. Pesquisa origem e destino do Metrô

A primeira pesquisa OD de São Paulo, realizada em 1967, nasceu da necessidade de obtenção de dados dos padrões de mobilidade da população para os estudos e projetos da rede básica do Metrô da cidade. A pesquisa OD procura identificar e quantificar as viagens diárias da população, conforme o motivo da viagem e o modo de transporte utilizado. Além dos dados relativos a mobilidade, a pesquisa também levanta dados sobre as características socioeconômicas da população, procurando captar desta forma as necessidades de viagens cotidianas dos habitantes da cidade (METRÔ,1997). A abrangência da pesquisa e a competência com que é realizada torna a Pesquisa OD do Metrô de São Paulo uma importante fonte de dados, já utilizada em vários estudos sobre a metrópole paulistana durante estas quatro décadas.

Para a realização das análises apresentadas nesta dissertação, foram acessados os dados OD relativos aos anos de 1987 e 1997. Os dados foram fornecidos diretamente pelo METRÔ e são apresentados em tabelas associadas às subdivisões geográficas conhecidas como zonas OD.

Pode-se dizer que a pesquisa OD como ela é realizada em São Paulo é uma fonte direta de dados socioeconômicos *populacionais censitários* e indireta de *territoriais cadastrais*. No primeiro caso, porque são levantadas informações socioeconômicas como a escolaridade, a renda, faixas etárias sobre a população residente por domicílios. Já os dados de categoria territorial cadastral podem ser estimados pelas informações sobre as necessidades de viagem que procuram classificar as viagens intra-urbanas por seu motivo (e.g. localização dos empregos).

Os dados OD são coletados através de pesquisa amostral domiciliar e abrangem os 39 municípios da Região Metropolitana de São Paulo (RMSP). Na pesquisa de 1997 cerca de 30.000 domicílios, distribuídos em 389 zonas OD97 (Fig.4.1d), foram levantados. O Município de São Paulo compõe-se de 270 dessas zonas. As 389 zonas OD97 foram projetadas de forma a respeitar algumas hierarquias superiores de áreas, dessa forma é possível apresentar os dados em bases distritais e municipais a partir da

agregação de zonas OD97. Devido às alterações na subdivisão territorial para adequação a divisão distrital, perdeu-se a seqüência histórica entre 87 e 97. A pesquisa OD87, por sua vez, é apresentada em 254 zonas OD87 (Fig.4.1c), em 1987 a população da região metropolitana foi estimada em aproximadamente 14,3 milhões de habitantes. Em 1997 a população total estimada foi de 16,8 milhões de habitantes. As zonas OD87 não podem recompor as divisões distritais, apenas respeitam as fronteiras municipais.

Os dados tabulares da pesquisa OD foram fornecidos em planilhas formato Microsoft Excel. A base territorial georreferenciada para as zonas OD87 e zonas OD97 fornecidas em formato *shape_file* (.shp) do ArcView.3.1. As tabelas foram importadas para Microsoft Access onde foi estabelecida como chave primária o código da zonas OD (COD_ZON). A geometria das bases territoriais poligonais OD87 e OD97 e seus identificadores foram convertidas para o formato ASCII/SPRING pelo aplicativo *shp2spr* disponível no SPRING3.5.

4.1.2. Mapa da exclusão/inclusão social da cidade de São Paulo

O mapa da E/I social da cidade de São Paulo, diferentemente da pesquisa OD do METRÔ, não produz dados primários. No mapa, as variáveis apresentadas são coletadas de fontes primárias, sendo a principal o IBGE. Esses dados depois de coletados são integrados a um BDG e então manipulados para a construção de indicadores compostos como o índice de discrepância (Idi) e o índice de exclusão social (Iex). Este trabalho coordenado por Sposati (1996) tornou-se referência em análises sociais intra-urbana por trazer um embasamento crítico inovador quanto a utilização de indicadores sociais quantitativos. Também quanto à maneira de comunicar estes dados, adotando a linguagem cartográfica através de mapas digitais em SIG. Em 2000, algumas variáveis do mapa foram atualizadas (Sposati,2000) com os dados da contagem de 1996 do IBGE e com alguns dados da pesquisa OD de 1997.

No mapa da E/I social as informações *populacionais censitárias*, predominantemente do IBGE, também são complementadas por informações *territoriais cadastrais* levantadas principalmente em cadastros de departamentos

governamentais. A tabela com as variáveis primárias e compostas do mapa da E/I social de 1996 e 2000 e das pesquisas OD 87 e 97 pode ser consultada no Anexo II.

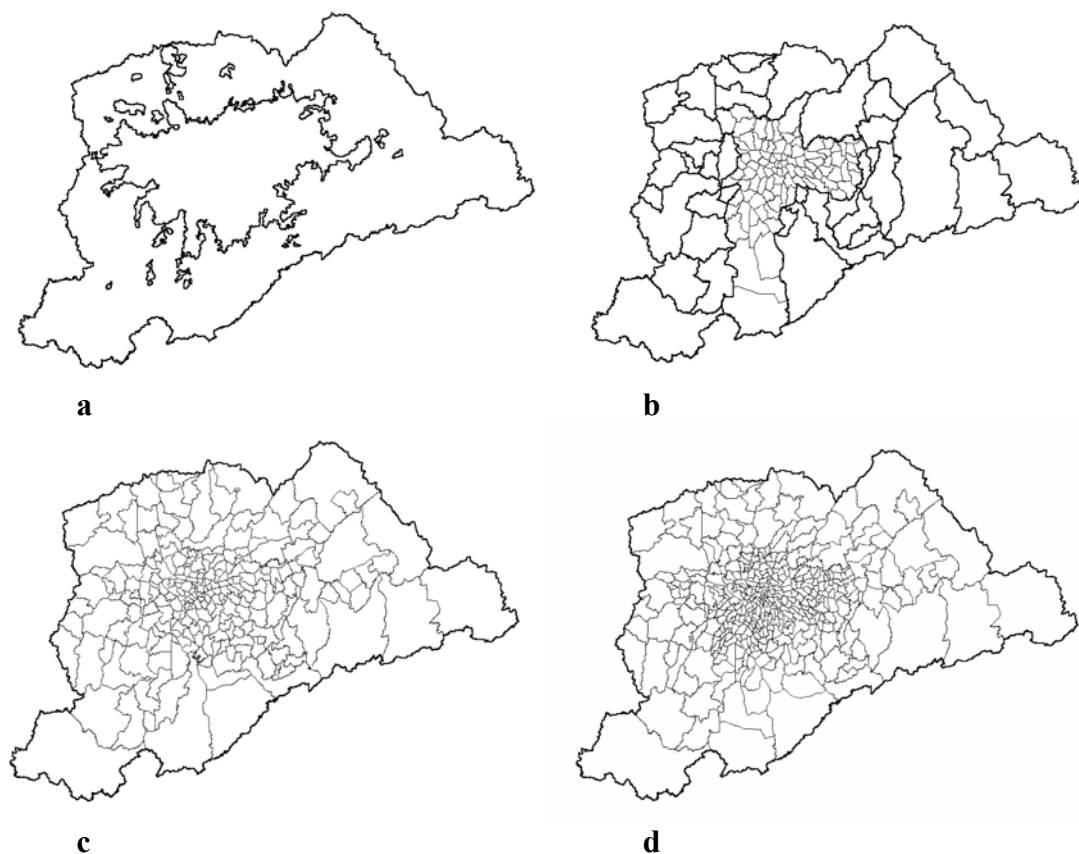


FIGURA.4.1. Bases territoriais dos dados acessados: (a)mancha urbana extraída da imagem Landsat5, (b)Distritos de SP e municípios da RMSP, (c)O/D87 e (d)O/D97.

A base territorial do mapa da E/I social da cidade de São Paulo é a divisão distrital. Ele não abrange toda a região metropolitana, apenas os 96 distritos do município de São Paulo que foram definidos por lei municipal em 1991. Por este motivo, em algumas análises aqui aplicadas foram incorporadas informações a nível municipal dos 38 municípios provenientes da sumário de dados da grande São Paulo de 1999 da Emplasa (Empresa Metropolitana de Planejamento da Grande São Paulo S.A.) (Emplasa,1999) para recompor com as informações distritais do município de São Paulo a RMSP (Fig.4.1b).

Os dados tabulares do mapa da E/I social foram fornecidos em planilhas formato Microsoft Excel. A base distrital georreferenciada fornecidas em formato *mif* do

Mapinfo. Estes dados também foram convertidos para formato SPRING e incorporados ao BDG através do aplicativo *mif2spr*.

Vale ressaltar que apesar da incorporação de todas as variáveis listadas, tanto da pesquisa OD quanto do mapa da E/I social, no BDG consolidado sobre a metrópole paulistana, nem todas foram utilizadas nos experimentos apresentados nos próximos capítulos. A incorporação de todas as variáveis teve como objetivo disponibilizar uma base georreferenciada a mais completa possível para trabalhos futuros.

4.1.3. *Imagens digitais de sensores remotos orbitais*

As imagens digitais de satélite foram as únicas informações de natureza *fisicoterritorial* incorporadas ao BDG. As imagens acessadas foram geradas pelos sensores SPOT-HRV em setembro de 1994, Landsat5-TM em agosto de 1997 e Landsat7-ETM+ em setembro de 1999. As imagens de maior resolução espacial, as pancromáticas SPOT e Landsat7, foram utilizadas para visualizações da textura intra-urbana, dos sistemas principais de transporte de superfície e localização de grandes feições construídas (marcos visuais). A imagem multispectral Landsat5 foi utilizada para visualização de massas de vegetação, sistema hídrico principal e extração da mancha urbana (Fig.4.1a). A tabela 4.1 apresenta o resumo das especificações das imagens utilizadas.

TABELA.4.1. Especificações das imagens incorporadas ao BDG.

SENSOR	SPOT-HRV	Landsat5-TM	Landsat7-ETM+
bandas espectrais (comprimento de onda)	pancromática (0,45µm-0,90µm)	3 (0,63µm-0,69µm) 4 (0,76µm-0,90µm) 5 (1,55µm-1,75µm)	pancromática (0,05µm-0,90µm)
data de aquisição	08/1994	08/1997	09/1999
Resolução espacial	10 metros	30 metros	15 metros

4.2. O Projeto do BDG

A etapa subsequente à pesquisa e coleta de dados é o projeto do banco de dados. Esta seção apresenta o modelo de dados desenhado para o estudo de caso de São Paulo utilizando o modelo OMT-G apresentado na última seção do capítulo anterior.

No alto do modelo está a superclasse RMSP. Ela é representada por coleções de geo-campos e geo-objetos que se especializam em categorias do Modelo de Dados do SPRING. Os dados de entrada são de duas categorias: *imagem* (geo-campo) e *cadastral* (geo-objetos). Dessa forma o primeiro degrau na escala hierárquica do modelo estabelece as subclasses imagem e cadastral (Fig.4.2).

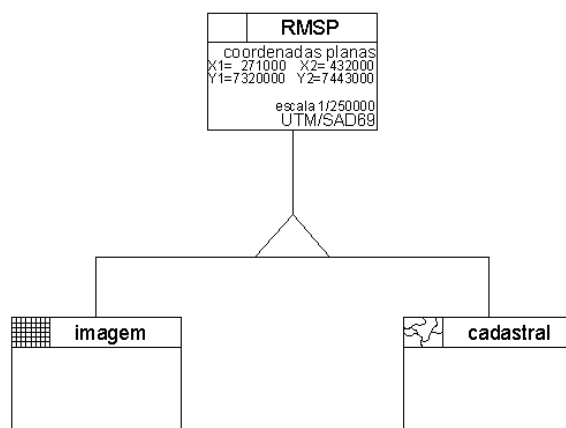


FIGURA.4.2. Primeiro nível hierárquico no modelo OMT-G. Superclasse RMSP especializa-se em subclasses imagem e cadastral.

A classe imagem especializa-se nas três imagens apresentadas anteriormente (Fig.4.3). A classe cadastral especializa-se nas diferentes subdivisões territoriais e organiza-se hierarquicamente conforme as possibilidades de agregação das unidades em divisões superiores. Assim, a classe cadastral RMSP é subdividida em municípios, estes são subdivididos em zonas OD87 e em zonas OD97. Analogamente, a classe município contém o MSP e este é subdividido em seus 96 distritos que também se subdividem em 270 zonas OD97 (Fig.4.4).

O modelo OMT-G aqui apresentado até o segundo nível hierárquico refere-se a modelagem dos dados de entrada. Todas as operações e processamentos que foram

feitos a partir destes dados geraram novos dados que estão em níveis hierárquicos inferiores e por vezes em diferentes categorias formais. O desenho final do esquema OMT-G completo representando todos os dados utilizados durante o desenvolvimento deste trabalho pode ser consultado no Anexo III desta dissertação.

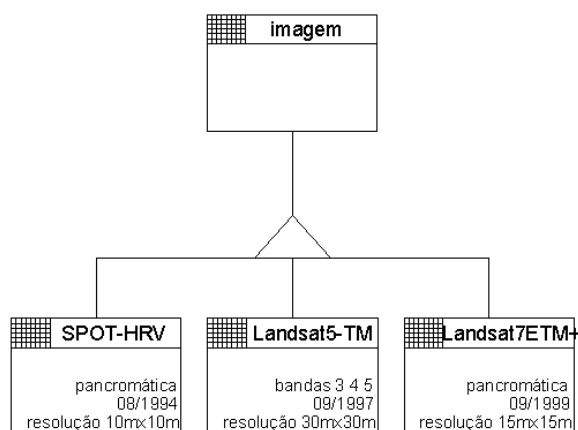


FIGURA.4.3 - Segundo nível hierárquico no modelo OMT-G, categoria imagem.

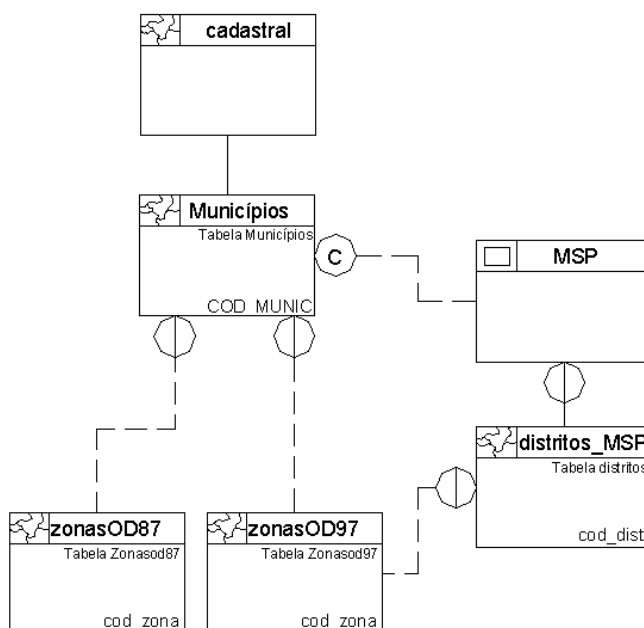


FIGURA.4.4. Segundo nível hierárquico no modelo OMT-G, categoria cadastral.

CAPÍTULO 5

ANÁLISE ESPACIAL I: O PROBLEMA DAS UNIDADES DE ÁREA MODIFICÁVEIS

5.1. Contextualização

Grande parte dos geodados socioeconômicos, mesmo tratando-se de informações essencialmente relativas a indivíduos, são disponibilizados como características estatísticas de grupos de indivíduos associados a parcelas do território na qual eles habitam. Isso porque muitas vezes é imperativo que se preserve a privacidade das pessoas ou empresas que fornecem estes dados ao responderem às pesquisas e recenseamentos. Ocorre que os resultados de qualquer análise quantitativa que nestes se aplique estarão diretamente influenciados pela definição desses grupos, ou espacialmente condicionados pela definição das fronteiras das subdivisões territoriais a que estes grupos de indivíduos estiverem associados. Convencionou-se chamar esta interferência de *Problema das Unidades de Área Modificáveis* (MAUP). Desde os primeiros trabalhos em geografia quantitativa no início do século passado já se reconhecia a influência do MAUP sobre as análises espaciais (Gehlke e Biehl, 1934 citado por Openshaw, 1984) (Robinson, 1950 citado por Openshaw, 1984). Porém, nas últimas décadas, com a difusão do geoprocessamento, retomou-se a discussão sobre os efeitos do MAUP nas análises geográficas quantitativas aplicadas aos geodados associados a subdivisões territoriais.

Segundo Wrigley et al. (1996), o MAUP é composto de efeitos de *zoneamento e escala*. Ele os define como: “O *efeito de escala* é a tendência, dentro de um sistema de unidades de áreas modificáveis, de se obter diferentes resultados estatísticos para um mesmo conjunto de dados quando a informação é agrupada em diferentes níveis de resolução espacial (e.g. setores censitários, distritos, municípios). O *efeito de zoneamento* é a variabilidade dos resultados estatísticos obtida dentro de um conjunto de unidades de áreas modificáveis em função das várias possibilidades de agrupamentos em uma *dada escala*, e não em função da variação do tamanho dessas áreas - i.e. a

diferença nos resultados devido a simples alteração das fronteiras ou configurações das zonas dentro de uma mesma escala de análise”.

A significância do MAUP, enquanto efeito endêmico a dados censitários associados a áreas, deve ser considerada frente ao propósito do estudo. É necessário que se estabeleça o objetivo das inferências, se são estas relativas às áreas ou relativas aos indivíduos que ali vivem. Caso o objetivo seja inferências às características dos indivíduos que ali habitam, não haverá garantias teóricas de que os resultados obtidos sejam boas estimativas dessas características individuais (Openshaw,1984). Convencionou-se chamar este efeito de *falácia ecológica*. Para Wrigley et al.(1996), “a falácia ecológica envolve a inferência inadequada de relações em nível de indivíduo a partir de resultados obtidos em nível de áreas ou grupos. Devido aos efeitos de escala e zoneamento do MAUP, as relações medidas em nível de unidades de áreas a partir de coeficientes de correlação tendem, em geral, a apresentar valores absolutos maiores que as correlações desconhecidas em nível de indivíduos.” A questão que se coloca então é como tratá-lo dentro das análises socioeconômicas intra-urbanas.

5.2. Abordagens ao MAUP

O primeiro ponto a destacar é que o processo de agregação dos dados individuais, seja qual for a resolução espacial de comunicação destes, impõem inevitavelmente os efeitos do MAUP, assim, este é inerente a todas as variáveis associadas a divisões territoriais e nunca poderá ser totalmente removido. Todos os trabalhos pesquisados sobre o tema preocupam-se em quantificar a interferência desses agrupamentos para estabelecer, a partir daí, técnicas capazes de minimizar os efeitos do MAUP no conjunto de variáveis disponíveis.

Frente a este desafio, duas abordagens distintas podem ser encontradas na literatura: (a) uma reconhece o problema como intrinsecamente relacionado ao suporte zonal e objetiva a formulação de critérios ótimos de projetos de zoneamento territorial (Openshaw,1978) (Openshaw,1996) (Martin,1995) (Martin,2000). Destes os esforços

teóricos resultaram as iniciativas inéditas do Censo 2000 do Reino Unido, onde já é realidade o acesso a dados em suportes territoriais gerados interativamente de acordo com a finalidade do estudo, as *Output Areas* (OAs) (Martin,2000). As OAs são subdivisões territoriais geradas segundo critérios de forma, homogeneidade e tamanho da população. (b) A outra abordagem procura o desenvolvimento de operadores algébricos para se estimar e reduzir as instabilidades que o MAUP impõem ao conjunto de variáveis dentro do suporte zonal original. Neste sentido, destacam-se os esforços teóricos de Openshaw (1984), Wrigley et al. (1996), Green e Flowerdew (1996) e Bailey e Gattrel (1995) que se concentraram no desenvolvimento de técnicas estatísticas para este fim. As próximas seções deste capítulo tratarão de apresentar um estudo de caso aplicado sobre os dados de São Paulo sob cada uma destas abordagens.

5.3. O PROJETO DO ZONEAMENTO TERRITORIAL COMO ANÁLISE ESPACIAL

5.3.1. Fundamentação teórica

A idéia de que um projeto de uma determinada subdivisão territorial observe critérios ótimos em relação à consistência estatística dos dados e assim, se configure como uma ferramenta teórica de análise espacial, vem sendo continuamente defendida por Openshaw em sua produção acadêmica. Em 1996, ele reafirma esta necessidade e fundamenta: “Apenas quando as zonas forem completamente homogêneas em relação ao micro-dado que representam, não haverá possibilidade de erros de representação espacial. Na prática, a heterogeneidade do padrão do micro-dado interage com as fronteiras e com o tamanho destas zonas gerando todas as formas de complexidade.” No mesmo sentido, Wrigley et al. (1996) diz: “Os efeitos de escala e zoneamento, e os associados problemas da falácia ecológica, emergem em parte devido ao que se pode referir como *homogeneidade da área* ou *efeitos de agrupamento* – ao fato de áreas geográficas serem construídas por agrupamentos aleatórios de indivíduos/domicílios, porém estes tendem a ser mais semelhantes em áreas próximas do que em distantes.” O diagrama apresentado a seguir (Fig.5.1) explicita a lógica desta interferência no valor do coeficiente de correlação.

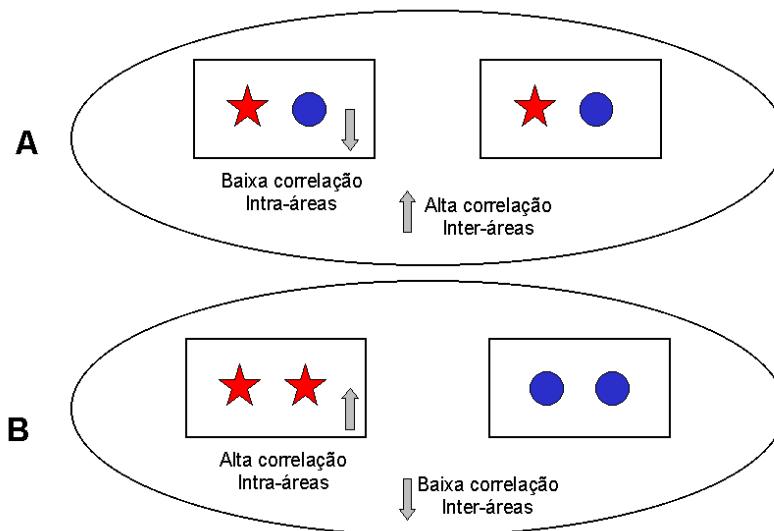


FIGURA.5.1. Esquema ilustrativo dos efeitos das diferentes possibilidades de agrupamento de um mesmo grupo de indivíduos nos coeficientes de correlação intra-áreas (desconhecido) e inter-áreas (conhecido). Na situação **A**, a heterogeneidade interna acarreta uma alta correlação externa. Na situação **B**, o efeito se inverte. Através deste esquema, compreende-se a interferência da *homogeneidade de área* nas análises estatísticas sobre uma mesma população.

Dentro da complexidade espacial do espaço intra-urbano e da escala de observação necessária ao seu estudo, a questão sobre a homogeneidade intra-área poucas vezes se resume a uma lógica espacial simples, tema já discutido anteriormente no capítulo 2 deste trabalho. Nas cidades é comum que rios, avenidas, pontes, córregos e muros definam fronteiras territoriais. Claras e evidentes por vezes, sutis e nebulosas em outras, as fronteiras intra-urbanas existem e estão diretamente relacionadas à configuração espacial do assentamento das diferentes atividades e grupos populacionais no território. No caso das metrópoles brasileiras, e especificamente em São Paulo, o significado dessas fronteiras está fortemente relacionado ao padrão socialmente segregatório inerente ao processo histórico de urbanização destas cidades (Villaça,1998).

Dentro deste universo teórico, refletir sobre o que diz Openshaw sobre critérios de zoneamento, remete a questões urbanísticas fundamentais sobre o significado das fronteiras no conceito de estrutura intra-urbana. Para Openshaw (1996): “Um sistema de zonas é uma forma de descrição muito simples. As bordas da zona devem conter em si significado substantivo, como, por exemplo, o lugar onde elas representem fronteiras ou

transições segundo algum padrão específico. As áreas delimitadas devem, também, conter significado geográfico. Por exemplo, unidades de espaço devem possuir algum propósito descritivo, seja devido a um critério de similaridade (e.g. totais populacionais iguais) ou por serem dominadas por diferentes características (e.g. tipo de regiões).” Refletindo sobre estas afirmações no contexto dos estudos de estruturas intra-urbanas, pode-se dizer que o significado das fronteiras está estreitamente relacionado a configuração espacial dos territórios intra-urbanos e toda a complexidade dimensional que nestes está embutida. Portanto, tratar com dados associados a subdivisões zonais neste tipo de análise deve permitir a maior flexibilidade possível na construção destes limites, já que segundo o ponto de vista, diferentes territórios podem se configurar e não necessariamente sob o mesmo arranjo espacial. Esta flexibilidade só é atingida quando são acessados os dados a nível mais desagregado possível, possibilitando a construção interativa destas fronteiras em função da análise ou problema que se queira observar.

Em Openshaw (1978), o autor apresenta uma série de avaliações sobre os efeitos do MAUP em um conjunto de dados a partir de mensurações sob diferentes escalas de agregações. Na segunda parte do artigo, é apresentado o que ele chama de critérios ótimos de zoneamento. Para ele, um sistema de zonas deve possuir um certo grau de uniformidade em critérios como igual população, tamanho, densidade, compacidade e homogeneidade. Baseando-se nestes critérios, ele apresenta a formulação de uma série de constritores que operaram sobre um conjunto de geodados zonais desagregados, gerando assim agregações controladas. Basicamente estes algoritmos buscam minimizar a soma dos desvios absolutos em relação à média da variável adotada como critério em todas as novas zonas definidas ou reagrupadas. Esta variável pode ser a população total, a densidade ou a área destas zonas. Com relação ao parâmetro homogeneidade, os algoritmos buscam a maximização da correlação intra-áreas (Martin,2000). Todo este esforço metodológico está consolidado no AZP (*Automatic Zoning Procedure*), uma ferramenta computacional desenvolvida por Openshaw em 1977 e que hoje vem sendo aplicada para a geração das OAs do Censo do Reino Unido.

5.3.2. Experimento

Já foi dito que o MAUP é inerente aos geodados associados a subdivisões zonais e que, portanto, não pode ser removido. Frente a este fato, o primeiro procedimento que se pode adotar quando se acessam dados desta natureza é analisar a subdivisão territorial em si, antes mesmo de qualquer manipulação com variáveis. A partir dos critérios já expostos na seção anterior, pode-se avaliar a representatividade destas subdivisões. Um procedimento simples e que pode ser útil neste ponto é construir um gráfico cruzando a dimensão territorial das zonas com a população que ali reside.

Nas zonas OD97 torna-se evidente a divisão em dois regimes representacionais diversos: as zonas com grandes extensões territoriais e pouca população (áreas rurais) e pequenas zonas com alta densidade populacional (área urbana) (Fig.5.2). No caso do conjunto das zonas OD97, a necessidade de recobrimento de toda a região metropolitana geopolítica, que vai além da área urbanizada, abrangendo também áreas rurais, a relação entre o tamanho da zona e sua população total é inversa. Na figura 5.3 apresenta-se a mancha urbana extraída de imagem TM de 1997 sobreposta às zonas classificadas por sua população total, nesta também pode-se perceber este efeito.

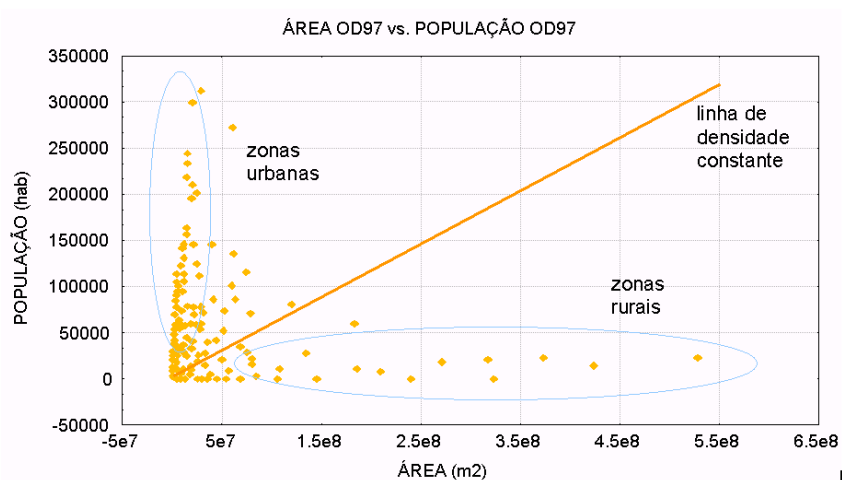


FIGURA.5.2 - Gráfico de espalhamento das 389 zonas OD97 cruzando as populações totais e o tamanho das zonas.

É certo que muitas vezes em geodados zonais urbanos apresentem-se as áreas rurais que envolvem a área urbanizada e que certamente respondem a padrões espaciais

não intra-urbanos. Nestas zonas poderia ser aplicado um processo de “filtragem” a partir de análises complementares com auxílio de imagens de satélite capazes de identificá-las, podendo-se aplicar outros critérios de análise. Porém, dentro da mancha urbana contínua, os critérios de zoneamento deveriam ser aplicados a fim de que se possam controlar os efeitos do MAUP neste ambiente, o qual na realidade, representa o verdadeiro objeto das análises intra-urbanas.

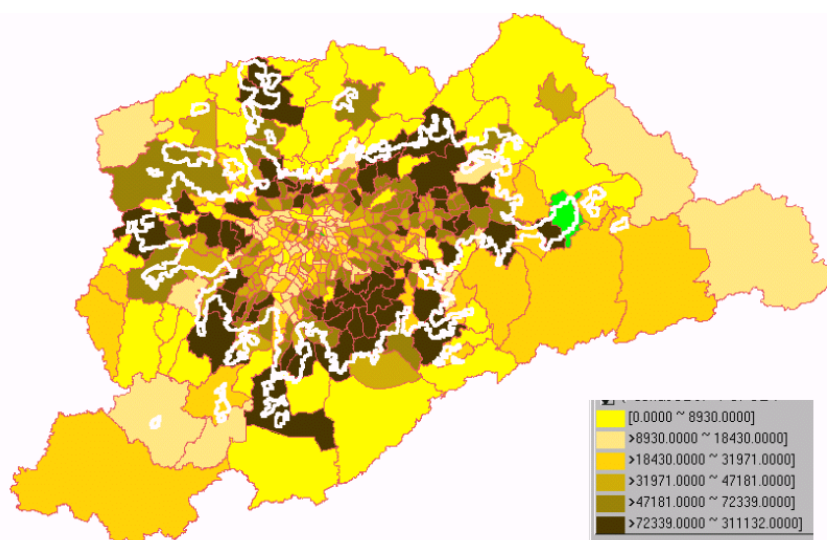


FIGURA.5.3 - As 389 zonas OD97 classificadas de acordo com sua população total e a mancha urbana extraída de imagem TM de 1997 sobreposta.

5.3.2.1. *Quantificação do efeito de escala*

A fim de ilustrar o MAUP dentro do estudo de caso apresentado neste trabalho, foram comparados os coeficientes de correlação entre três variáveis da Pesquisa OD97 em nível de zonas OD97 e de distritos do município de São Paulo. Assim, o experimento permite avaliar o efeito de escala subjacente ao agrupamento de zonas OD97 em distritos. A escolha da base distrital do município de São Paulo para a comparação se deu em função desta ser uma divisão geopolítica consolidada dentro do município e muito utilizada em diversos estudos para a apresentação de dados (e.g. mapa da E/I social). As variáveis selecionadas e aplicadas no estudo foram: a) população com 60 anos ou mais sobre a população total; b) Total de habitantes não

alfabetizados sobre a população total; c) renda individual per capita. Os mapas coropléticos são apresentados na figura.5.4.

A análise visual dos mapas coropléticos permite identificar o *efeito de escala* no agrupamento das 270 zonas OD97 nos 96 distritos do município de São Paulo. Nota-se uma significativa perda de informação quando se agregam os polígonos menores em unidades de área maiores. Em ambientes intra-urbanos, caracterizados por alta complexidade espacial, essa perda de informação interfere substancialmente na identificação e análise dos diversos territórios intra-urbanos e seus inter-relacionamentos. Nestes ambientes, a granulosidade da representação espacial é importante para que se possa ter uma leitura mais fiel do mosaico territorial que o compõe.

Mesmo as zonas OD97 apresentando uma resolução espacial mais fina que os distritos, dessa forma se configurando como subdivisão mais adequada à análise do que a distrital, é importante realçar que a representatividade populacional destas zonas dentro da área urbanizada é de 30 mil habitantes em média, o que certamente ainda oculta uma complexidade intra-zona OD97. Porém, neste caso, devido ao fato da pesquisa OD ser amostral e não um levantamento universal, sua confiabilidade estatística está associada a esta representatividade populacional, tornando-se este o fator chave definidor da resolução espacial dos dados. Como nem sempre é possível o acesso aos dados com a maior resolução espacial possível, como os setores censitários, a falácia ecológica surge então como um problema comum em análises urbanísticas que visam a formulação políticas públicas claramente relacionadas aos problemas individuais de sua população. Muitas dessas análises focalizam estudos descritivos, nos quais é comum a confusão entre as características das áreas com as características da população que ali vive.

A comparação das matrizes de correlação (Tab.5.1) entre as três variáveis selecionadas para os dois conjuntos de unidades de área revela o grau de interferência do MAUP nas análises quantitativas aplicadas sobre estes dois conjuntos de unidades de

área. No exemplo, foram utilizadas três variáveis comumente utilizadas a caracterizações de indivíduos, exemplos de variáveis populacionais censitárias utilizadas para formulação de planos de intervenção e políticas públicas. Tomando-se essas variáveis, “não alfabetizados”, “renda individual per capita” e “população com 60 anos ou mais”, poder-se-iam realizar afirmações sobre a inter-relação delas na população do município de São Paulo. Porém, como já foi colocado, estas afirmações estão necessariamente relacionadas à partição territorial do município, e isoladas deste contexto se configurariam como afirmações errôneas sobre estas inter-relações. Assim, não se pode afirmar que a correlação entre a renda per capita da população e o número de pessoas não alfabetizadas em São Paulo seja de $-0,59$ ou $-0,78$. Na realidade, não se pode inferir qual é a verdadeira correlação em nível de indivíduos entre estas variáveis a partir desses dados pois o processo de agrupamento introduz irreversivelmente os efeitos do MAUP em qualquer sistema de unidades de área. Neste caso, somente pode-se dizer que para os distritos do município esta correlação é de $-0,78$ e para as zonas OD97 de $-0,59$. O resultado corrobora as teorias sobre o efeito de escala, ou seja, quanto menor a resolução espacial das subdivisões espaciais maior será o coeficiente de correlação medido entre as variáveis a estas associadas.

Variável **A** = População com 60 anos ou mais / população total
 Variável **B** = Total de habitantes não alfabetizados / população total
 Variável **C** = Renda individual per capita (em reais)

Distritos do município de São Paulo (n=96) Zonas OD97 do município de São Paulo (n=270)

TABELA.5.1 - Matrizes de correlação para as variáveis selecionadas para as zonas OD97 e os distritos do município de São Paulo (Fonte: pesquisa OD97 METRÔ).

	A	B	C
A	1	-0,81	0,65
B	-0,81	1	-0,78
C	0,65	-0,78	1

	A	B	C
A	1	-0,53	0,53
B	-0,53	1	-0,59
C	0,53	-0,59	1

270 zonasOD97 do município de São Paulo



96 distritos do município de São Paulo

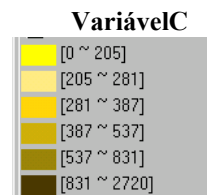
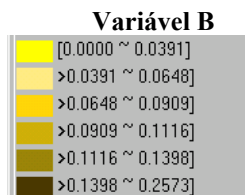
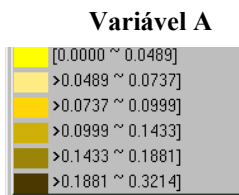
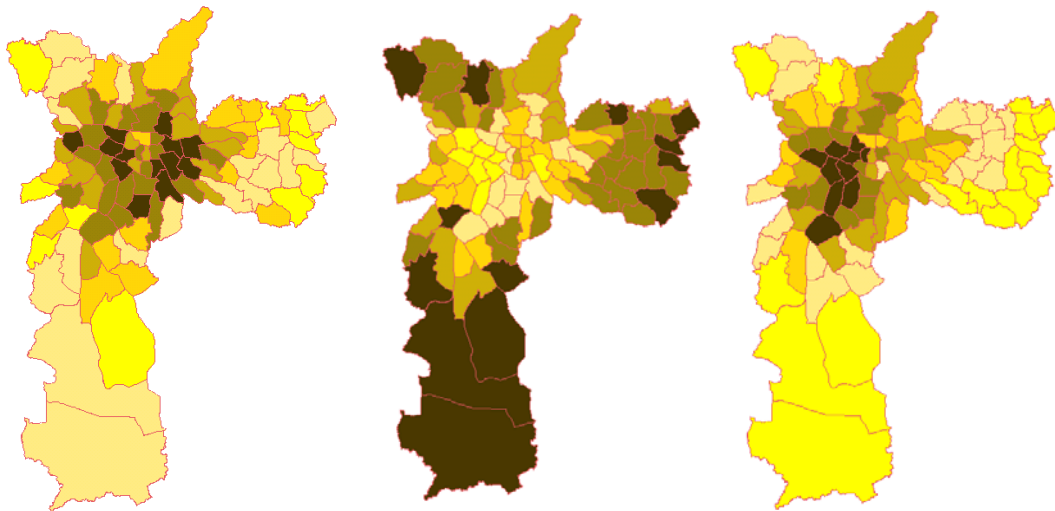


FIGURA.5.4 - Acima: as 270 zonas OD97 do município de São Paulo agrupadas pelas variáveis selecionadas. Abaixo: os 96 distritos do município de São Paulo classificados segundo os mesmos grupos aplicados nas zonas OD97 (Fonte: pesquisa OD97 METRÔ).

5.3.2.2. *Quantificação do efeito de zoneamento*

A segunda parte do experimento está interessada em analisar o chamado efeito de zoneamento dentro de um conjunto de zonas. Para isto, foi necessário reagrupar as 270 zonas OD97 que compõem o município de São Paulo em 96 novas zonas tornando possível assim a comparação entre as correlações deste novo conjunto com os 96 distritos. Na construção das 96 novas zonas procurou-se aplicar algum critério de zoneamento para que se pudesse avaliar também a diferença nos coeficientes de correlação medido para as variáveis utilizadas no experimento. No caso, apesar de não se ter uma ferramenta computacional como o AZP capaz de gerar automaticamente novos agrupamentos, foi possível estabelecer um critério interativo de homogeneidade intrazona através da classificação das zonas OD97 em grupos de acordo com alguma variável. A partir daí, utilizou-se estas zonas classificadas como pano de fundo durante a edição vetorial do novo plano de informação com as 96 novas zonas. É claro que este processo estará sujeito a uma variedade muito grande de arranjos espaciais de saída, uma vez que os critérios de agrupamento podem se alterar de usuário a usuário.

No caso aqui apresentado, os critérios adotados foram por ordem de prioridade os seguintes: (a) a maior homogeneidade intra-área possível com relação à variável renda individual per capita; (b) compor novas zonas a partir do agrupamento de três zonas OD97; (c) sempre que um distrito apresentar estes requisitos preservá-lo como zona; (d) tentar compor zonas mais compactas possível. A escolha da variável “renda individual per capita” como critério de homogeneização foi devido à alta correlação que esta variável tende a apresentar com outras variáveis socioeconômicas populacionais censitárias. Assim, esperava-se captar um padrão de homogeneidade também nas outras duas variáveis utilizadas na análise, “população maior que 60 anos” e “população não alfabetizada”. A nova subdivisão territorial é apresentada na figura 5.5. Nesta, é possível perceber que o critério forma das áreas foi o que menos interferiu no processo de reagrupamento, neste critério percebe-se que a divisão distrital tende a apresentar zonas mais compactas, apesar de heterogêneas. Parece claro que ao responder mais a este critério, a divisão distrital partiu da subdivisão espacial do espaço intra-urbano

levando mais em conta aspectos físicos e morfológicos e menos aspectos socioeconômicos.

Na análise da matriz de correlação (Tab.5.2) pode-se visualizar o efeito de zoneamento, uma vez que manteve-se o mesmo número de zonas que nos distritos, sobre os índices de correlação das variáveis. Percebe-se que houve uma diminuição dos valores absolutos entre todas as variáveis quando comparada com as correlações medidas em nível de distritos. Isso era esperado, pois quando se atinge uma maior homogeneidade intra-área a tendência é aumentar a heterogeneidade interáreas e assim diminuir os índices de correlação. Além destas três variáveis, avaliou-se também o efeito sobre a variável “total de empregos”, porém nesta as correlações ao invés de diminuir, aumentaram. Neste caso, acredita-se que isto se deva ao fato da natureza da variável não ser a mesma das outras três. O total de empregos em uma zona está mais relacionado às características territoriais cadastrais das zonas e não necessariamente apresentam o mesmo comportamento que a população que ali reside, não se configurando assim como um bom parâmetro de avaliação do processo de reagrupamento.

TABELA.5.2 - Matriz de correlação entre as variáveis do sistema de zonas gerado por critério de homogeneidade intra-zona.

Variável **A** = Pop. >60 anos / população total

Variável **B** = Total de habitantes não alfabetizados / população total

Variável **C** = Renda individual per capita (em reais)

	A	B	C
A	1	-0,52	0,57
B	-0,52	1	-0,72
C	0,57	-0,72	1

96 zonas definidas a partir da variável renda per capita

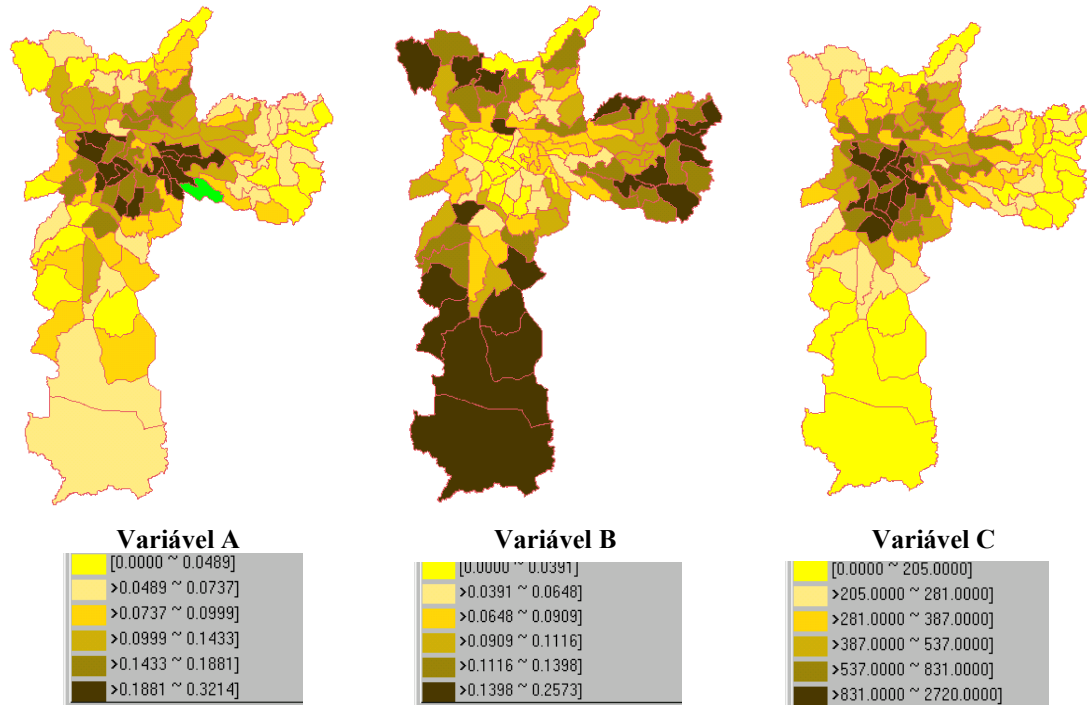


FIGURA.5.5 - Novas 96 zonas construídas interativamente pelo critério de homogeneidade da variável renda individual per capita classificadas segundo os mesmos grupos da figura 5i.3. e pelas mesma variáveis.

5.4. Estimação empírica de Bayes

Na seção anterior foram apresentadas evidências do MAUP dentro do conjunto de dados acessado. Também foi mostrado como o projeto das subdivisões territoriais interfere nas análises quantitativas em função da presença do MAUP no conjunto de dados. Nesta seção, apresenta-se um método estatístico para o tratamento das instabilidades que o MAUP impõem aos dados socioeconômicos do tipo taxas e proporções quando as zonas a que as variáveis estão associadas foram projetadas sem respeitar os critérios anteriormente expostos. Estas instabilidades estão associadas àquelas zonas que, por possuírem um baixo número de habitantes ou indivíduos, estão sujeitas a uma flutuação maior das taxas e proporções nelas medidas.

5.4.1. Fundamentação teórica

É comum que os geodados socioeconômicos, além de associados a subdivisões zonais, sejam disponibilizados e comunicados como taxas e porcentagens. Alguns cartógrafos chegam a dizer que mapas coropléticos somente podem ser aplicados sobre valores relativos e nunca sobre números absolutos. Porém, sempre que um dado é comunicado como taxa ou proporção é importante que se conheça a representatividade da população das n zonas sobre a população total para que se possa compreender melhor o fenômeno, evitando incorrer aos mascaramentos que as taxas e proporções por vezes acarretam. Segundo Bailey e Gattrel(1995): “Quando o atributo de interesse for uma taxa ou proporção, o mapeamento dessas taxas para a visualização da variabilidade geográfica é o primeiro passo óbvio em qualquer análise. Entretanto, o uso das taxas brutas observadas pode ser ilusivo, já que a variabilidade de tais taxas será função do valor da ‘população’ a que estas estão relacionadas, e esta pode diferir muito de área para área.”

Uma abordagem possível ao tratamento destas instabilidades inerentes aos dados relativos a taxas e proporções é aplicar um fator de correção às zonas que considere a representatividade de sua população, assim restringindo a alta variabilidade que as zonas de baixa ‘população’ tendem a apresentar. Para tanto, Bailey e Gattrel (1995) sugerem a aplicação da *Estimação Empírica de Bayes* que consiste na estimação de novas taxas ou proporções a partir de um conhecimento a priori sobre o conjunto de variáveis observado. No caso, o conhecimento a priori considerado é baseado no comportamento global da variável sobre todo o conjunto de zonas, assim é razoável dizer que aquelas zonas que apresentarem uma taxa muito distante da média global observada e que possuam baixa população estarão sujeitas a um fator de restrição maior.

Segundo Bailey e Gattrell (1995), a *Estimação Empírica de Bayes* pode ser formalizada da seguinte maneira: “Suponha que a taxa real desconhecida para cada zona seja θ_i e que $r_i=y_i/n_i$ seja a taxa observada (onde y_i é a população associada a determinado atributo dentro de uma população total n_i). Em uma abordagem não

bayesiana a melhor estimativa $\hat{\theta}_i^L$, de θ_i , é apenas r_i . Entretanto, suponha agora que se conheça a distribuição de probabilidade a priori de cada θ_i , com média γ_i e variância ϕ_i . Pode-se demonstrar que a melhor estimativa bayesiana de θ_i , baseada na combinação da distribuição a priori com as taxas observadas é dada por:

$$\hat{\theta}_i = w_i r_i + (1 - w_i) \gamma_i$$

onde

$$w_i = \frac{\phi_i}{(\phi_i + \gamma_i / n_i)}$$

O fator de ajuste, w_i , é função da população na i -ésima zona e da variância, ϕ_i , da distribuição a priori. Ainda que seja simples o entendimento do processo de estimação empírica de Bayes, sua aplicabilidade está restringida ao fato de que não se conhece a média e a variância da distribuição a priori da verdadeira e desconhecida taxa na i -ésima zona. Assim, sua aplicabilidade está condicionada a algumas decisões reducionistas como por exemplo supor que a média e a variância seja igual para todo o conjunto de zonas. Apesar de um procedimento de estimação aproximado, a aplicação deste método pode representar uma alternativa simples de se aplicar e útil como um parâmetro complementar à análise crítica do usuário.

Considerando-se a estimação de γ e ϕ pelo método dos momentos, ou método direto, a média pode ser estimada a partir da média global,

$$\hat{\gamma} = \frac{\sum y_i}{\sum n_i}$$

e a variância em função da variância amostral ponderada das taxas observadas em relação a esta média,

$$\hat{\phi} = \frac{\sum n_i (r_i - \hat{\gamma})^2}{\sum n_i} - \frac{\hat{\gamma}}{\bar{n}}$$

sendo \bar{n} a população média entre todas as zonas. A partir das estimativas da média e variância, pode-se formular a estimativa bayesiana das taxas em cada zona como:

$$\hat{\theta}_i = \hat{\gamma} + \frac{\hat{\phi}(r_i - \hat{\gamma})}{(\hat{\phi} + \hat{\gamma}/n_i)}$$

5.4.2. Experimento

Esta metodologia foi aplicada para a RMSP, baseada nos dados disponíveis da EMPLASA (1999) sobre a taxa de mortalidade infantil para os 38 Municípios da RMSP e os 96 distritos do MSP no ano de 1996. O índice de mortalidade infantil, tão comumente aplicado como indicador de desenvolvimento humano, é um caso típico de geodados do tipo taxa associado a subdivisões territoriais. Por se aplicarem apenas a uma parcela da população (infantil) tende a ser mais sensível ao efeito de variabilidade estatística acima descrito. Assim, para efeito de experimentação, esta metodologia foi aplicada aos dados referentes às taxas de mortalidade infantil¹ referentes aos distritos do município de São Paulo e aos outros 38 municípios da RMSP. A base geográfica utilizada neste estudo foi escolhida em função da disponibilidade de dados. A EMPLASA disponibiliza informações para a RMSP agregadas nos 39 municípios e para os 96 distritos do MSP. Assim, criou-se uma base de 134 polígonos compostos pelos 96 distritos do MSP e 38 municípios restantes da RMSP. Foram utilizadas as taxas referentes ao ano de 1996, uma vez que as taxas são calculadas sobre o número de nascidos vivos, porém este último dado não era acessível, assim para fins de cálculos, foi adotada a população com idade inferior a um ano de idade disponível nos dados da contagem 1996 (IBGE), como estimativa do total de nascidos vivos.

Analisando-se o gráfico de espalhamento da população com idade inferior a 1 ano e a taxa de mortalidade infantil (Fig.5.6) percebe-se a maior variabilidade nas taxas referentes aos polígonos de menor população. Este comportamento tipo funil, é

¹ Óbitos de menores de 1 ano, por mil nascidos vivos.

conhecido como *heteroscedasticidade* (Neter e Wasserman, 1974), e está relacionado à não constância da variância sobre todas as observações.

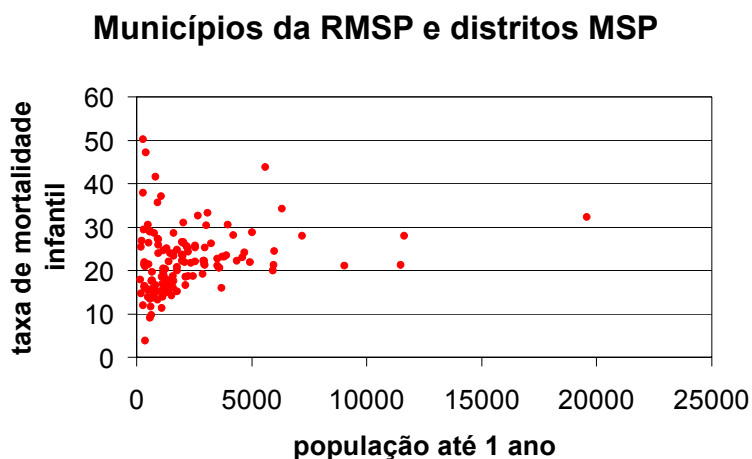


FIGURA.5.6- Gráfico de espalhamento da taxa de mortalidade infantil e a população até 1 ano de idade para os 96 distritos do município de São Paulo e do 38 municípios restantes da RMSP.

5.4.3. Resultados

Os resultados obtidos com a aplicação da metodologia ficaram dentro do esperado. A análise do gráfico de espalhamento das novas taxas de mortalidade infantil estimadas e sua relação com a população de até um ano de idade (Fig.5.7) revela que houve uma restrição da variância maior no grupo de polígonos que apresentam baixa população. Nas áreas de maior população as taxas se mantiveram quase que inalteradas. A transformação aplicada nos dados conseguiu eliminar a heteroscedasticidade observada nas taxas brutas.

É importante destacar que este método permite a quantificação da confiabilidade associada a um conjunto de dados em função da representatividade populacional do suporte territorial. No caso extremo aqui computado, tem-se associada ao município de Salesópolis uma taxa de mortalidade infantil de 50,31 para uma população com idade inferior a 1 ano de 268 crianças em 1996. A taxa estimada pelo método bayesiano foi reduzida para 30,15, diferença significativa de 2/5 e que deve ser considerada quando

afirmações sobre as condições de vida da população daquele município forem analisadas com base nesta variável.

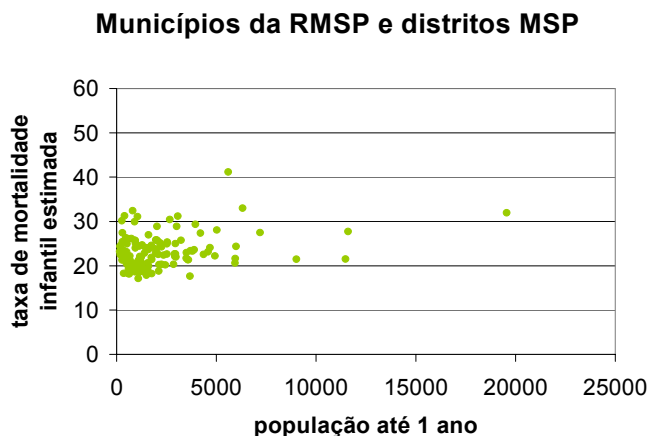


FIGURA.5.7- Gráfico de espalhamento da taxa de mortalidade infantil estimada e a população até 1 ano de idade para os 96 distritos do município de São Paulo e do 38 municípios restantes da RMSP.

A comparação visual dos mapas coropléticos (Fig.5.8) das taxas brutas e nas taxas estimadas mostra que os polígonos fora da mancha urbanizada foram aqueles onde as maiores taxas foram encontradas. Da mesma forma, nestes geralmente com baixa densidade populacional que o restritor mais operou. Outra observação refere-se a perda da variabilidade espacial da taxa bruta quando comparada com a taxa estimada. Não há dúvidas de que quando se aplicam estes filtros matemáticos introduz-se um ponderador que atrairá o valor das novas taxas em torno de um valor médio global. Como consequência, a visualização destas novas taxas, classificadas sob os mesmos intervalos para as taxas brutas, resultará muito mais homogênea, o que acarreta, num primeiro momento, uma perda de informação.

A solução para este efeito homogeneizador pode estar na incorporação de padrões espaciais na formulação. Assim, as taxas estimadas seriam aproximadas em função do comportamento dessa variável em sua vizinhança. A incorporação da informação da vizinhança está relacionada às propriedades topológicas dos polígonos, geralmente essa informação está associada à adjacência e por esta se definem os vizinhos. A incorporação desta informação ao método empírico de Bayes é direta,

supondo-se agora que a média e variância da distribuição a priori desconhecidas é igual e constante para todas as zonas vizinhas. No estudo de caso realizado, não foi aplicada a metodologia bayesiana localmente condicionada pela impossibilidade do cálculo da variância local pelo sistema.

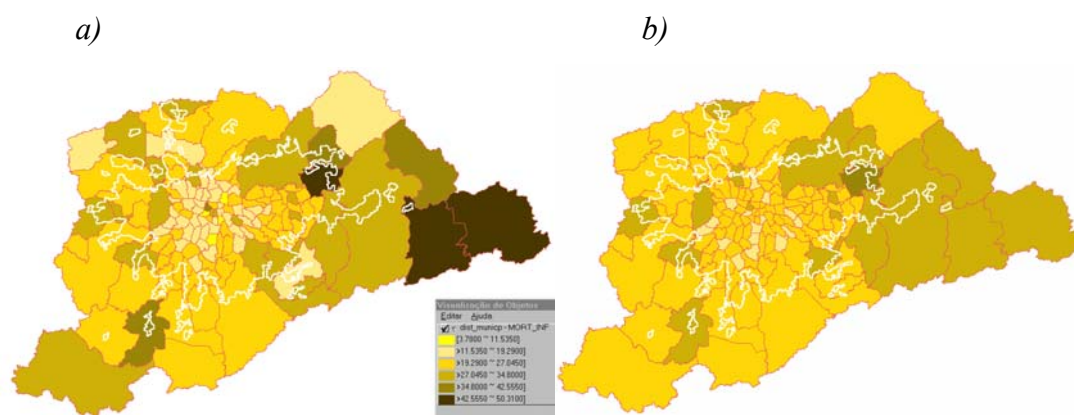


FIG.5.8 - Mapas coropléticos com a taxa de mortalidade infantil observada (a) e estimada (b) para os 96 distritos do município de São Paulo e dos 38 municípios restantes da RMSP. (Fonte: Seade, IBGE).

5.5. Conclusão

Neste capítulo foi abordado como os geodados associados a subdivisões territoriais são sensíveis a fatores relacionados com a representatividade do suporte territorial. Em análises de estruturas intra-urbanas, interessadas no processo de revelação dos territórios intra-urbanos, ou seja, dos elementos urbanísticos relevantes que articulados configuram esta estrutura, é fundamental o reconhecimento do analista sobre as interferências quantitativas a que suas conclusões estão sujeitas. Aqui, o ‘quebra-cabeças’ trará uma imagem diferente do mesmo fenômeno em função do conjunto de peças que se utilize. Portanto, quanto mais criteriosa for a interpretação do analista frente ao fenômeno observado e quantificado em dados, melhor a identificação da estrutura territorial oculta nos mapas coloridos.

CAPÍTULO 6

ANÁLISE ESPACIAL II: INDICADORES LOCAIS DE ASSOCIAÇÃO ESPACIAL

Na seção anterior, discutiu-se o efeito da definição nem sempre criteriosa das subdivisões territoriais sobre os dados agregados por áreas. Nesta, discute-se a *dependência espacial* como uma característica inerente a este tipo de representação. A dependência, ou interdependência espacial, pode ser entendida como a tendência a que o valor de uma ou mais variáveis associadas a uma determinada localização assemelhe-se mais ao valor de suas amostras vizinhas do que ao restante das localizações do conjunto amostral. Nas análises computacionais de estruturas intra-urbanas, a noção de dependência espacial relaciona-se diretamente com os três conceitos de escala intra-urbana, território intra-urbano e topologia intra-urbana. A família de ferramentas estatísticas apresentada a seguir tem por finalidade quantificar a dependência espacial e assim, trazer uma expressão computacional do arranjo espacial das semelhanças e discrepâncias entre os polígonos que compõem o mosaico colorido freqüentemente utilizados na representações de geodados socioeconômicos. Neste capítulo, discutem-se e exemplificam-se algumas possibilidades de aplicação destas ferramentas neste tipo de análise através de experimentos realizados sobre os dados OD97.

6.1. Fundamentação Teórica

6.1.1. *A estatística espacial*

A necessidade de quantificação da dependência espacial presente num conjunto de geodados levou ao desenvolvimento da chamada *estatística espacial*. Segundo Anselin (1992), “a característica que distingue a análise estatística dos dados espaciais é que seu foco principal está em inquirir *padrões espaciais* de lugares e valores, a *associação espacial* entre eles e a variação sistemática do fenômeno por localização.” Ele segue, “além disso, afora a importância do significado geográfico que esta medida traz em si, estes padrões espaciais causam problemas de mensuração, conhecidos como

efeitos espaciais, tais como *dependência espacial* e *heterogeneidade espacial*, que afetam a validade dos métodos estatísticos tradicionais”.

A dependência espacial pode ser medida de diferentes formas. O índice de Moran (I) é a estatística mais difundida e mede a autocorrelação espacial a partir do produto dos desvios em relação à média. Há outras medidas que apesar de similares, medem a dependência espacial a partir de outras operações, diferença simples como no índice de Geary ou soma simples como nas estatísticas G (Anselin,1992). Formalmente, I é escrito como:

$$I = \frac{n}{W} \left(\frac{\sum_i \sum_j w_{ij} z_i z_j}{\sum_i z_i^2} \right) \text{ para } i \neq j$$

onde: n é o número de observações
 w_{ij} é o elemento na matriz de vizinhança para o par i e j
 W é a soma dos ponderadores da matriz
 z_i e z_j são desvios em relação à média

A informação espacial é incorporada no modelo a partir da matriz de vizinhança, W . Nesta, a vizinhança pode ser definida de diferentes maneiras, *e.g.*- distância euclidiana, tempo de viagem ou acessibilidade. Em Gattrel (1991), discutem-se algumas possibilidades de formalizações de relações espaço-tempo como medida de proximidade. A mais comumente utilizada para geodados zonais define vizinhança a partir da propriedade topológica de contiguidade, assim W é uma matriz binária (0,1), onde 1 está associado às zonas com fronteiras em comum e 0 àquelas sem esta propriedade (Fig.6.1). Como a matriz de vizinhança é utilizada em cálculos de indicadores de análise exploratória, por conveniência, ela é muitas vezes utilizada normalizada por linha, ou seja, com a soma dos ponderadores de cada linha igual a 1 (Câmara et al.,2002).

O índice de Moran acima apresentado é uma medida global da autocorrelação espacial, ou seja, mede a dependência espacial baseado em observações simultâneas no

conjunto de n localizações. É uma medida importante, pois indica o grau de associação espacial presente no conjunto de dados, porém sintetiza inúmeras possibilidades de padrões de associação local entre as amostras georreferenciadas. São nestes padrões locais que em grande parte reside a informação relevante para análises geográficas. Especificamente em análises intra-urbanas é fundamental que se investiguem as configurações locais de associação espacial, pois, o foco da análise é justamente a observação da organização territorial interna, ou seja, importa saber como determinadas características geográficas se distribuem espacialmente, se há concentrações ou tendências de determinadas características que possam revelar os elementos territoriais estruturais.

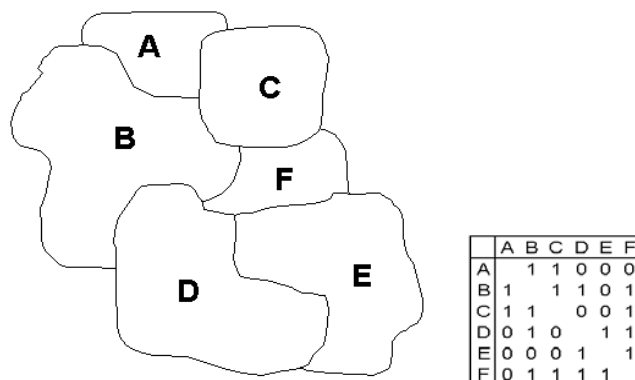


FIGURA.6.1 - Exemplo de divisão zonal com matriz de vizinhança associada definida por propriedade de contigüidade.

6.1.2. Estatística espacial local

A *estatística espacial local* foi desenvolvida para quantificar o grau de associação espacial a que cada localização do conjunto amostral está submetida em função de um modelo de vizinhança preestabelecido. Convencionou-se chamar de *Indicadores Locais de Associação Espacial* (LISA) os operadores estatísticos com esta finalidade. Anselin (1995) aponta que existe uma proporcionalidade direta entre o valor da autocorrelação global e os valores das autocorrelações locais, ele demonstra que os LISAs permitem a decomposição dos indicadores globais em contribuições individuais, indicando porções territoriais de não *estacionariedade* e identificando aglomerados (*clusters*) significativos de valores semelhantes em torno de determinadas localizações.

Similarmente aos índices globais, nos índices locais, a interdependência ou associação espacial é quantificada como *autocorrelação espacial*, porém nos últimos, o universo amostral restringe-se a uma determinada localização e sua vizinhança. Dentre os LISAs mais difundidos estão o Índice Local de Moran (I_i) e as Estatísticas G_i e G_i^* . Estes podem ser escritos como:

$$I_i(d) = \frac{(x_i - \bar{x})}{s^2} \sum_j w_{ij}(d)(x_j - \bar{x}) \quad \text{para } j \neq i$$

e

$$G_i(d) = \frac{\sum_j w_{ij}(d)x_j}{\sum_j x_j} \quad \text{para } j \neq i$$

$$G_i^*(d) = \frac{\sum_j w_{ij}(d)x_j}{\sum_j x_j} \quad \text{para todos os } j$$

onde:

$w_{ij}(d)$ é o ponderador na matriz de vizinhança W para o par i e j
 d é a medida de distância estabelecida pelo modelo de vizinhança
 x_i e x_j são valores encontrados na posição i e suas vizinhas js
 \bar{x} é a média amostral global
 s^2 é a variância amostral global

Estas duas medidas de autocorrelação espacial local, apesar de similares, devem ter seus resultados interpretados de maneiras diferentes. No índice local de Moran a autocorrelação espacial é calculada a partir do produto dos desvios em relação à média como uma medida de covariância, dessa forma, valores significativamente altos indicam altas probabilidades de que haja locais de associação espacial tanto de polígonos com altos valores associados como com baixos valores associados. Por outro lado, baixos valores apontam para um padrão que pode ser entendido como pontos de comportamento mais errático da variável observada entre um polígono e seus vizinhos. Nas estatísticas G_i e G_i^* diferentemente, o cálculo é realizado pela soma simples das amostras vizinhas em relação a uma determinada posição i , sendo que no caso do índice G_i , o valor da localização em estudo i não é incluído na somatória e, no caso do índice

G_i^* , sim. Portanto nestas, valores significativamente altos indicam locais de associação espacial de amostras de alto valor e valores significativamente baixos indicam agrupamentos de amostras de baixos valores. A significância referida em ambos os casos dependerá da suposição de alguma distribuição estatística ou da construção de uma pseudo-distribuição empírica através de permutação (Getis e Ord,1992).

A técnica de permutação permite estabelecer significância estatística empírica do resultado obtido. Esta significância empírica é obtida a partir de uma distribuição gerada pela repetida substituição aleatória dos n valores pertencentes ao conjunto amostral nas i -ésimas localizações e calculando-se novos resultados para cada novo arranjo. Assim, tem-se a possibilidade de comparação do resultado obtido sob arranjo espacial original e os n resultados obtidos nos arranjos gerados aleatoriamente (Anselin,1992) (Câmara et al.,2002).

Além dos mapas de significância, há uma outra ferramenta utilizada na exploração de padrões de associação espacial conhecida como mapa de espalhamento de Moran (*Moran Scatterplot Map*). Este, apesar de não medir diretamente a autocorrelação espacial, é uma importante ferramenta de visualização de padrões desta autocorrelação. A idéia é comparar a distribuição espacial de uma variável com a média local de sua vizinhança, estes valores são normalizados, então com média global igual a zero. A partir daí, o mapa de espalhamento de Moran classifica as zonas segundo quatro regimes espaciais: áreas com associação espacial positiva, Q1-valor positivo e médias locais positiva (alto-alto) e Q2-valor negativo e média local negativa (baixo-baixo); e áreas com associação espacial negativa, Q3-valor positivo e média local negativa (alto-baixo) e Q4-valor negativo e média local positiva (baixo-alto) (Fig.6.2). As zonas classificadas como Q3 e Q4 podem ser interpretadas como locais que não seguem o mesmo processo de dependência espacial, indicando pontos de transições entre diferentes padrões espaciais ou pontos de não estacionariedade do atributo.

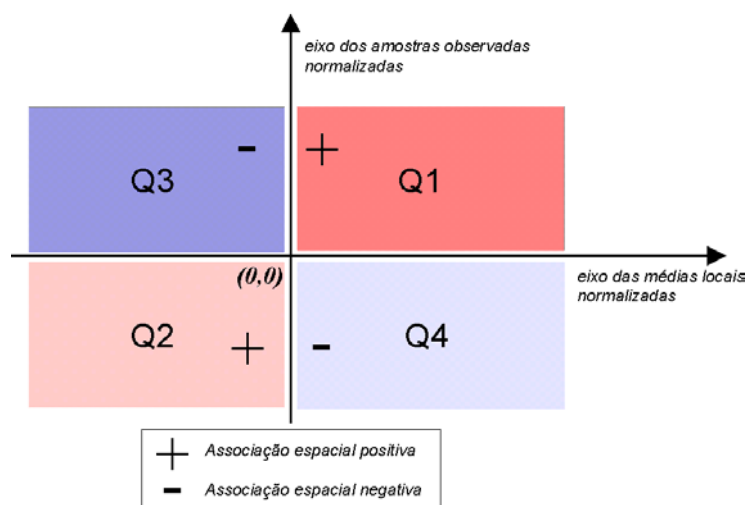


FIGURA.6.2 - Esquema do gráfico de espalhamento de Moran.

6.2. As Estatísticas Locais nas Análises de Estruturas Intra-Urbanas

Conforme colocado no segundo capítulo, o objetivo principal das análises de estruturas intra-urbanas é compreender as localizações relativas dos seus elementos componentes, os territórios intra-urbanos, suas características e seus inter-relacionamentos espaciais e funcionais. No contexto desta dissertação, pode-se dizer que as análises de estruturas intra-urbanas estão interessadas na *distribuição espacial* dos eventos e objetos que, em última análise, remetem aos territórios que inter-relacionados configuram esta estrutura. Sob esta perspectiva, as estatísticas espaciais locais se colocam como ferramentas importantes por sua capacidade de inferir padrões locais e globais da distribuição espacial das variáveis georreferenciadas.

Em Getis e Ord (1996) são identificadas três aplicações principais dos LISAs. A primeira é a detecção de *clusters* significativos de variáveis de valores similares em torno de determinadas localizações (*'hot-spots'*). A segunda é inferir sobre as condições de estacionariedade do conjunto de amostras georreferenciadas. Por fim, como ferramenta auxiliar na determinação da áreas de influência de variáveis e conseqüente determinação da escala de observação. Esses objetivos podem ser relacionados a exploração de características distributivas como *concentrações*, *persistências* e *transições* que podem ser consideradas expressões quantitativas da estrutura territorial.

À luz dessas considerações, é possível explorar algumas formas onde a análise da distribuição espacial de variáveis georreferenciadas possa expressar conceitos urbanísticos relacionados com a noção de estrutura intra-urbana. Uma aplicação direta dos LISAs é inferir sobre as *áreas de influência* de determinadas variáveis, indicando a extensão, dentro de um conjunto de localizações, de padrões espaciais locais onde há uma tendência em se encontrar valores similares para essas amostras. A apreensão desta estrutura de correlação espacial não é direta. Muitas vezes ela é ocultada pela excessiva fragmentação imposta pelo suporte zonal aos quais os dados estão associados. A análise das semelhanças e diferenças dentro do conjunto de valores através das ferramentas estatísticas pode realçar os diferentes arranjos espaciais encobertos por essa fragmentação. Estes filtros matemáticos, ao medirem a autocorrelação espacial, revelam também as extensões territoriais onde determinados padrões de associação espacial tendem a persistir. Dessa forma, a interpretação está diretamente relacionada com a noção de escala intra-urbana, pois ao indicar estas extensões, indicam áreas de influência de possíveis territórios e apontam para um grau de desagregação da informação compatível com a análise daquela determinada característica.

Similarmente, indicações sobre os territórios intra-urbanos podem ser associadas às análises das *concentrações espaciais* de determinadas características socioeconômicas. A capacidade de detecção de agrupamentos significativos de amostras de valores próximos em torno de cada localização permite identificar os pontos onde tais características predominam, apontando assim, potenciais territórios desenhados a partir destas características. Por exemplo, pode-se procurar por agrupamentos significativos de habitantes com altos níveis de escolaridade, apontando para os territórios dos mais instruídos, ou agrupamentos significativos de altas taxas de homicídio como indicativo para os territórios mais violentos. Sabendo-se que a segregação socioterritorial é uma das características essenciais das estruturas intra-urbanas das cidades brasileiras, tais ferramentas se colocam como particularmente importantes dentro do repertório analítico do urbanista como potencialmente reveladoras dessas estruturas.

Há ainda o terceiro conceito relacionado à estrutura intra-urbana que é a noção de topologia intra-urbana. Associadas a este conceito estão as características distributivas identificadas como *persistências* e *transições*, ou seja, expressões do desenho distributivo global da variável no espaço que revelam, além das concentrações, locais destacados a partir das diferenças, informação muitas vezes relevante na análise. Estes pontos de autocorrelação espacial negativa podem ser interpretados como polígonos estranhos a sua vizinhança. A possibilidade de identificação destes padrões de persistência e transição das características associadas às variáveis georreferenciadas traz uma expressão possível da topologia intra-urbana, relacionada às localizações relativas dos elementos componentes da estrutura e suas articulações espaço-funcionais.

A próxima seção deste capítulo pretende ilustrar as possibilidades de aplicação dos LISAs em análises urbanas com alguns experimentos realizados sobre os dados OD97.

6.3. Exemplos de Aplicações dos LISAs em Análises Intra-Urbanas

Para a realização dos experimentos em estatísticas locais foram utilizados os software ArcView3.0 e SpaceStat1.9. A base de dados utilizada foi a OD97 e dela foram selecionadas duas variáveis, a primeira foi o total de empregos por zona dividido por sua área total, ou seja, a densidade de empregos (**emp_area**), a segunda foi a porcentagem da população com até o primeiro grau incompleto, (**low_edu**). A seleção dessas variáveis teve como objetivo estudar tanto características socioeconômicas territoriais cadastrais quanto populacionais censitárias. No caso do número de empregos, esperava-se detectar agrupamentos de zonas com altas densidades de empregos que pudessem apontar para o centro e subcentros da estrutura intra-urbana de São Paulo. Na segunda variável analisada esperava-se obter indicações de parcelas do território intra-urbano com elevada exclusão social, uma vez que a baixa escolaridade é forte indicador de realidades sociais de alto risco. A figura 6.3 apresenta as zonas OD97 intra-urbanas classificadas por quantís destas duas variáveis.

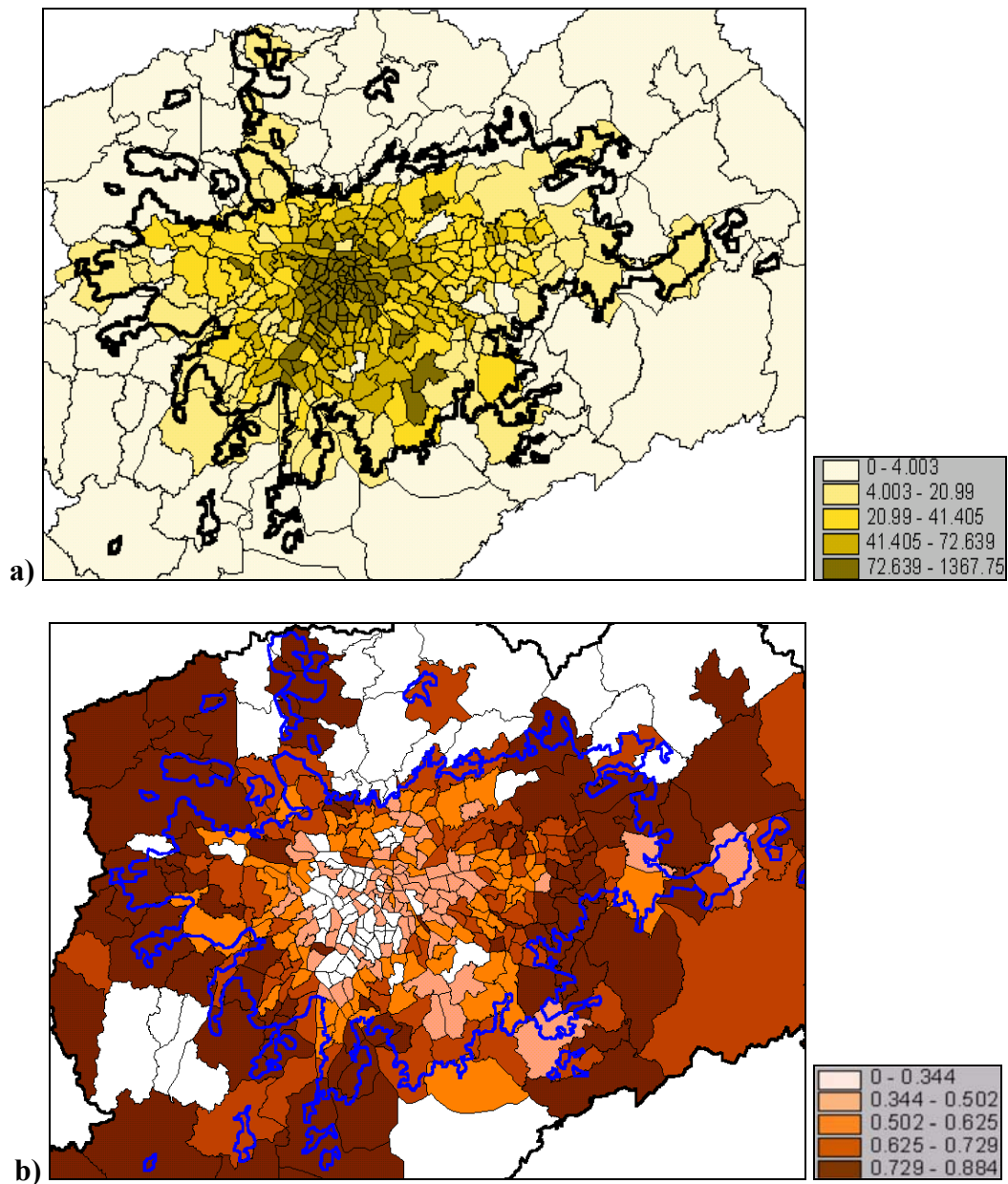


FIGURA.6.3 - a) Densidade de empregos classificados por quintis; b) Porcentagem da população com baixa escolaridade classificadas por quintis.

Três ferramentas exploratórias foram aplicadas: os mapas de significância do índice local de Moran, os resultados das estatística G_i^* classificados por desvios padrões, e o mapa de espalhamento de Moran. No experimento, a matriz de vizinhança foi definida pela propriedade topológica de contiguidade e utilizada normalizada nos cálculos processados. Todas as visualizações são apresentadas sobrepostas ao contorno

da mancha urbana extraído da imagem landsat-TM gerada em 1997, assim direciona-se a interpretação aos polígonos urbanizados, mais densamente ocupados. Esta forma de apresentação deve-se ao fato de que os polígonos mais externos da RMSP são pouco representativos da estrutura intra-urbana, eles podem ser afetados por instabilidades decorrentes da baixa densidade demográfica como é sugerido no experimento realizado no capítulo anterior. A decisão de não aplicar filtros de estimação bayesiana antes do cálculo dos LISAs foi em função do reconhecimento da influência que estes estimadores, baseados em médias globais ou locais conhecidas a priori, exercem nas medidas de autocorrelação espacial sobretudo naqueles polígonos com baixa representatividade populacional. Assim, optou-se por preservar as autocorrelações espaciais originais observadas para o cálculo dos LISAs e direcionar a análise para os polígonos mais representativos da estrutura intra-urbana, ou seja, os internos à mancha.

6.3.1. A autocorrelação espacial para a densidade de empregos

As análises preliminares realizadas através do índice de Moran global sobre a variável **emp_area** apontaram com significância de 99,9% após 999 permutações a existência de um padrão de autocorrelação espacial. De fato, as análises de autocorrelação espacial local efetuadas revelam um padrão de extrema concentração da densidade de emprego nas zonas mais centrais da cidade. O mapa de significância (Fig.6.4) do índice local de Moran calculado indica que este agrupamento é altamente significativo e bastante compacto.

Correspondem às zonas OD97 com maior significância, 99,9%, a área mais central da cidade estendendo-se da região da Luz, ao norte, até a região da Avenida Paulista, a sudoeste. Com significância menor, 95%, um anel de polígonos circunda este núcleo e abrange no eixo leste-oeste desde a zona leste próxima, Brás e Bresser, até Pinheiros; e no eixo noroeste-sudeste, de Perdizes a Vila Mariana. Afora este agrupamento, quase todos os outros polígonos internos a mancha urbana não apresentam significativa autocorrelação espacial, indicando não haver um padrão específico da densidade de emprego nestas áreas. Nos polígonos mais externos, a

significativa autocorrelação foi atribuída devido à persistência do padrão de baixas densidades de emprego relacionadas às áreas rurais externas a estrutura.

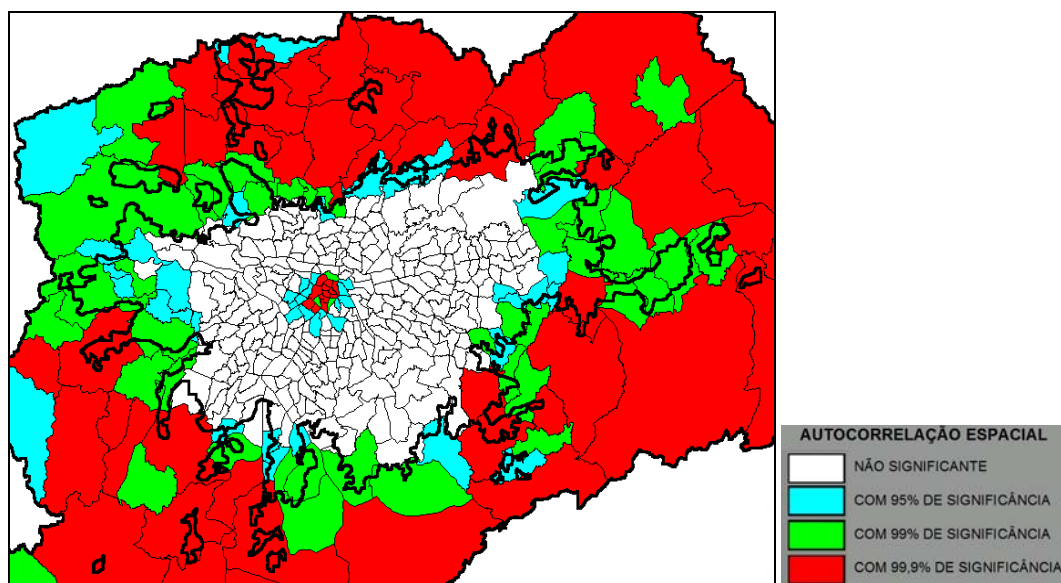


FIGURA.6.4 - Mapa de significância dos resultados obtidos para índice local de Moran para a variável emp_area após 999 permutações.

O mapa da estatística G_i^* classificada por desvios também aponta para os polígonos centrais como positivamente autocorrelacionados e apresenta um terceiro anel mais externo positivamente correlacionados que pode ser associado a extensão do centro expandido definido pelo anel viário (Marginais - Bandeirantes - Tancredo Neves - Salim F. Maluf), sendo que na direção da Zona Norte e de Santo Amaro este limite é ultrapassado. O cálculo do G_i^* é uma soma simples de vizinhos normalizada, assim, sua visualização classificada por desvios em relação a média permite uma interpretação direta da intensidade e do sinal da associação espacial. Na variável normalizada, a média tende a zero, assim a representação se faz por números de desvios padrões em relação ao zero em valor positivo, indicando os locais onde esta soma foi maior que zero e negativos em casos contrários. A figura 6.5 apresenta um mapa deste tipo onde se nota uma predominância da variável se espalhar, a partir do núcleo compacto central, preponderantemente no sentido sudoeste. O mapa dos índices G_i^* normalizados é uma

ferramenta interessante pela simplicidade de construção e interpretação, é uma maneira direta de se avaliar o comportamento global das variáveis sobre o espaço.

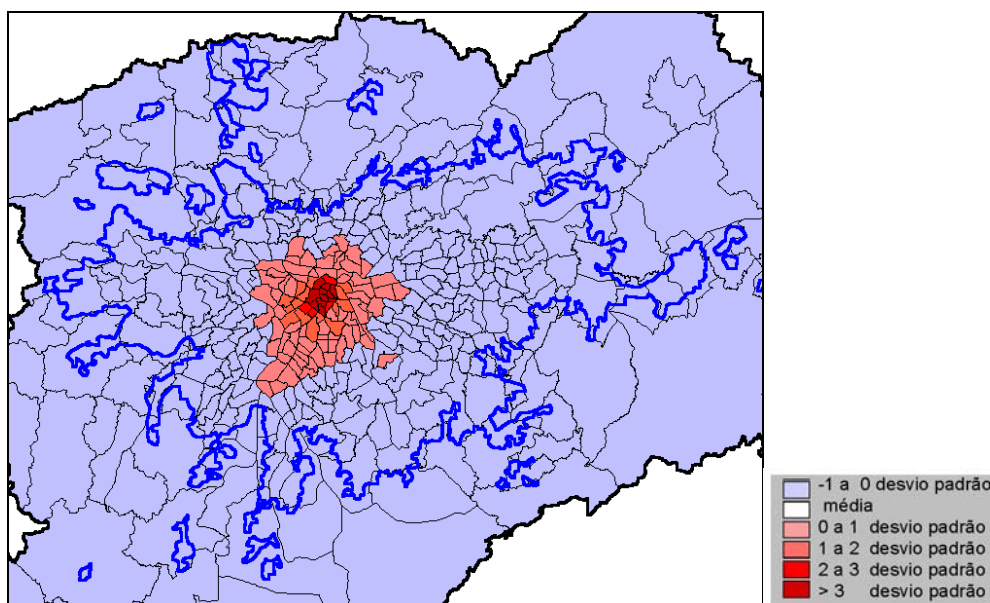


FIGURA.6.5 - Os índices G_i^* normalizados calculados para a variável emp_area classificados por desvios em relação à média.

O mapa de espalhamento de Moran construído para a variável **emp_area** (Fig.6.6) permite a visualização dos polígonos relacionados aos pontos de transição de padrão. Estes podem indicar as áreas onde diferentes padrões se encontram, imprimindo uma instabilidade na correlação espacial local identificada no gráfico de espalhamento. Chamam a atenção os polígonos identificados como Q3, valor alto inserido em vizinhança com média baixa, fora do núcleo central compacto. Uma análise mais atenta revelou que estes polígonos estão associados a importantes subcentros na estrutura territorial da metrópole, caso do centro de Guarulhos, São Caetano, Santo André, São Bernardo do Campo e Osasco, além das zonas OD97 correspondentes à São Miguel Paulista e a cidade Nitro-Operária também indicadas como pólos de empregos inseridos na vizinhança de baixa densidade de empregos na zona leste.

Em conjunto, estas três ferramentas exploratórias apontaram evidências de que os territórios do trabalho, componentes essenciais da estrutura intra-urbana, se configuram como um núcleo compacto com alta densidade de empregos nas zonas

centrais, circundado por anéis de densidade decrescente desta atividade. Este núcleo destaca-se do padrão predominante nas outras zonas OD97 internas a mancha urbana, apontado como áreas de baixas densidades pelo mapa Gi*. Dentro destes territórios com poucos empregos algumas instabilidades pontuais foram detectadas pelo mapa de espalhamento de Moran, revelando-se como os subcentros da estrutura intra-urbana. Desenha-se assim, a expressão matemática da configuração territorial das atividades empregatícias intra-urbanas da cidade de São Paulo. A extrema concentração dos empregos na cidade é reconhecidamente uma das principais causas dos grandes problemas de acessibilidade que a cidade vem enfrentando, pois ao gerar enormes deslocamentos diários para uma grande faixa da população, impõe perdas na sua produtividade e qualidade de vida.

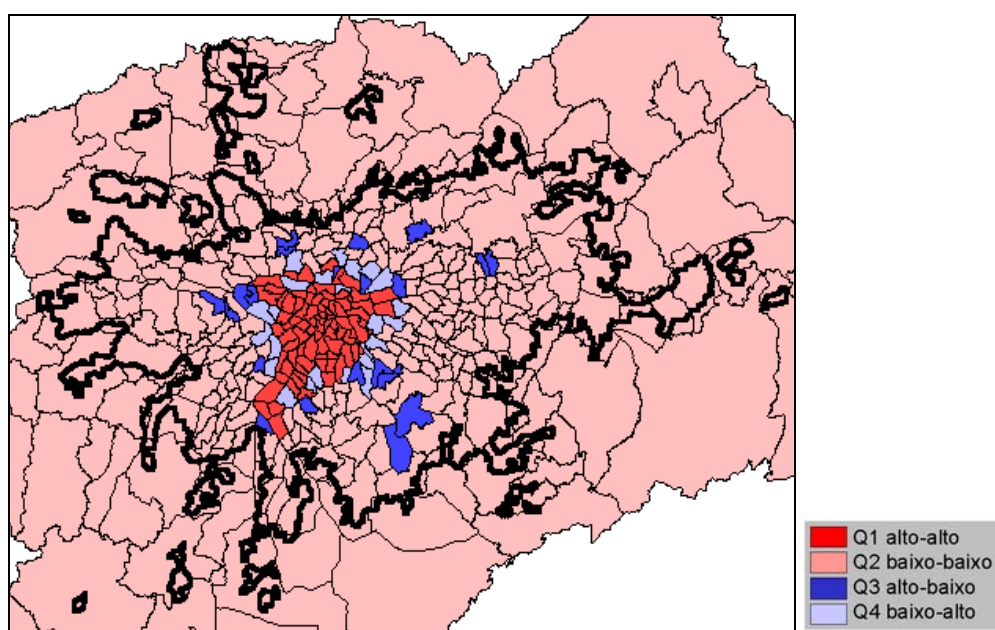


FIGURA.6.6 - Mapa de espalhamento de Moran obtido para a variável densidade de empregos.

6.3.2. A autocorrelação espacial para a variável baixa escolaridade

Assim como para a variável **emp_area**, para a variável **low_edu** também o valor do índice global de Moran apontou para a existência de padrões de autocorrelação com significância de 99,9% após 999 permutações. Porém, nas análises de autocorrelação espacial local realizadas sobre a variável **low_edu** foram identificados

agrupamentos significativos de dois tipos: um de baixos valores e outros de altos valores.

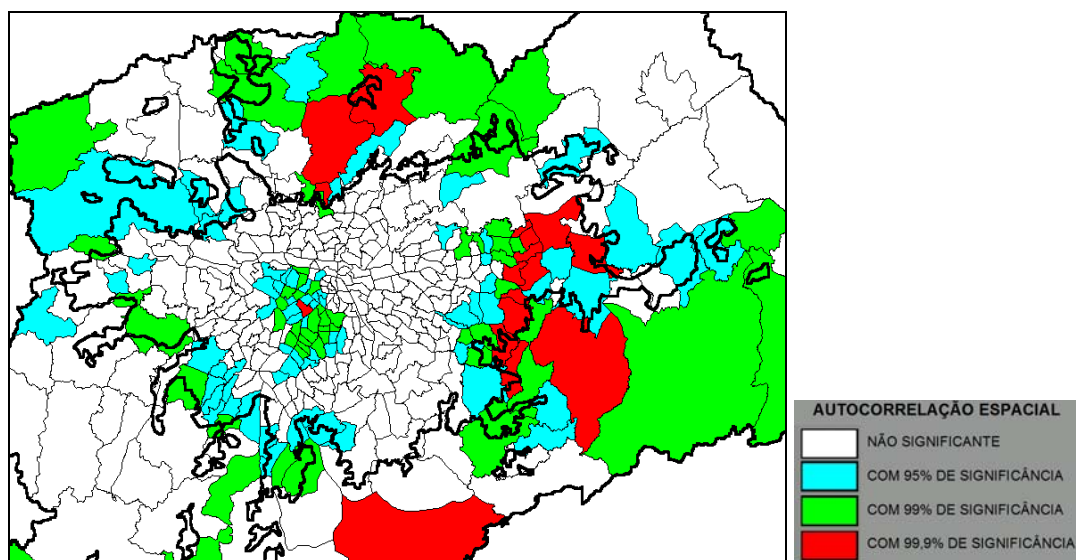


FIGURA.6.7- Mapa de significância do índice Moran local padronizado obtido para a variável low_edu após 999 permutações.

O mapa de significância de Moran local (Fig.6.7) evidencia a associação espacial de população de baixas taxas de população com escolaridade precária nas zonas do quadrante centro-sudoeste. Este padrão predomina na área circunscrita a um raio de aproximadamente 6 Km em torno dos Jardins, abrangendo a Vila Mariana, a Consolação, o Sumaré, o Alto de Pinheiros, o Jóquei Club, o Brooklin e a Chácara Flora. Além deste, agrupamentos de zonas com altas taxas de população com escolaridade precária foram identificados. Os mais relevantes na nossa análise são aqueles identificados na área mais externa da mancha urbana. Entre estes, o agrupamento mais significativo situa-se no extremo da zona leste e estende-se de norte a sul. Neste estão compreendidas as zonas OD97 de Vista Alegre, Iguatemi, Cidade Tiradentes, Juscelino Kubitschek, Santos Dumond, Fazenda Itaim até Itaquaquecetuba, com significância de 99,9% após 999 permutações, e também apresentando zonas circundantes com significância de 95% e 99%, abrangendo da zona OD97 de Ribeirão Pires até Jardim Helena e Limoeiro. Já no quadrante sul-sudoeste, agrupamentos significantes foram apontados nas zonas do Grajaú, Cocaia, Jardim Presidente, Pedreira

e Eldorado. Ao lado desta observa-se o agrupamento das zonas do M'Boi Mirim, Jardim Ângela, Guarapiranga, Capão Redondo até Parque Pinheiros. Todas estas regiões são reconhecidas como extensões territoriais com altas taxas de moradores com alfabetização precária, e estão associadas às periferias tantas vezes referidas como os territórios da exclusão social na cidade.

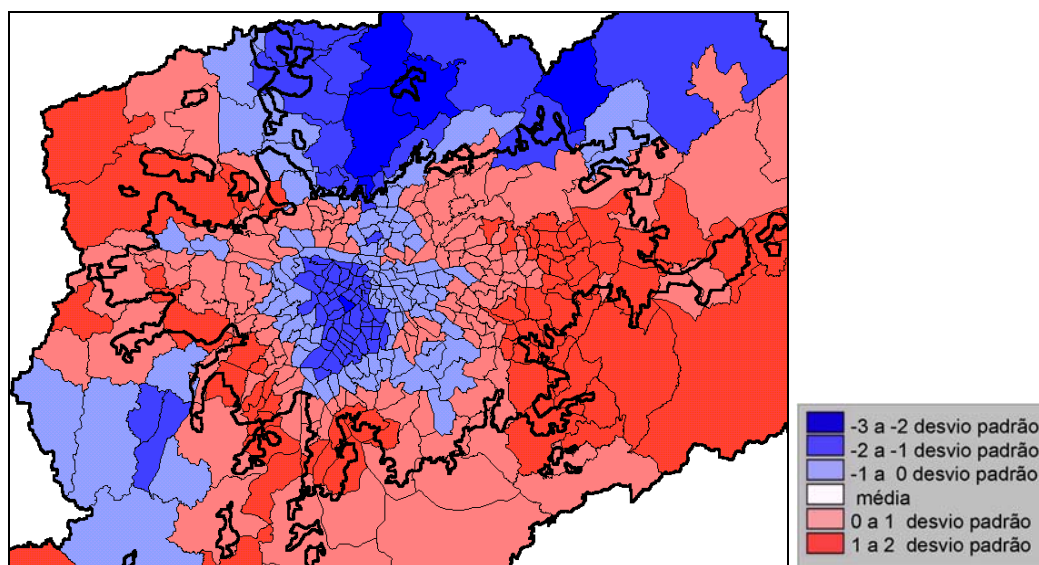


FIGURA.6.8 -Os índices G_i^* normalizadas calculados para a variável `low_edu` classificados por desvios em relação à média.

A visualização dos resultados de G_i^* para **low_edu** classificados por desvios em relação à média (Fig.6.8) revelam as tendências globais da distribuição da população com baixa escolaridade dentro da estrutura intra-urbana. O mapa aponta para o padrão centro-periferia com a área central definida como no mapa de significância de Moran local. No mapa G_i^* é possível observar a transição gradual dos valores das somas das áreas mais centrais de baixos valores até os pontos de maiores valores da soma das amostras `low_edu` vizinhas. Chama especial atenção o agrupamento de polígonos vermelhos inseridos no quadrante sudeste da área azul relacionado às zonas da grande favela de Heliópolis, e as duas zonas azuis inseridas no extremo noroeste da mancha urbanizada relacionadas à Tamboré e Barueri, zonas conhecidas por seus condomínios residenciais fechados de alto padrão.

O mapa de espalhamento de Moran obtido para a variável **low_edu** (Fig.6.9) destaca as zonas correspondentes às zonas de transições entre os dois padrões dominantes de associação espacial positiva como na região do Morumbi, onde a zona de Paraisópolis se apresenta como negativamente autocorrelacionada com seu entorno ao lado do portal do Morumbi, também associação negativa porém em quadrante oposto. As áreas de instabilidades mais evidentes foram, no quadrante sudeste, em torno da favela do Heliópolis, apontando sua importância e proximidade com áreas de zonas com baixa densidade de população com escolaridade precária como São Caetano do Sul, Jardim da Saúde e Ipiranga e no quadrante oeste nas zonas pertencentes ao eixo da Rodovia Castelo Branco que se apresentam como uma seqüência de associação espacial negativa.

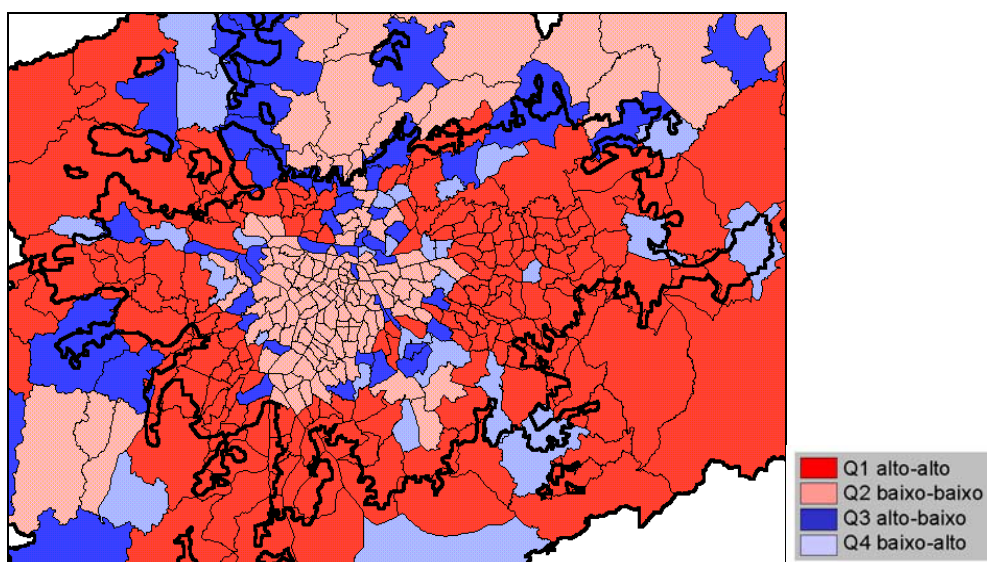


FIGURA.6.9 - Mapa de espalhamento de Moran obtido para a variável **low_edu**.

Assim como na análise da variável **emp_area**, a aplicação destas três ferramentas exploratórias em conjunto para a variável **low_edu** traz informações quantitativas sobre a distribuição espacial dos valores observados. Prevalece um arranjo centro-periferia, com núcleo caracterizado por concentração de pequenas taxas de escolaridade precária, circundada por zonas de densidade crescente de população nesta situação, até se configurarem novos agrupamentos significativos nas periferias mais externas da faixa urbanizada relacionadas às periferias da cidade. Este padrão

dominante é rompido em alguns pontos de não estacionariedade como no entorno da favela de Heliópolis e Paraisópolis e nas zonas OD97 de Tamboré e Barueri.

6.4. Conclusão

Os LISAs apresentaram-se como ferramentas eficazes nas análises de estrutura intra-urbana por sua capacidade de revelar padrões de distribuição espacial como concentrações, transições, persistências e áreas de influência nas variáveis georreferenciadas. A simplicidade do modelo de vizinhança adotado nos exemplos colocou-se como fator limitante a sua aplicação, para a plena utilização destas ferramentas em estudos de estruturas intra-urbanas novos modelos de vizinhança devem ser formulados baseados não só na propriedade de proximidade como também na conectividade, medida pelo tempo de viagem (*e.g.*). Também há que se destacar que as aglomerações identificadas como significativas estão necessariamente associadas à resolução espacial dos dados de entrada, e que uma resolução mais fina poderia revelar outros padrões espaciais.

Utilizados em associação, o mapa de espalhamento e significância de Moran, e o mapa dos resultados G_i^* , representam uma poderosa ferramenta na interpretação de geodados socioeconômicos apresentados por unidades de área. A capacidade de estimar a significância da associação espacial dos valores observados em função de sua vizinhança, permite revelar os arranjos ocultos pela fragmentação que este tipo de suporte impõe a leitura dos mapas.

CAPÍTULO 7

ANÁLISE ESPACIAL III: DE UNIDADES DE ÁREA A SUPERFÍCIES CONTÍNUAS

Os geodados socioeconômicos geralmente são disponibilizados como características estatísticas de grupos associados a unidades de área. Como já foi colocado no capítulo 5, há problemas envolvendo este tipo de representação, pois os critérios de zoneamento nem sempre estão relacionados ao fenômeno modelado. Isto tem levado ao desenvolvimento de representações alternativas para este tipo de dado. Nesta perspectiva, a modelagem dos dados socioeconômicos como geo-campos tem despertado particular interesse. Um dos motivos é que este tipo de representação permite uma leitura mais adequada da distribuição contínua da densidade de uma variável ou atributo sobre o território, o outro motivo deve-se ao fato do conceito de superfície numérica ampliar o espectro de aplicações de técnicas de análise espacial sobre dados socioeconômicos (Martin,1996a). A questão fundamental relacionada à aplicação desta representação deve-se a impossibilidade de se conhecer ou coletar os valores das variáveis em todos os pontos do território. Assim, estas superfícies devem ser construídas a partir de dados pontuais ou zonais através de processos de interpolação, estimando-se os valores nos pontos do território onde se desconhece esta informação. Neste capítulo, explora-se a aplicação de técnicas geoestatísticas de interpolação para a construção de superfícies para dados socioeconômicos intra-urbanos.

7.1. Fundamentação Teórica

Dentro do conjunto de técnicas de interpolação para a construção de superfícies computacionais contínuas a partir de amostras irregularmente espaçadas, duas categorias podem ser identificadas, os interpoladores determinísticos e os estocásticos. Convencionou-se denominar de determinísticos os interpoladores cujos contradomínios das funções de interpolação são determinados por operações que levam em conta apenas as configurações espaciais das amostras, desconsiderando-se a estrutura de variação do atributo no espaço. Assim, o processo de interpolação determinística envolve quatro

passos: (a) definição do raio de busca ou vizinhança em torno do ponto a ser estimado, (b) identificação dos dados pontuais presentes nesta vizinhança, (c) determinação da função matemática que representará a variação entre os pontos deste subconjunto e (d) estimação dos valores desconhecidos para os pontos na grade contínua. Este procedimento repete-se até que todos os pontos da grade sejam estimados. Dentre as funções determinísticas, as mais comumente utilizadas são as ponderadoras por inverso da distância, vizinho mais próximo e as funções locais de média, máximo e mínimo (Burrough e McDonnell,1998).

Os métodos geoestatísticos de interpolação foram desenvolvidos a partir do reconhecimento das limitações presentes nos métodos determinísticos. Estas limitações referem-se principalmente a incapacidade de entendimento da estrutura de correlação espacial a priori do fenômeno, informação necessária para a definição de funções ideais de ponderação (Burrough e McDonnell,1998). Nos métodos geoestatísticos, as amostras georreferenciadas pontuais são consideradas realizações de *variáveis aleatórias regionalizadas* e o processo de interpolação baseia-se no estudo da estrutura de variabilidade espacial destas. Neste contexto, os valores dos atributos são representados como *superfícies estocásticas* ou *campos aleatórios*, cujas funções de probabilidade modelam a incerteza dos valores estimados. Mais formalmente, diz-se que dentro de uma região \mathbf{A} da superfície terrestre, para cada posição $\mathbf{u} \in \mathbf{A}$, o valor do atributo $z(\mathbf{u})$ é modelado como uma variável aleatória $Z(\mathbf{u})$ (Felgueiras,1999) (Isaaks e Srivastava,1989).

Convencionou-se chamar de *krigeagem* as técnicas geoestatísticas de interpolação em reconhecimento aos estudos geológicos pioneiros de Daniel G. Krige (Krige,1951 *citado por* Camargo e Fuks,2002) a partir da segunda metade do século passado. Segundo Camargo e Fuks (2002), o que diferencia a krigeagem de outros métodos de interpolação é a estimação de uma matriz de covariância espacial que determina os pesos atribuídos às diferentes amostras, o tratamento de redundância dos dados, a vizinhança a ser considerada no procedimento inferencial e o erro associado ao valor estimado. Resumidamente, os procedimentos geoestatísticos de interpolação

envolvem três passos: (a) análise exploratória das amostras georreferenciadas, (b) análise da variabilidade estrutural do conjunto de dados através do cálculo e modelagem do variograma e, (c) realização de inferências ou a interpolação por krigagem propriamente dita.

Sendo os valores do atributo nos pontos $\mathbf{u} \in A$ modelados como variáveis aleatórias regionalizadas, $Z(\mathbf{u})$, dois enfoques para a determinação do modelo de distribuição destas variáveis são possíveis: (a) o *paramétrico*, que supõe um modelo de distribuição a priori que é definido completamente por um conjunto limitado de parâmetros, como no modelo de distribuição normal determinado pelos valores de média e variância; (b) o *não paramétrico*, que se caracteriza por não estabelecer nenhum modelo de distribuição a priori, mas sim construí-lo a partir de uma aproximação discretizada obtida sob um conjunto de valores estimados. Em ambos os casos, os modelos de distribuição probabilísticos são usados para a estimativa de valores do atributo em posições não conhecidas e modelagem da incerteza dos valores desses atributos (Felgueiras, 1999).

Neste capítulo serão apresentados dois métodos de estimação por krigagem, um paramétrico conhecido como *krigagem ordinária*, outro não paramétrico denominado *krigagem ordinária por indicação*.

7.1.1. Krigagem ordinária

A krigagem ordinária parte da hipótese de que não há variação significativa no fenômeno modelado em larga escala, assim supõe-se que há *estacionariedade de segunda ordem* e que a média de um atributo é constante e não depende da localização dentro de uma região de interesse. Dessa forma, $\mu(u) = \mu \forall u \in A$. A estacionariedade de segunda ordem implica que a covariância entre duas variáveis aleatórias regionalizadas depende apenas da distância que as separa representada pelo vetor $\overset{\frown}{h}$, assim $C(u, u + \overset{\frown}{h}) = C(\overset{\frown}{h})$. Felgueiras (1999) demonstra que partindo-se das premissas de não tendenciosidade do estimador, ou seja, que o erro ou resíduo de estimação deva ser zero,

e que a variância deste erro seja mínima, é possível calcular os valores de ponderação de krigeagem referente à amostra α e à posição u , $\lambda_\alpha(u)$. O estimador de krigeagem ordinária não requer o conhecimento a priori da média estacionária μ pois parte da condição de que a somatória dos ponderadores é igual a 1, assim $\sum_{\alpha=1}^{n(u)} \lambda_\alpha(u) = 1$. Com isso, explica-se a extrema robustez do algoritmo de krigeagem ordinária pois este opera com médias, ou tendências, estimadas localmente e não com uma única média estacionária preestabelecida como ocorre com outros estimadores de krigeagem como a krigeagem simples (Felgueiras,1999).

Estabelecido o arcabouço teórico da krigeagem ordinária, parte-se para o cálculo das covariâncias $C(u, u + \mathbf{h})$ necessário para a determinação dos ponderadores $\lambda_\alpha(u)$ utilizados no processo de estimação. As covariâncias são calculadas a partir de um *modelo teórico de semivariograma*, $\gamma(\mathbf{h})$, ajustado sobre o *semivariograma experimental*, $\hat{\gamma}(\mathbf{h})$, obtido interativamente a partir do conjunto de amostras georreferenciadas. O cálculo do semivariograma experimental se dá pela fórmula abaixo:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [z(u_\alpha) - z(u_\alpha + \mathbf{h})]^2$$

onde:

$N(\mathbf{h})$ é o número de pares de valores medidos, $z(u_\alpha)$ e $z(u_\alpha + \mathbf{h})$, separados pelo vetor distância \mathbf{h} .

O semivariograma é a ferramenta básica de suporte às técnicas de krigeagem pois quantifica a variação de um fenômeno regionalizado no espaço. As hipóteses de estacionariedade e média constante levam a postular um comportamento idealizado para o semivariograma experimental mostrado na figura 7.1. Neste, espera-se que as observações mais próximas geograficamente, tenham comportamento mais semelhante entre si do que aquelas separadas por distâncias maiores. Assim, a diferença

$[z(u_\alpha) - z(u_\alpha + \frac{1}{h})]$ deveria crescer à medida que aumenta a distância $\frac{1}{h}$ que as separa (Camargo e Fuks,2002).

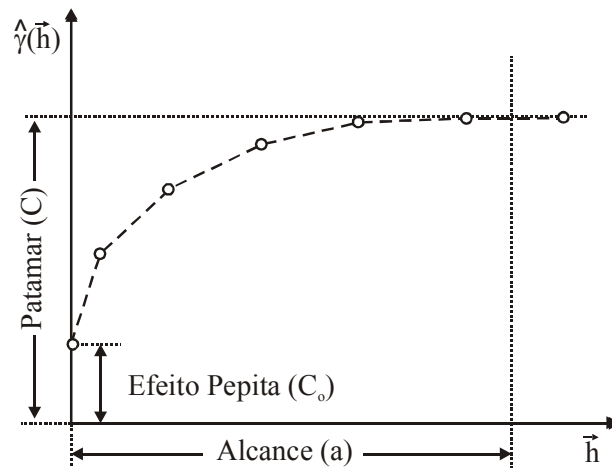


FIGURA.7.1 - Semivariograma experimental idealizado e seus parâmetros utilizados na definição de um modelo teórico de ajuste. Fonte:Camargo(2002)

Os modelos teóricos mais utilizados como modelos de ajuste sobre o semivariograma experimental são os modelos esférico, exponencial, gaussiano e potência. Os valores de efeito pepita, alcance e patamar obtidos na variografia experimental das amostras são utilizados na definição dos parâmetros do modelo teórico de ajuste. O procedimento de ajuste não é direto e automático, como em uma regressão (e.g.), mas sim interativo, nesse processo o usuário faz um primeiro ajuste e verifica a adequação do modelo teórico, deste primeiro semivariograma é possível uma percepção inicial da estrutura de variabilidade espacial, principalmente em termos de seu alcance, valor relacionado a distância dentro da qual as amostras apresentam-se espacialmente correlacionadas. Parte-se então para um refinamento interativo dos parâmetros até que se consiga um modelo de ajuste satisfatório. É justamente neste processo que reside a principal dificuldade de aplicação de ferramentas geoestatísticas em comparação aos métodos determinísticos, pois esta etapa exige do usuário a correta interpretação dos semivariogramas e do significado de seus parâmetros para que o melhor modelo transitivo seja aplicado. Em Camargo e Fuks (2002) são apresentados os principais modelos de ajuste e suas equações normalizadas.

7.1.2. *Krigeagem ordinária por indicação*

A principal crítica que se faz ao uso da krigeagem ordinária é em relação a utilização de sua medida de variância como medida de incerteza de estimação. Isso porque na maioria das vezes a hipótese de distribuição gaussiana da distribuição local dos erros é difícil de se constatar e a variância obtida através da krigeagem linear depende unicamente da configuração geométrica das amostras e não do valor do seu atributo fornecendo apenas indicações sobre áreas onde a amostragem poderia ser melhorada (Isaaks e Srivastava,1989) (Felgueiras,1999) (Felgueiras,2002).

A krigeagem por indicação, também conhecida como krigeagem probabilística, é uma forma não linear de krigeagem na qual o conjunto original de dados são transformados, ou codificados, de uma escala contínua em uma escala binária (i.e.- 1 se $z(u_\alpha) \leq K_n$, e 0 em caso contrário, onde K_n é o valor de corte). A krigeagem por indicação, diferentemente da krigeagem ordinária onde o interesse é obter a melhor estimativa de $z(u)$, fornece a *probabilidade* que o valor do atributo em questão exceda determinado limiar (Burrough e McDonnell,1998) (Felgueiras,2002). Ela é indicada em muitos casos onde interessa definir áreas com maior ou menor probabilidade de que o evento ocorra. Neste sentido, o trabalho de Bönisch (2001) é exemplar na aplicação desta informação em zoneamentos pedoclimáticos.

A principal vantagem deste método geoestatístico de interpolação, além da medida de incerteza estimada a partir dos valores do conjunto amostral, é sua aplicabilidade mais geral. A krigeagem por indicação é uma técnica não paramétrica, portanto não restrita a modelagem de atributos com distribuições simétricas como a gaussiana; permite modelar tanto atributos de natureza numérica como temática e; pode ser usada para modelar atributos com alta variabilidade espacial sem necessidade de filtragem de “outliers” (Felgueiras,2002). As limitações desta metodologia estão associadas principalmente ao alto grau de interatividade do usuário para a definição do número de valores de corte (codificação) e na definição de um variograma para cada um dos conjuntos codificados, demandando assim um tempo de trabalho significativamente maior. Portanto, a decisão sobre a aplicação deste método deve estar condicionada a

relevância da estimativa da incerteza ao estudo e a análise da distribuição de suas amostras.

Está fora do escopo deste trabalho o detalhamento dos conceitos que formalizam o estimador de krigeagem por indicação, para isto recomenda-se consultar o trabalho de (Felgueiras,1999) que é minucioso na apresentação deste estimador tanto para dados de natureza numérica quanto temática.

7.2. A REPRESENTAÇÃO DA ESTRUTURA INTRA-URBANA COMO GEO-CAMPOS

7.2.1. A continuidade espacial intra-urbana

A representação de dados como geo-campos parte da abstração de que o fenômeno analisado é, em essência, espacialmente contínuo. Quando aplicado na adjetivação das representações do tipo geo-campos, o termo *contínuo* está relacionado a duas idéias que se originam em contextos conceituais diferentes: o matemático associado à *variação numérica contínua* de um atributo sobre o espaço-tempo e o geográfico associado a *recobrimento espacial ininterrupto* desta variável sobre uma determinada região da superfície terrestre. A continuidade espacial intra-urbana aqui discutida está relacionada ao contexto geográfico, e neste sentido é possível dizer que a distribuição da população tende a ser contínua, já que a concentração de atividades e a alta densidade populacional são aspectos característicos destes ambientes. Assim, a escala intra-urbana se contrapõe a escala regional caracterizada por agrupamentos populacionais que, mesmo interconectados por redes de infra-estrutura, estão separados por extensões territoriais de baixa densidade populacional. Entretanto, esta continuidade espacial intra-urbana depende, primeiro, da configuração física do sítio que condiciona o assentamento da população a determinadas áreas e, segundo, do arranjo espacial das atividades e das diferentes parcelas da população estabelecidas no território que, em última análise, remete ao conceito de estrutura intra-urbana.

Nos aspectos fisicoterritoriais, a continuidade da ocupação do território está intrinsecamente condicionada ao suporte original, à geografia física do sítio. Neste

sentido, pode-se dizer que a cidade do Rio de Janeiro, em comparação a cidade de São Paulo, é menos contínua devido a presença dos grandes maciços rochosos e corpos d'água que determinaram sua ocupação. O recobrimento territorial da área urbanizada pode ser revelado através da utilização de imagens de satélite e fotos aéreas, constituindo-se como importantes fontes de dados auxiliares na construção e interpretação das superfícies numéricas.

A continuidade espacial da representação das atividades e características de grupos populacionais estabelecidos no espaço intra-urbano também está condicionada ao que se pode chamar genericamente *uso do solo intra-urbano*. Tomando como exemplo a modelagem de geodados populacionais censitários, relacionados principalmente a levantamentos domiciliares, a continuidade espacial destes dados está comprometida à distribuição das áreas residenciais. Da mesma forma, os geodados socioeconômicos territoriais cadastrais apresentam descontinuidades espaciais que devem ser consideradas quando representados como superfícies contínuas. Estas questões de recobrimento territorial de atributos são relevantes tanto para representações zonais como para contínuas, porém nestas últimas colocam-se como pontos fundamentais uma vez que estão na origem desta representação e determinam os alcances e limitações subjacentes a sua aplicação, sobretudo em dados socioeconômicos.

7.2.2. A interpolação a partir de dados zonais

A principal diferença entre os processos de interpolação de dados ambientais e populacionais, além da natureza do fenômeno, é o formato dos dados de entrada. A interpolação de dados ambientais geralmente é aplicada em dados pontuais relacionados a pontos de coleta ou medição de propriedades físicas ou químicas em determinadas localizações (e.g.- teor de argila no solo, temperatura, altitude, etc.). Dados populacionais, apesar de coletados individualmente, ou pontualmente, dificilmente são acessados neste formato. Assim, o processo de construção de superfícies contínuas populacionais geralmente é realizado a partir de dados zonais. As conseqüências disto são: (a) estes dados estão sujeitos às interferências relativas ao MAUP como visto no

capítulo 5 e; (b) dificilmente pode-se localizar o *centro de massa* do atributo a um ponto exato dentro da zona a qual ele está associado. Desta forma, toda a superfície que se construa a partir de dados zonais estará influenciada por estes fatores e qualquer análise que sobre elas se aplique deverá considerá-los.

O procedimento mais aplicado para a geração de superfícies populacionais consiste em atribuir o valor do atributo associado a uma unidade de área, onde admite-se uma distribuição de densidade constante por todo interior da área (Fig.7.2a), a um ponto interno desta área (geralmente seu centróide) (Fig.7.2b), e assim obter um conjunto de pontos no espaço aos quais são aplicados os procedimentos de interpolação que geram as superfícies numéricas. Em Martin (1996a) expõem-se as possibilidades de representações de dados populacionais e a função de densidade correspondente a cada uma delas (Fig.7.2).

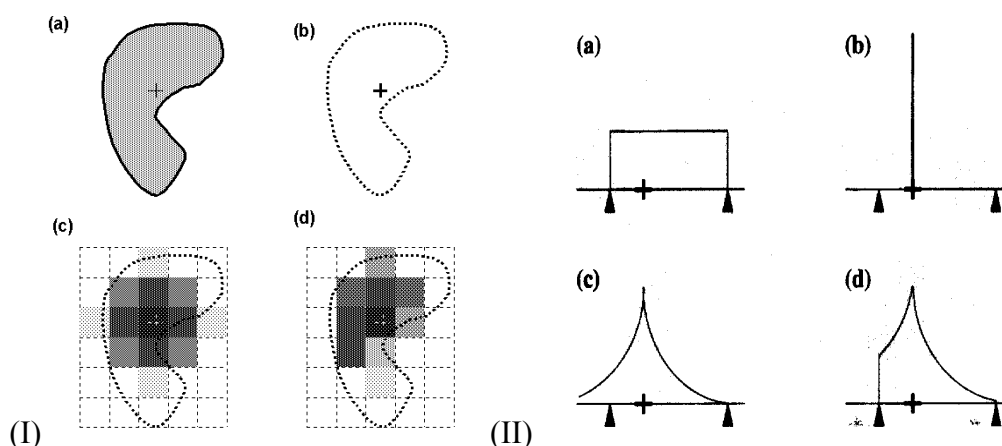


FIGURA.7.2. Os métodos de representação cartográfica para dados populacionais (I), e as funções de densidade associadas a cada um deles (II). a) representação zonal (coroplético); b) pontual (centróides); c) modelo de superfície sem restrição; d) modelo de superfície com restrição zonal.

FONTE: Martin (1996).

Segundo Martin (1995), os métodos de transição de dados em modelos zonais para modelos de superfícies baseados em técnicas tradicionais de interpolação apresentam duas limitações. A primeira é a incapacidade de preservar o volume de população correto sob cada uma das unidades de área (Fig.7.2c), havendo o que se pode definir como o transbordamento da informação de uma zona para as outras. A segunda

está relacionada à incapacidade de reconstituir as discontinuidades espaciais do atributo, como as regiões despovoadas, por exemplo.

A partir destas limitações, diferentes técnicas de interpolação foram desenvolvidas. O *método picnofilático* de interpolação, desenvolvido por Tobler (Burrough,1998) (Martin,1995) (Pina,2001), garante que o volume de um atributo sob uma unidade de área seja preservado (Fig7.2d). A aplicação deste método exige que o atributo seja transformado em uma função de densidade. Os valores máximo e mínimo obtidos por interpolação picnofilática são geralmente muito maiores em módulo do que os valores medidos, o que em certas situações pode acarretar problemas de interpretação. A questão da representação de áreas descontínuas foi enfrentada por Martin (1996a) através de uma adaptação do método do estimador Kernel de densidade de pontos (Bailey e Gattrel,1995), associado às idéias de preservação de volume desenvolvidas por Tobler, gerando o que se convencionou denominar de *método dos centróides populacionais* ou Método de Martin. Apesar de tratar as principais dificuldades dos métodos tradicionais de interpolação para dados populacionais, os resultados obtidos por este método dependem de um componente interativo que é a definição por parte do usuário do raio de busca que determina quais centróides populacionais serão considerados na estimativa de cada ponto da superfície interpolada. Uma avaliação deste método, e dos resultados sob diferentes especificações de raios de busca, é apresentada em Pina (2001) em experimento realizado para a Cidade do Rio de Janeiro sobre os dados censitários do IBGE.

7.2.3. A krigagem na interpolação de geodados zonais socioeconômicos

Nos exemplos apresentados a seguir, a krigagem é aplicada na interpolação de dados zonais socioeconômicos. O procedimento adotado é semelhante aos anteriormente apresentados onde, primeiro associam-se os valores dos atributos zonais a um conjunto de pontos (no caso, os centróides dos polígonos) para, posteriormente, aplicarem-se os procedimentos de interpolação por krigagem (Fig7.3). A aplicabilidade deste procedimento dependerá do objetivo da análise. Caso este for a observação de tendências globais do atributo sobre o território, a técnica se coloca como uma boa

opção devido a capacidade de análise da estrutura de variabilidade espacial a partir da variografia. Caso o objetivo da análise seja a estimação do valor do atributo em nível local, ou pontual, a aplicação deste procedimento apresenta fortes limitações.

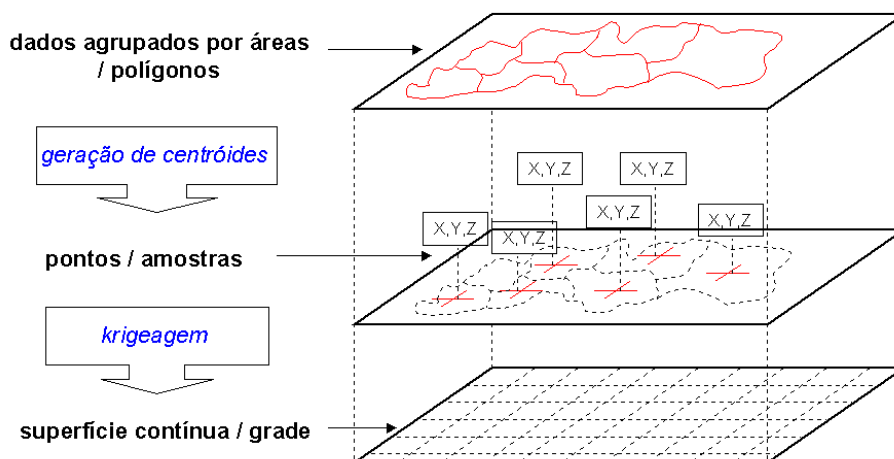


FIGURA.7.3. Esquema ilustrativo das etapas do processo de construção de superfícies contínuas a partir de dados zonais por krigagem.

As principais limitações associadas à aplicação da krigagem na interpolação de dados zonais são: (a) é um método de interpolação sem restrição zonal, ou seja, o valor do atributo associado a uma unidade de área não é mantido sob a zona após a interpolação; (b) não considera as discontinuidades espaciais do atributo e; (c) parte de hipóteses estatísticas, como normalidade da distribuição, nem sempre podem ser comprovadas. Entretanto, a krigagem também apresenta vantagens na sua utilização quando comparada às outras técnicas. As ferramentas de exploração estatísticas como o variograma e correlograma podem ser utilizadas mesmo quando o objetivo da análise não seja a interpolação como indicador da estrutura de variabilidade dos dados zonais. A krigagem é a única das técnicas descritas que considera esta informação no processo de interpolação. Na próxima seção, serão apresentados alguns experimentos realizados para a RMSP aplicando krigagem ordinária e por indicação em interpolações de dados zonais.

7.3. Experimentos

Os métodos de krigagem ordinária e probabilística foram aplicados aos dados OD87 e OD97 da RMSP com o objetivo de analisar o potencial destas ferramentas em um experimento concreto sobre dados intra-urbanos. A idéia foi utilizar a modelagem como geo-campos de um mesmo atributo das pesquisas OD de 87 e 97 para estudar a evolução deste atributo neste período a partir da comparação das superfícies construídas com a mesma resolução espacial, rompendo assim com a impossibilidade de comparação que a mudança do suporte zonal entre estas duas pesquisas acarretou (figura 4.1.c e 4.1.d). Para a realização deste estudo foi selecionada a variável número total de automóveis por zona OD em relação à população total desta zona (variável auto_pop). O objetivo específico foi avaliar a expansão da presença da população motorizada nos territórios da cidade e obter indicações sobre as tendências espaciais desta expansão. Os procedimentos descritos a seguir foram realizados sobre os dados OD87 e replicados aos OD97.

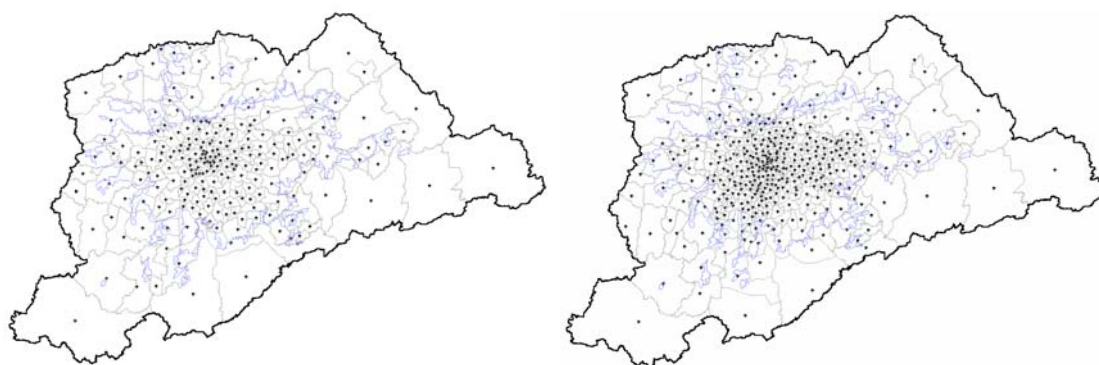


FIGURA.7.4 - À esquerda, conjunto amostral de 254 pontos gerado a partir dos centróides dos polígonos OD87; à direita, os 389 pontos gerados pelo mesmo método para os polígonos OD97 e mancha urbana extraída de imagem landsat de 1997.

O primeiro passo no processo de interpolação de dados zonais é a geração do conjunto de pontos, ou amostras, aos quais são associados o valor do atributo em estudo. No caso do conjunto de dados OD, os valores dos atributos zonais foram associados aos centróides dos polígonos. Assim, 254 pontos configuraram o plano amostral da OD87 e 389 pontos na OD97 (Fig.7.4). Os conjuntos amostrais apresentam

uma distribuição desigual de pontos, havendo uma concentração de amostras na parte central com uma diminuição gradativa desta densidade de pontos diretamente proporcional a distância em relação ao centro. Como as amostras mais espaçadas relacionam-se às zonas de baixa densidade populacional, é esperado que sua representatividade no processo de interpolação seja insuficiente, assim a análise das superfícies geradas a partir deste conjunto amostral também será acompanhada da informação da extensão da mancha urbana.

7.3.1. Krigagem ordinária

A partir da geração dos conjuntos amostrais, partiu-se para as análises estatísticas de suas distribuições para a determinação da aplicabilidade ou não do método de krigagem ordinária. Nos casos analisados, o conjunto de amostras da variável `auto_pop OD97` apresentou um comportamento coerente à suposição de normalidade de sua distribuição, desta forma não apresentando grandes impedimentos na aplicação da krigagem ordinária no processo de interpolação. Já para as amostras relativas à OD87, esta suposição de normalidade foi feita apesar da clara assimetria apresentada pela análise do histograma (Fig.7.5).

Outro aspecto investigado antes da variografia foi a condição isotrópica ou anisotrópica do conjunto amostral. Diz-se que o conjunto amostral de um atributo é anisotrópico quando a variabilidade ou distribuição espacial deste atributo ocorre mais intensamente numa direção e menos em outra. Quando a variabilidade tende a ser semelhante para todas as direções, diz-se que este conjunto amostral apresenta um comportamento isotrópico (Camargo,2002). As análises de detecção de anisotropia realizadas através do semivariograma de superfície (Fig.7.6) apontaram para um comportamento quase isotrópico, com uma pequena anisotropia no eixo nordeste-sudoeste desconsiderada nos procedimentos de variografia.

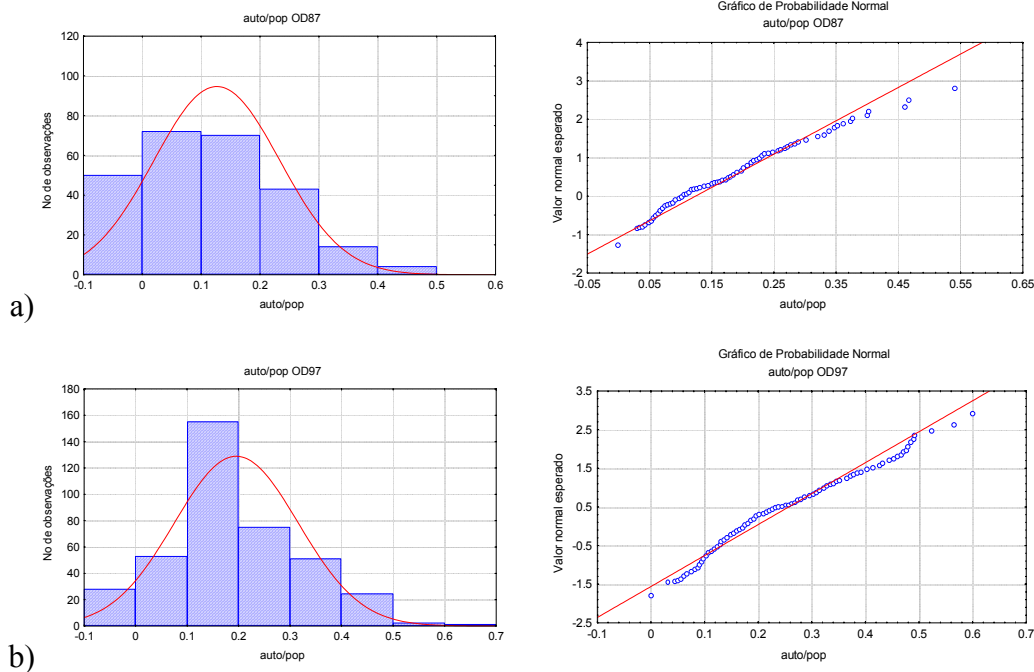


FIGURA.7.5 - Análise das distribuições dos conjuntos amostrais de auto_pop OD87 (a); e auto_pop OD97 (b)

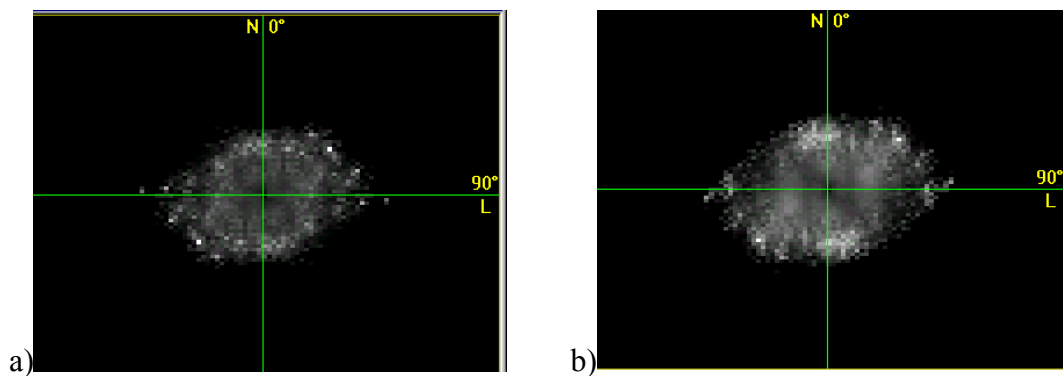


FIGURA.7.6 - Superfícies de semivariograma referentes aos campos de amostras para auto_pop OD87 (a) e auto_pop OD97 (b).

Assim, sob a suposição de normalidade do conjunto amostral e admitindo-se um comportamento isotrópico do atributo, partiu-se para a determinação interativa do semivariograma experimental para os conjuntos de dados. Esperava-se que a suposição de estacionariedade e média constante subjacente aos procedimentos de krigagem ordinária impusessem dificuldades para a determinação de um semivariograma experimental para este tipo de dado. Porém, tanto na variografia para OD87 quanto para

OD97, lograram-se determinar semivariogramas bastante próximos das funções exponenciais ajustada a eles (Fig.7.7).

Os alcances dos semivariogramas apontam para a existência de correlação espacial dentro de distâncias da ordem de 25Km. Esta é praticamente a distância entre a fronteira norte e a sul da mancha urbana em 1997. No sentido leste-oeste, esta distância aproxima-se de três quartos da distância que separa as fronteiras da mancha urbana. Uma interpretação que se pode dar a este alcance é que ele está relacionado ao padrão de correlação intra-urbana da cidade de São Paulo, e que a partir desta distância as amostras não mais tenderiam a apresentar este padrão de correlação, caracterizando-se como amostras não relacionadas à estrutura intra-urbana. Isto também é determinado pela resolução espacial na qual foram calculados os semivariogramas experimentais. A complexidade espacial intra-urbana dificilmente poderia ser modelada por um semivariograma calculado em resolução espacial muito alta, pois diferentemente dos dados geofísicos, nas cidades é comum que mesmo dados espacialmente próximos apresentem uma grande variabilidade. Neste caso, a variografia consegue capturar um comportamento coerente com as possibilidades de ajustes de funções transitivas ideais apenas quando os *lags*, ou as distâncias $\frac{1}{h}$, são suficientemente grandes para que o semivariograma experimental não seja afetado pelo ruído que a variabilidade local intra-urbana tende a imprimir. Nas análises aqui apresentadas, os semivariogramas foram determinados para *lags* de 3Km para as amostras OD97 e 2.9Km para OD87. Assim, definiu-se como resolução da grade interpolada unidades espaciais com lados de 500x500 metros, ou *pixels* de 0,25Km², dimensão aproximada de quatro quadras agrupadas.

Através das superfícies de tendências geradas (Fig.7.9) é possível visualizar o incremento da população motorizada em todos os pontos da cidade no período compreendido entre 1987 e 1997. Nota-se a expansão da área mais motorizada do quadrante sudoeste já consolidada em 87, e também crescimento expressivos da taxa de motorização no subcentro do ABC, no ponto correspondente ao Jardim Anália

Franco na região leste e na zona norte. No subcentro de Mogi das Cruzes também é possível visualizar uma tendência do incremento do número de automóveis.

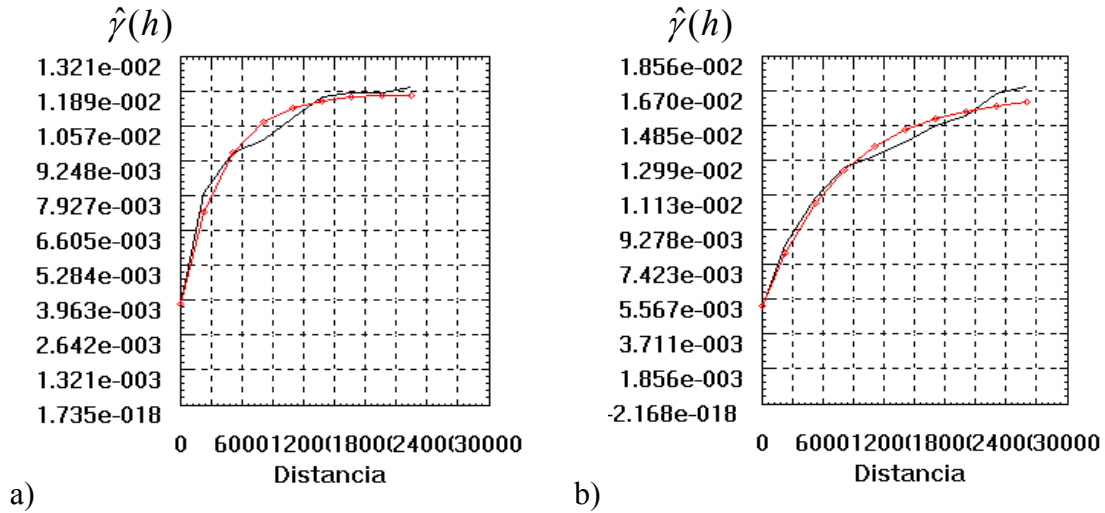


FIGURA.7.7 - Semivariogramas experimentais calculados para as amostras auto_pop OD87 (a) e auto_pop OD97 (b) (em preto); e respectivas funções exponenciais ajustadas a eles (em vermelho).

A análise dos histogramas dos erros de estimação indica um bom desempenho dos procedimentos adotados. As distribuições são simétricas, apresentam baixa variância e médias próximas a zero, apontando para uma estimação não tendenciosa nos dois casos (Fig.7.8).

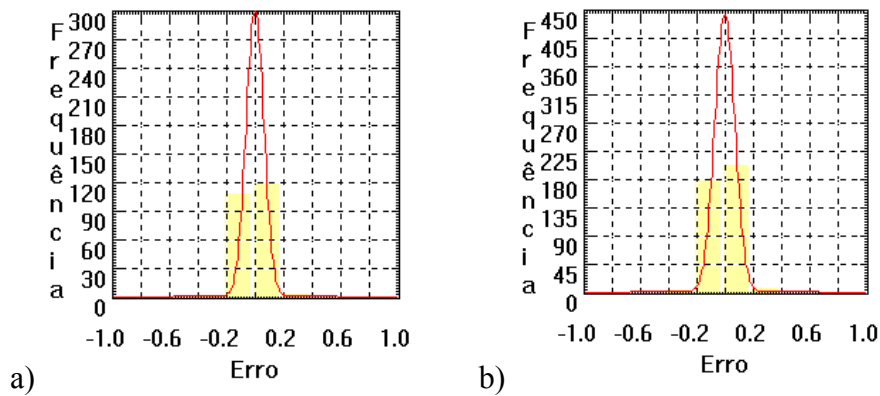


FIGURA.7.8 - Histogramas dos erros de estimação para as funções de interpolação por krigagem ordinária para auto_pop OD87 (a) e auto_pop OD97 (b).

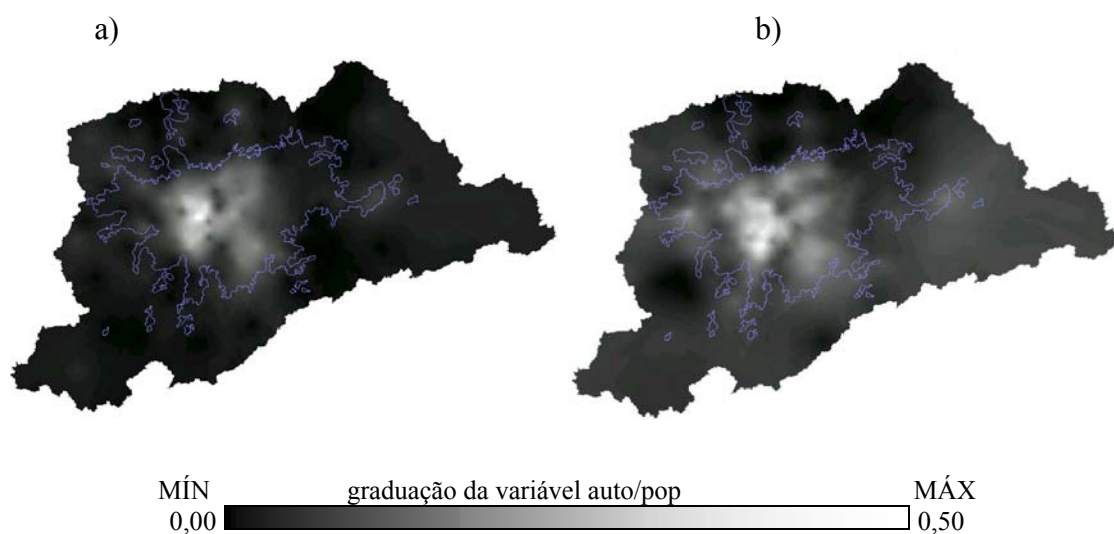


FIGURA.7.9 - Superfícies contínuas geradas por krigagem ordinária para a variável auto_pop OD87(a) e OD97(b) representadas por variações em níveis de cinza.

Apesar da representação por geo-campos ser uma boa alternativa para a visualização de tendências, a excessiva suavização que o processo de krigagem ordinária impõem aos dados ainda representa limitações na sua utilização em alguns procedimentos de análises intra-urbanas. Como alternativa, a aplicação de métodos geoestatísticos não paramétricos, como a krigagem probabilística, podem representar uma opção, pois tendem a representar melhor as variações mais abruptas dos valores e associar a estes locais uma medida de incerteza indicadora desta variabilidade.

7.3.2. Krigagem ordinária por indicação

Os procedimentos de krigagem por indicação partem do estabelecimento dos valores de corte que definirão os conjuntos amostrais binários para os quais são calculados os semivariogramas experimentais. No estudo de caso aqui apresentado, foram definidos para os conjuntos de amostras, OD97 e OD87, nove conjuntos binários cujos valores de corte, K_n , corresponderam aos decís de cada conjunto. A utilização dos decís como valores de corte permitiu que um bom número de valores de probabilidade acumulada fossem estimados melhorando a precisão da aproximação discretizada da real distribuição de probabilidade do atributo analisado. A tabela 7.1. apresenta os parâmetros dos modelos de semivariogramas teóricos ajustados aos semivariogramas

experimentais para os conjuntos amostrais binários definidos pelos decís para a OD87 e OD97.

A análise dos resultados numéricos da krigagem probabilística aplicada sobre as amostras OD87 e OD97 de auto/pop aponta para uma dificuldade na modelagem da variabilidade espacial das amostras pertencentes aos últimos decís do conjunto de valores georreferenciados nos dois anos. Percebe-se que o valor do efeito pepita predomina sobre o valor da contribuição indicando ineficácia na determinação de uma correlação entre as amostras, acarretando assim uma imprevisibilidade nos resultados inferidos a partir dos valores observados (Felgueiras,1999). Como o efeito pepita é o valor da semivariância para a distância zero e representa a componente da variabilidade espacial que não pode ser relacionado com uma causa específica (variabilidade ao acaso) (Camargo,2002), é provável que esta variabilidade modelada seja mais função de um ruído aleatório local do que a estrutura de covariância do conjunto amostral. Pode-se observar que o alcance obtido nestes variogramas são muito pequenos, isso reduz o universo de localizações, ou de pares de amostras com correlações computadas, a um número pouco representativo em relação à extensão do recobrimento territorial da variável analisada.

Como a krigagem por indicação estima os valores a partir de uma função de distribuição acumulada (fdac) aproximada, é de se esperar que esta incapacidade de capturar uma estrutura de correlação nos últimos decís interfere na aproximação da distribuição acumulada da variável nas n localizações da grade a ser estimada. Isto impõe sérias limitações à utilização deste procedimento quando o objetivo da análise é a melhor estimativa numérica dos valores pontuais desconhecidos. Porém, em análise de tendências como no caso aqui analisado, esta limitação não impede sua aplicação, desde de que a análise dos resultados seja criteriosa, sobretudo quanto a interpretação da variabilidade local nas grades computadas.

Auto_pop OD87 (N° amostras=254)					
	Valor	Modelo	Efeito Pepita	Contribuição	Alcance
1º corte*	0,000	-	-	-	-
2º corte	0,031	Gaussiano	0,052	0,189	20544,025
3º corte	0,060	Gaussiano	0,076	0,231	18842,785
4º corte	0,076	Gaussiano	0,069	0,264	17571,813
5º corte	0,103	Gaussiano	0,095	0,235	16006,201
6º corte	0,137	Esférico	0,132	0,167	13782,514
7º corte	0,181	Esférico	0,151	0,093	10791,801
8º corte	0,210	Esférico	0,125	0,081	4634,581
9º corte	0,271	Gaussiano	0,109	0,032	3916,190

Auto_pop OD97 (N° amostras=389)					
	Valor	Modelo	Efeito Pepita	Contribuição	Alcance
1º corte	0,064	Gaussiano	0,028	0,107	17190,549
2º corte	0,098	Gaussiano	0,055	0,193	17381,024
3º corte	0,125	Gaussiano	0,084	0,215	17702,533
4º corte	0,146	Gaussiano	0,114	0,196	15653,839
5º corte	0,173	Esférico	0,139	0,161	14090,654
6º corte	0,196	Esférico	0,128	0,157	12125,414
7º corte	0,244	Esférico	0,123	0,116	9216,619
8º corte	0,300	Esférico	0,067	0,127	4678,365
9º corte	0,368	Esférico	0,079	0,047	5426,495

TABELA.7.1 - Parâmetros dos modelos de semivariogramas teóricos ajustados aos semivariogramas experimentais para os conjuntos amostrais binários definidos pelos decis para a OD87 e OD97. (* não se estabeleceu variografia para o primeiro decil pois o conjunto com valores nulos apresentava número maior que 26 elementos)

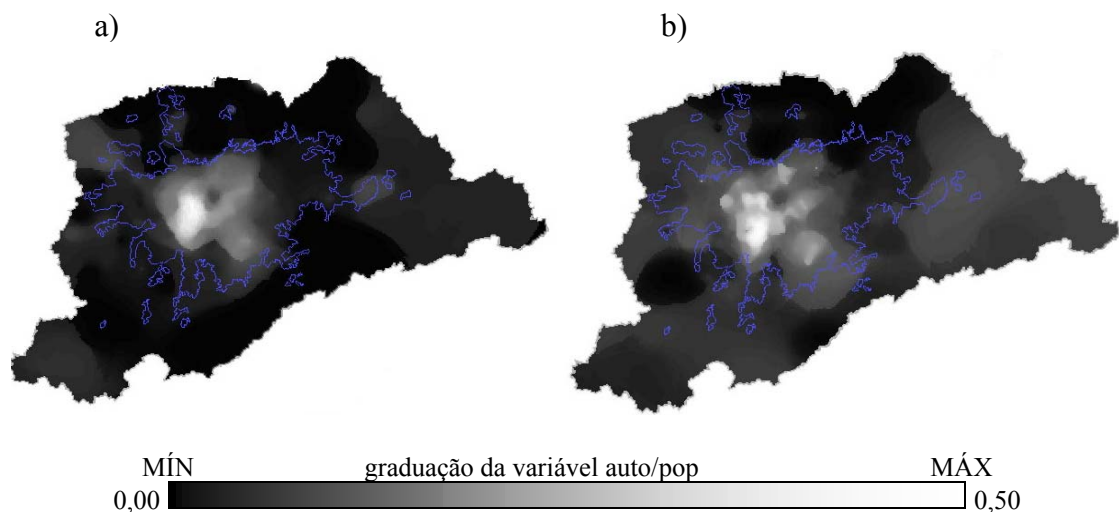


FIGURA.7.10 - Superfícies contínuas geradas por krigagem probabilística para a variável auto_pop OD87(a) e OD97(b) representadas por variações em níveis de cinza.

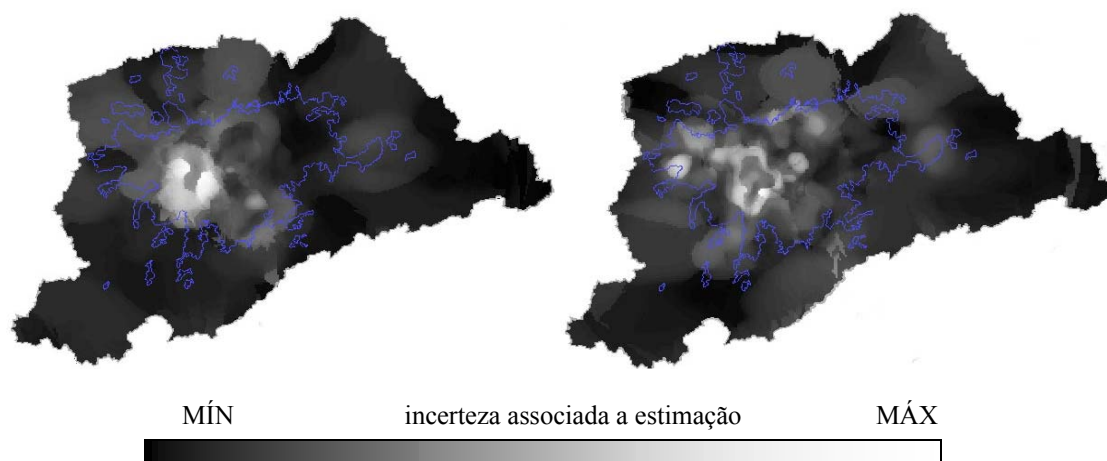


FIGURA.7.11 - Grade de incerteza associada as grades estimadas por krigeagem probabilística definida por distâncias interdecís. (a) OD87 e (b) OD97.

As superfícies geradas por krigeagem probabilística (Fig.7.10) lograram, como esperado, preservar melhor as variações abruptas dos valores do atributo nos conjuntos amostrais. Se comparados aos resultados obtidos pela krigeagem ordinária (Fig.7.9), percebe-se que o espalhamento do valor do atributo é significativamente menor nas superfícies geradas por krigeagem por indicação. Outro aspecto positivo desta metodologia é a informação de incerteza associada que permite uma interpretação simultânea dos resultados obtidos e do grau de confiabilidade destes sobre o território. No caso dos dados OD, percebe-se que o locais de maior incerteza são aqueles onde os valores das amostras georreferenciadas próximas tem maior variabilidade, indicando pontos de transições de padrões ou locais de não estacionariedade (Fig.7.11). Pode-se notar na grade de incerteza associada aos dados OD97 que o centro da mancha de alta concentração de população motorizada no quadrante sudoeste indica baixo grau de incerteza, porém circundada por um cinturão de alta incerteza, local associado a transição do padrão espacial.

7.4. Conclusões

Os métodos de construção de campos numéricos a partir de dados zonais aqui apresentados mostraram-se muito eficientes para a representação de tendências de variáveis, sobretudo quando o interesse da análise é a evolução do padrão de distribuição desta variável no espaço-tempo. A comparação das grades computadas a

partir da variável `auto_pop` das pesquisas OD87 e OD97 permite uma leitura eficiente da evolução da distribuição espacial da variável no espaço intra-urbano, e rompe assim com a leitura fragmentada que a representação zonal impõe ao analista. A figura 7.12 apresenta as grades geradas para os dois conjuntos de dados por krigagem ordinária e probabilística classificadas por faixas de valores onde é possível observar a maior suavização que o procedimento paramétrico imprime ao resultado.

Os resultados obtidos no experimento realizado para São Paulo foram satisfatórios frente ao objetivo de representar a evolução da população motorizada dentro da estrutura intra-urbana. A leitura que este tipo de representação propicia é clara e direta, percebe-se que a área sudoeste de concentração da taxas de motorização mais altas expandiu-se em todas as direções neste período e que as áreas secundárias de concentração de altas taxas de motorização que já se desenhavam em 87, consolidaram-se e expandiram-se, e em 97, já se colocavam como evidentes pontos de concentração de população motorizada. Apesar da expansão de algumas áreas como o ABC, o leste próximo (Tatuapé e Jardim Anália Franco) e Santana, o Centro e as periferias ainda são os pontos mais baixos da grade. Este padrão que se revela em 87, mantém-se em 97, apontando para uma certa estabilização e consolidação da estrutura espacial desta variável.

Outro aspecto importante na utilização de interpoladores estocásticos é a introdução da variografia como ferramenta exploratória dos dados georreferenciados capaz de revelar sua estrutura de correlação espacial, informação relevante para determinação de áreas de influência, transições e comportamento errático que podem ser entendidas como expressões computacionais possíveis de conceitos relacionados com a noção de estrutura territorial. A geoestatística ainda apresenta fortes restrições na sua utilização em construção de geo-campos a partir de dados zonais quando o objetivo é recuperar o valor desconhecido da variável pontualmente. Isso se deve ao próprio processo de geração do conjunto amostral, no caso atribuição do valor zonal a um ponto interno a zona, onde muitas vezes estas não são pontos representativos do fenômeno analisado. Os pontos característicos do atributo, como valores máximos e mínimos,

assim como os pontos de inflexão são importantes para que o interpolador possa representar adequadamente a variação do atributo (Felgueiras,1999).

Neste sentido, dois caminhos apresentam-se como pontos de interesse de trabalhos futuros, um é a formulação de métodos mais eficazes de geração do conjunto amostral pontual envolvendo dados auxiliares como imagens e levantamentos em campo. Outro é a formulação de uma metodologia específica de aplicação de técnicas geoestatística na interpolação de dados de natureza socioeconômica, talvez incorporando aspectos contemplados pelos métodos de construção de superfícies de densidade já existentes.

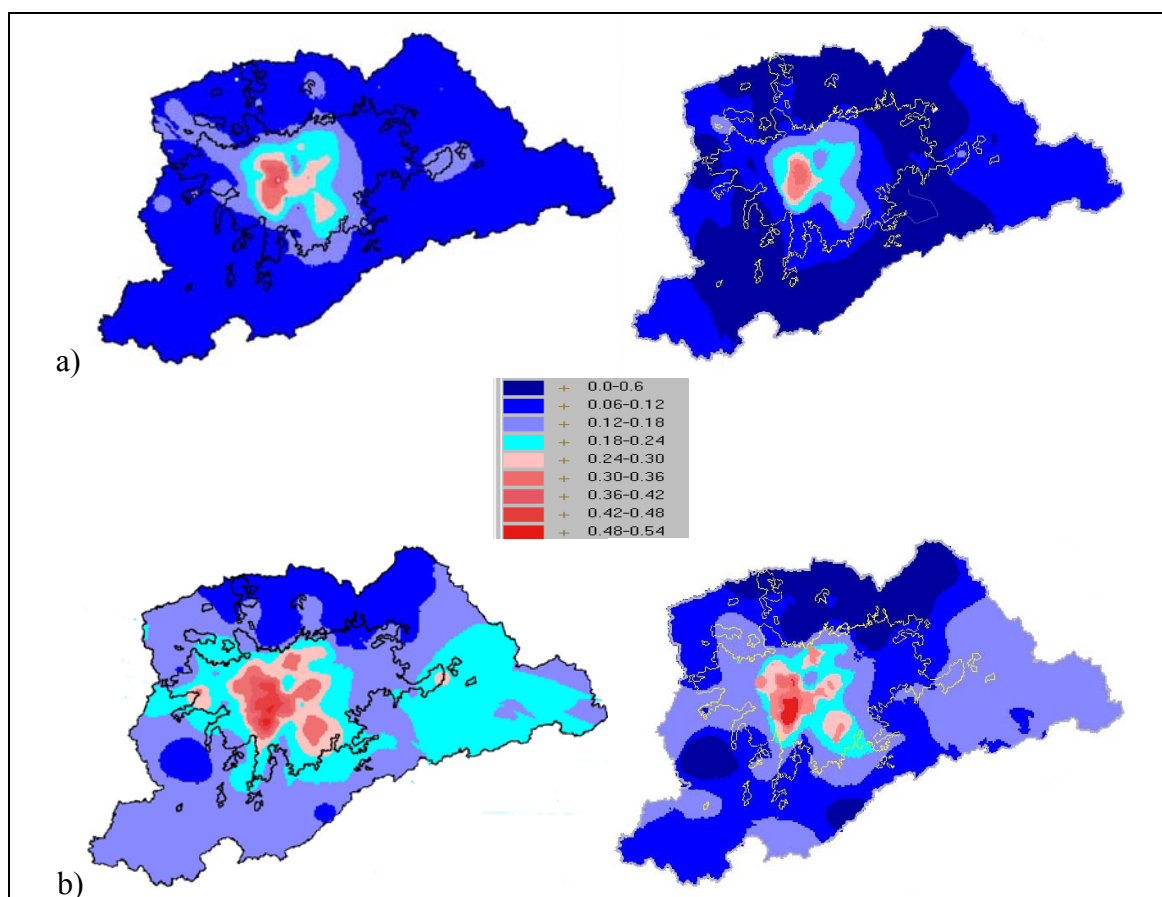


FIGURA.7.12 - Representação temática por faixas de valores de auto_pop das superfícies contínuas geradas por krigagem para OD87(a) e OD97(b). À direita: método de krigagem por indicação; à esquerda: método de krigagem ordinária.

CAPÍTULO 8

CONCLUSÕES

8.1. Síntese das Contribuições do Trabalho

A principal contribuição deste trabalho foi trazer um novo posicionamento crítico no trato da informação geográfica quantitativa dentro da área de urbanismo. Este novo posicionamento nasce no reconhecimento da natureza representacional dos geodados digitais e se consolida nas possibilidades de manipulações matemático-computacionais que o ambiente digital propicia. Para que estas possibilidades se efetivem, faz-se necessário que se estabeleça a ponte entre as referências dos universos teórico-conceitual do geoprocessamento e do urbanismo. A transposição destas referências para a representação digital impõem uma redução de dimensionalidade do fenômeno que deve ser controlada. O controle sobre este processo se fundamenta no domínio das técnicas matemáticas subjacentes a este ambiente. A aplicação da tecnologia do geoprocessamento amplia as possibilidades analíticas em estudos intra-urbanos. Entretanto, estas possibilidades somente se viabilizarão sob o domínio de um conjunto de conhecimentos que, num primeiro momento, são estranhos ao urbanismo. Este trabalho, ao explorar as possíveis expressões computacionais dos conceitos urbanísticos em ambiente SIG, contribuiu para a consolidação desta ponte conceitual entre os dois universos de conhecimento apostando numa crescente utilização desta tecnologia em estudos desta natureza.

Este trabalho percorreu as etapas da aplicação do geoprocessamento em estudos intra-urbanos, a partir de estudos de caso sobre os dados da pesquisa OD de São Paulo. A figura 8.1. apresenta, de forma esquemática, a seqüência de etapas organizadas em dois grandes grupos relacionados a momentos distintos do processo: a modelagem de dados e a análise espacial. O primeiro momento se caracteriza pela identificação dos conceitos e entidades relevantes para o estudo, dentro do universo conceitual do urbanismo e nas possibilidades de representação destes em ambiente computacional. No

capítulo 3 deste trabalho, procurou-se definir e caracterizar os universos abstracionais envolvidos na modelagem de estruturas intra-urbanas nos diferentes níveis, partindo da discussão sobre o domínio ontológico e chegando ao desenho do banco em modelo OMT-G.

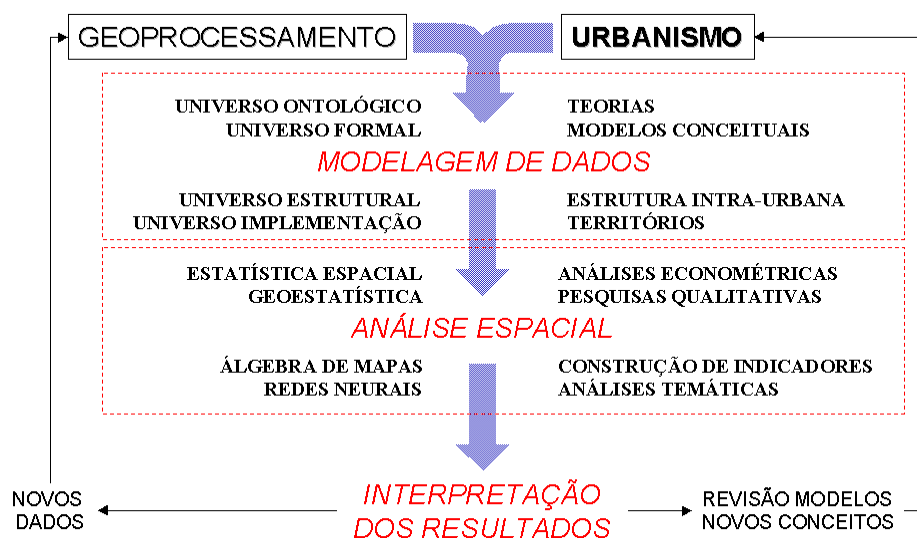


FIGURA.8.1 - Descrição esquemática das etapas envolvidas no processo de aplicação de análises urbanísticas em ambiente SIG.

O segundo momento trata da exploração das possibilidades de manipulações matemático-computacionais que esta tecnologia incorpora, demonstrando que a utilização de SIGs para dados intra-urbanos pode e deve ir além das funções de armazenamento e apresentação destes dados em forma de cartografias digitais. A possibilidade de incorporação da componente espacial aos modelos matemáticos permite dar um passo além na direção do entendimento do vínculo que se estabelece entre as mais diversas situações da vida cotidiana e o território. Três temas relacionados às possibilidades de aplicação de ferramentas de Análise Espacial foram abordados no capítulo 5: o MAUP, os LISAs e a krigeagem. Em estudos intra-urbanos, onde é bastante difundida a utilização de informações sob forma de taxas e indicadores, saber reconhecer as interdependências que estas medidas sofrem em função de sua natureza geográfica é fundamental para sua correta interpretação. Sempre que o dado é apresentado como característica estatística de agregações de indivíduos deve-se

considerar o MAUP. Saber reconhecê-lo e controlá-lo foi o tema do capítulo 5. No capítulo 6, discutiu-se a autocorrelação espacial como característica inerente aos dados geográficos. Os LISAs se mostraram bastante eficazes em análises de padrões de distribuições espaciais, como concentrações, persistências e áreas de influência de variáveis georreferenciadas intra-urbanas. O capítulo 7 apresenta ferramentas geoestatísticas para a geração de superfícies contínuas a partir de dados socioeconômicos. Estas técnicas se mostraram importantes em análises de tendências, revelando padrões encobertos pela fragmentação que a representação zonal impõe aos dados.

8.2. O estado atual da tecnologia e seu horizonte

Em seu estado atual, a principal limitação que a tecnologia enfrenta está relacionada ao alcance dos modelos espaço-temporais na representação da realidade geográfica. O arcabouço representacional que a dicotomia geo-campos e geo-objetos configura, apesar de sólido e difundido, ainda se coloca como um fator limitante na representação de fenômenos socioespaciais. As dimensões mais subjetivas das relações sociais dificilmente podem ser capturadas como camadas matriciais sobrepostas ou a conjuntos de geometrias arranjadas no espaço euclidiano. Hoje, pode-se dizer que as representações computacionais de fenômenos socioespaciais se apoiam em modelos inicialmente desenvolvidos para dados de natureza fisicoterritorial. O desafio é formular novos modelos capazes de integrar e relacionar as múltiplas dimensões que configuram o universo do discurso urbano. Tais modelos passarão pela formulação de novas abstrações que tragam expressões computacionais possíveis a estes territórios. Nesta direção, uma promissora linha de pesquisa é aquela que incorpora a tradição quantitativa da construção de indicadores sociais ao repertório da análise espacial apontando para novas perspectivas na construção de *indicadores sociais espacialmente*

sensíveis que vem sendo reforçada em outros países como demonstram os estudos recentes no Reino Unido¹.

Nas representações temporais, aspecto também fundamental neste tipo de análise, há uma distância ainda maior a ser percorrida. Até hoje, pouco se conseguiu introduzir da informação temporal nos modelos, sendo esta usualmente apresentada como seqüências estáticas de eventos sucessivos e não incorporada como informação relevante. O desafio da incorporação de modelos dinâmicos em análises geográficas quantitativas está além da esfera mais restrita da modelagem de dados em geoprocessamento. Na realidade este desafio está colocado nas mais diversas áreas do conhecimento preocupadas com a previsibilidade em sistemas complexos, e neste campo há ainda mais indagações do que conclusões. A meteorologia é uma das ciências mais interessadas em estudar padrões temporais em sistemas complexos e muito de suas inferências são realizadas em ambiente SIG. Talvez novos modelos representacionais para dados socioeconômicos, como campos vetoriais e mapeamentos em espaço fase, possam ser desenvolvidos para ampliar as possibilidades da representação temporal. Espera-se daí, que a dimensão temporal possa, cada vez mais, ser incorporada nas inferências baseadas em geoprocessamento abrindo novas perspectivas ao uso desta tecnologia.

8.3. Trabalhos Futuros

A seqüência imediata dos estudos apresentados nesta dissertação deverá estar direcionada para experimentação de outras ferramentas analíticas não contempladas aqui. Dentre estas, aquelas baseadas em modelos de redes, fundamentadas na propriedade topológica de conectividade, podem revelar novas dimensões sobre este conjunto de dados ainda não exploradas neste trabalho. A pesquisa Origem-Destino, como o próprio nome revela, traz informação relevante para os estudos intra-urbanos

¹ como no *indices of deprivation 2000 (DETR- Department of environment, 2000)* no Reino Unido e as iniciativas do censo inglês e suas OAs (Martin, 2000)

relacionada aos fluxos, aos deslocamentos da população dentro da cidade. Uma idéia inicial simples é relacionar a oferta de determinado serviço existente em uma zona com o fluxo por ela atraído em função deste serviço. Este tipo de informação é relevante para a compreensão da dinâmica intra-urbana, permitindo mapear acessos e impedimentos aos serviços que localizados, compõem também a estrutura intra-urbana.

Há também que se avançar nos estudos apresentados nos três capítulos de Análise Espacial. Com relação aos experimentos sobre os efeitos do MAUP, a utilização de algoritmos de agregação automática em processos ótimos de agregação zonal podem representar novas perspectivas no controle destes efeitos. Ao mesmo tempo, pode-se avançar nas técnicas bayesianas incorporando-se a medida de vizinhança ao modelo, restringindo assim sua ponderação à sua média local. Dentro das experiências com os LISAs, novas medidas de proximidade como definidoras de vizinhança podem ser avaliadas, baseadas não só em contiguidade mas também em medidas de acessibilidade, como tempo de viagem por exemplo, introduzindo o conceito de autocorrelação espacial também nas inferências baseadas em redes. Por fim, relacionados aos modelos de superfícies, experimentos com os métodos picnofiláticos e de Martin poderão ser realizados e comparados com os resultados obtidos por krigeagem com operações de álgebra de mapas sobre as grades resultantes. Enfim, há na literatura muitas ferramentas analíticas que podem ser aplicadas sobre os dados da pesquisa OD e do mapa da Exclusão/Inclusão de São Paulo e a cada resultado novos dados e outras possibilidades podem se abrir, instaurando ciclos de análises realimentados pela oferta crescente de geodados digitais intra-urbanos.

8.4. Considerações Finais

Os resultados apresentados e discutidos no decorrer desta dissertação são frutos de um longo e intenso processo de aprendizagem onde muitas barreiras e preconceitos foram superados. Este processo teve início na inquietação causada pela dificuldade de compreensão da totalidade de uma cidade como São Paulo. As cidades se configuram neste início de século como o hábitat de preferência de 80% dos brasileiros, hoje no

Brasil somos 136 milhões de habitantes urbanos. Porém, conquistamos isto percorrendo claramente a *rota do desastre*. Os urbanistas se colocam hoje diante de um quadro onde sequer a clareza do objeto de análise existe. O que se vê é a perplexidade diante de um fenômeno incontrolável, cenário das mais graves mazelas da população.

Atuar nestas cidades exige que as conheçamos e as entendamos e este trabalho espera haver contribuído em demonstrar que muitas são as possibilidades que a tecnologia hoje oferece. Porém, somente o posicionamento crítico em relação a elas, baseado no reconhecimento de seus alcances e limitações, pode assegurar o êxito deste encontro. O encontro do urbanista e dos novos instrumentos matemático-computacionais a sua disposição para auxiliá-lo na tarefa cotidiana da construção de um melhor entendimento das cidades, sem, no entanto, jamais esquecer, como nos ensinou o professor Milton Santos que “geometrias não são geografias”.

REFERÊNCIAS BIBLIOGRÁFICAS

- Agre, P. E. **Toward a critical technical practice: lessons learned in trying to reform AI** [on-line],2000. <<http://dlis.gseis.ucla.edu/people/pagre/critical.html>>, set2000
- Anselin, L. Local indicators of spatial association-LISA. **Geographical Analysis**, v.27, n.2, 93/115, 1995.
- Anselin, L. **SpaceStat tutorial** [on line],1992. <<http://www.spacestat.com>>,1999
- Argan, G. C. **Arte moderna**. São Paulo: Companhia das Letras, 1995. 709p.
- Augé, M. **Não-lugares: Introdução a uma antropologia da supermodernidade**. Campinas: Papirus, 1994. 111p.
- Bailey, T.; Gattrel, A. C. **Interactive spatial data analysis**. London: Longman, 1995.
- Batty, M. Visualizing urban dynamics, In:P. B. Longley, M. ed. **Spatial analysis: modelling in a GIS Environment**, New York:John Wiley & Sons, 1996,p. 297-320
- Bönisch, S. **Geoprocessamento ambiental com tratamento de incerteza: o caso do zoneamento pedoclimático para a soja no Estado de Santa Catarina**. Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2001.
- Borges, K. **Modelagem de Dados Geográficos: Uma Extensão do Modelo OMT para Aplicações Geográficas**. Dissertação (Mestrado, Departamento de Ciência da Computação) - UFMG, Belo Horizonte, 1997.
- Brissac, N.; Dercon, C. **Como intervir em grande escala?** projeto artecidade brasmitte,2001. <<http://www.uol.com.br/artecidade/urbanismo.htm>>, jul. 2001
- Burrough, P.; McDonnell, R. A. **Principles of Geographical Information Systems**. Oxford: Oxford University Press, 1998. 333p.
- Caldeira, T. **Cidade de muros: crime, segregação e cidadania em São Paulo**. São Paulo: Edusp, 2000.
- Câmara, G.; Casanova, M.A.; Hemerly, A.S.; Magalhães,G.C.; Medeiros, C.M.B. **Anatomia de sistemas de informação geográfica**. Campinas: Instituto de Computação-UNICAMP, 1996. 193p.
- Câmara, G.; Barbosa, C. C. F.; Davis, C.; Fonseca, F. **Conceitos Básicos em Geoprocessamento**,1999. [on-line] <<http://www.dpi.inpe.br/gilberto/livro/cap2-conceitos.pdf>>, fev.2000
- Câmara, G.; Monterio, A.M.V., Medeiros, J.S. **Representações computacionais do espaço: um diálogo entre a geografia e a ciência da geoinformação**. In: Workshop sobre Novas Tecnologias em Ciências Geográficas, 2000, Rio Claro. **Anais...** Rio Claro: UNESP, 2000.
- Câmara, G.; Monteiro, A.M.V. Geocomputation Technics for Spatial Analysis: Is It the case for Health Data? **Cadernos de saúde pública**, v.17, n.5, p.1059-1081, 2001a.

- Câmara, G.; Monteiro, A.M.V. **Introdução ao geoprocessamento: Aula 1- conceitos básicos**,2001b. [on-line].
<<http://www.dpi.inpe.br/cursos/ser300/aulas/aula1-definicoes>>, jun 2001
- Câmara, G.; Monteiro, A.M.V. **Introdução ao geoprocessamento: Aula 2- modelos de dados geográficos** 2001c. [on-line].
<<http://www.dpi.inpe.br/cursos/ser300/aulas/aula2-modelos>>, jun 2001
- Câmara, G.; Carvalho, M.S.; Cruz, O.G.; Correa, V. **Análise espacial de áreas**. São José dos Campos: INPE, 2002. [on-line]
<<http://www.dpi.inpe.br/gilberto/tutoriais/analise/cap2.pdf>> jun. 2004.
- Camargo, E. C. G.; Fuks, S.D. Análise espacial de superfícies por geoestatística, In: Fuks, Suzana Druck; Carvalho, Marília Sá; Câmara, Gilberto; Monteiro, Antonio Miguel Vieira. eds. **Análise espacial de dados geográficos**. [on-line] www.inpe.br/gilberto/livro/analise/. Publicado como (INPE-8558-PRE/4302)
- Campos Filho, C. M. **Cidades brasileiras: seu controle ou o caos**. São Paulo: Nobel, 1989. 143p.
- Castells, M. **A questão urbana**. São Paulo: Editora Paz e Terra, 2000. 590p.
- Cauvin, C. Cartographier l'invisible de la ville. **La Recherche**, n. 337, p.48, dez. 2000.
- Companhia do metropolitano de São Paulo - METRÔ **Manual da Pesquisa O/D Domiciliar 1997**. São Paulo: METRÔ/STM, 1997.
- Corbusier, L. **A carta de Atenas**. São Paulo: Hucitec USP, 1989.
- Couclelis, H. People manipulate objects (but cultivate fields): beyond the raster-vector debate in GIS., In:A. U. Frank, Campari, I., Formentini, U. ed. **Theories and methods of spatio-temporal reasoning in geographic space**, Berlin:Springer, 1992, p. 65-77.
- Crosta, A. P. **Processamento digital de imagens de sensoriamento remoto**. Campinas: IG/UNICAMP, 1992. 170p.
- Davis, C. Modelagem de Dados Geográficos (II). **InfoGEO Set/Out**. p38-46. 1998
- Davis, C.; Borges,K. **Modelagem semântica em Geoprocessamento**,1999.[online] <<http://www.dpi.inpe.br/gilberto/livro/cap18-modelagem.pdf>>, fev.2000.
- Department of environment - DETR, Measuring multiple deprivation at the small area level: the indices of deprivation 2000. **Regeneration Research Summary**, n.37, 2000. [online]. <<http://www.regeneration.detr.gov.uk>> abr. 2002
- Echenique, M. El concepto de sistemas, modelos y teorías en los estudios urbanos, In:M. Echenique ed. **Modelos matemáticos de la estructura espacial urbana: aplicaciones en América Latina**. Buenos Aires:Ediciones Nueva Visión S.A., 1975.
- Empresa Paulista de Planejamento Metropolitano SA - EMPLASA. **Sumário de dados da grande São Paulo**. São Paulo: EMPLASA, 1999. 1CD-ROM.
- Felgueiras, C. A. **Modelagem ambiental com tratamento de incertezas em sistemas de informações geográficas: paradigma geoestatístico por**










- indicação.** Tese (Doutorado em Computação Aplicada) - Instituto Nacional de Pesquisas Espaciais. São José dos Campos, 1999.
- Felgueiras, C. A.; Fuks, S.D.; Monteiro, A.M.V. Análise de superfícies por geoestatística indicadora. In: Fuks, S. D.; Carvalho, M. S; Câmara, G; Monteiro, A. M.V. **Análise espacial de dados geográficos.** [on-line]. www.inpe.br/gilberto/livro/analise/. Publicado como: INPE-8559-PRE/4303.)
- Fonseca, F.; Egenhofer, M.; Davis, C.; Borges, K. **Ontologies and knowledge sharing in urban GIS** [on-line],2000.
<<http://www.spatial.maine.edu/~fred/#Publications>>, jul.2001
- Gattrel, A. C. Autocorrelation in spaces. **Environment and Planning A**, v.11, p. 507-516, 1979.
- Gattrel, A. C. Concepts of space and geographical data, In: Maguire, D; Goodchild, M.; Rhind, D. ed. **Geographical information systems: principles and applications.** Essex:Longman Scientific & Technical, 1991. v.1, p. 119-134
- Gehlke, C. E.; Biehl, K. Certain effects of grouping upon the size of correlation coefficient in census tract material. **Journal of American Statistical Association**, n.29, p. 169-170, 1934.
- Getis, A.; Ord, J.K. The analysis of spatial association by use of distance statistics. **Geographical Analysis**, 24, n.3, p.189-206, 1992.
- Getis, A.; Ord, J.K. Local spatial statistics: an overview, In: Longley, P.; Batty, M. ed. **Spatial analysis: modelling in a GIS environment**, New York:John Wiley & Sons, Inc., 1996,p. 390p
- Gomes, J. M.; Velho, L. **Computação visual: imagens.** Rio de Janeiro: SMB, 1995.
- Green, M.; Flowerdew, R. New evidence on the modifiable areal unit problem, In: Longley, P.; Batty, M. ed. **Spatial analysis: modelling in a GIS environment**, New York:John Wiley & Sons, 1996, p. 41-55
- Guez, A.; Stabilini, S.; Zedda, R. Les Italiens à l'heure des villes. **La Recherche**, n.337, p.30, dez. 2000.
- Hillier, B. **The common language of space: a way of looking at the social, economic and environmental functioning of cities on a common basis.** [on-line],1998.
<www.spacesyntax.com/publications/commonlang>. mar.2001
- Isaaks, E. H.; Srivastava, R.M. **An introduction to applied geostatistics.** New York: Oxford University Press, 1989. 561p.
- Koga, D. H. U. **Cidades Territorializadas entre enclaves e potências.** Tese (Doutorado em Serviço Social), Pontifícia Universidade Católica de São Paulo - PUCSP. São Paulo, 2001.
- Koolhaas, R.; Mau, B. **S,M,L,XL.** Rotterdam: 010 Publishers, 1995. 1344p.
- Kowarick, L. **Escritos urbanos.** São Paulo: Editora 34, 2000. 143p.
- Krige, D. G. A statistical approach to some basic mine evaluation problems on the Witwatersrand. **Johannesburg Chemistry Metallurgy Mining Society South African**, v6, n.52, p.119-139, 1951.
- Longley, P. B.; Batty, M. **Spatial analysis: modelling in a GIS environment.** New York: John Wiley & Sons, 1996.

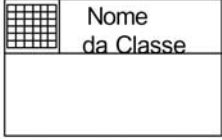
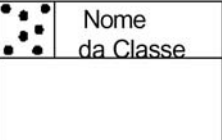
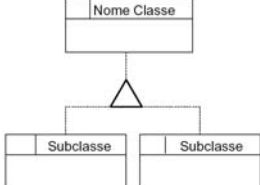
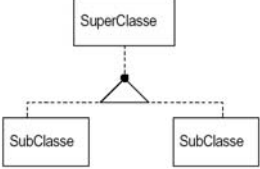
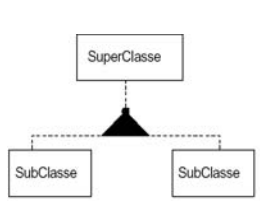
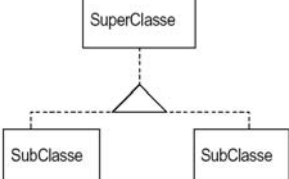
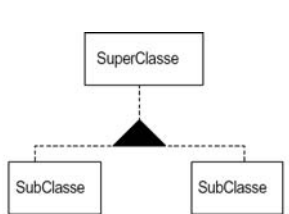
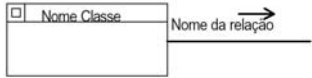
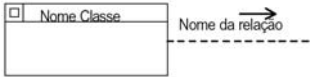
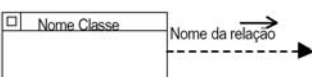
- Marcuse, P. The enclave, the citadel, and the guetto. What has changed in the Post-Fordist U.S. City. **Urban affairs review**, 33, n.2, p.228-264, 1997.
- Maricato, E. **Metrópole na periferia do capitalismo: ilegalidade, desigualdade e violência**. São Paulo: HUCITEC, 1996. 141p.
- Martin, D. **Geographic information systems: socioeconomic applications**. London: Routledge, 1995.
- Martin, D. An assessment of surface and zonal models of population. **International journal of geographical information systems**, v. 10, p. 973-989, 1996a.
- Martin, D. Depicting changing distributions through surface estimation, In: Longley, P.; Batty, M. ed. **Spatial analysis: modelling in a GIS environment**, New York: John Wiley & Sons, p. 105-122, 1996b
- Martin, D. **Census 2001: making the best of zonal geographies**. The Census of Population: 2000 and Beyond, University of Manchester, 2000.
- Neter, J.; Wasserman, W. **Applied linear statistical models**. Georgetown: Richard D. Irwin, 1974.
- Openshaw, S. An empirical study of some zone-design criteria. **Environment and Planning A**, v. 10, p. 781-794, 1978.
- Openshaw, S. Ecological fallacies and the analysis of areal census data. **Environment and Planning A**, v. 16, p. 17-31, 1984.
- Openshaw, S. Developing GIS-relevant zone-based spatial analysis methods, In: Longley, P.; Batty, M. ed. **Spatial analysis: modelling in a GIS environment**, New York: John Wiley & Sons, 1996, p. 390
- Paiva, J. A.; Casanova, M.A.; Câmara, G. **Arquiteturas para bancos de dados geográficos**, [on-line], 1998. <www.dpi.inpe.br/gilberto/livro> set. 2001
- Peuquet, D. J. A conceptual framework and comparison of spatial data models. **Cartographica**, v. 21, p. 66-113, 1984.
- Pina, M. F. **Aplicação e avaliação de técnica de interpolação espacial para a geração de superfícies de densidade a partir de dados populacionais em uma região do Município do Rio de Janeiro**. Tese de Doutorado, COPPE-UFRJ. Rio de Janeiro, 2001.
- Richards, J. A. **Remote Sensing Digital Image Analysis**. Berlim: Springer-Verlag, 1995. 340p.
- Robinson, A. H. Ecological correlation and the behaviour of individuals. **American Sociological Review**, n.15, p. 351-357, 1950.
- Rolnik, R. **A Cidade e a Lei**. São Paulo: Studio Nobel, 1999. 242p.
- Rolnik, R. **Reestruturação urbana da metrópole paulistana: análise de territórios em transição** [CD-ROM]. São Paulo: Instituto Pólis, 2000.
- Rufino, I. A. A.; Trigueiro, E.B.F.; Medeiros, V.A.S. **Geoprocessamento e análise sintática do espaço: estudo das relações entre vitalidade urbana e preservação arquitetural no centro histórico de Natal**. GISBrasil2001, 2001.
- Rumbaugh, J. **OMT insights: perspectives on modelling from the Journal of Object-Oriented Programing**. New York: SIGS Books, 1996. 390p.


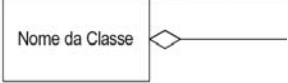
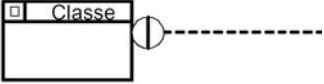
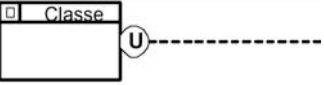
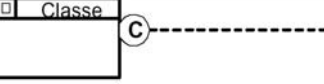
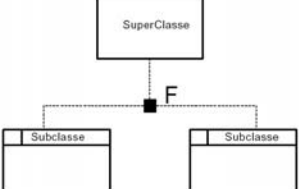
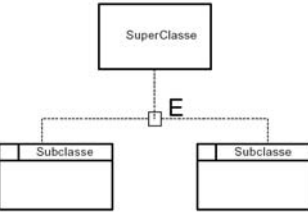
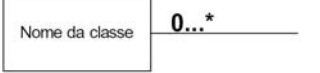
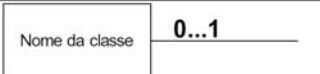
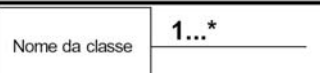

- Saboya, R. T. Análises Espaciais em Planejamento Urbano. **Revista Brasileira de Estudos Urbanos e Regionais**, n.3, 61-79, 2000.
- Santos, M. **A urbanização brasileira**. São Paulo: Editora Hugitec, 1993. 157p.
- Santos, M. **Metamorfozes do espaço habitado**. São Paulo: Editora Hucitec, 1997. 124p.
- Santos, M. **Territorio e Sociedade: entrevista com Milton Santos**. São Paulo: Fundação Perseu Abramo, 2000.
- Schowengert, R. A. **Remote sensing: models and methods for image processing**. New York: Academic Press, 1997.
- Schuurman, N. Critical GIS: Theorizing an Emerging Science. **Cartographica**, 36, n.4, 1999.
- Sposati, A. **Mapa da Exclusão/Inclusão Social da Cidade de São Paulo**. São Paulo: EDUC, 1996.
- Sposati, A. **Mapa da exclusão/inclusão social da cidade de São Paulo/2000** [CD-ROM]. São Paulo: Instituto Pólis, 2000.
- Tufte, E. R. **Envisioning Information**. Cheshire: Graphic Press, 1994. 126p.
- Villaça, F. **Espaço Intra-Urbano no Brasil**. São Paulo: Studio Nobel, 1998. 373p.
- Wrigley, N.; Holt, T.; Steel, D.; Trammer, M. Analysing, modelling, and resolving the ecological fallacy, In: Longley, P.; Batty, M. ed. **Spatial analysis: modelling in a GIS environment**, New York: John Wiley & Sons, 1996, p. 25-41

APÊNDICE A

LEGENDA- PICTOGRAMAS UTILIZADOS NO MODELO OMT-G

REPRESENTAÇÃO GRÁFICA	SIGNIFICADO
 <div style="border: 1px solid black; padding: 2px; display: inline-block; margin-left: 5px;">Nome da Classe</div>	<p style="text-align: center;">CLASSE GEORREFERENCIADA</p> <p>Geo-objeto: Linha <i>Ex.: Muro</i></p>
 <div style="border: 1px solid black; padding: 2px; display: inline-block; margin-left: 5px;">Nome da Classe</div>	<p style="text-align: center;">CLASSE GEORREFERENCIADA</p> <p>Geo-objeto: Ponto <i>Ex.: Árvore</i></p>
 <div style="border: 1px solid black; padding: 2px; display: inline-block; margin-left: 5px;">Nome da Classe</div>	<p style="text-align: center;">CLASSE GEORREFERENCIADA</p> <p>Geo-objeto: Polígono <i>Ex.: Lote</i></p>
 <div style="border: 1px solid black; padding: 2px; display: inline-block; margin-left: 5px;">Nome da Classe</div>	<p style="text-align: center;">CLASSE GEORREFERENCIADA</p> <p>Geo-objeto: Linha Unidirecionada <i>Ex.: Trecho da rede de esgoto</i></p>
 <div style="border: 1px solid black; padding: 2px; display: inline-block; margin-left: 5px;">Nome da Classe</div>	<p style="text-align: center;">CLASSE GEORREFERENCIADA</p> <p>Geo-objeto: Linha Bidirecionada <i>Ex.: Trecho da rede elétrica</i></p>
 <div style="border: 1px solid black; padding: 2px; display: inline-block; margin-left: 5px;">Nome da Classe</div>	<p style="text-align: center;">CLASSE GEORREFERENCIADA</p> <p>Geo-objeto: Nó <i>Ex.: Poço de visita</i></p>
 <div style="border: 1px solid black; padding: 2px; display: inline-block; margin-left: 5px;">Nome da Classe</div>	<p style="text-align: center;">CLASSE GEORREFERENCIADA</p> <p>Geo-campo: Rede triangular irregular <i>Ex.: TIN</i></p>
 <div style="border: 1px solid black; padding: 2px; display: inline-block; margin-left: 5px;">Nome da Classe</div>	<p style="text-align: center;">CLASSE GEORREFERENCIADA</p> <p>Geo-campo: Isolinha <i>Ex.: Curvas de nível</i></p>
 <div style="border: 1px solid black; padding: 2px; display: inline-block; margin-left: 5px;">Nome da Classe</div>	<p style="text-align: center;">CLASSE GEORREFERENCIADA</p> <p>Geo-campo: Polígonos Adjacentes <i>Ex.: Divisão de regiões administrativas</i></p>

 <p>Nome da Classe</p>	<p>CLASSE GEORREFERENCIADA</p> <p>Geo-campo: Grade Regular <i>Ex.: Imagem de Satélite</i></p>
 <p>Nome da Classe</p>	<p>CLASSE GEORREFERENCIADA</p> <p>Geo-campo: Amostras <i>Ex.: Pontos de coleta (altimetria-e.g.)</i></p>
 <p>Nome Classe</p> <p>Subclasse</p> <p>Subclasse</p>	<p>GENERALIZAÇÃO ESPACIAL</p> <p>As subclasses possuem atributos gráficos diferentes da superclasse no entanto, herdam os atributos alfanuméricos</p>
 <p>SuperClasse</p> <p>SubClasse</p> <p>SubClasse</p>	<p>Generalização Espacial Disjunta/total</p> <p>A superclasse é a união de subclasses disjuntas</p>
 <p>SuperClasse</p> <p>SubClasse</p> <p>SubClasse</p>	<p>Generalização Espacial Sobreposta/total</p> <p>A união de instâncias das subclasses equivalem ao conjunto das instâncias da superclasse, existindo instâncias que pertencem a mais de uma subclasse</p>
 <p>SuperClasse</p> <p>SubClasse</p> <p>SubClasse</p>	<p>Generalização Espacial Disjunta/parcial</p> <p>Nem todas instâncias da superclasse estão representadas nas subclasses disjuntas</p>
 <p>SuperClasse</p> <p>SubClasse</p> <p>SubClasse</p>	<p>Generalização Espacial Sobreposta/parcial</p> <p>Nem todas as instâncias da superclasse estão representadas nas subclasses e as instâncias existentes nas subclasses podem pertencer simultaneamente a mais de uma subclasse</p>
 <p>Nome Classe</p> <p>Nome da relação</p>	<p>ASSOCIAÇÃO SIMPLES</p>
 <p>Nome Classe</p> <p>Nome da relação</p>	<p>Relacionamento Espacial</p>
 <p>Nome Classe</p> <p>Nome da relação</p>	<p>HIERARQUIA ESPACIAL</p>

	<p>RELACIONAMENTO EM REDE</p>
	<p>AGREGAÇÃO</p>
	<p>AGREGAÇÃO ESPACIAL Subdivisão Espacial (subdividido em)</p>
	<p>AGREGAÇÃO ESPACIAL União Espacial (união de)</p>
	<p>AGREGAÇÃO ESPACIAL Contém</p>
	<p>Generalização Cartográfica pela Forma SOBREPOSTA/TOTAL</p>
	<p>Generalização Cartográfica pela escala DISJUNTA/TOTAL</p>
	<p>CARDINALIDADE Zero ou mais</p>
	<p>CARDINALIDADE Zero ou um</p>
	<p>CARDINALIDADE Um ou mais</p>
	<p>CARDINALIDADE Exatamente um</p>

Fonte: Davis, 1999

APÊNDICE B

Modelo OMT-G do BDG

