



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

INPE-8556-PRE/4300

ANÁLISE ESPACIAL E GEOPROCESSAMENTO

Gilberto Câmara
Antônio Miguel Monteiro
Suzana Druck Fucks
Marília Sá Carvalho

ANÁLISE ESPACIAL DE DADOS GEOGRÁFICOS □
Instituto Nacional de Pesquisas Espaciais – INPE, São José dos Campos, SP, Brazil.

INPE
São José dos Campos
2002

ANÁLISE ESPACIAL E GEOPROCESSAMENTO

Gilberto Câmara

Antônio Miguel Monteiro

Suzana Druck Fucks

Marília Sá Carvalho

1.1 INTRODUÇÃO

Compreender a distribuição espacial de dados oriundos de fenômenos ocorridos no espaço constitui hoje um grande desafio para a elucidação de questões centrais em diversas áreas do conhecimento, seja em saúde, em ambiente, em geologia, em agronomia, entre tantas outras. Tais estudos vem se tornando cada vez mais comuns, devido à disponibilidade de sistemas de informação geográfica (SIG) de baixo custo e com interfaces amigáveis. Estes sistemas permitem a visualização espacial de variáveis como população de indivíduos, índices de qualidade de vida ou vendas de empresa numa região através de mapas. Para tanto, basta dispor de um banco de dados e de uma base geográfica (como um mapa de municípios), e o SIG é capaz de apresentar um mapa colorido permitindo a visualização do padrão espacial do fenômeno.

Além da percepção visual da distribuição espacial do problema, é muito útil traduzir os padrões existentes com considerações objetivas e mensuráveis, como nos seguintes casos:

- Epidemiologistas coletam dados sobre ocorrência de doenças. A distribuição dos casos de uma doença forma um padrão no espaço? Existe associação com alguma fonte de poluição? Evidência de contágio? Variou no tempo?
- Deseja-se investigar se existe alguma concentração espacial na distribuição de roubos. Roubos que ocorrem em determinadas áreas estão correlacionados com características sócio-econômicas dessas áreas?
- Geólogos desejam estimar a extensão de um depósito mineral em uma região a partir de amostras. Pode-se usar essas amostras para estimar a distribuição do mineral na região?
- Deseja-se analisar uma região para fins de zoneamento agrícola. Como escolher as variáveis explicativas – solo, vegetação, geomorfologia – e

determinar qual a contribuição de cada uma delas para definir em que local o tipo de cultura é mais adequado?

Todos esses problemas fazem parte da *análise espacial de dados geográficos*. A ênfase da Análise Espacial é mensurar propriedades e relacionamentos, levando em conta a localização espacial do fenômeno em estudo de forma explícita. Ou seja, a idéia central é incorporar o espaço à análise que se deseja fazer. Esse livro apresenta um conjunto de ferramentas visando responder a essas questões. Pretende-se auxiliar os interessados a estudar, explorar e modelar processos que se expressam através de uma distribuição no espaço, aqui chamados de fenômenos geográficos.

Um exemplo pioneiro, onde intuitivamente se incorporou a categoria espaço às análises realizadas foi realizado no século XIX por John Snow. Em 1854, ocorria em Londres uma das várias epidemias de cólera trazidas das Índias. Pouco se sabia então sobre os mecanismos causais da doença. Duas vertentes científicas procuravam explicá-la: uma relacionando-a aos miasmas, concentrados nas regiões baixas e pantanosas da cidade, e outra à ingestão de água insalubre. O mapa (Figura 1-1) localiza a residência dos óbitos ocasionados pela doença e as bombas de água que abasteciam a cidade, permitindo visualizar claramente uma destas – em Broad Street – como o epicentro da epidemia. Estudos posteriores confirmaram esta hipótese, corroborada por outras informações tais como a localização do ponto de captação de água desta bomba a jusante (rio abaixo) da cidade, em local onde a concentração de dejetos, inclusive de pacientes coléricos era máxima. Essa é uma situação típica onde a relação espacial entre os dados contribuiu significativamente para o avanço na compreensão do fenômeno, sendo um dos primeiros exemplos da análise espacial.

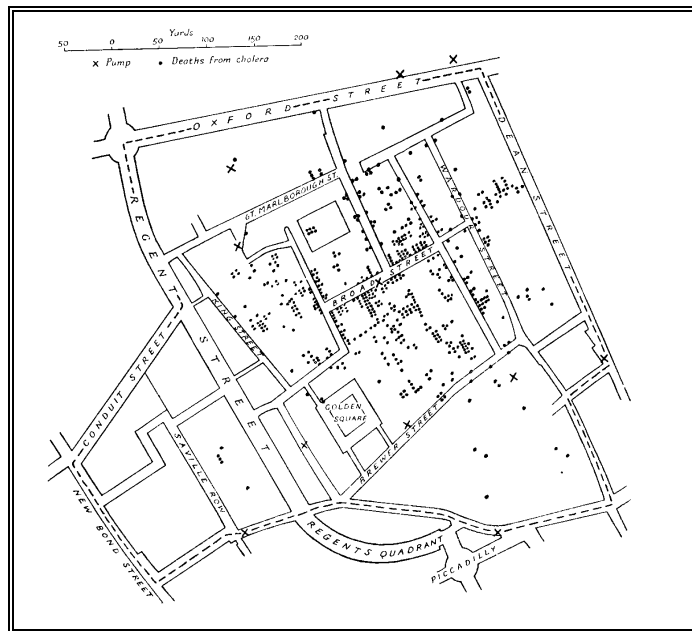


Figura 1-1 - Mapa de Londres com óbitos por cólera identificados por pontos e poços de água representados por cruzes.

1.2 TIPOS DE DADOS EM ANÁLISE ESPACIAL

A taxonomia mais utilizada para caracterizar os problemas de análise espacial considera três tipos de dados:

- *Eventos ou Padrões Pontuais* - fenômenos expressos através de ocorrências identificadas como pontos localizados no espaço, denominados processos pontuais. São exemplos: localização de crimes, ocorrências de doenças, e localização de espécies vegetais.
- *Superfícies Contínuas* - estimadas a partir de um conjunto de amostras de campo, que podem estar regularmente ou irregularmente distribuídas. Usualmente, este tipo de dados é resultante de levantamento de recursos naturais, e que incluem mapas geológicos, topográficos, ecológicos, fitogeográficos e pedológicos.
- *Áreas com Contagens e Taxas Agregadas* - tratam-se de dados associados a levantamentos populacionais, como censos e estatísticas de saúde, e que originalmente se referem a indivíduos localizados em pontos específicos do espaço. Por razões de confidencialidade, estes dados são agregados em unidades de análise, usualmente delimitadas por polígonos fechados (setores censitários, zonas de endereçamento postal, municípios).

A partir da divisão acima, verifica-se que os problemas de análise espacial lidam com dados *ambientais* e com dados *socioeconômicos*. Em ambos os casos, a análise espacial é composta por um conjunto de

procedimentos encadeados cuja finalidade é a escolha de um *modelo inferencial* que considere explicitamente os relacionamentos espaciais presentes no fenômeno. Em geral, o processo de modelagem é precedido de uma fase de *análise exploratória*, associada à apresentação visual dos dados sob forma de gráficos e mapas e a identificação de padrões de dependência espacial no fenômeno em estudo.

No caso de *análise de padrões de pontos*, o objeto de interesse é a própria localização espacial dos eventos em estudo. Como na situação analisada por Snow, o objetivo é estudar a distribuição espacial destes pontos, testando hipóteses sobre o padrão observado: se é aleatório, ou ao contrário se apresenta-se em aglomerados ou regularmente distribuído. É também o caso dos estudos visando estimar o sobre-risco de doenças ao redor de usinas nucleares. Outro caso é estabelecer o relacionamento de ocorrência de eventos com características do indivíduo, incorporando a possibilidade de haver algum fator ambiental, do qual não se dispõe de dados. Por exemplo, será que a mortalidade por tuberculose, mesmo considerando os fatores de risco conhecidos, varia com o local de residência do paciente? As técnicas usuais no tratamento deste tipo de problema são abordadas no Capítulo 2.

Como exemplo, a Figura 1-2 ilustra a aplicação das análise de padrões pontuais para o caso de mortalidade por causas externas em Porto Alegre, com os dados de 1996, realizada por Simone Santos e Christovam Barcellos, da FIOCRUZ. A localização dos homicídios (vermelho), acidentes de trânsito (amarelo) e suicídios (azul) esta mostrada na Figura 1-2 (à esquerda). À direita, apresenta-se uma superfície para a intensidade estimada, que pode ser pensada como a “temperatura da violência”. A superfície interpolada mostra um padrão de distribuição de pontos com uma forte concentração no centro da cidade e decrescendo em direção aos bairros mais afastados.

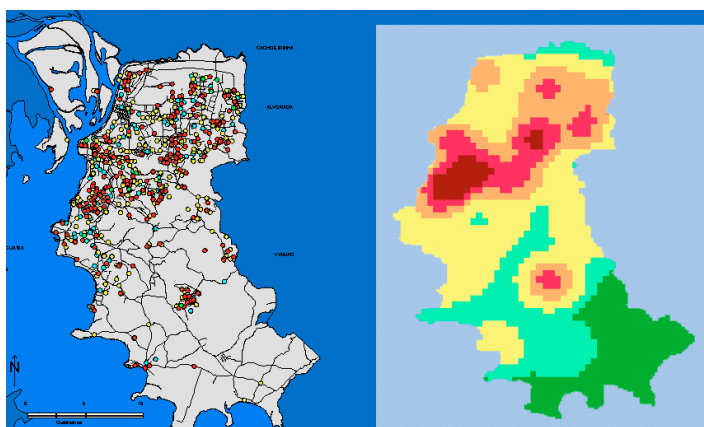


Figura 1-2 Distribuição de casos de mortalidade por causas externas em Porto Alegre em 1996 e estimador de intensidade.

Para a *análise de superfícies*, o objetivo é reconstruir a superfície da qual se retirou e mediu as amostras. Como exemplo, considere-se a distribuição de perfis e amostras de solo para o estado de Santa Catarina e áreas próximas, e o mapa de distribuição espacial da variável *saturação por bases*, produzidos por Simone Bönisch, do INPE, e apresentados na Figura 1-3..

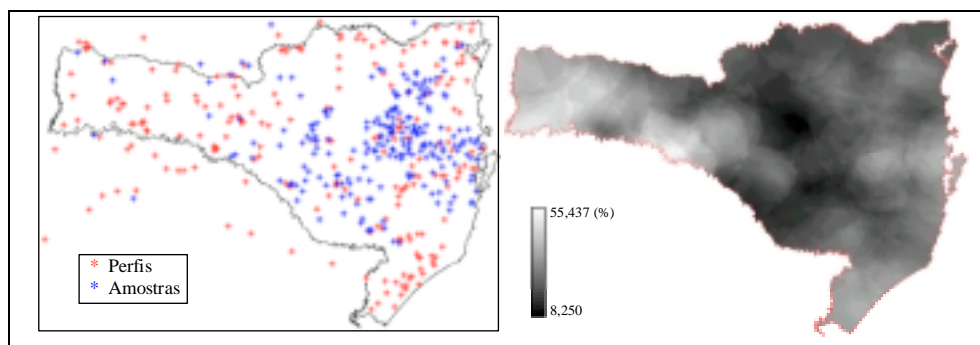


Figura 1-3 - Distribuição de perfis e amostras de solo em Santa Catarina (esquerda) e distribuição contínua estimada para a variável saturação por bases (direita).

Como foi construído este mapa? As cruzes destacadas indicam a localização dos pontos de coleta de amostras do solo; a partir destas medidas, foi estimado um modelo de dependência espacial, que permitiu a interpolação da superfície apresentada no mapa. O modelo inferencial, discutido em maior detalhe nos capítulos 3 e 4, tem por objetivo quantificar a dependência espacial entre os valores das amostras. Este modelo utiliza as técnicas da *geoestatística*, cuja hipótese central é o conceito de *estacionariedade*, que supõe um comportamento homogêneo da estrutura de correlação espacial na região de estudo, e será discutido na seção 1.4 deste capítulo. Como dados ambientais são resultantes de fenômenos naturais de longa e média duração (como os processos geológicos), as hipóteses de estacionariedade é decorrente da relativa estabilidade destes processos; na prática, isto implica que a estacionariedade está presente num grande número de situações. Deve ser observado que a estacionariedade é uma hipótese de trabalho não restritiva na abordagem de problemas não-estacionários. Métodos como krigeagem universal, *fai-k*, deriva externa, krigeagem colocada, krigeagem disjuntiva destinam-se ao tratamento de fenômenos não estacionários.

No caso de *análise de áreas*, desenvolvida no Capítulo 5, os dados são, em grande parte, oriundos de levantamentos populacionais tais como censos, estatísticas de saúde e cadastramento de imóveis. Estas áreas são usualmente delimitadas por polígonos fechados onde se supõe haver homogeneidade interna, ou seja, mudanças importantes só ocorrem nos limites. Evidentemente, esta é uma premissa nem sempre verdadeira, dado que freqüentemente as unidades de levantamento são definidas por critérios

operacionais (setores censitários) ou políticos (municípios) e não há qualquer garantia que a distribuição do evento seja homogênea dentro destas unidades. Em países com grandes contrastes sociais como o Brasil, é freqüente que grupos sociais distintos estejam agregados em uma mesma região de coleta – favelas e áreas nobres – resultando em indicadores calculados que representam a média entre populações diferentes. Em diversas regiões, as unidades amostrais apresentam ainda diferenças importantes em população e área. Neste caso, tanto a apresentação em mapas coropléticos como o cálculo simples de indicadores populacionais pode levar a distorções nos indicadores obtidos e será preciso utilizar técnicas de ajuste de distribuições.

Como exemplo de dados agregados por áreas, considere-se a Figura 1-4 (esquerda), que apresenta a distribuição espacial do índice de exclusão/inclusão social de São Paulo, produzido pela equipe liderada pela prof. Aldaíza Sposati (PUC/SP). Os indicadores de exclusão/inclusão social foram produzidos a partir de dados coletados nos 96 distritos de São Paulo, com base no censo de 1991. A partir deste mapa, foi possível extrair agrupamentos de exclusão e inclusão social, mostrados na Figura 1-4 (direita), que indicam os extremos de exclusão e inclusão social na cidade.

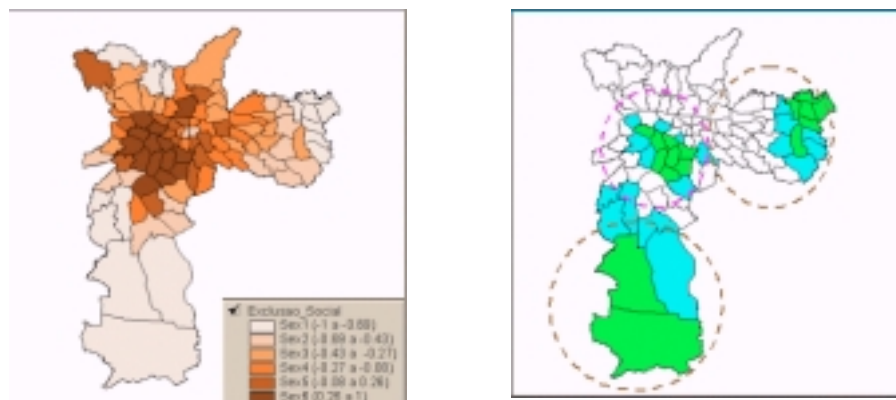


Figura 1-4- Mapa de Exclusão/Inclusão Social de São Paulo (1991) e agrupamentos de exclusão social (Zonas Leste e Sul) e inclusão social (centro).

1.3 REPRESENTAÇÃO COMPUTACIONAL DE DADOS GEOGRÁFICOS

O termo *Sistemas de Informação Geográfica* (SIG) é aplicado para sistemas que realizam o tratamento computacional de dados geográficos e armazenam a geometria e os atributos dos dados que estão *georeferenciados*, isto é, localizados na superfície terrestre e representados numa projeção cartográfica. Numa visão abrangente, pode-se indicar que um SIG tem os seguintes componentes, como mostrado na Figura 1-5:

- Interface com usuário;
- Entrada e integração de dados;
- Funções de processamento gráfico e de imagens;
- Visualização e plotagem;
- Armazenamento e recuperação de dados (organizados sob a forma de um banco de dados geográficos).

Estes componentes se relacionam de forma hierárquica. A *interface homem-máquina* define como o sistema é operado e controlado. No nível intermediário, um SIG deve ter mecanismos de processamento de dados espaciais (*entrada, edição, análise, visualização e saída*). Internamente ao sistema, um *banco de dados geográficos* armazena e recupera os dados espaciais. Cada sistema, em função de seus objetivos e necessidades, implementa estes componentes de forma distinta, mas todos os subsistemas citados estão presentes num SIG.

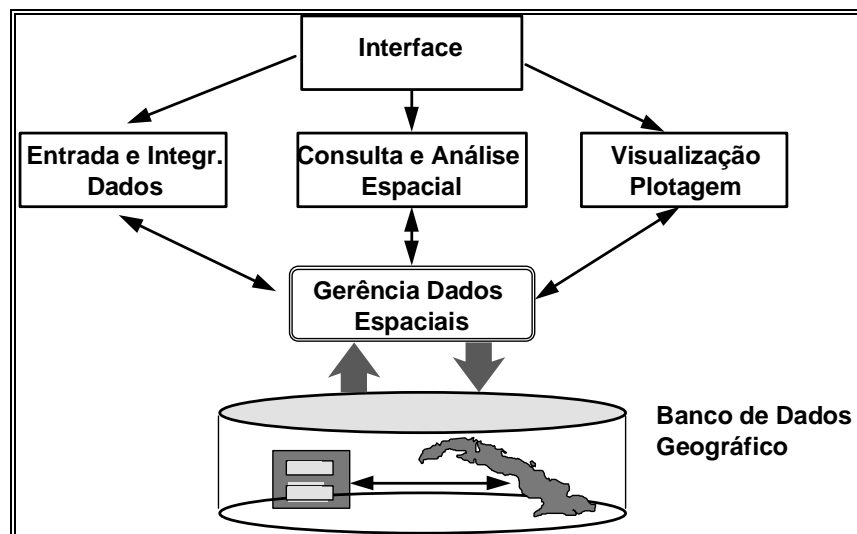


Figura 1-5 - Arquitetura de Sistemas de Informação Geográfica.

A organização de bancos de dados geográficos mais utilizada é o *modelo geo-relacional* (ou *arquitetura dual*), que utiliza um sistema gerenciador de bancos de dados (SGBD) relacional, como o DBASE ou ACCESS, para armazenar em suas tabelas os atributos dos objetos geográficos, e arquivos gráficos separados para guardar as representações geométricas destes objetos.

A principal vantagem do *modelo geo-relacional* é poder utilizar os SGBDs relacionais de mercado. Do ponto de vista do usuário, esta organização permite que aplicações convencionais, concebidas e desenvolvidas dentro do ambiente do SGBD relacional, compartilhem os

atributos dos objetos geográficos. No entanto, como o SGBD relacional não conhece a estrutura gráfica externa, existe o sério risco de se introduzir inconsistências no banco de dados geográfico. Imagine-se, por exemplo, que um usuário de aplicação exclusivamente alfanumérica possa excluir um registro alfanumérico, mas que compõe um conjunto de atributos para uma determinada entidade geográfica. Esta entidade geográfica passa a não ter mais atributos, tornando-se inconsistente. Assim, o acesso a atributos alfanuméricos de dados geográficos só pode ser feito de maneira criteriosa, dentro de controles rígidos que precisam ser implementados pela aplicação, uma vez que o modelo geo-relacional não oferece qualquer recurso para a garantia automática da integridade dos dados.

As representações geométricas utilizadas incluem as seguintes alternativas:

- *Pontos 2D*: Um=ponto2D é um par ordenado (x, y) de coordenadas espaciais. Um ponto indica um local de ocorrência de um evento, como no caso da mortalidade por causas externas, mostrada na Figura 1-2.
- *Polígonos*: Um=polígono é um conjunto de pares ordenados $\{(x, y)\}$ de coordenadas espaciais, de tal forma que o último ponto seja idêntico ao primeiro, formando uma região fechada do plano. Na situação mais simples, cada polígono delimita uma objeto individual (como no caso dos distritos de São Paulo na Figura 1-4); no caso mais geral, uma região individual de interesse pode ser delimitada por vários polígonos.
- *Amostras*: consistem de pares ordenados $\{(x, y, z)\}$ nos quais os pares (x, y) indicam as coordenadas geográficas e z indica o valor da fenômeno estudado para essa localização. Usualmente as amostras estão associadas as levantamentos de campo, como no caso de dados geofísicos, geoquímicos e oceanográficos. O conceito de amostra pode ser generalizado para o caso de múltiplas medidas em uma mesma localidade.
- *Grade regular*: é uma matriz onde cada elemento está associado a um valor numérico. Esta matriz está associada a uma região da superfície terrestre, a partir de coordenada inicial, normalmente referida ao canto inferior esquerdo da matriz, e de espaçamentos regulares nas direções horizontal e vertical.
- *Imagem*: é uma matriz onde cada elemento está associado a um valor inteiro (usualmente na faixa entre 0 e 255), utilizada para visualização. Esta matriz é utilizada para apresentação gráfica de uma grade regular. Os valores numéricos da grade são escalonados para o

intervalo de apresentação da imagem; os maiores valores serão mostrados em níveis de cinza mais claros, e os menores em níveis de cinza mais escuros. Na maior parte dos SIG, oferecem a possibilidade de apresentar uma grade regular na forma de imagem (em preto e branco ou em cores), com conversão automática ou controlada pelo usuário. A Figura 1-3 (direita) mostra a imagem da distribuição da variável saturação por bases em Santa Catarina.

As geometrias associadas a pontos, amostras e polígonos estão apresentadas na Figura 1-6 e a grade regular está mostrada na Figura 1-8. Usualmente, a referência geográfica dos dados está guardada nas coordenadas das estruturas de dados, que está associada a uma projeção cartográfica planar, ou a valores de latitude (coordenada Y) e longitude (coordenada X).

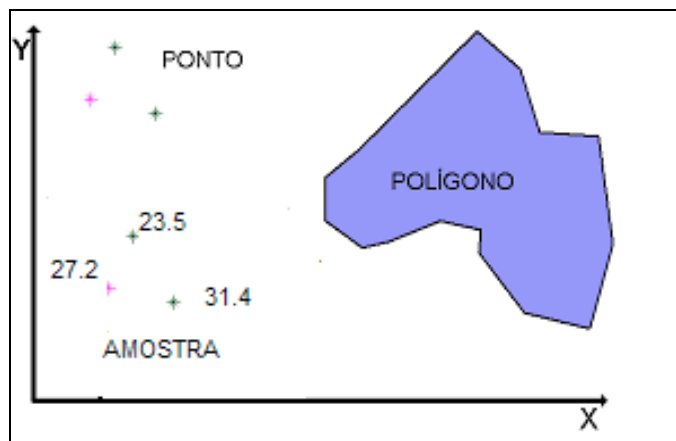


Figura 1-6 – Geometrias: Ponto2D, Amostra e Polígono

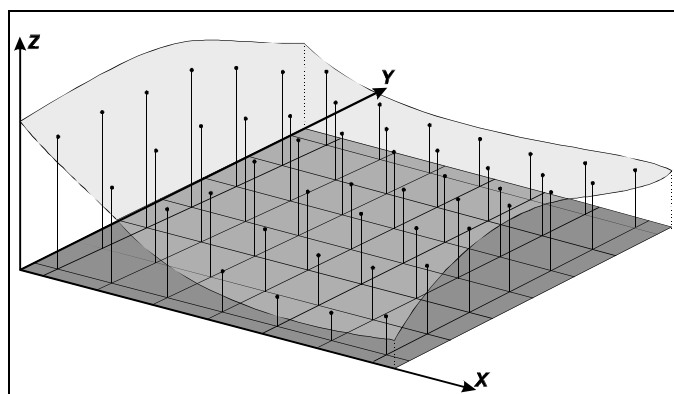


Figura 1-7 – Representação Geométrica de Grade Regular

No *modelo geo-relacional*, os atributos descritivos de cada objeto são organizados na forma de uma *tabela*, onde as *linhas* correspondem aos *dados*

e as *nomes das colunas* correspondem aos *nomes dos atributos*. Cada linha da tabela corresponde aos valores associados a um objeto geográficos; a cada objeto geográfico está associado a um identificador único ou *rótulo*, através do qual é feita uma *ligação lógica* entre seus atributos e sua representação geométrica.

Com relação aos três tipos básicos de dados utilizados em análise espacial, as *áreas* são armazenadas num SIG com estratégia dual na forma apresentada na Figura 1-8. Cada área, que pode ser um setor censitário, distrito de saúde ou município, é representada graficamente por um polígono fechado e seus atributos são guardados numa tabela de um SGBD relacional. A Figura 1-8 mostra uma fazenda de uma empresa florestal, dividida em talhões, para efeitos de cultivo. Cada talhão recebe um identificador que está associado ao mesmo tempo ao polígono que o delimita e à linha da tabela que contém seus atributos. No exemplo, a ligação é feita através dos registros no campo TALHÃO. O mesmo tipo de relacionamento lógico é ser feito em todos os outros casos, como por exemplo: moradores em um lote, lotes em uma quadra, quadras em bairro, bairros em uma cidade; hidrantes de segurança ou telefones públicos ao longo de uma avenida; postos de serviço e restaurantes ao longo de uma rodovia.

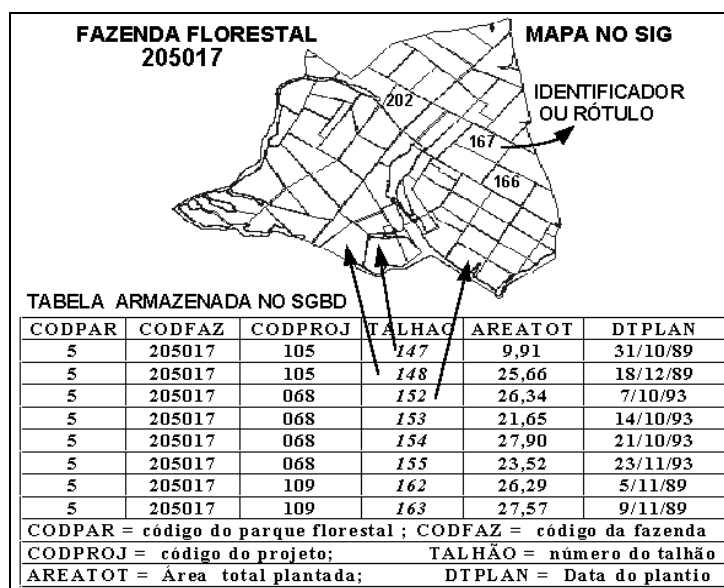


Figura 1-8 - Estratégia dual para bancos de dados geográficos.

No caso de *eventos*, estes também podem ser associados a um SGBD relacional, por exemplo para armazenar o endereço da ocorrência de um homicídio e a sua causa. Aplica-se o mesmo princípio para o caso de áreas: cada evento está associado a um identificador, que é a ligação entre o arquivo de coordenadas geográficas e a tabela no banco de dados.

Para as *superfícies*, a situação mais comum é tratar apenas com arquivos gráficos, sem o armazenamento dos resultados em um SGBD relacional. Neste caso, a situação mais usual é que os dados de entrada são armazenados como *amostras*, adicionadas a um *polígono* com os limites da região de estudo. O processo de estimação produz uma *grade regular* que descreve de forma aproximada o fenômeno na região de estudo. Esta grade pode ser transformada numa imagem para fins de apresentação (como na Figura 1-3).

1.4 CONCEITOS BÁSICOS EM ANÁLISE ESPACIAL

Dependência Espacial

Um conceito chave na compreensão e análise dos fenômenos espaciais é a *dependência espacial*. Essa noção parte do que Waldo Tobler chama de primeira lei da geografia: “*todas as coisas são parecidas, mas coisas mais próximas se parecem mais que coisas mais distantes*”. Ou, como afirma Noel Cressie, “*a dependência [espacial] está presente em todas as direções e fica mais fraca à medida em que aumenta a dispersão na localização dos dados*”.

Generalizando, pode-se afirmar que a maior parte das ocorrências, sejam estas naturais ou sociais, apresentam entre si uma relação que depende da distância. O que quer nos dizer este princípio? Se encontramos poluição num trecho de um lago, é provável que locais próximos a esta amostra também estejam poluídos. Ou que se a presença de uma árvore adulta inibe o desenvolvimento de outras, esta inibição diminui com a distância, e após determinado raio outras árvores grandes serão encontradas.

Autocorrelação Espacial

A expressão computacional do conceito de dependência espacial é a *autocorrelação espacial*. Este termo foi derivado do conceito estatístico de *correlação*, utilizado para mensurar o relacionamento entre duas variáveis aleatórias. A preposição “auto” indica que a medida de correlação é realizada com a mesma variável aleatória, medida em locais distintos do espaço. Para medir a autocorrelação espacial, pode-se utilizar diferentes indicadores, todos baseados na mesma idéia: verificar como varia a dependência espacial, a partir da comparação entre os valores de uma amostra e de seus vizinhos. Os indicadores de autocorrelação espacial são casos particulares de uma estatística de produtos cruzados do tipo

$$\Gamma(d) = \sum_{i=1}^n \sum_{j=1}^n w_{ij}(d) \xi_{ij} \quad (1-1)$$

Este índice expressa a relação entre diferentes variáveis aleatórias como um produto de duas matrizes. Dada uma distância d , a matriz w_{ij} fornece uma medida de contigüidade espacial entre as variáveis aleatórias z_i e z_j , por

exemplo, informando se são separadas de distância menor que d . A matriz ξ_{ij} fornece uma medida de correlação entre estas variáveis aleatórias, que pode ser o produto destas variáveis, como no caso do índice de Moran para áreas, discutido no capítulo 5 do livro, cuja expressão é

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

onde w_{ij} é 1 se as áreas geográficas associadas a z_i e z_j se tocam, e 0 caso contrário. Outro exemplo de indicador é o variograma, discutido no capítulo 3, onde se computa o quadrado da diferença dos valores, como no caso da expressão a seguir

$$\hat{\gamma}(d) = \frac{1}{2N(d)} \sum_{i=1}^{N(d)} [z(x_i) - z(x_i + d)]^2 \quad (1-3)$$

onde $N(d)$ é o número de amostras separadas pela distância d .

Em ambos os casos, os valores obtidos devem ser comparados com os valores que seriam produzidos no caso de não haver associação espacial entre as variáveis. Valores significativos de índices de autocorrelação espacial são evidências de dependência espacial e indicam que o postulado de independência das amostras, base da maior parte dos procedimentos de inferência estatística, é inválido e que os modelos inferenciais para estes casos devem levar explicitamente o espaço em conta em suas formulações.

Inferência Estatística para Dados Espaciais

Uma conseqüência importante da dependência espacial é que as inferências estatísticas neste tipo de dados não serão tão eficientes quanto no caso de amostras independentes do mesmo tamanho. Em outras palavras, a dependência espacial leva a uma perda de poder explicativo. De forma geral, isto se reflete em variâncias maiores para as estimativas, níveis menores de significância em testes de hipóteses e um ajuste pior para os modelos estimados, comparados a dados de mesma dimensão que exibam independência.

Na maior parte dos casos, a perspectiva mais apropriada é considerar os dados espaciais não como um conjunto de amostras independentes, mas como uma única realização de um *processo estocástico*. À diferença da visão amostral tradicional, em que cada observação traz uma informação independente, no caso de um processo estocástico todas as observações são utilizadas de forma conjunta para descrever o padrão espacial do fenômeno estudado. A hipótese feita é que, para cada ponto u de uma região A

contínua em \mathfrak{R}^2 , os valores inferidos de um atributo $z - \hat{z}(u)$ - são realizações de um processo $\{Z(u), u \in A\}$. Neste caso, é preciso fazer hipóteses sobre a estabilidade do processo estocástico, ao supor - por exemplo - que o mesmo seja *estacionário* e/ou *isotrópico*, conceitos discutidos a seguir.

Estacionariedade e Isotropia

Os principais conceitos estatísticos que definem a estrutura espacial dos dados relacionam-se aos efeitos de 1ª e 2ª ordem. *Efeito de 1ª ordem* é o valor esperado, isto é, a média do processo no espaço. *Efeito de 2ª ordem* é a covariância entre as áreas s_i e s_j . Um conceito importante neste tipo de estudo é o de estacionariedade. O processo é considerado *estacionário* se os efeitos de 1ª e 2ª ordem são constantes, em toda a região estudada, ou seja, não há tendência. Um processo é *isotrópico* se, além de estacionário, a covariância depende somente da distância entre os pontos e não da direção entre eles.

Um processo estocástico Z é dito ser *estacionário de segunda ordem* se a esperança de $Z(\mathbf{u})$ é constante em toda a região de estudo A , ou seja não depende da sua posição

$$E\{Z(\mathbf{u})\} = m \quad (1-4)$$

e a estrutura de covariância espacial depende unicamente do vetor relativo entre pontos $\mathbf{h} = \mathbf{u} - \mathbf{u}'$

$$C(\mathbf{h}) = E\{Z(\mathbf{u}) \cdot Z(\mathbf{u} + \mathbf{h})\} - E\{Z(\mathbf{u})\}E\{Z(\mathbf{u} + \mathbf{h})\} \quad (1-5)$$

Dado um processo espacial específico, a hipótese da estacionariedade pode ser corroborada a partir de procedimentos de análise exploratória e estatísticas descritivas, cujo cálculo deve considerar explicitamente a localização espacial. Na covariância espacial $C|\mathbf{h}|$, o vetor \mathbf{h} compreende a distância $|\mathbf{h}|$ e a direção. Quando a estrutura de covariância, além de variar com a distância, varia simultaneamente em função da direção, ela é dita ser *anisotrópica*. No caso em que a dependência espacial é a mesma em todas as direções, diz-se que o fenômeno é *isotrópico*. A modelagem da estrutura de covariância espacial é melhor detalhada nos capítulos que se seguem. Por ora é importante salientar as características básicas de uma estrutura de covariância espacial de forma a tornar compreensível os conceitos utilizados no livro.

1.5 O PROCESSO DA ANÁLISE ESPACIAL

A análise espacial é composta por um conjunto de procedimentos encadeados cuja finalidade é a escolha de um modelo inferencial que considere explicitamente o relacionamento espacial presente no fenômeno. Os procedimentos iniciais da análise incluem o conjunto de métodos genéricos de análise exploratória e a visualização dos dados, em geral através de mapas. Essas técnicas permitem descrever a distribuição das variáveis de estudo, identificar observações atípicas (*outliers*) não só em relação ao tipo de distribuição, mas também em relação aos vizinhos, e buscar a existência de padrões na distribuição espacial. Através desses procedimentos é possível estabelecer hipóteses sobre as observações, de forma a selecionar o modelo inferencial melhor suportado pelos dados.

Os modelos inferenciais espaciais são usualmente apresentados em três grandes grupos: variação contínua, variação discreta e os processos pontuais. A resolução de um problema espacial pode envolver a utilização de um deles ou a interação de alguns ou mesmo de todos. O exemplo abaixo ilustra as diferenças entre esses modelos, como podem ser utilizados e como interagem dentro de um mesmo processo em que questões, baseadas em fatos reais, devem ser respondidas.

A Leishmaniose visceral é uma doença principalmente de animais, mas que também atinge o homem. O principal reservatório doméstico da doença urbana são os cães, não havendo tratamento para esses. A doença é transmitida por mosquitos, que se reproduzem no solo e em matéria orgânica em decomposição, como pés de bananeira e folhas caídas. Nos últimos anos foram detectados alguns surtos epidêmicos em cidades brasileiras como Belo Horizonte, Araçatuba, Cuiabá, Teresina e Natal. O controle da doença está fundamentado no combate ao inseto e na eliminação de cães doentes da área de foco, definido em 200 metros em torno do caso humano ou canino. Entretanto, a intensiva aplicação das medidas preconizadas não vem obtendo o resultado desejado, mantendo-se a endemia. Por outro lado, a população, embora coopere no primeiro momento, quando da descoberta de casos humanos graves, depois de meses de levantamentos começa de não aceitar mais a eliminação dos cães. O problema é grave, e ainda sem solução, sendo necessário avaliar a eficácia das estratégias de controle no contexto urbano. Utilizando as ferramentas de análise espacial, algumas questões podem acumular subsídios para responder a esse problema. Como por exemplo:

Qual o raio de dispersão do mosquito em torno de seu habitat?

Na modelagem da dispersão do vetor da Leishmaniose, essencial para estimar o raio de dispersão do mosquito que define a área de borrifação ao redor de casos da doença, dois modelos podem ser utilizados:

- Os de variação contínua, onde o objetivo é gerar superfícies contínuas determinando as áreas de maior risco a partir de uma amostra de locais onde se fez a coleta dos mosquitos (amostra de pontos descontínuos).
- Os processos pontuais, onde o objetivo é modelar a probabilidade de captura de mosquitos. Nesse caso, a variável aleatória não é o valor de um atributo (presença ou ausência de mosquito) mas o local onde foi capturado.

Em área urbana, qual é o ambiente preferencial de reprodução do mosquito?

Para estimar os locais criadouros de mosquitos é necessário identificar numa determinada região as áreas de concentração de alguns atributos ambientais que propiciam o aparecimento do mosquito como por exemplos relativos a matéria orgânica e condições dos solos solo. Nesse caso os modelos de variação continua poderiam ser utilizados para inferir superfícies com os valores desses atributos.

Existe relação entre prevalência canina e condições socioeconômicas da população?

Apenas os mosquitos, isoladamente, não perpetuam a epidemia. É necessário que haja animais doentes dos quais eles se alimentem como, por exemplo, os cães. Entretanto, é sabido que tanto a presença e resistência dos cães à doença depende do estado nutricional e conseqüentemente da situação socioeconômica, como aceitação da eliminação dos animais doentes também é relacionada à renda. Assim é necessário estudar conjuntamente a incidência da doença em cães, o perfil socioeconômico da população e também a prevalência de casos humanos. O tipo de análise neste caso envolve contagens por áreas, por exemplo, indicadores socioeconômicos. Isto é, a informação disponível é completa sobre a região, em dados agrupados por área. Assim o que se objetiva é estudar a relação entre os diferentes indicadores considerando sua estrutura espacial. Nesses casos, utiliza-se o modelo de variação discreta.

Uma vez exemplificado a utilização dos modelos inferenciais básicos e como esses procedimentos podem ou não interagir na resolução de determinada questão, os conceitos básicos de cada um deles serão apresentados a seguir.

Modelos Inferenciais

Motivados por diferentes áreas de aplicações, os modelos inferenciais foram desenvolvidos separadamente para cada uma das situações acima descritas. A unificação deste campo ainda não está totalmente definida, e freqüentemente é possível aplicar mais de um tipo de modelagem ao mesmo conjunto de dados, como se pode ver no exemplo acima. Quais seriam então as vantagens de uma forma sobre a outra? Claro que algumas vezes o fenômeno em estudo apresenta variação espacial discreta, isto é um pontos isolados no espaço, mas freqüentemente os modelos discretos são usados por razões de ordem prática, tais como a disponibilidade dos dados exclusivamente por áreas. Uma das vantagens dos modelos contínuos é que a inferência não se limita a áreas arbitrariamente definidas. Por outro lado, modelos discretos permitem, mais facilmente estimar parâmetros de associação entre variáveis. A escolha final será do pesquisador, que sabe não existir o “modelo certo”, mas que busca um modelo que melhor se ajuste aos dados e tenha maior potencial de contribuir para a compreensão do fenômeno em estudo.

Processo pontual

Processos pontuais são definidos como um conjunto de pontos irregularmente distribuídos em um terreno, cuja localização foi gerada por um mecanismo estocástico. A localização dos pontos é o objeto de estudo, que tem por objetivo compreender seu mecanismo gerador. Considera-se um conjunto de pontos (u_1, u_2, \dots) numa determinada região A onde ocorreram eventos. Por exemplo, sendo o fenômeno em estudo homicídios ocorridos em uma determinada região, deseja-se verificar se existe um padrão geográfico para esse tipo de crime, ou seja, encontrar sub-regiões em A com maior probabilidade de ocorrência.

O processo pontual é modelado considerando subregiões S em A através de sua esperança $E[N(S)]$ e a covariância $C[N(S_i), N(S_j)]$, onde $N(S)$ denota o número de eventos em S . Sendo o objetivo da análise estimar as localizações prováveis de ocorrência de determinados eventos, essas estatísticas devem ser inferidas considerando o valor limite da quantidade de eventos por área. Este valor limite corresponde à esperança de $N(S)$ para uma pequena região du em torno do ponto u , quando essa tende a zero. Essa esperança é denominada *intensidade* (propriedade de primeira ordem), sendo definida como;

$$\lambda(u) = \lim_{|du| \rightarrow 0} \left\{ \frac{E[N(du)]}{|du|} \right\}, \quad (1-6)$$

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

$$\lambda(d(u_i), d(u_j)) = \lim_{du_i, du_j \rightarrow 0} \left\{ \frac{C[N(du_i), N(du_j)]}{du_i, du_j} \right\} \quad (1-7)$$

Quando o processo é *estacionário*, $\lambda(u)$ é uma constante $\lambda(u) = \lambda$; se também é *isotrópico*, $\lambda(u_i, u_j)$ se reduz à $\lambda(|h|)$, sendo $|h|$ a distância entre os dois pontos. Quando o processo é não estacionário, ou seja, a intensidade média varia na região A , a modelagem da estrutura de dependência $\lambda(u_i, u_j)$ deve incorporar a variação de $\lambda(u)$. Os processos pontuais são abordados no capítulo 2.

Varição contínua

Os modelos inferenciais de variação contínua consideram um processo estocástico $\{Z(u), u \in A, A \subset \mathfrak{R}^2\}$, cujos valores podem ser conhecidos em todos os pontos da área de estudo. A partir de uma amostra de um atributo z , coletada em vários pontos u contidos em A , $\{z(u_\alpha), \alpha = 1, \dots, n\}$, objetiva-se inferir uma superfície contínua dos valores de z . A estimação deste processo estocástico pode ser feita de forma completamente não-paramétrica ou a partir de estimadores de krigagem, como os descritos nos capítulos 3 e 4 deste livro. Esses modelos inferenciais clássicos de estimação de superfícies são denominados de *geoestatística*.

A geoestatística utiliza dois tipos de procedimentos de estimação: a krigagem e a simulação estocástica. Na *krigagem*, em cada ponto u_o , um valor da variável aleatória Z é estimado, $\hat{z}(u_o)$ utilizando-se um estimador, $\hat{Z}(u_o)$, que é a função dos dados e da estrutura de covariância espacial $\hat{Z}(u_o) = f(C, (n))$. Esses estimadores apresentam algumas propriedades importantes: são não tendenciosos e ótimos no sentido de que minimizam funções dos erros inferenciais.

Na *simulação estocástica*, os procedimentos reproduzem imagens da função aleatória Z através de realizações equiprováveis do modelo do processo estocástico estabelecido. Cada realização também chamada de imagem estocástica reflete as propriedades consideradas no modelo de função aleatória considerado. Geralmente as realizações devem honrar os dados e reproduzir a função de distribuição acumulada univariada, $F(z)$, e a estrutura de covariância espacial considerada.

A krigagem tem portanto como objetivo compor a superfície z através de estimativas pontuais ótimas, $\hat{z}(u)$, enquanto que a simulação objetiva

reproduzir a variabilidade espacial dessa superfície através de possíveis representações globais do modelo de função aleatória. Para que os processos inferenciais, da krigeagem e da simulação, possam ser realizados, é necessário fazer a hipótese que o processo estocástico seja *estacionário de segunda ordem*, isto é, um processo cuja média é constante no espaço e cuja covariância depende apenas do vetor distância entre as amostras. Neste livro, o foco será somente nos procedimentos de krigeagem, apresentados no capítulo 3 e 4.

Variação discreta

Os modelos inferenciais de variação discreta dizem respeito à distribuição de eventos cuja localização está associada a áreas delimitadas por polígonos. Este caso ocorre com muita frequência quando lidamos com fenômenos agregados por municípios, bairros ou setores censitários, como população, mortalidade e renda. Neste caso, não dispomos da localização exata dos eventos, mas de um valor agregado por área. O objetivo é modelar o padrão de ocorrência espacial do fenômeno geográfico em estudo. Os modelos de variação discreta serão estudados no capítulo 5 do livro.

Neste tipo de modelagem considera-se que o espaço geográfico em estudo, a região A , é um conjunto fixo de unidades espaciais. O modelo de distribuição mais utilizado considera um processo estocástico $\{Z_i : i = 1, \dots, n\}$, composto por um conjunto de variáveis aleatórias. Busca-se construir uma aproximação para a distribuição conjunta dessas variáveis $Z = \{Z_1, \dots, Z_n\}$, onde cada variável aleatória está associada a uma das áreas e possui uma distribuição a ser estimada. Se o processo é estacionário, o valor esperado de Z_i é a média global da região e a estrutura de covariância depende unicamente da distância, ou da estrutura de vizinhança entre as áreas.

1.6 CONCLUSÕES

Este capítulo apresentou os principais conceitos de análise espacial de dados geográficos e os principais tipos de dados e suas representações computacionais. Os diferentes tipos e problemas de Análise Espacial de dados geográficos são resumidos na Tabela 1-1, que serviu para a organização para deste livro.

Tabela 1-1

Tipos de Dados e Problemas em Análise Espacial

	Tipos de Dados	Exemplo	Problemas Típicos
Análise de Padrões Pontuais	Eventos Localizados	Ocorrência de Doenças	Determinação de Padrões e Agregamentos
Análise de Superfícies	Amostras de Campo e Matrizes	Depósitos Minerais	Interpolação e Medidas de Incerteza
Análise de Áreas	Polígonos e Atributos	Dados Censitários	Regressão e Distribuições Conjuntas

Para resumir a discussão, é importante considerar o problema conceitual da Análise Espacial do ponto de vista do usuário, que pode ser resumido na Figura 1-9. Os especialistas dos domínios do conhecimento (como Ciência dos Solos, Geologia e Saúde Pública) desenvolvem teorias sobre os fenômenos, com suporte das técnicas de visualização dos SIG. Estas teorias incluem hipóteses gerais sobre o comportamento espacial dos dados. A partir destas teorias, é necessário que o especialista formule modelos inferenciais quantitativos, que podem ser submetidos a testes de validação e de corroboração, através dos procedimentos de Análise Espacial. Os resultados numéricos podem então dar suporte ou ajudar a rejeitar conceitos qualitativos das teorias de domínio.

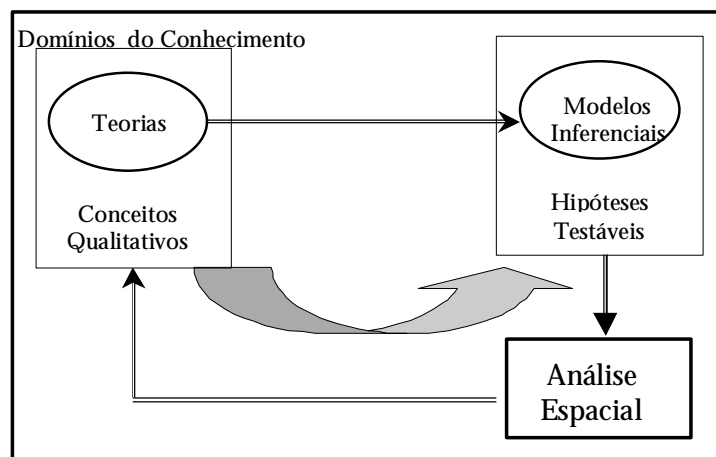


Figura 1-9 – Relação entre análise espacial e as teorias disciplinares.

Como discutido neste capítulo, e exemplificado com o caso da leishmaniose visceral, não há um “modelo certo” para cada problema. Os modelos inferenciais são úteis sobretudo para ganharmos maior

conhecimento do problema. Muitas vezes será preciso combinar as diferentes abordagens (processos pontuais, variação contínua e variação discreta) para agregar informação ao problema estudado. Nesse caso, não há “receita de bolo” e seja qual for o domínio do conhecimento, os especialistas irão se beneficiar em conhecer *todas* as técnicas aqui apresentadas.

Esta visão expressa ao mesmo tempo o potencial e as limitações da Análise Espacial. As técnicas quantitativas de Análise Espacial devem sempre estar a serviço do conhecimento dos especialistas e nunca ser utilizadas como um fim em si. Seu uso consistente requer que duas pré-condições sejam satisfeitas: o domínio dos fundamentos teóricos de Geoprocessamento e Estatística Espacial e uma metodologia de trabalho sólida, resultado da associação de modelos matemáticos (necessariamente reducionistas) com a interpretação (necessariamente subjetiva) do especialista.

A necessidade de combinar diferentes modelos inferenciais e de dispor de um conhecimento sólidas das diferentes técnicas decorre da própria natureza do espaço geográfico. Para usar a formulação de Milton Santos, o espaço é uma *totalidade*, expressa pelas dualidades entre *forma* e *função* e entre *estrutura* e *processo*; estas polaridades são evidenciadas quanto utilizamos ferramentas analíticas. Com o uso de SIG e de análise espacial, podemos caracterizar adequadamente a *forma* de organização do espaço, mas não a *função* de cada um de seus componentes; podemos ainda estabelecer qual a *estrutura* do espaço, ao modelar o fenômeno em estudo, mas dificilmente poderemos estabelecer a natureza dinâmica dos *processos*, sejam naturais ou sociais. A relação entre *estrutura* e *processo* apenas poderá se resolver quando da combinação entre as técnicas analíticas (que descrevem a estrutura de organização do espaço) e o especialista (que compreende a dinâmica do processo).

Esta abordagem nos permite construir uma visão não maniqueísta da tecnologias de Análise Espacial e Geoprocessamento. Nem panacéia com procedimentos de aplicação universal, nem mero instrumento de automação de técnicas estabelecidas, requerem de seus usuários uma postura ativa e crítica. Este equilíbrio entre *forma* e *função* e entre *estrutura* e *processo* está na essência do uso adequado dos conceitos apresentados neste livro.

1.7 REFERÊNCIAS

O livro-texto básico sobre análise espacial, escrito de maneira didática e com grande quantidade de exemplos, é “Spatial Data Analysis by Example” (Bailey and Gattrel, 1995). Seu conteúdo e as discussões com o prof. Trevor Bailey foram a influência principal para os autores. Outro livro texto introdutório de caráter geral é Fotheringham et al. (2001), que, embora

menos didático que o livro de Bailey e Gattrel, tem resultados mais recentes. Para os dados socioeconômicos, o livro de Martin (1995) ainda representa uma boa introdução, embora com muitas limitações na parte de estatística espacial. Em Português, o livro recente de Renato Assunção (2001) representa uma fonte atualizada e bem-escrita de referência, especialmente no tocante a estimadores bayesianos e a testes de conglomerados para áreas e eventos.

Para os estudiosos com formação matemática mais sólida, o texto de Cressie (1991) apresenta uma fundamentação sobre o tema, com ênfase nos modelos de variação contínua. Uma referência básica sobre geoestatística, com um conjunto extensivo de exemplos é o livro de Issaks e Srivastava (1989). A descrição da GSLIB, uma das bibliotecas mais utilizadas para o desenvolvimento de programas em geoestatística, pode ser encontrada no livro de Deutsch e Journel (1992).

Para uma introdução geral ao Geoprocessamento, o leitor pode consultar Câmara et al. (2001) ou Burrough e McDonnell (1998). Com relação à integração entre geoestatística e SIGs, o leitor deve referir-se a Camargo (1997), que descreve o desenvolvimento de um módulo geoestatístico no ambiente SPRING. O exemplo de Santa Catarina está baseado no trabalho de Bönisch (2001). Aplicações de Análise Espacial para problemas de saúde pública são discutidas em Carvalho (1997).

Assunção, R. (2001). *Estatística Espacial com Aplicações em Epidemiologia, Economia, Sociologia*. Belo Horizonte, UFMG. (disponível em <www.est.ufmg.br/~assuncao>)

Bailey, T. and A. Gattrel (1995). *Spatial Data Analysis by Example*. London, Longman.

Bönisch, S. (2001) *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.

Burrough, P.A.; McDonell, R.; *Principles of Geographical Information Systems*. Oxford, Oxford University Press, 1998.

Câmara, G.; Davis, C.; Monteiro, A.M.; D'Alge, J.C. *Introdução à Ciência da Geoinformação*. São José dos Campos, INPE, 2001 (2a. edição, revista e ampliada, disponível em www.dpi.inpe.br/gilberto/livro).

Camargo, E. (1997). *Desenvolvimento, Implementação e Teste de Procedimentos Geoestatísticos (Krigagem) no Sistema de Processamento de Informações Georreferenciadas (SPRING)*.

Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos.

Carvalho, M.S. (1997) *Aplicação de Métodos de Análise Espacial na Caracterização de Áreas de Risco à Saúde*. Tese de Doutorado em Engenharia Biomédica, COPPE/UFRJ. (Internet: <www.procc.fiocruz.br/~carvalho>).

Cressie, N. (1991) *Statistics for Spatial Data*. Chichester, John Wiley.

Deutsch, C. e A. Journel (1992). *GSLIB: Geostatistical Software Library and user's guide*. New York, Oxford University Press.

Fotheringham, A.S., C. Brunson And M.E. Charlton (2000), *Quantitative Geography*, London: Sage.

Issaks, M. e E. Srivastava (1989). *An Introduction to Applied Geostatistics*. New York, Oxford University Press, 1989.

Martin, D. (1995). *Geographic Information Systems: Socioeconomic Applications*. London, Routledge.

Tufte, E. (1983). *The Visual Display of Quantitative Information*. Cheshire, CT, Graphics Press.

APÊNDICE

SOFTWARE PARA ANÁLISE ESPACIAL

A popularidade dos sistemas de informação geográfica e o desenvolvimento e validação das técnicas de estatística espacial, descritas neste livro, tem motivado empresas e instituições envolvidas no desenvolvimento de software, a buscar formas de unificar estas abordagens. Até pouco tempo, era muito difícil encontrar SIGs com funções de análise espacial. Mais recentemente, esta situação está mudando rapidamente e boa parte das técnicas descritas neste livro já está integrada a algum dos SIGs disponíveis no Brasil. Em função da abrangência das técnicas descritas, nem todas estão integradas num único software e o especialista pode necessitar de combinar diferentes sistemas.

Para informação do leitor, incluímos a seguir uma descrição de bibliotecas e softwares especializados em análise espacial e de sistemas de informação geográfica que dispõem de funções de análise espacial. Dadas as rápidas mudanças, pede-se ao leitor que considere ser esta uma lista parcial e necessariamente incompleta. Para uma visão atualizada, recomenda-se uma visita ao sítio www.ai-geostats.org, mantido por Gregorie Dubois, e que representa um excelente portal sobre o tema.

Além dos programas citados a seguir, deve-se ressaltar que o IDRISI e o GRASS, dois SIG muito populares, tem interface com o ambiente GSTAT e portanto podem realizar análises geoestatísticas. Veja-se o conteúdo da Tabela 1-3.

TABELA 1-2

GSLIB – Biblioteca para Geoestatística

Descrição	Biblioteca para desenvolvimento de programas em geoestatística, escrita em Fortran 90
Autores	Clayton Deutsch e André Journel
Disponibilidade	Software livre em <www.gslib.com>
Funções	<i>Análise Exploratória:</i> estatísticas descritivas, cálculo de variograma (2D e 3D). <i>Estimação:</i> krigeagem simples e ordinária, com modelo de tendência, co-krigagem, krigeagem por indicação, simulação seqüencial (gaussiana e por indicação), com suporte a variáveis contínuas ou categóricas.
Aplicabilidade	Geoestatística Linear (cap 3) e por Indicação (cap 4)

TABELA 1-3

GSTAT – Software para Geoestatística

Descrição	Ambiente para desenvolvimento de programas em geoestatística, escrito em C. Possui interface com IDRISI e GRASS.
Autores	Edsger Pebesma
Disponibilidade	Software livre em <www.gstat.org>
Funções	<i>Análise Exploratória:</i> estatísticas descritivas, cálculo de variograma (2D e 3D). <i>Estimação:</i> krigeagem simples, ordinária e universal (com modelo de tendência), co-krigagem, krigeagem por indicação, simulação seqüencial (gaussiana e por indicação), com suporte a variáveis contínuas ou categóricas.
Aplicabilidade	Geoestatística Linear (cap 3) e por Indicação (cap 4)

TABELA 1-4

ClusterSeer – Clustering de Processos Pontuais

Descrição	Programa para detecção de clusters (conglomerados) associados a eventos
Autores	Godfrey Jacquez
Disponibilidade	Software comercial em < www.terraseer.com >
Funções	<i>Detecção de Conglomerados Espaciais</i> : testes focados (Diggle, Bithell, Besag e Newell, Turnbull) e globais (Besag e Newell, função K de Ripley). <i>Detecção de Conglomerados Espaço-Temporais</i> (Kuldorff)
Aplicabilidade	Análise de Eventos (cap 2)

TABELA 1-5

CrimeStat – Análise de Estatísticas Criminais

Descrição	Software para análise de eventos associados a criminalidade
Autores	
Disponibilidade	Software livre em < www.icpsr.umich.edu/NACJD/crimestat.html >
Funções	<i>Estatísticas descritivas</i> : centro médio, elipse dos desvios padrões, índice I de Moran. <i>Detecção de conglomerados</i> : função K de Ripley, k-médias e índices locais de Moran. <i>Estimador de densidade</i> : "kernel estimator".
Aplicabilidade	Análise de Eventos (cap 2)

TABELA 1-5

SpaceStat – Análise Espacial de Áreas

Descrição	Software para análise espacial de áreas, com ênfase em técnicas de regressão espacial. Possui interface com ArcView.
Autor	Luc Anselin
Disponibilidade	Comercial em http://www.spacestat.com/
Funções	<p><i>Análise Exploratória:</i> estatísticas descritivas, índice I de Moran (global e local), mapa de Moran, índice C de Geary, com testes de hipóteses sobre autocorrelação espacial.</p> <p><i>Estimação:</i> Regressão por mínimos quadrados, e regressão espacial com várias técnicas: modelos SAR (<i>spatial lag e spatial error</i>), com inclusão de heterocedasticidade.</p>
Aplicabilidade	Análise de Áreas (capítulo 5)

TABELA 1-6

SPRING

Descrição	Software de geoprocessamento de propósito geral, com funções de processamento de imagens, modelagem de terreno, álgebra de mapas e consulta a bancos de dados geográficos. Possui interface com SpaceStat e suas funções de geoestatística utilizam a GSLIB.
Autores	Equipe da Divisão de Processamento de Imagens do INPE
Disponibilidade	Software livre em < www.dpi.inpe.br/spring >
Funções de Análise Espacial	<p><i>Análise Exploratória:</i> estatísticas descritivas, cálculo de variograma (2D e 3D), índice I de Moran (global e local), mapa de Moran, índice C de Geary, com testes de hipóteses sobre autocorrelação espacial.</p> <p><i>Deteção de conglomerados:</i> função K de Ripley, vizinho mais próximo e índices locais de Moran.</p> <p><i>Estimador de densidade:</i> "kernel estimator".</p> <p><i>Estimação:</i> krigeagem simples e ordinária, krigeagem por indicação, simulação seqüencial (gaussiana e por indicação), com suporte a variáveis contínuas ou categóricas.</p>
Aplicabilidade	Análise de eventos (cap 2), geoestatística Linear (cap 3) e por Indicação (cap 4), análise de áreas (cap 5).

TABELA 1-7

ArcGIS Geostatistical Analyst

Descrição	Extensão do ArcGIS (software de geoprocessamento de propósito geral)
Autores	Konstantin Krivoruchko e equipe da ESRI
Disponibilidade	Comercial em <www.esri.com>
Funções de Análise Espacial	<i>Análise Exploratória</i> : estatísticas descritivas, cálculo de variograma (2D e 3D), análise de tendências <i>Estimação</i> : krigagem simples e ordinária, krigagem por indicação, co-krigagem e krigagem disjuntiva
Aplicabilidade	Geoestatística Linear (cap 3) e por Indicação (cap 4)