

PREFÁCIO

A abordagem tradicional para o tratamento de dados sociais, econômicos e ambientais utiliza técnicas estatísticas bem estabelecidas, como testes de hipóteses, análise de variância e modelos lineares. Estas, embora relevantes e úteis, não permitem considerar o espaço geográfico. Uma das motivações do livro que ora entregamos ao leitor é que “onde” ocorrem os fenômenos é muito importante, e em alguns casos essencial à compreensão do problema. Assim, este livro apresenta um conjunto de técnicas estatísticas que fazem uso explícito da localização dos fenômenos. Tais técnicas partem do princípio que “há algo de especial com dados espaciais”, e que “o local faz a diferença”. Originalmente desenvolvidas em ambientes especializados, tais técnicas estão se tornando disponíveis a um público cada vez mais amplo, seja pela inclusão de funções de estatística espacial nos SIGs, seja pelo desenvolvimento de interfaces eficientes para combinar a capacidade de visualização dos SIGs com o poder analítico dos pacotes de estatística, mas também pela disseminação dos softwares de domínio público, democratizando o acesso à ferramenta.

Imaginamos nosso leitor como um especialista em disciplinas como Ciência dos Solos, Geologia, Oceanografia, Saúde Coletiva ou Políticas Públicas, que possui noções básicas de estatística multivariada e deseja utilizar as técnicas de análise espacial em seus dados georeferenciados. Buscamos produzir um livro texto que pudesse ser acessível a este leitor, limitando as questões teóricas de estatística ao mínimo indispensável. Para tanto, adotamos uma abordagem *prática*, focando nossa discussão em técnicas de eficácia comprovada e disponíveis de forma ampla. O caráter introdutório e (esperamos) didático deste livro também decorre do fato de ser utilizado como suporte a cursos de pós-graduação ministrados pelos autores no INPE, na USP e na FIOCRUZ . Este material também é utilizado em tutoriais e cursos de curta duração que vem sendo ministrados pelos autores em diversas instituições no Brasil. A seleção dos métodos apresentados no livro foi relativamente arbitrária, baseada na experiência dos autores no ensino e em trabalhos de pesquisa, parte das quais utilizadas nos exemplos ao longo do livro. Sendo um livro introdutório, optamos por omitir algumas técnicas importantes, mas que iriam requerer um conhecimento prévio diferenciado, como é o caso dos modelos que utilizam inferência bayesiana

Como material complementar, tornamos disponível os dados utilizados nos exemplos do livro, com sugestões de exercícios e procedimentos operacionais para que o leitor possa reproduzi-los e amplia-los, no sítio www.dpi.inpe.br/gilberto/livro/analise. A maior parte das técnicas pode ser

executada pelo software SPRING ou a partir da ligação entre o SPRING e softwares especializados, como o SpaceStat. O SPRING é um desenvolvimento do INPE e está acessível livremente em <www.dpi.inpe.br/spring>. Além disso, o pacote estatístico R <www.r-project.org>, de domínio público, também permite aplicar a maior parte das técnicas apresentadas.

O que os esperamos do leitor? Além de uma leitura participativa, recomendamos o empenho em reproduzir as técnicas apresentadas, tanto nos dados de exemplo, como em seus próprios problemas. Gostaríamos ainda de imaginar que pelo menos parte de nossos leitores venha a se interessar por questões mais avançadas de estatística espacial, para as quais apontamos a bibliografia relevante.

Muito devemos a nossos colaboradores científicos Corina Costa Freitas, Eduardo Camargo, Carlos Felgueiras e João Argemiro de Carvalho Paiva (INPE), Eduardo Assad (EMBRAPA), Maria Leonor Assad (UnB), Flávio Nobre (COPPE/UFRJ), Oswaldo Gonçalves Cruz, Christovam Barcellos, Tiago Maria Lapa, Wayner Vieira de Souza (FIOCRUZ), Renato Assunção (UFMG) e Aldaíza Sposati e Dirce Koga (PUC/SP), Paulo Justiniano Ribeiro Jr. e Silvia Shimakura (UFPR), Trevor Bailey (Universidade de Exeter, Inglaterra) pelas frutíferas discussões sobre o tema de análise espacial. Ressaltamos ainda que parte considerável dos exemplos apresentados é resultante de teses, dissertações e trabalhos desenvolvidos por nossos alunos de pós-graduação do INPE e da ENSP. Portanto nossos agradecimentos a: Marcos Neves, Simone Bönisch, Silvana Amaral Kampel, José Luiz Rodriguez Yi, Waldiza Brandão, Frederico Roman Ramos, Patrícia Genovez e Virgínia Ragoni Corrêa, alunos do INPE, e Tatiana Campos, Eleonora D’Orsi, Mirian Carvalho de Souza, Simone Maria dos Santos, Rejane Sobrino Pinheiro, alunos da ENSP/FIOCRUZ. Finalmente, a edição deste livro contou com a fundamental participação da Embrapa-Cerrados, e o apoio institucional de Carlos Magno, seu chefe. Nosso obrigado também a Ricardo Cartaxo Modesto de Souza, representando os agradecimentos a todos os engenheiros da Divisão de Processamento de Imagens do INPE, que tem tornado possível, ao longo destes muitos anos, o sonho de materializar as idéias e artigos científicos em produtos e tecnologia para o uso coletivo.

Como leitura complementar a este livro, sugerimos ao leitor duas especiais: o livro *Interactive Spatial Data Analysis*, de Trevor Bailey e Anthony Gatrell, obra de grande valor didático, que serviu de inspiração para os autores, e a monografia *Estatística Espacial com Aplicações em Epidemiologia, Economia e Sociologia*, de Renato Assunção (UFMG), disponível em sua página www.est.ufmg.br/~assuncao.

Após mais de uma década da implantação dos primeiros sistemas de Geoprocessamento no Brasil, a comunidade precisa dar um “salto de qualidade” em sua capacidade de bem utilizar os dados espaciais. Nossa intenção foi contribuir para este necessário aprimoramento. Considerando a grande diversidade

dos profissionais de Geoprocessamento, os autores têm consciência que a leitura deste texto irá exigir um esforço considerável por parte de uma grande parcela de seu público-alvo. Acreditamos que o conhecimento a ser adquirido será plena recompensa do esforço.

São José dos Campos, Rio de Janeiro, Brasília, Fortaleza, Praia das Fontes,
Junho de 2002

Suzana Druck
Marília Sá Carvalho
Gilberto Câmara
Antônio Miguel Vieira Monteiro

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.

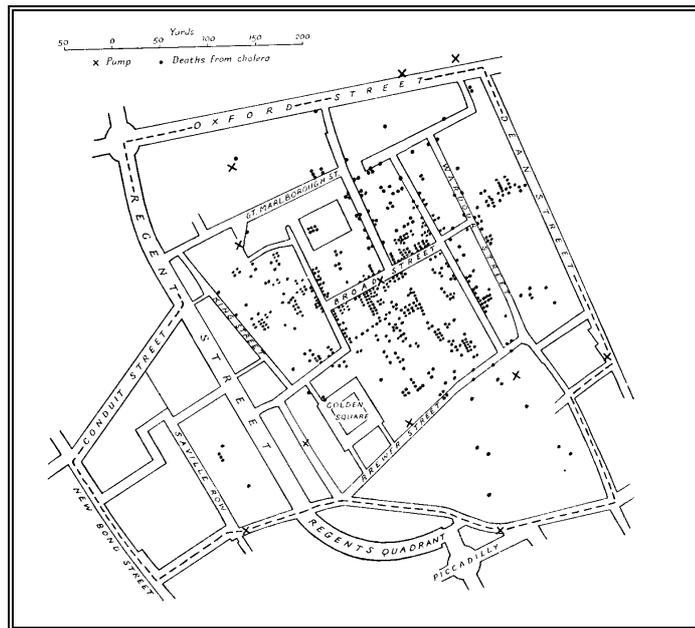


Figura 1-1 - Mapa de Londres com óbitos por cólera identificados por pontos e poços de água representados por cruzes.

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.

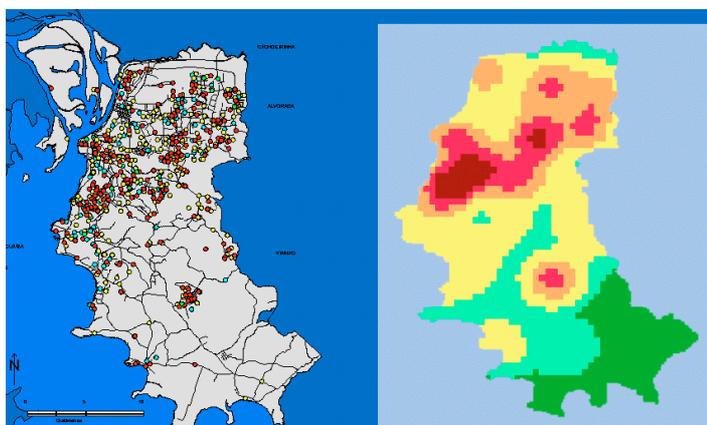


Figura 1-2 Distribuição de casos de mortalidade por causas externas em Porto Alegre em 1996 e estimador de intensidade.

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..

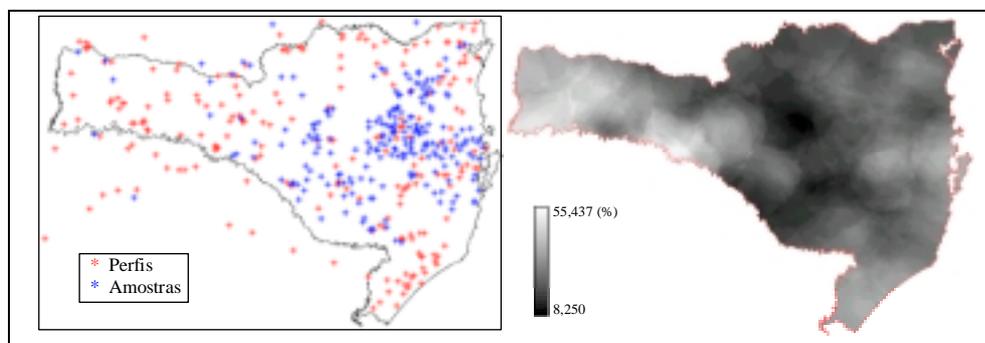


Figura 1-3 - Distribuição de perfis e amostras de solo em Santa Catarina (esquerda) e distribuição contínua estimada para a variável saturação por bases (direita).

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 krigagem universal, *fai-k*, deriva externa, krigagem colocada, krigagem 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.

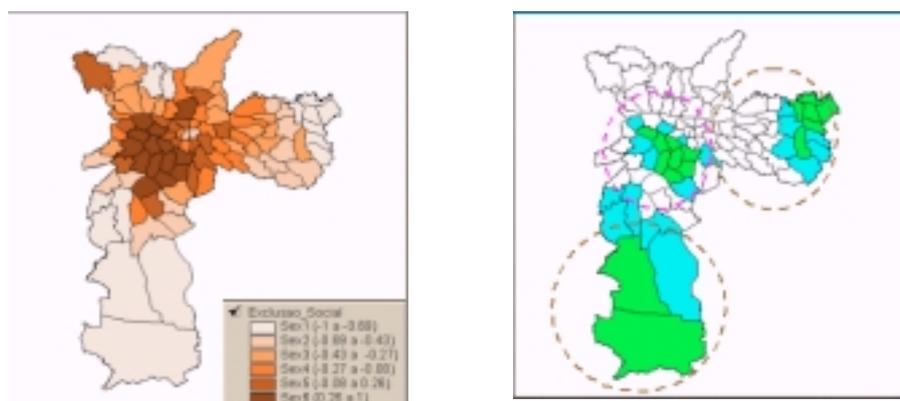


Figura 1-4- Mapa de Exclusão/Inclusão Social de São Paulo (1991) e agrupamentos de exclusão social (Zonas Leste e Sul) e inclusão social (centro).

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.

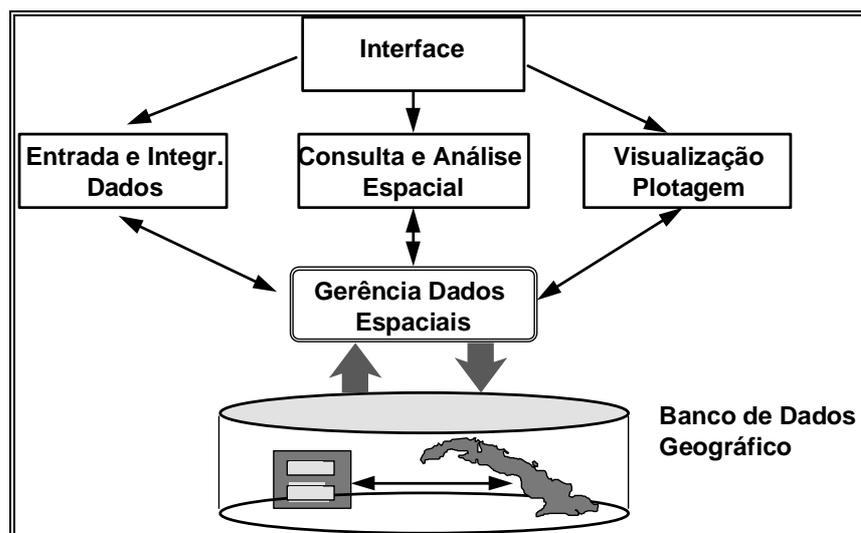


Figura 1-5 - Arquitetura de Sistemas de Informação Geográfica.

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).

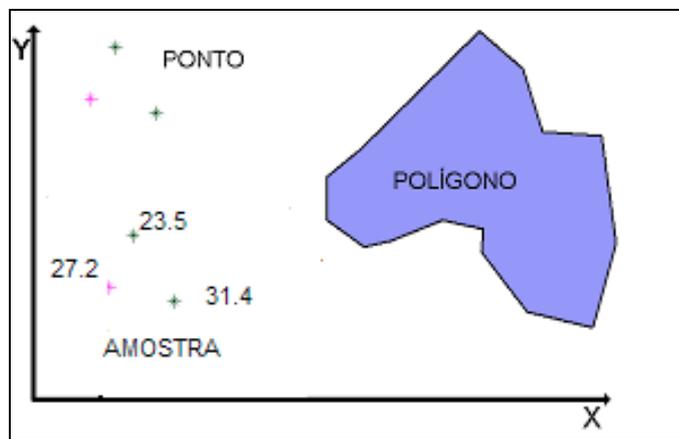


Figura 1-6 – Geometrias: Ponto2D, Amostra e Polígono

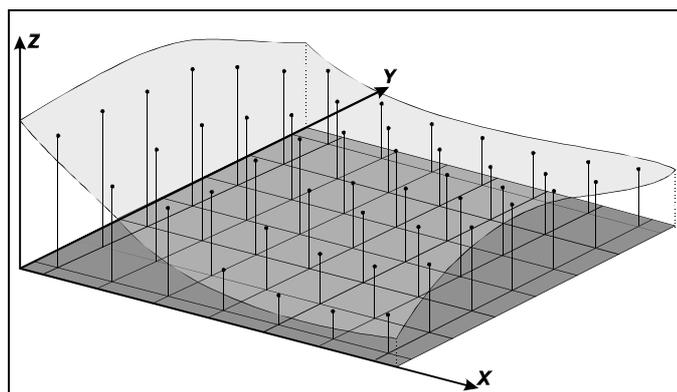


Figura 1-7 – Representação Geométrica de Grade Regular

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.

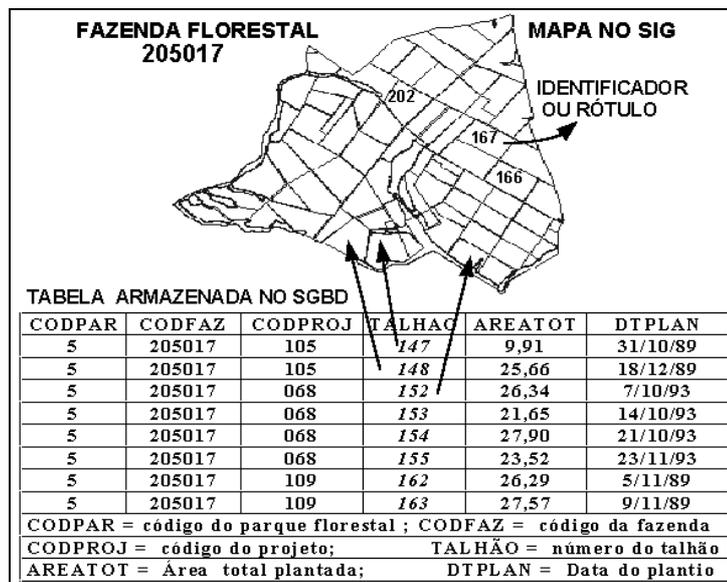


Figura 1-8 - Estratégia dual para bancos de dados geográficos.

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 consequência 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.

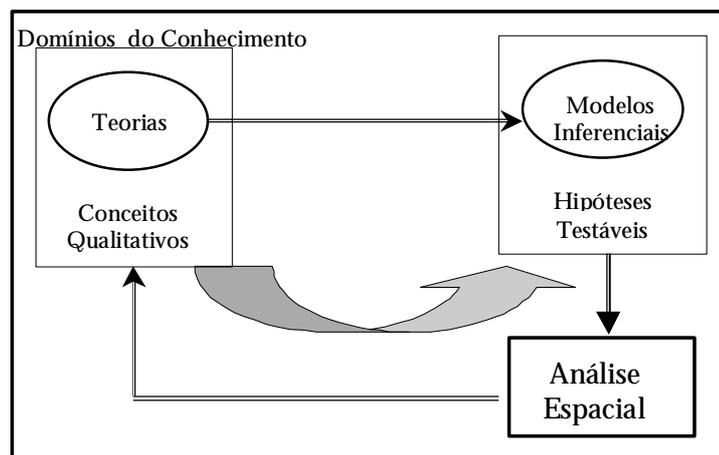


Figura 1-9 – Relação entre análise espacial e as teorias disciplinares.

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)

2 ANÁLISE ESPACIAL DE EVENTOS

Gilberto Câmara
Marília Sá Carvalho

2.1 INTRODUÇÃO

Neste capítulo serão estudados os 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. O objetivo destas análises é estudar a distribuição espacial destes pontos, testando hipóteses sobre o padrão observado: se é aleatório, se apresenta-se em aglomerados ou se os pontos estão regularmente distribuídos. O objeto de interesse é a própria localização espacial dos eventos em estudo.

O tipo de dado nestes estudos consiste em uma série de coordenadas de pontos (P_1, P_2, \dots) dos eventos de interesse dentro da área de estudo. O termo *evento* refere-se a qualquer tipo de fenômeno localizável no espaço que, dentro de nossa escala de investigação, possa estar associado a uma representação pontual. Exemplos incluem:

- Epidemiologia: residência de casos de doenças
- Sociologia: local de ocorrência de ofensas criminais
- Demografia: localização de cidades
- Biologia: localização de espécies vegetais de interesse

Para ilustrar estes conceitos, considere a figura 2.1, que apresenta a distribuição de 299 óbitos de menores de um ano, registrados no ano de 1998, de crianças nascidas no mesmo ano na cidade de Porto Alegre, Rio Grande do Sul, divididos em neonatais (menores de 28 dias de nascidos) e posneonatais (entre 28 dias e um ano). A análise de padrões neste tipo de dado pode ser utilizada como uma forma de identificação de possíveis áreas com maior concentração de mortes infantis, de comparação entre os óbitos nos dois grupos de idade, e de identificação de fatores de risco associados a esta ocorrência.

Os dados de distribuições pontuais têm as seguintes características:

- A área dos eventos não é uma medida válida apesar de em muitos casos ocuparem espaço. Mesmo na análise do padrão de distribuição de cidades estas são consideradas como um ponto no espaço do estudo.

- Os pontos em geral não estão associados a valores, mas apenas à ocorrência dos eventos considerados.
- Em alguns estudos os pontos podem estar associados a atributos de identificação, como no exemplo acima, em óbitos neonatais e posneonatais. Quando este atributo é elemento do estudo, através da comparação da distribuição espacial destes atributos, denomina-se processo pontual marcado.

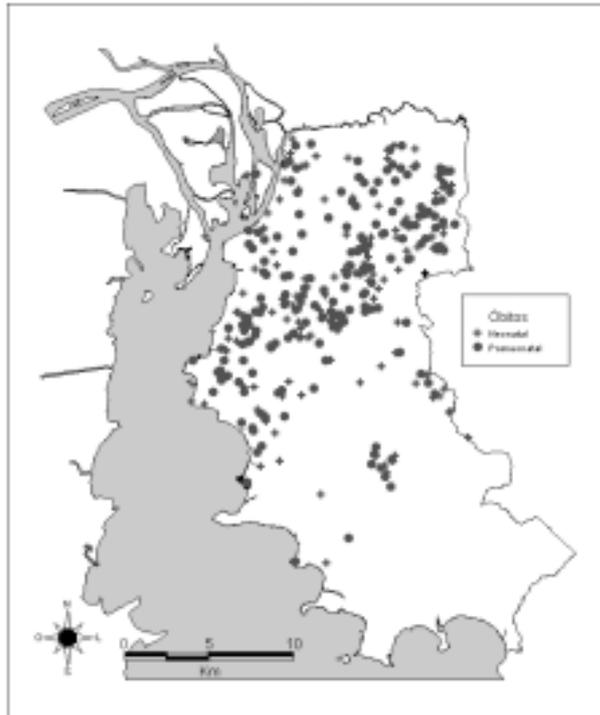


Figura 2-1 - Distribuição espacial de mortalidade infantil – neonatal e posneonatal - em Porto Alegre em 1998.

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

2.2 CARACTERIZAÇÃO DE DISTRIBUIÇÕES DE PONTOS

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

Os efeitos de primeira ordem, considerados globais ou de larga escala, correspondem a variações no valor médio do processo no espaço. Neste caso, estamos interessados na *intensidade* do processo, isto é, no número de eventos por unidade de área. Efeitos de segunda ordem, denominados locais ou de pequena escala, representam a *dependência espacial* no processo, proveniente da estrutura de correlação espacial. Para medir a dependência espacial, procuramos estimar o relacionamento entre pares de eventos (por unidade de área) no espaço, o que corresponde a uma aproximação do cálculo da covariância entre as variáveis aleatórias que representam cada evento¹.

Considera-se um conjunto de pontos (u_1, u_2, \dots) numa determinada região A onde ocorreram eventos. 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

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

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

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

Quando o processo é *estacionário*, $\lambda(u)$ é uma constante, ou $\lambda(u) = \lambda$; se também é *isotrópico*, $\lambda(u_i, u_j)$ se reduz à $\lambda(|h|)$, sendo $|h|$ a distância entre os dois pontos. Quando o processo é não estacionário, ou seja, a intensidade média varia

¹ Vale lembrar a discussão do seção 1, onde caracterizamos os eventos no espaço por um processo estocástico, onde cada ocorrência é uma realização de uma variável aleatória distinta.

na região A , a modelagem da estrutura de dependência $\lambda(u_i, u_j)$ deve incorporar a variação de $\lambda(u)$. A maior parte das técnicas de análise de distribuição de pontos supõe, explícita ou implicitamente, um comportamento estacionário e isotrópico do processo aleatório subjacente aos eventos analisados.

No exemplo acima da mortalidade infantil, a ocorrência dos óbitos está condicionada pela distribuição dos nascimentos. Além disso, características individuais da criança, tais como prematuridade e peso, são importantes condicionantes do óbito. É possível, entretanto, modelar estes eventos e detectar áreas de sobre-risco, considerando simultaneamente o padrão de distribuição dos nascimentos e óbitos, e verificando a variação da intensidade do evento na região e a estrutura de correlação local.

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

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

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

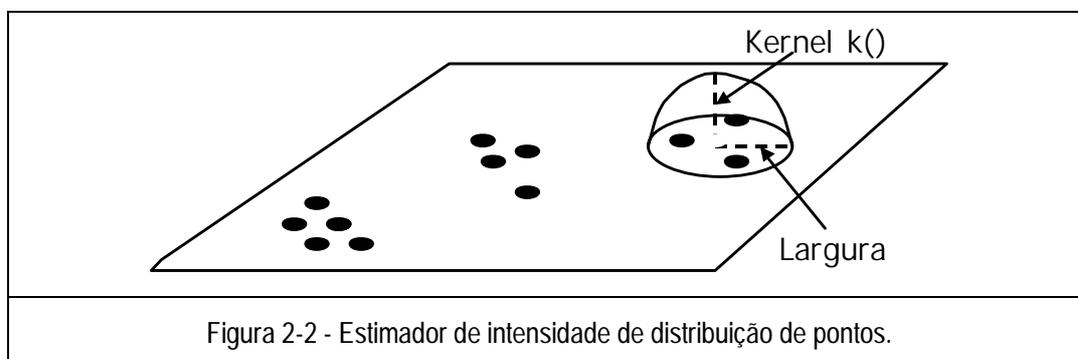
Esta formulação nos permite estabelecer uma base de comparação entre uma distribuição completamente aleatória (que seria gerada por um processo de Poisson) e os dados coletados em campo. O procedimento mais usual para estimar a probabilidade associada ao padrão encontrado será produzir uma *simulação*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

Podemos gerar um conjunto de simulações, para que possamos obter uma base de comparação entre o comportamento de um processo aleatório e a distribuição dos eventos medidos. Os conceitos de CSR são utilizados para

caracterizar os efeitos de segunda ordem em distribuição de pontos, utilizando os métodos do *vizinho mais próximo* e da função K, descritos a seguir. São também utilizados para avaliação em vários métodos de detecção de aglomerados (*clusters*).

2.3 ESTIMADOR DE INTENSIDADE ("KERNEL ESTIMATION")

Uma alternativa simples para analisar o comportamento de padrões de pontos é a estimar a intensidade pontual do processo em toda a região de estudo. Para isto, pode-se ajustar uma função bi-dimensional sobre os eventos considerados, compondo uma superfície cujo valor será proporcional à intensidade de amostras por unidade de área. Esta função realiza uma contagem de todos os pontos dentro de uma região de influência, ponderando-os pela distância de cada um à localização de interesse, como mostrado na Figura 2-2.



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

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

Este estimador é chamado *kernel estimator* e seus parâmetros básicos são: (a) um raio de influência ($\tau \geq 0$) que define a vizinhança do ponto a ser interpolado e controla o "alisamento" da superfície gerada; (b) uma função de estimação com propriedades de suavização do fenômeno. O *raio de influência* define a área centrada no ponto de estimação u que indica quantos eventos u_i contribuem para a estimativa da função intensidade λ . Um raio muito pequeno irá gerar uma superfície muito descontínua; se for grande demais, a superfície poderá ficar muito amaciada. No caso da função de interpolação $k()$, é comum usar funções de terceira ou quarta ordem, como

$$k(h) = \frac{3}{\pi}(1-h^2) \quad (2.5)$$

ou o *kernel gaussiano*

$$k(h) = \frac{1}{2\pi\tau} \exp\left(-\frac{h^2}{2\tau^2}\right) \quad (2.6)$$

Nestes estimadores, h representa a distância entre a localização em que desejamos calcular a função e o evento observado. Com o uso desta função de quarta ordem (equação 2.5), o estimador de intensidade pode ser expresso como:

$$\hat{\lambda}_\tau(u) = \sum_{h_i \leq \tau} \frac{3}{\pi\tau^2} \left(1 - \frac{h_i^2}{\tau^2}\right)^2 \quad (2.7)$$

O estimador de intensidade é muito útil para nos fornecer uma visão geral da distribuição de primeira ordem dos eventos. Trata-se de um indicador de fácil uso e interpretação. A figura 2.3 ilustra a aplicação do estimador de intensidade para o caso de mortalidade por causas externas em Porto Alegre, com os dados de 1996. A localização dos homicídios (vermelho), acidentes de trânsito (amarelo) e suicídios (azul) esta mostrada na figura 2.3 à esquerda e o estimador de intensidade dos homicídios é apresentado na figura 2.3. 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.

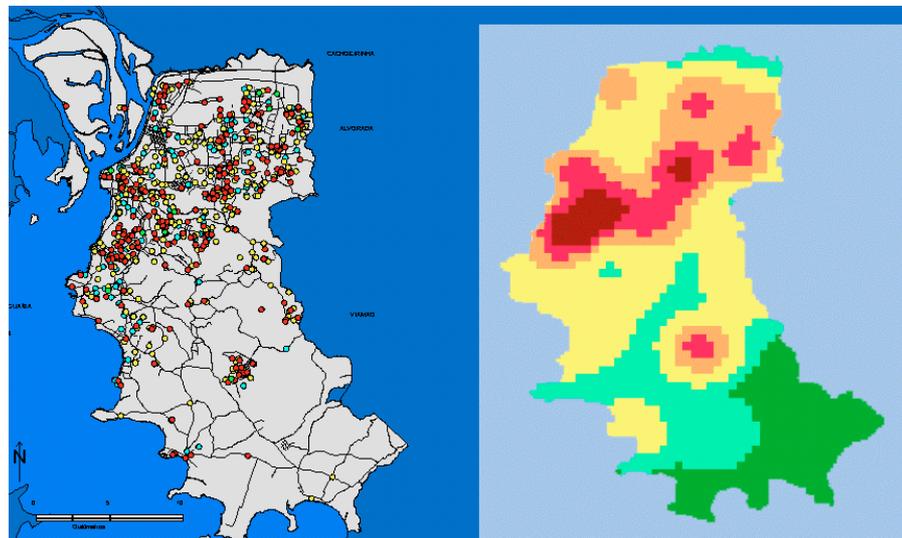


Figura 2.3: Distribuição de casos de mortalidade por causas externas em Porto Alegre em 1996 e estimador de intensidade.

2.4 ESTIMADORES DE DEPENDÊNCIA ESPACIAL

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

Método do Vizinho Mais Próximo

O método do vizinho mais próximo estima a função de distribuição cumulativa $\hat{G}(h)$ baseado nas distâncias h entre eventos em uma região de análise. Esta função de distribuição pode ser estimada empiricamente da seguinte forma:

$$\hat{G}(h) = \frac{\#(d(u_i, u_j) \leq h)}{n} \quad (2.8)$$

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

A plotagem dos resultados desta função de distribuição cumulativa empírica $\hat{G}(h)$ pode ser usada como um método exploratório para se verificar se existe evidência de interação entre os eventos. Se esta plotagem apresentar um crescimento rápido para pequenos valores de distância, esta situação aponta para interação entre os eventos caracterizando agrupamentos nestas escalas. Se esta plotagem apresentar valores pequenos no seu início, e só crescer rapidamente para valores maiores de distância, esta situação aponta para uma distribuição mais regular. A Figura 2-4 mostra a função $\hat{G}(h)$ para os dados de mortalidade infantil de Porto Alegre (figura 2.1), com distância mínima de 0 km e distância máxima de 1 km. Verifica-se que a curva mostra um crescimento acentuado para distâncias até 500 m para depois se estabilizar, o que caracteriza agrupamento nesta faixa de distâncias.

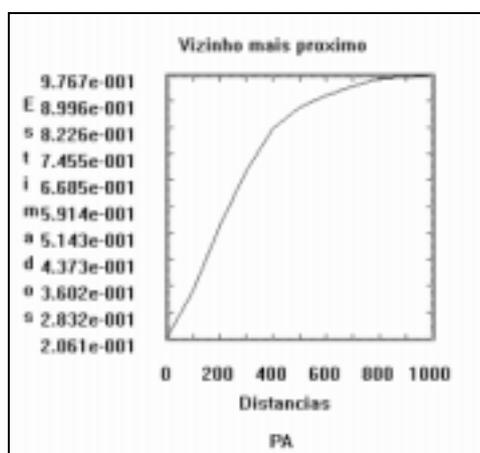


Figura 2-4 – Função vizinho-mais-próximo para mortalidade infantil neonatal em Porto Alegre.

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

$$G(h) = 1 - e^{-\lambda m^2} \quad h \geq 0 \quad (2.9)$$

A estimação simulada para a distribuição $G(w)$ assumindo-se CSR é calculada como

$$\bar{G}(h) = \frac{\sum_i^k \hat{G}_i(h)}{k} \quad (2.10)$$

onde $\hat{G}_i(h)$, $i=1,2,..,k$ são funções de distribuição empíricas, estimadas a partir de k simulações independentes dos n eventos, na hipótese de CSR (n eventos independentes e uniformemente distribuídos). Para verificar a condição de aleatoriedade, calculamos ainda os envelopes de simulação superior e inferior, definidos como se segue:

$$\begin{aligned} U(h) &= \max\{\hat{G}_i(h)\}, \quad i=1,..,k \\ L(h) &= \min\{\hat{G}_i(h)\}, \quad i=1,..,k \end{aligned} \quad (2.11)$$

A plotagem da distribuição estimada $\hat{G}(h)$ versus a distribuição simulada $\bar{G}(h)$, com a adição dos envelopes inferior e superior, permite medir a significância dos desvios relativo a aleatoriedade. Se a condição CSR for válida para os dados observados, o gráfico da curva de $\hat{G}(h)$ versus $\bar{G}(h)$ deve ser praticamente linear com um ângulo de 45 graus. Se o dado apresenta tendências para agrupamentos, os traçados no gráfico estarão acima da linha de 45 graus, ao passo que para padrões de regularidade os traçados ficarão abaixo da linha de 45 graus.

A Figura 2-5 mostra um exemplo de gráfico mostrando o posicionamento da distribuição e dos envelopes com relação a linha de 45 graus, para os dados referentes mortalidade infantil neonatal em Porto Alegre. Neste caso percebe-se a posição dos envelopes e da distribuição acima da linha de 45 graus, o que caracteriza agrupamento para as distâncias em análise.

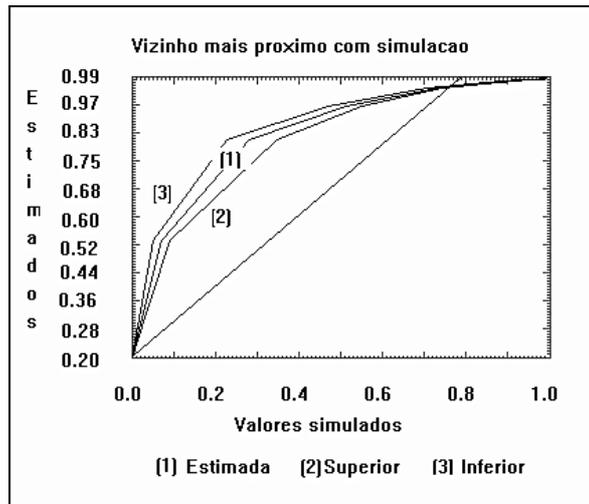


Figura 2-5 – Gráfico de $\hat{G}(h)$ (estimado) versus $\bar{G}(h)$ (CSR), com envelopes superior e inferior, para os dados de mortalidade neonatal em Porto Alegre

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

Função K

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

$$\lambda K(h) = E(\# \text{ eventos contidos a uma distância } h \text{ de um evento arbitrário}) \quad (2.12)$$

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

$$\hat{K}(h) = \frac{A}{n^2} \sum_i^n \sum_{j, i \neq j}^n \frac{I_h(d_{ij})}{w_{ij}} \quad (2.13)$$

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

A função K é usada como ferramenta exploratória na comparação entre estimativa empírica — $\hat{K}(h)$ — e a resultante de um processo de padrão de pontos espacial aleatório — $\bar{K}(h)$. Para um processo aleatório $K(h)$ seria πh^2 . Portanto, uma forma de comparar a estimativa K de um conjunto de dados observados com πh^2 seria plotar a função $\hat{L}(h)$ definida como:

$$\hat{L}(h) = \sqrt{\frac{\hat{K}(h)}{\pi}} - h \quad (2.14)$$

O gráfico de $\hat{L}(h)$ em função da distância h indica atração espacial entre eventos ou agrupamentos para valores positivos, sendo o agrupamento mais forte em picos positivos, e indica repulsão espacial ou regularidade em pontos de valores negativos. Uma abordagem similar à do vizinho mais próximo pode ser feita para se estimar a significância dos desvios da distribuição $\hat{L}(h)$ em relação à condição de aleatoriedade (CSR). Os envelopes inferior e superior são construídos a partir de k simulações independentes de n eventos na região A . Na análise do gráfico com a distribuição e os envelopes, picos positivos na função estimada $\hat{L}(h)$ que estão acima do envelope superior evidenciam ocorrência de agrupamento na escala considerada, portanto, se todos os valores da função $\hat{L}(h)$ estiverem acima do envelope superior e com valores positivos, teremos agrupamentos em todas as escalas. Depressões negativas na função estimada $\hat{L}(h)$ que estiverem abaixo do envelope inferior, evidenciam regularidade nessa escala, portanto, se todos os valores de $\hat{L}(h)$ estiverem abaixo do envelope inferior e com valores negativos, tem-se regularidade em todas as escalas.

A Figura 2-6 mostra o gráfico da função $\hat{L}(h)$ e dos envelopes de simulação para o dado de Porto Alegre (Figura 2-1). Verifica-se valores positivos para a função L , estando os mesmos acima dos envelopes, o que caracteriza agrupamento em todas as escalas de distância.

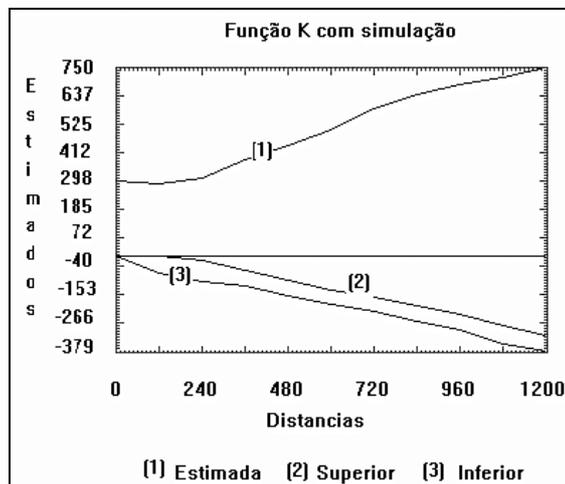


Figura 2-6 – Função K com simulação para os dados de mortalidade neonatal em Porto Alegre.

2.5 PROCESSO PONTUAL MARCADO

Um das situações mais importantes na análise espacial de pontos é a possibilidade de comparação entre dois processos espaciais. Tipicamente, um dos processos representa os casos em estudo, e o outro os casos de um processo pontual que representa um conjunto de casos de controle. Por exemplo, num estudo realizado por Peter Diggle na Inglaterra sobre câncer de laringe, foi utilizado dados de câncer de pulmão como indicadores da distribuição espacial da população. Esta situação pode ser generalizada supondo dois processos pontuais, o primeiro cujos casos localizam-se em (u_1, u_2, \dots, u_l) e o segundo cujos casos estão nos pontos $(u_{n+1}, u_{n+2}, \dots, u_{n+m})$. Cada tipo de evento pode ser modelado como uma distribuição de Poisson, I e II, com intensidades $\lambda_1(u)$ e $\lambda_2(u)$. Define-se o risco na localidade u como uma função da razão entre λ_1 e λ_2 . O objetivo da análise é investigar a variação espacial desta razão na região.

Se estimarmos a intensidade de cada processo através de uma função *kernel*, a razão entre as duas funções será a intensidade do risco. E cada uma das funções estudadas anteriormente pode ser adaptada para verificar a relação entre os pontos do processo I com os pontos do processo II. Por exemplo, visando estudar a dispersão de duas espécies vegetais pode-se verificar a relação de cada ponto com o vizinho mais próximo da outra espécie.

2.6 ESTUDOS CASO-CONTROLE

Considere-se um tipo de estudo onde temos dois tipos de eventos, por exemplo recém-natos que morrem antes de completar um ano e os que sobrevivem a esta idade. Sendo esta variável do tipo binomial a resposta do estudo, dependente de diversas covariáveis tais como prematuridade, existência de doenças na gravidez, escolaridade da mãe, e incluindo sua localização no espaço, pode-se modelar o processo utilizando o método clássico de regressão logística, próprio para este tipo de distribuição. O que particulariza o contexto espacial é a forma de se incluir a localização dos pontos no modelo. Diversas formas de estimar este risco em cada localidade são possíveis, entre as quais utilizar o mesmo kernel da razão como um dos termos da regressão, que toma uma forma semi-paramétrica abaixo:

$$\text{logit}(y_i) = \beta x_i + g(s_i), \quad (2.15)$$

onde:

- y_i é a variável resposta, e tem a forma sim/não, zero/um (óbitos/nascimentos),
- a função de ligação da regressão é o logit, como usual para dados binomiais,
- x_i é o vetor de covariáveis,

- β é o vetor de parâmetros estimado pelo modelo, que no caso da regressão logística é a razão de chances (odds ratio) relacionada a cada covariável,
- $g(s_i)$ é a razão do estimador de intensidade *kernel* de casos e controles.

O ganho deste tipo de modelagem é possibilitar a estimativa da variação espacial do risco, controlando pelos fatores conhecidos de variação de risco. Os procedimentos de estimação dos parâmetros destes modelos baseia-se em métodos iterativos usuais de modelos aditivos generalizados, onde se estima os parâmetros da regressão, e sobre os resíduos estima-se a função *kernel*, e assim sucessivamente até que as estimativas não mais se alterem. O método permite identificar áreas de sobre ou sub risco significativamente diferente da média global. A largura de banda a ser utilizada é importante, e pode ser definida através de métodos automáticos ou selecionada pelo pesquisador visando ajustar a uma conhecida estrutura espacial. No estudo da mortalidade infantil em Porto Alegre (figura 2-1) os dados foram analisados segundo esta proposta, incluindo como fatores de risco individuais: (a) peso ao nascer, (b) semanas gestacionais, (c) sexo da criança, (d) (e) idade da mãe, (f) grau de instrução da mãe, (g) tipo de gravidez e (h) tipo de parto, numa regressão logística cuja expressão é:

$$\log \left\{ \frac{p(s, \mathbf{x})}{1 - p(s, \mathbf{x})} \right\} = \beta_0 + \beta_1 \text{ sexo} + \beta_2 \text{ peso} + \beta_3 \text{ idade} + \beta_4 \text{ inst} + \beta_5 \text{ ges} + \beta_6 \text{ grav} + \beta_7 \text{ parto} + g(s). \quad (2.16)$$

A interpretação dos resultados é razoavelmente direta: os parâmetros β indicam a razão de chances estimada pelo modelo (Quadro 2-1), da forma usual da regressão logística, e no mapa são apresentadas as áreas onde a probabilidade de obter o valor do *kernel* estimado está “significativamente” diferente da intensidade média do processo. O algoritmo para estimar a largura de banda ótima para os dados utiliza validação cruzada de mínimos ponderados para o passo de regressão não-paramétrica. No passo de suavizamento (Eq. 2.15) escolhe-se o valor de h que minimiza:

$$CV(h) = \frac{\sum_{i=1}^n w_i \{z_i - \hat{g}^{-1}(s_i)\}^2}{n}, \quad (2.17)$$

onde $\hat{g}^{-1}(s_i)$ é a estimativa de $g(s_i)$ construída com o valor de banda h usando todos os dados com exceção do par (s_i, z_i) . Testa-se diferentes valores de h , sendo escolhido o que minimiza o somatório.

Quadro 1: Estimativas dos efeitos de covariáveis utilizando o valor da banda obtido por validação cruzada			
Fator	Estimativa	Erro padrão	P-valor
Intercepto	4,0717	0,9487	0,0000
Sexo	-0,3674	0,2713	0,1761
Peso ao nascer	-0,0018	0,0002	0,0000
Idade da mãe	-0,0131	0,0197	0,5059
Instrução da mãe	0,0718	0,2753	0,7942
Duração da gestação	1,1685	0,3737	0,0018
Tipo de gravidez	-0,2006	0,6558	0,7598
Tipo de parto	-0,5320	0,2838	0,0613

A figura 2-7 mostra os mapas de risco para a mortalidade infantil após, incluídas as co-variáveis individuais da criança e da mãe. É interessante observar que no centro da cidade de Porto Alegre existe uma região onde o risco da mortalidade é significativamente menor e outra onde é maior. Quanto às variáveis individuais, somente foram significativas o peso ao nascer, que é reconhecidamente a variável mais associada à mortalidade neo-natal, e a duração da gestação, indicativo de prematuridade. Além de mapeamento do risco, é importante avaliar se a superfície estimada varia significativamente na região, ou seja, se existem evidências estatísticas suficientes para rejeitar a hipótese nula de risco constante na região, tendo-se controlado os fatores individuais de risco. Em termos do modelo, isso equivale ao teste da hipótese $H_0: g(s)=0$. Também é de interesse a construção de contornos de tolerância que auxiliam na identificação de áreas onde o risco é significativamente superior (ou inferior) à média global. Ou seja, reconhecendo o papel de um dado fator como um preditor importante da mortalidade infantil e controlando-o, deseja-se identificar áreas onde o risco é significativamente mais elevado, buscando orientar a intervenção.

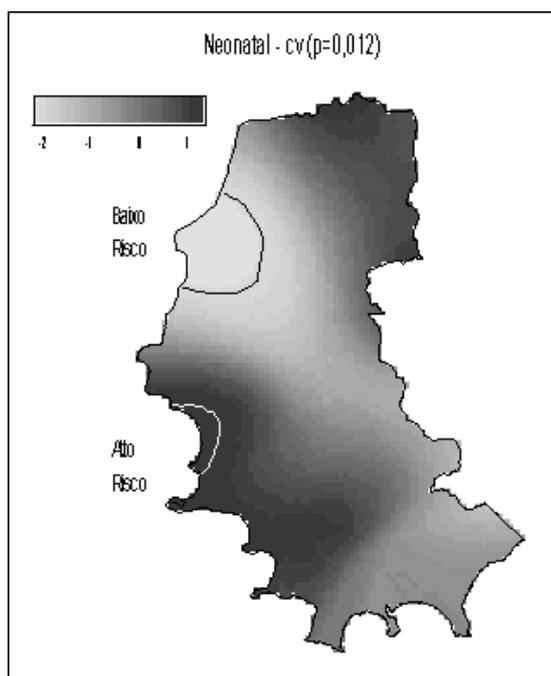


Figura 2-7. Mapas de risco para a mortalidade infantil, controlando para fatores individuais, com a largura de banda estimada por validação cruzada, Porto Alegre, 1998

O teste global do risco e a identificação de áreas de baixo e alto risco podem ser feitos utilizando o método de simulação Monte Carlo, seguindo os passos do algoritmo abaixo:

1. Ajustando-se um modelo de regressão logística convencional, para cada evento – caso ou controle – estima-se a probabilidade ajustada \hat{p}_i . Ou seja, dadas as covariáveis daquele registro, qual é a probabilidade ser um caso.
2. Fixando-se as localizações de cada ponto, amostra-se m dos n indivíduos (sem reposição) com probabilidade proporcional a \hat{p}_i e estes são rotulados como casos e os $n-m$ restantes como controles.
3. Calcula-se uma nova estimativa de $g(s)$, $\hat{g}_1(s)$, a estimativa centralizada em torno da média $\tilde{g}_1(s) = \hat{g}_1(s) - \bar{g}_1$, onde $\bar{g}_1 = n^{-1} \sum_{i=1}^n \hat{g}_1(s_i)$ e a estatística
$$t_1 = n^{-1} \sum_{i=1}^n (\tilde{g}_1(s_i))^2.$$
4. Repete-se os passos 1 e 2 m vezes.
5. Constrói-se uma superfície de p-valores que para cada s fornece a proporção dos valores de $\tilde{g}_j(s)$, $j=1, \dots, m$, menores do que a estimativa original, digamos $\tilde{g}_0(s)$.

6. Adiciona-se os contornos de 0.05 e 0.95 da superfície de p-valores ao mapa de $\tilde{g}_0(s)$ como contornos de 90% de confiança para indicar áreas de alto/baixo risco.
7. Para o teste de hipótese, define-se k o número de $t_j > t_0$ (obtida a partir de $\tilde{g}_0(s)$) e o nível de significância correspondente por $p = (k + 1)/(m + 1)$.

2.7 REFERÊNCIAS

A referência das técnicas mais básicas apresentadas neste capítulo é o livro de Trevor Bailey, “*Spatial Data Analysis by Example*” (Bailey and Gattrel, 1995). As técnicas de caso-controle espacial foram desenvolvidas por Peter Diggle e colaboradores, e a maior parte das rotinas e algoritmos está disponível na página da do Departamento de Matemática e Estatística da Universidade de Lancaster (<http://www.maths.lancs.ac.uk>). O relatório técnico “An S+ library on risk estimation and cluster detection in case-control studies”, de Jarner, M. F. and Diggle, P. J., mostra as funções desenvolvidas e como usá-las. Está disponível em <http://www.maths.lancs.ac.uk/dept/stats/techabstracts02.html>.

Os modelos aditivos generalizados, que servem de base para a extensão espacial podem ser melhor estudados em HASTIE, T. J.; TIBSHIRANI, R. J., 1990, *Generalized Additive Models*. London:Chapman and Hall. Um excelente livro para estudar modelos de regressão é o HOSMER, D. W.; LEMESHOW, S., 1989, *Applied Logistic Regression*. New York:Wiley.

Os trabalho sobre mortalidade infantil em Porto Alegre foi publicado no número especial dos Cadernos de Saúde Pública sobre o tema de estatísticas espaciais em saúde (volume 17(5), outubro-novembro 2001, 1251-1261), disponível na Internet (www.scielo.br).

1. DIGGLE, P. J., 1992. **Point process modelling in environmental epidemiology. Relatório Técnico** MA92/70, Lancaster: Department of Mathematics and Statistics, Lancaster University.
2. KELSALL, J. E.; DIGGLE, P. J. , 1995b. Non-parametric estimation of spatial variation in relative risk. **Statistics in Medicine**, 14:2335-2342.
3. KELSALL, J. E.; DIGGLE, P. J., 1998. Spatial variation in risk of disease: a nonparametric binary regression approach. **Applied Statistics**, 47:559-573.

ANÁLISE ESPACIAL DE SUPERFÍCIES

Eduardo Celso Gerbi Camargo

Suzana Druck Fucks

Gilberto Câmara

3.1 INTRODUÇÃO

No capítulo anterior, apresentamos técnicas de Análise Espacial para eventos discretos, associados a ocorrências pontuais. Neste capítulo, apresentamos técnicas para tratamento e análise de dados de superfícies. De uma forma geral, estes dados estão disponíveis na forma de amostras pontuais, e para utilizá-los de forma efetiva em um ambiente de Geoprocessamento, necessitamos de um procedimento de interpolação, para gerar uma representação na forma de grade regular, como ilustrado na Figura 3-1.

As amostras são valores representativos do fenômeno estudado, usualmente obtidas a partir de levantamento de campo, e que apresentam consistência de metodologia e unidade. Conforme explicado no capítulo 1, essas amostras podem representar tanto variáveis naturais (como teor de argila no solo) como socioeconômicas (como taxa de homicídios).

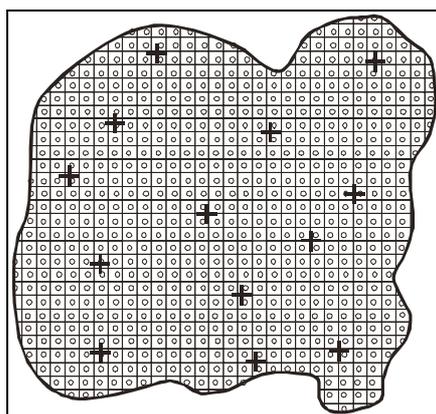


Figura 3-1 – Ilustração do processo de interpolação: amostras (cruzes) e aproximação da superfície por uma grade regular (círculos).

Para gerar superfícies que aproximem o fenômeno estudado de forma realista, é necessário modelar sua variabilidade espacial. Os modelos que objetivam gerar superfícies a partir de procedimentos de interpolação, de forma geral, representam a variável em estudo como uma combinação da variabilidade em larga e pequena

escala. Esse enfoque, entretanto, não é único. Assim, pode-se tomar três grandes abordagens: Para tanto, pode-se tomar três grandes abordagens:

- *Modelos determinísticos de efeitos locais*: cada ponto da superfície é estimado apenas a partir da interpolação das amostras mais próximas, utilizando funções como *inverso do quadrado da distância*. A suposição implícita é que predominam os efeitos puramente locais. Neste caso, não é feita qualquer hipótese estatística sobre a variabilidade espacial. Estes interpoladores serão apresentados na seção 3.2 deste capítulo.
- *Modelos determinísticos de efeitos globais*: a suposição implícita nesta classe de interpoladores é que, para a caracterização do fenômeno em estudo, predomina a variação em larga escala, e que a variabilidade local não é relevante. Este é caso do interpoladores por superfícies de tendência, apresentados na seção 3.3 deste capítulo.
- *Modelos estatísticos de efeitos locais e globais (krigagem)*: cada ponto da superfície é estimada apenas a partir da interpolação das amostras mais próximas, utilizando um estimador estatístico. Esses procedimentos requerem que a variabilidade local e global sejam modelada através de modelos apresentados como

$$Z(x) = \sum_{j=1}^p \beta_j f_j + \varepsilon(x)$$

Nesse caso $E\{Z(x)\} = \sum_{j=1}^p \beta_j f_j$ aonde β_j é um conjunto de parâmetros desconhecidos e f_j um conjunto de funções básicas, em geral, polinomiais.

Esses estimadores apresentam propriedades de não ser tendenciosos e de procurar minimizar os erros inferenciais. Eles podem ser estimados através de procedimentos como a krigagem universal e as funções intrínsecas de ordem k não abordadas nesse capítulo.

Neste capítulo, iremos dar ênfase ao uso de técnicas de krigagem ordinária, ou seja a um caso particular desse modelo global em que $p=1$ e $k=0$, aonde k representa a ordem da função f_j , e β_1 igual a média local. A ênfase nesse procedimento é devido às suas propriedades, sua grande importância na modelagem de fenômenos naturais e também porque esse capítulo objetiva procedimentos que priorizam a interpolação espacial (predição). A modelagem de tendências ou variação em larga escala se faz necessária quando a etiologia de um fenômeno deve ser estudada e aonde a estimação da tendência é importante na compreensão do fenômeno. As técnicas da krigagem são discutidas a partir da seção 3.4. Para a comparação entre os interpoladores, foram utilizados dados da EMBRAPA – Solos,

obtidos na Fazenda Canchim, em São Carlos - SP. Trata-se de amostragem de 85 observações georreferenciadas coletadas no horizonte Bw (camada do solo com profundidade média de 1m), conforme ilustra a Figura 3-2. Dentre as variáveis disponíveis, selecionou-se para estudo o teor de argila, cujas estatísticas básicas amostrais são apresentadas na Tabela 3.1.

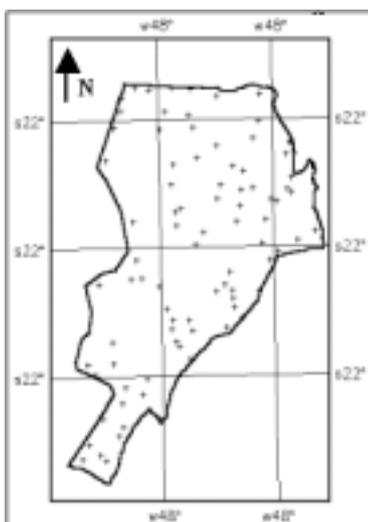
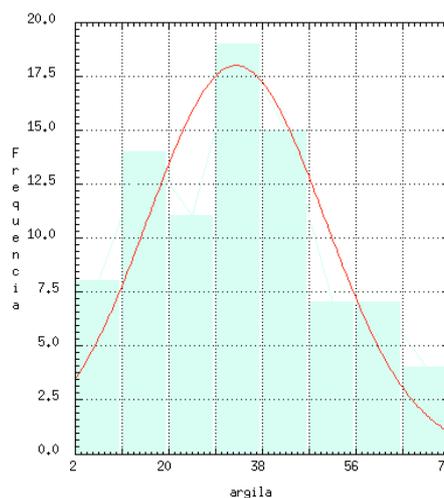


Figura 3-2- Disposição das amostras de teor de argila da Fazenda Canchim (EMBRAPA).

Tabela 3-1 - ESTATÍSTICAS DA AMOSTRA.

Número de observações	85
Média	33,035
Variância	288,034
Desvio Padrão	16,972
Coefficiente de variação	0,514
Coefficiente de assimetria	0,214
Coefficiente de curtose	2,344
Quartil Inferior	10
Mediana	33
Quartil superior	43



O histograma das amostras mostra que a distribuição do teor de argila é levemente alongada à direita. Neste caso, a distribuição é dita ser positivamente assimétrica, com coeficiente de assimetria de 0,214. Quanto ao grau de achatamento, o coeficiente de curtose (2,344) indica que a distribuição é

ligeiramente platicúrtica. Dentre outros valores apresentados na Tabela 3-1, nota-se que a média e a mediana, medidas que procuram caracterizar o centro da mesma distribuição de frequências, possuem valores próximos (33,035 e 33,0), respectivamente. Assim sendo, a distribuição da variável em estudo, pode ser considerada aproximadamente simétrica.

3.2 MODELOS DETERMINÍSTICOS LOCAIS

Uma alternativa simples para gerar uma superfície bidimensional a partir de amostras pontuais é ajustar uma função bidimensional sobre os amostras considerados, compondo uma superfície cujo valor será proporcional à local intensidade de amostras. A formulação geral para este tipo de interpolação é:

$$\hat{z}_i = \frac{\sum_{j=1}^n w_{ij} z_j}{\sum_{j=1}^n w_{ij}}, \quad (3.1)$$

onde: z_i é o valor de cota de um ponto i qualquer da grade, z_j é a cota de uma amostra j vizinha do ponto i da grade e w_{ij} é um fator de ponderação. A Figura 3-3 ilustra o procedimento de estimação.

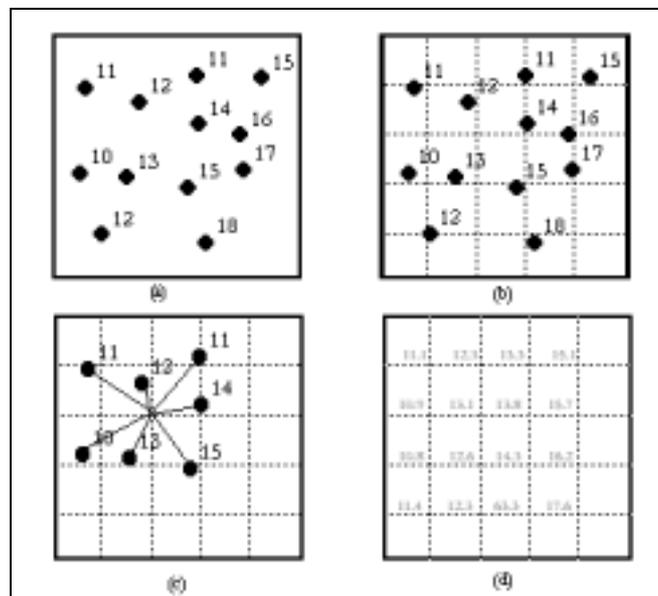


Figura 3-3 Ilustração do processo de interpolação por estimador local: (a) configuração original de amostras; (b) grade regular superposta às amostras; (c) interpolação de um valor a partir dos vizinhos; (d) grade regular resultante

Variações desse esquema básico são os interpoladores: (a) por vizinho mais próximo; (b) por média simples; (c) por média ponderada; Nos três primeiros casos, considera-se uma região em torno do ponto a ser interpolado como contendo os pontos que influenciam na interpolação. A *interpolação por vizinho mais próximo* é definida pela escolha de apenas uma amostra vizinha para cada ponto da grade. Este interpolador deve ser usado quando se deseja manter os valores de cotas das amostras na grade, sem gerar valores intermediários. A *interpolação por média simples* considera o valor de cota z do elemento da grade igual a média aritmética dos valores de cota das amostras vizinhas. Neste caso considera-se que o fator de ponderação w_{ij} é igual a $1/n$ para qualquer amostra considerada. Na *interpolação por média ponderada* o valor de cota de cada elemento da grade é definido pela média ponderada dos valores de cota das amostras vizinhas. A ponderação mais usada na prática é o inverso da distância euclidiana do ponto da grade à amostra considerada ou seja:

$$w_{ij} = 1/d_{ij}^k, \quad (3.2)$$

onde: k é o expoente da distância, geralmente igual a 1 ou 2 e; d_{ij} é o valor de distância da amostra j ao ponto i da grade, expresso por:

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (3.3)$$

Uma comparação visual entre os resultados desses interpoladores é mostrada na Figura 3-4 para os dados do teor de argila da Fazenda Canchim. Os mapas ilustram os defeitos típicos dessas funções simples: as funções de vizinho mais próximo e média simples tendem a produzir superfícies com variações abruptas; no caso do inverso do quadrado da distância, os máximos locais tendem a ser muito acentuados, formando “picos” artificiais.

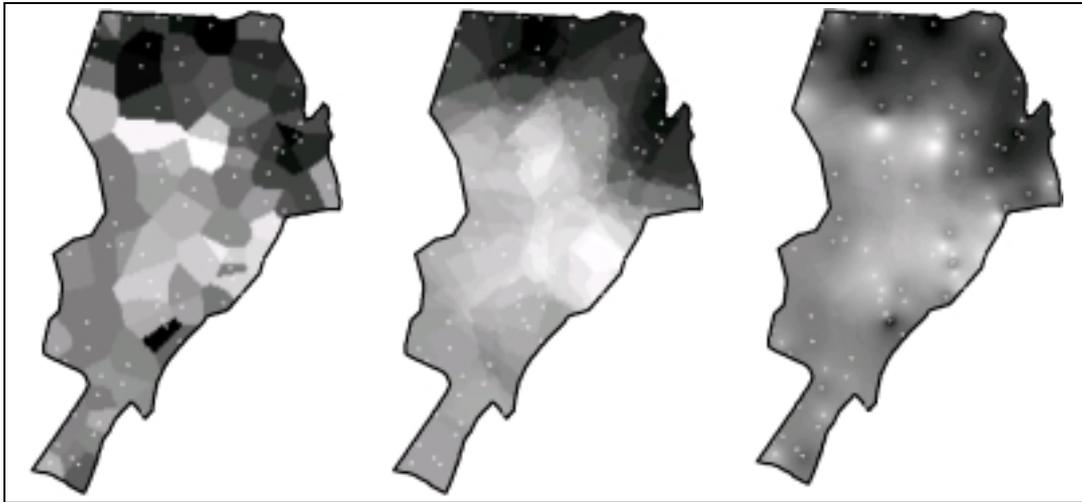


Figura 3-4 - Comparação entre interpoladores de média móvel, para o mesmo conjunto de amostras. À direita, inverso do quadrado da distância; no centro, média simples; à esquerda, vizinho mais próximo. Regiões mais claras representam alto valores e vice-versa.

Um refinamento desses estimadores é o uso de uma função de ponderação mais complexa que a *média simples* ou o *inverso do quadrado da distância*. Esta classe de estimadores é descrita na literatura como *kernel estimators*, ou *estimadores de densidade não-paramétricos*. Estes estimadores generalizam a idéia de média móvel local, ao supor que a densidade do fenômeno varia localmente de forma suave, sem “picos” nem “descontinuidades”. Seu objetivo é produzir superfícies mais suaves, que se espera mais representativas de fenômenos naturais e socioeconômicos. Estes estimadores são do mesmo tipo que os discutidos no capítulo 2 para o caso de eventos pontuais, agora generalizados para o caso de amostras.

Um *kernel estimator* é um estimador cujos parâmetros básicos são: (a) um raio de influência que define a vizinhança do ponto a ser interpolado; (b) uma função de estimação com propriedades “convenientes” de suavização do fenômeno. Para toda posição z_i cujo valor queremos estimar, o estimador de intensidade será computado a partir dos valores das amostras $\{z_1, \dots, z_n\}$ contidos num raio de tamanho τ , e da distância euclidiana d_{ij} entre a i -ésima posição e a j -ésima amostra (como expresso na equação 3.3), a partir de funções do tipo

$$\hat{z}_i = \frac{\sum_{j=1}^n k\left(\frac{d_{ij}}{\tau}\right) z_j}{\sum_{j=1}^n k\left(\frac{d_{ij}}{\tau}\right)}, \quad d_{ij} \leq \tau \quad (3.4)$$

Esta fórmula é uma generalização da equação 3.1, na qual o cômputo dos pesos w_{ij} foi substituído por uma função generalizada dependente da distância. Exemplos destas funções incluem o *kernel* gaussiano

$$k(x, y, \tau) = \frac{1}{2\pi\tau} \exp\left(-\frac{d_{ij}^2}{2\tau^2}\right), \quad (3.5)$$

ou o *kernel* de quarta ordem

$$k(x, y, \tau) = \frac{3}{\pi\tau^2} \left(1 - \frac{d_{ij}^2}{\tau^2}\right)^2 \quad (3.6)$$

Para ilustrar esta classe de estimadores, foram geradas duas superfícies a partir das mesmas amostras usadas para produzir os mapas da Figura 3-4. A partir de um *kernel* de quarta ordem (equação 3.6), foram gerados dois mapas mostrados na Figura 3-5, com raios de busca de 500 e 1500 metros. A comparação entre os mapas mostra a grande importância de uma seleção apropriada do raio de busca no uso de *kernel estimators*. No primeiro mapa predominam os efeitos locais, pelo uso de um raio de busca reduzido; o segundo mapa evidencia melhor a distribuição do fenômeno, pelo uso de um raio mais apropriado aos dados.

Em resumo, os *kernel estimators* são uma alternativa viável a métodos mais sofisticados de interpolação, pois não requerem a parametrização da estrutura de correlação espacial (como no caso da geoestatística). As superfícies interpoladas são suaves e aproximam muitos fenômenos naturais e socioeconômicos. As desvantagens destes estimadores são a forte dependência no raio de busca e a excessiva suavização da superfície, que pode em alguns casos esconder variações locais importantes.

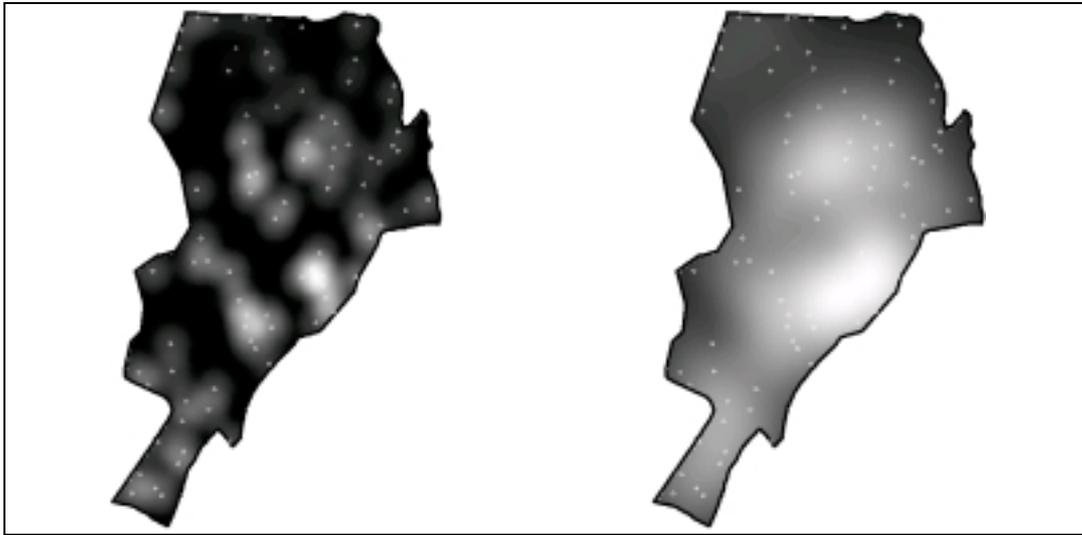


Figura 3-5- Superfícies de teor de argila interpoladas por kernel de quarta ordem. À esquerda, raio de busca de 500m; à direita, raio de busca de 1500m.

3.3 SUPERFÍCIES DE TENDÊNCIA

As superfícies de tendência são interpoladores determinísticos globais. A superfície é aproximada por um ajuste polinomial aos dados, através de um processo de regressão múltipla entre os valores do atributo e as localizações geográficas. Essa função polinomial é então utilizada para estimar os valores dos pontos em todas as localizações de uma grade regular que aproxima a superfície.

As superfícies de tendência buscam modelar a variação espacial em larga escala através de uma regressão múltipla entre os valores de atributo e as localizações geográficas. A saída é uma função polinomial na qual o valor do atributo é expresso em função das coordenadas da superfície, expressas em duas ou três dimensões. Exemplos incluem equações lineares do tipo:

$$z = \alpha_1 + \alpha_2 x + \alpha_3 y \quad (3.7)$$

e equações quadráticas como:

$$w = \alpha_1 + \alpha_2 x + \alpha_3 y + \alpha_4 xy + \alpha_5 x^2 + \alpha_6 y^2 \quad (3.8)$$

A suposição implícita nos interpoladores por superfícies de tendência é que, para a caracterização do fenômeno em estudo, predomina a variação em larga escala, e que a variabilidade local não é relevante. Neste modelo, a função de autocorrelação continua decaindo mesmo após ultrapassar a distância onde há influências locais; a covariância não se estabiliza com a distância e assim o fenômeno analisado é não-estacionário.

Para o caso dos dados de teor de argila da Fazenda Canchim (acima descritos), foi realizada uma análise de tendência usando uma regressão linear. Os ajustes indicaram um coeficiente de determinação (R^2 ajustado) de apenas 17,3%, o que indica não haver efeitos espaciais significativos de larga escala. Deste modo, pode-se esperar que estes dados sejam modeláveis por interpoladores locais, sejam determinísticos (seção 3.2) ou estocásticos (seção 3.4 e seguintes).

Um exemplo típico de superfícies de tendência é o uso de dados de longitude, latitude e altitude para estimar a distribuição de temperatura. Neste caso, o objetivo foi estimar a distribuição de temperatura para o estado de Santa Catarina, para a época do plantio de soja, em intervalos de 10 dias (decêndios). Partindo da época recomendada para semeadura e do ciclo de diferentes cultivares de soja, determinou-se um período de análise compreendido entre 11/10 e 20/05 (22 decêndios), permitindo que cultivares com ciclos diferentes, semeadas dentro da época recomendada, tivessem todo o seu ciclo avaliado neste estudo. Foram coletados dados de temperatura média diária e precipitação diária de 27 estações meteorológicas monitoradas pela Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina S. A. – Epagri, com uma série histórica de aproximadamente cinco anos, mostrados na Figura 3-6.

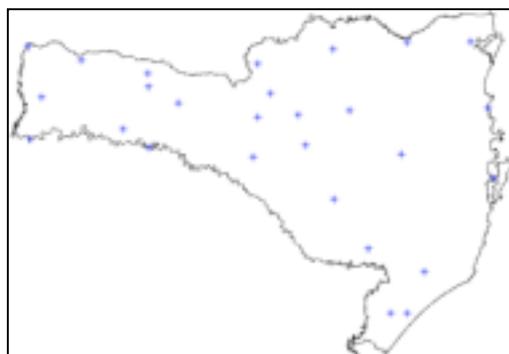


Figura 3-6– Distribuição espacial das estações monitoradas pela Epagri.

A partir dos dados diários, foi calculada a média decendial. Esta média das 27 estações foi utilizada no cálculo de superfícies de tendência a partir de uma equação do tipo:

$$z(x, y, h) = \alpha_1 x + \alpha_2 y + \alpha_3 h + \alpha_4 \quad (3.9)$$

onde z é a temperatura calculada a partir da longitude (x), latitude (y) e altitude (h). Para o primeiro decêndio (11/10 a 20/10), os resultados estão mostrados na Tabela 3.1. Na análise dos coeficientes da regressão, mostrada na Tabela 3.2, a relação entre as variáveis independentes com a variável dependente (temperatura média decendial) foi verificada, inicialmente, pelo teste “F” e, depois,

pelo teste “t” de Student. Esta análise indicou todos os coeficientes como significativos. A normalidade dos resíduos foi avaliada pelo teste de Keifer-Salmon, e aceita a hipótese.

Tabela 3-2 - Coeficientes para Estimativa de Temperatura em Santa Catarina (Decêndio de 11/10 a 20/10).=

	Valor	Teste F	Teste T	p-valor	Comentários
Intercepto	9,475	7,169			Significativo
Latitude	-0,447	0,169	-2,637		(idem)
Longitude	0,466	0,085	5,488		(idem)
Altitude	-0,005	0,000	-16,162		(idem)
R ² ajustado	0,909				

A grande vantagem das superfícies de tendência é sua simplicidade e facilidade de cálculo. No entanto, a suposição implícita do modelo, em negligenciar a variabilidade local, não é realista para a maior parte dos dados naturais. Adicionalmente, os parâmetros estimados são muito sensíveis a valores extremos (*outliers*). Apesar destes problemas, as superfícies de tendência são úteis para remover efeitos de primeira ordem, quando a média varia de forma consistente no espaço. Outros usos importantes são a análise dos resíduos de estimação; tais resíduos também são bastante informativos, pois mostram a existência de sub-regiões que apresentam diferenças significativas na tendência geral.

No exemplo apresentado, trata-se de uma situação favorável, em que, em função do comportamento da temperatura, da época do ano e das características do estado de Santa Catarina, apenas a variação em larga escala foi capaz de produzir estimativas acuradas. Esta situação não é a mais usual. Na prática, na maior parte das vezes as variações locais não podem ser ignoradas. Neste caso, será preciso modelar o comportamento da variável e para isto, utiliza-se a abordagem geoestatística, descrita a seguir.

3.4 MODELOS ESTATÍSTICOS DE EFEITOS LOCAIS E GLOBAIS: KRIGAGEM

3.1.1 FUNDAMENTAÇÃO TEÓRICA

A krigagem compreende um conjunto de técnicas de estimação e predição de superfícies baseada na modelagem da estrutura de correlação espacial. A hipótese implícita no procedimento geoestatístico é que o processo estudado é estacionário (veja-se a definição de estacionariedade no capítulo 1 do livro). Os passos num estudo empregando técnicas de krigagem incluem:

- (a) análise exploratória dos dados;
- (b) análise estrutural (modelagem da estrutura de correlação espacial);
- (c) interpolação estatística da superfície.

O procedimento de interpolação é chamado de *krigagem* em honra a Daniel Krige, o pioneiro em introduzir o uso de médias móveis para evitar a superestimação sistemática de reservas em mineração. O que diferencia a krigagem de outros métodos de interpolação é a estimação de uma matriz de covariância espacial que determina os pesos atribuídos às diferentes amostras, o tratamento da redundância dos dados, a vizinhança a ser considerada no procedimento inferencial e o erro associado ao valor estimado. Além disso, a krigagem também fornece estimadores com propriedades de não tendenciosidade e eficiência.

A estrutura teórica da krigagem está baseada no conceito de *variável regionalizada*, desenvolvida por Georges Matheron. Uma variável regionalizada é uma variável distribuída no espaço (ou tempo) cujos valores são considerados como realizações de uma função aleatória (ou processo aleatório, ou campo aleatório, ou processo estocástico). Esta teoria permite incluir hipóteses estatísticas em processos espaciais locais. A variação espacial de uma variável regionalizada pode ser expressa pela soma de três componentes: a) uma componente estrutural, associada a um valor médio constante ou a uma tendência constante; b) uma componente aleatória, espacialmente correlacionada; e c) um ruído aleatório ou erro residual. Se o vetor x representa uma posição em uma, duas ou três dimensões, então o valor da função aleatória Z , em x , é dada por:

$$Z(x) = \mu(x) + \varepsilon'(x) + \varepsilon'' \quad (3.10)$$

onde:

- $\mu(x)$ é uma função determinística que descreve a componente estrutural de Z em x ;
- $\varepsilon'(x)$ é um termo estocástico correlacionado, que varia localmente;

- ε'' é um ruído aleatório não correlacionado, com distribuição normal com média zero e variância σ^2 .

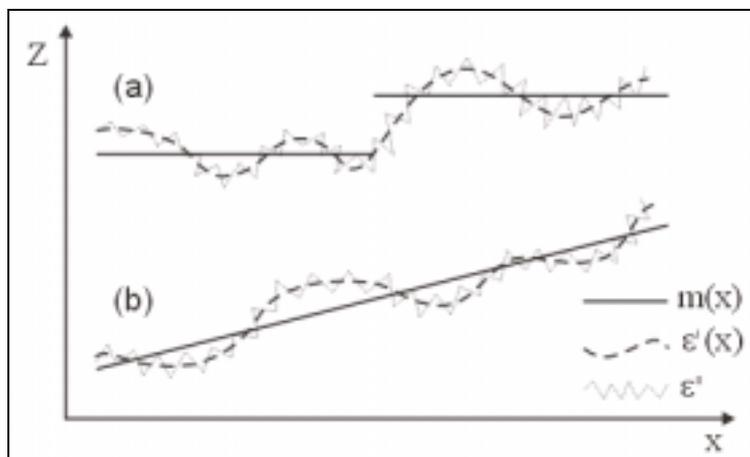


Figura 3-7- Componentes de uma variável regionalizada.

As Figura 3-7(a) e (b) ilustram as três componentes principais da variação espacial. A Figura 3.8(a) apresenta uma componente determinística que possui um comportamento regular (diferença entre os níveis médios), enquanto a componente determinística na Figura 3.8(b) apresenta uma tendência constante.

A hipótese mais simples sobre o comportamento da variável regionalizada é que a média do fenômeno, $\mu(\mathbf{x})$, seja constante na região de estudo, o que implica em não haver variação significativa na larga escala. Esta hipótese dá origem aos interpoladores de *Krigagem ordinária*, discutida a seguir. No caso de se querer modelar uma tendência, há vários métodos disponíveis: Krigagem Universal, Funções Aleatórias Intrínsecas de Ordem k, não discutidos neste capítulo.

Na hipótese da Krigagem ordinária, $\mu(x)$ é constante e denotada por m . Deste modo, o valor esperado da função aleatória Z nas posições x e $x + \mathbf{h}$ são iguais a m . Isto implica que o valor esperado da diferença entre os valores observados em x e $x + \mathbf{h}$, separados por um vetor de distância \mathbf{h} , é nulo:

$$E [Z(x) - Z(x+\mathbf{h})] = 0 \quad (3.11)$$

Admite-se também que o fenômeno considerado seja *estacionário de segunda ordem*, isto é, a covariância entre dois pares quaisquer $Z(x)$ e $Z(x + \mathbf{h})$, separados por um vetor distância \mathbf{h} , existe e depende somente de \mathbf{h} . Então:

$$C(\mathbf{h}) = COV [Z(x), Z(x+\mathbf{h})] = E[Z(x).Z(x+\mathbf{h})] - m^2 \quad (3.12)$$

Adicionalmente, a estacionariedade da covariância implica na estacionariedade da variância:

$$\text{Var}(Z(\mathbf{x})) = E [Z(\mathbf{x}) - m]^2 = E[Z^2(\mathbf{x})] - 2E[Z(\mathbf{x})].m + m^2 \quad (3.13)$$

ou ainda

$$\text{Var}(Z(\mathbf{x})) = E[Z^2(\mathbf{x})] - 2m.m + m^2 = E[Z^2(\mathbf{x})] - m^2 = C(0) \quad (3.14)$$

Deste modo, verifica-se que as hipóteses de média constante e estacionariedade da covariância implicam que a determinação da função $C(\mathbf{h})$ é suficiente para caracterizar a *variável regionalizada*. Isto quer dizer que, com base na Equação 3.10, a função $C(\mathbf{h})$ permite caracterizar o termo estocástico $\varepsilon'(\mathbf{x})$. Para determinar $C(\mathbf{h})$, utiliza-se uma função auxiliar, chamada de função *variograma* $2\gamma(\mathbf{h})$, definida por:

$$2\gamma(\mathbf{h}) = E[Z(\mathbf{x}) - Z(\mathbf{x}+\mathbf{h})]^2 \quad (3.15)$$

que pode ser desenvolvida em:

$$2\gamma(\mathbf{h}) = E[Z^2(\mathbf{x}) - 2Z(\mathbf{x}).Z(\mathbf{x}+\mathbf{h}) - Z^2(\mathbf{x}+\mathbf{h})] \quad (3.16)$$

ou ainda

$$2\gamma(\mathbf{h}) = E[Z^2(\mathbf{x})] - 2E[Z(\mathbf{x}).Z(\mathbf{x}+\mathbf{h})] - E[Z^2(\mathbf{x}+\mathbf{h})] \quad (3.17)$$

Da equação (3.14), obtém-se

$$E[Z^2(\mathbf{x})] = E[Z^2(\mathbf{x}+\mathbf{h})] = C(0) + m^2 \quad (3.18)$$

e da equação (3.13) obtém-se

$$E[Z(\mathbf{x}).Z(\mathbf{x}+\mathbf{h})] = C(\mathbf{h}) + m^2 \quad (3.19)$$

Substituindo as equações (3.18) e (3.19) na equação (3.17), obtém-se:

$$2\gamma(\mathbf{h}) = 2C(0) - 2C(\mathbf{h}) \text{ ou } \gamma(\mathbf{h}) = C(0) - C(\mathbf{h}) \quad (3.20)$$

onde:

$\gamma(\mathbf{h})$ representa o *semivariograma*, que é metade do variograma. A relação em (3.20) indica que sob a hipótese de estacionariedade de 2ª ordem, que a covariância e o semivariograma são formas alternativas de caracterizar a autocorrelação dos pares $Z(\mathbf{x})$ e $Z(\mathbf{x}+\mathbf{h})$ separados pelo vetor \mathbf{h} .

3.1.2 DETERMINAÇÃO EXPERIMENTAL DO SEMIVARIOGRAMA

O semivariograma é uma ferramenta básica de suporte às técnicas de Krigagem, pois permite representar quantitativamente a variação de um fenômeno regionalizado no espaço. O semivariograma pode ser calculado experimentalmente, considerando o esquema de amostragem em duas dimensões mostrado na Figura 3-8, onde $z(\mathbf{x})$ denota o valor de uma posição cujos componentes são (x_1, y_1) , e $z(\mathbf{x}+\mathbf{h})$ o valor da amostra numa posição cujos componentes são (x_2, y_2) , sendo \mathbf{h} um vetor distância (módulo e direção) que separa os pontos.

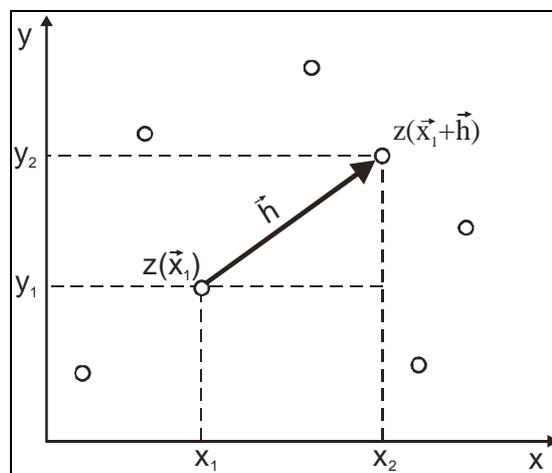


Figura 3-8 – Amostragem em duas dimensões.

A determinação experimental do semivariograma, para cada valor de \mathbf{h} , considera todos os pares de amostras $z(\mathbf{x})$ e $z(\mathbf{x}+\mathbf{h})$, separadas pelo vetor distância \mathbf{h} , a partir da equação:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2 \quad (3.21)$$

onde:

$\hat{\gamma}(h)$ é o semivariograma estimado e $N(h)$ é o número de pares de valores medidos, $z(\mathbf{x})$ e $z(\mathbf{x}+\mathbf{h})$, separados pelo vetor \mathbf{h} . Esta fórmula, entretanto, não é robusta. Podem existir situações em que variabilidade local não é constante e se modifica ao longo da área de estudo (heteroscedasticidade). Um caso particular desse fato (denominado *efeito proporcional*) ocorre quando as distribuições são assimétricas e a média se correlaciona com a variância. O estimador de semivariograma apresentado em (3.22) não é resistente a esse efeito e apresenta tendências que impedem a estimação correta de seus parâmetros. Para expressões alternativas, deve-se consultar Cressie (1993).

Na prática, pode-se fazer a hipótese adicional de que o fenômeno é *isotrópico* (com comportamento igual em todas as direções). Neste caso, a determinação experimental do semivariograma depende apenas da distância entre as amostras e não da direção relativa entre elas. O tratamento da anisotropia (caso em que a estrutura espacial do fenômeno varia conforme a direção) é discutido no Apêndice deste capítulo.

As hipóteses de estacionariedade e média constante levam a postular um comportamento idealizado para o semivariograma experimental, mostrado na Figura 3-9. Espera-se que observações mais próximas geograficamente tenham um comportamento mais semelhante entre si do que aquelas separadas por maiores distâncias. Assim, o valor absoluto da diferença entre duas amostras $z(x)$ e $z(x+h)$ deveria crescer à medida que aumenta a distância entre elas, até um valor na qual os efeitos locais não teriam mais influência.

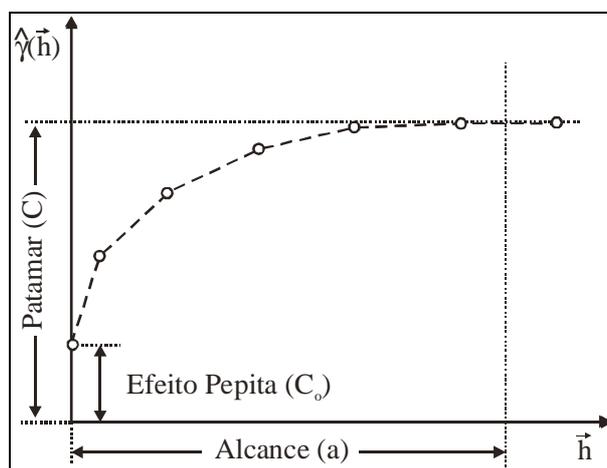


Figura 3-9 – Parâmetros do variograma.

Os parâmetros do semivariograma podem ser observados na Figura 3-9:

- *Alcance (a)*: distância dentro da qual as amostras apresentam-se correlacionadas espacialmente.
- *Patamar (C)*: é o valor do semivariograma correspondente a seu *alcance (a)*. Deste ponto em diante, considera-se que não existe mais dependência espacial entre as amostras, porque a variância da diferença entre pares de amostras ($Var[Z(x) - Z(x+h)]$) torna-se aproximadamente constante.
- *Efeito Pepita (C_0)*: idealmente, $\gamma(\mathbf{0})=0$. Entretanto, na prática, à medida que h tende para zero, $\gamma(h)$ se aproxima de um valor positivo chamado *Efeito Pepita (C_0)*, que revela a descontinuidade do semivariograma para distâncias menores do que a menor distância entre as amostras. 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). Parte desta descontinuidade pode ser também devida a erros de medição, sendo impossível quantificar se a maior contribuição provém dos erros de medição ou da variabilidade de pequena escala não captada pela amostragem.

3.1.3 *MODELOS TEÓRICOS*

O gráfico do semivariograma experimental, $\hat{\gamma}(h)$, calculado através da Equação (3.22), é formado por uma série de valores, conforme ilustra a Figura 3-9, sobre os quais se objetiva ajustar uma função. É importante que o modelo ajustado represente a tendência de $\hat{\gamma}(h)$ em relação a h . Deste modo, as estimativas obtidas a partir da krigagem serão mais exatas e, portanto mais confiáveis.

O procedimento de ajuste não é direto e automático, como no caso de uma regressão, por exemplo, mas sim interativo, pois nesse processo o intérprete faz um primeiro ajuste e verifica a adequação do modelo teórico. Dependendo do ajuste obtido, pode ou não redefinir o modelo, até obter um que seja considerado satisfatório.

Os modelos aqui apresentados são considerados modelos básicos, denominados modelos isotrópicos. Estão divididos em dois tipos: modelos com patamar e modelos sem patamar. Modelos do primeiro tipo são referenciados na geoestatística como modelos transitivos. Alguns dos modelos transitivos atingem o patamar (C) assintoticamente. Para tais modelos, o alcance (a) é arbitrariamente definido como a distância correspondente a 95% do patamar. Modelos do segundo tipo não atingem o patamar, e continuam aumentando enquanto a distância aumenta. Tais modelos são utilizados para modelar fenômenos que possuem capacidade infinita de dispersão. Os modelos transitivos mais utilizados são: modelo esférico (Sph), modelo exponencial (Exp) e modelo gaussiano (Gau). Estes modelos estão apresentados na Figura 3-10 com o mesmo alcance (a).

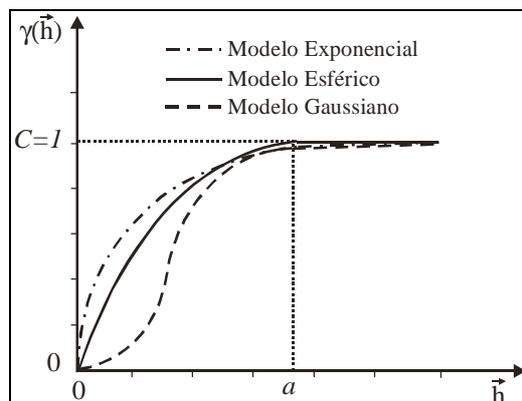


Figura 3-10 – Representação gráfica de modelos transitivos normalizados.

Modelo Esférico

O modelo esférico é um dos modelos mais utilizados e está representado na Figura 3-10. A equação normalizada deste modelo é:

$$\text{Sph}(\mathbf{h}) = \begin{cases} 0 & , \|\mathbf{h}\|=0 \\ 1,5\left(\frac{\|\mathbf{h}\|}{a}\right) - 0,5\left(\frac{\|\mathbf{h}\|}{a}\right)^3 & , 0 < \|\mathbf{h}\| \leq a \\ 1 & , \|\mathbf{h}\| > a \end{cases} \quad (3.22)$$

Modelo Exponencial

Um outro modelo bastante utilizado é o modelo exponencial, o qual é apresentado na Figura 3-10. A equação normalizada deste modelo é:

$$\text{Exp}(\mathbf{h}) = \begin{cases} 0 & , \|\mathbf{h}\|=0 \\ 1 - \exp\left(-\frac{\|\mathbf{h}\|}{a}\right) & , \|\mathbf{h}\| \neq 0 \end{cases} \quad (3.23)$$

Este modelo atinge o patamar assintoticamente, com o alcance prático definido como a distância na qual o valor do modelo é 95% do patamar.

Modelo Gaussiano

O modelo gaussiano é um modelo transitivo, muitas vezes usado para modelar fenômenos extremamente contínuos. Sua formulação é dada por:

$$\text{Gau}(\mathbf{h}) = \begin{cases} 0 & , \|\mathbf{h}\|=0 \\ 1 - \exp\left(-\frac{\|\mathbf{h}\|^2}{a}\right) & , \|\mathbf{h}\| \neq 0 \end{cases} \quad (3.24)$$

Semelhante no modelo exponencial, o modelo gaussiano atinge o patamar assintoticamente e o parâmetro a é definido como o alcance prático ou distância na qual o valor do modelo é 95% do patamar. O que caracteriza este modelo é seu comportamento parabólico próximo à origem, conforme a Figura 3-10 .

Até este ponto foram apresentados os principais modelos básicos normalizados, os quais são utilizados para ajustar o semivariograma experimental. Na prática, os semivariogramas experimentais possuem valores de efeito pepita (C_o) maior que zero e valores de patamar (C) maiores que a unidade, conforme ilustrado na Figura 3-11.

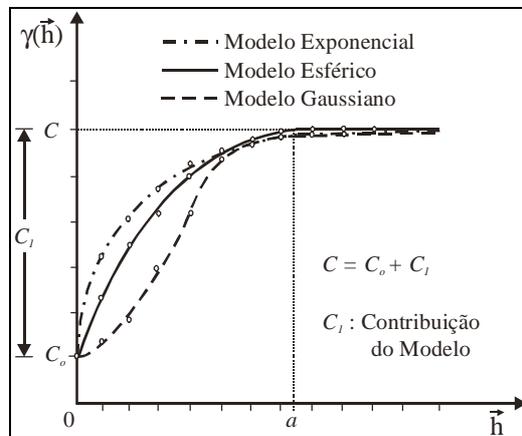


Figura 3-11 - Representação gráfica de semivariogramas experimentais e modelos teóricos.

Em resumo, os semivariogramas dos modelos transitivos básicos são assim definidos:

- *Modelo Esférico de Semivariograma:*

$$\gamma(h) = \begin{cases} 0 & , |h| = 0 \\ C_o + C_I \left[\frac{3}{2} \left(\frac{|h|}{a} \right) - \frac{1}{2} \left(\frac{|h|}{a} \right)^3 \right] = C_o + C_I [Sph(|h|)] & , 0 < |h| \leq a \\ C_o + C_I & , |h| > a \end{cases} \quad (3.25)$$

- *Modelo Exponencial de Semivariograma:*

$$\gamma(h) = \begin{cases} 0 & , |h| = 0 \\ C_o + C_I \left[1 - \exp\left(-\frac{|h|}{a}\right) \right] = C_o + C_I [Exp(|h|)] & , |h| \neq 0 \end{cases} \quad (3.26)$$

- *Modelo Gaussiano de Semivariograma:*

$$\gamma(h) = \begin{cases} 0 & , |h| = 0 \\ C_0 + C_1 \left[1 - \exp\left(-\frac{|h|}{a}\right)^2 \right] & , |h| \neq 0 \end{cases} = C_0 + C_1 [\text{Gau}(|h|)] , |h| \neq 0 \quad (3.27)$$

Modelos Aninhados

Existem determinados fenômenos em que são necessários modelos mais complexos de semivariograma para explicar suas variações espaciais. Estes modelos são combinações de modelos simples, denominados *aninhados*; em muitos casos, os modelos aninhados são necessários para explicar a variação de fenômenos decorrentes da combinação de fatores independentes de formação. Por exemplo, um modelo aninhado útil em estudos de mineração e pesquisa de solo é o duplo esférico, definido como:

$$\gamma(h) = \begin{cases} C_0 + C_1 \left[\frac{3}{2} \left(\frac{|h|}{a_1} \right) - \frac{1}{2} \left(\frac{|h|}{a_1} \right)^3 \right] = \gamma_1(h), & 0 < |h| \leq a_1 \\ C_0 + C_2 \left[\frac{3}{2} \left(\frac{|h|}{a_2} \right) - \frac{1}{2} \left(\frac{|h|}{a_2} \right)^3 \right] = \gamma_2(h), & a_1 < |h| \leq a_2 \\ C_0 + C_1 + C_2 & , |h| > a_2 \\ 0 & , |h| = 0 \end{cases} \quad (3.28)$$

onde,

- a_1 e C_1 correspondem aos parâmetros de alcance e contribuição, respectivamente, do primeiro modelo esférico ($\gamma_1(h)$).
- a_2 e C_2 correspondem aos parâmetros de alcance e contribuição, respectivamente, do segundo modelo esférico ($\gamma_2(h)$).

Este modelo é mostrado na Figura 3-12, onde as linhas sólida e pontilhada representam os modelos de ajuste teórico ao semivariograma experimental.

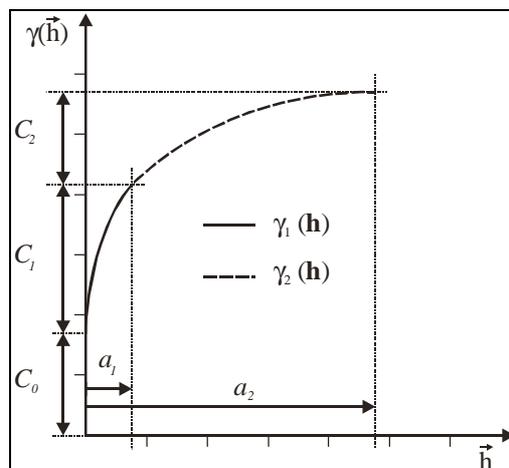


Figura 3-12 - Representação gráfica de um modelo duplo esférico.

Dependendo do fenômeno em estudo, outros modelos aninhados são necessários para caracterizar a variabilidade espacial.

3.5 KRIGAGEM

O termo *krigagem* é derivado do nome Daniel G. Krige, que foi o pioneiro a introduzir o uso de médias móveis para evitar a superestimação sistemática de reservas de mineração. Inicialmente, o método de krigagem foi desenvolvido para solucionar problemas de mapeamentos geológicos, mas seu uso expandiu-se com sucesso no mapeamento de solos, mapeamento hidrológico, mapeamento atmosférico e outros campos correlatos. A diferença entre a krigagem e outros métodos de interpolação é a maneira como os pesos são atribuídos às diferentes amostras. No caso de interpolação linear simples, por exemplo, os pesos são todos iguais a $1/N$ (N = número de amostras); na interpolação baseada no inverso do quadrado das distâncias, os pesos são definidos como o inverso do quadrado da distância que separa o valor interpolado dos valores observados. Na Krigagem, o procedimento é semelhante ao de interpolação por média móvel ponderada, exceto que aqui os pesos são determinados a partir de uma análise espacial, baseada no semivariograma experimental. Além disso, a krigagem fornece, em média, *estimativas não tendenciosas e com variância mínima*¹.

¹Estimativas não tendenciosas significam que, em média, a diferença entre valores estimados e observados para o mesmo ponto deve ser nula; e variância mínima significa que estes estimadores possuem a menor variância dentre todos os estimadores não tendenciosos.

A krigagem engloba um conjunto de métodos de estimação, incluindo procedimentos estacionários (krigagem simples e ordinária), não estacionários (krigagem universal, funções intrínsecas de ordem k), univariados e multivariados (co-krigagem etc). Este capítulo limita-se à apresentação da krigagem ordinária, descrita a seguir.

3.5.1. KRIGEAGEM ORDINÁRIA

Considere uma superfície sobre a qual se observe alguma propriedade do solo, Z , em n pontos distintos, com coordenadas representadas pelo vetor x . Assim, tem-se um conjunto de valores $\{z(x_i), i=1, \dots, n\}$, onde x_i identifica uma posição em duas dimensões representada pelos pares de coordenadas (x_i, y_i) . Suponha que se objetive estimar o valor de Z no ponto c . O valor desconhecido de $Z(x_0)$ pode ser estimado a partir de uma combinação linear dos n valores observados, adicionado a um parâmetro λ_0 :

$$Z^*(x_0) = \lambda_0 + \sum_{i=1}^n \lambda_i Z(x_i) \quad (3.29)$$

Deseja-se um estimador não tendencioso, isto é,

$$E[Z(x_0) - Z^*(x_0)] = 0 \quad \text{EPKPMF}$$

A relação acima impõe que as duas médias sejam iguais; assim aplicando-se a Equação 3.34 em 3.35, obtém-se:

$$E[Z(x_0)] = E\left[\lambda_0 + \sum_{i=1}^n \lambda_i Z(x_i)\right] \Rightarrow m = \lambda_0 + \sum_{i=1}^n \lambda_i m \quad (3.31)$$

A krigagem ordinária não requer o prévio conhecimento da média m . Neste caso, para que a igualdade da Equação 3.36 seja satisfeita é necessário que $\lambda_0 = 0$ e $\sum_{i=1}^n \lambda_i = 1$. Portanto, o estimador de Krigagem ordinária é:

$$Z^*(x_0) = \sum_{i=1}^n \lambda_i Z(x_i), \text{ com } \sum_{i=1}^n \lambda_i = 1 \quad (3.32)$$

Minimizando a variância do erro ($\text{Var}[Z(x_0) - Z^*(x_0)]$) na condição de $\sum_{i=1}^n \lambda_i = 1$, os pesos λ_i são obtidos a partir do seguinte sistema de equações, denominado sistema de krigagem ordinária:

$$\begin{cases} \sum_{j=1}^n \lambda_j C(\mathbf{x}_i, \mathbf{x}_j) - \alpha = C(\mathbf{x}_i, \mathbf{x}_0) & \text{para } i=1, \dots, n \\ \sum_{j=1}^n \lambda_j = 1 \end{cases} \quad (3.33)$$

onde,

- $C(\mathbf{x}_i, \mathbf{x}_j)$ e $C(\mathbf{x}_i, \mathbf{x}_0)$ são, respectivamente, a semivariância entre os pontos \mathbf{x}_i e \mathbf{x}_j e entre os pontos \mathbf{x}_i e \mathbf{x}_0 .
- α é o multiplicador de Lagrange necessário para a minimização da variância do erro.

A correspondente variância minimizada do erro, denominada variância de krigagem ordinária (σ_{ko}^2), é dada pela expressão

$\sigma_{ko}^2 = \text{Var}[Z(\mathbf{x}) - Z^*(\mathbf{x}_0)] = C(\mathbf{0}) - \sum_{i=1}^n \lambda_i C(\mathbf{x}_i, \mathbf{x}_0) - \alpha$	(3.34)
---	--------

A krigagem ordinária é um interpolador exato no sentido de que, quando as equações acima forem usadas, os valores interpolados irão coincidir com os valores dos pontos amostrais. Além disso, a variância da krigagem ordinária, indicada na equação (3.35), fornece informação importante sobre a confiabilidade dos valores interpolados.

3.6 ESTUDO DE CASO

Tomemos como exemplo a distribuição amostral apresentada na Figura 3-2, cuja as estatísticas descritivas estão sumarizadas na Tabela 3-1. A análise da variabilidade espacial, do teor de argila, é realizada com o auxílio do semivariograma. Esta é uma das etapas mais importantes, pois o modelo de semivariograma escolhido representa a estrutura de correlação espacial a ser utilizada nos procedimentos inferenciais de krigagem. O resultado apresentado na Figura 3-13, mostra o semivariograma omnidirecional (caso isotrópico) e seu modelo de ajuste.

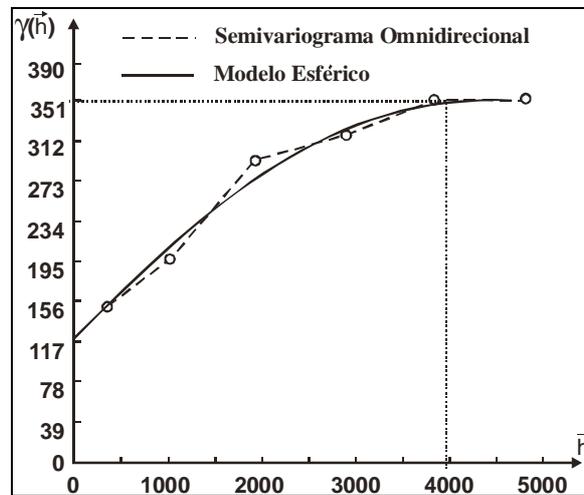


Figura 3-13 – Semivariograma omnidirecional e modelo esférico

O modelo de ajuste, mostrado na Figura 3-13, têm os seguintes parâmetros: Estrutura tipo Esférica, Efeito Pepita (C_0) = 118,85; Contribuição (C_1) = 230,89 e Alcance (a) = 3989,20. O modelo teórico, normalizado em relação ao alcance, leva a seguinte notação:

$$\gamma(h) = C_0 + C_1 \left[\text{Sph} \left(\frac{h}{a} \right) \right] = 118,85 + 230,89 \left[\text{Sph} \left(\frac{h}{3989,20} \right) \right] \quad (3.35)$$

Uma vez definido o modelo e validado o mesmo, a etapa seguinte refere-se à estimação de krigagem ordinária. Como resultado têm-se uma grade de valores estimados e uma outra que refere-se à variância de krigagem. Ambas são convertidas em superfícies e apresentadas na Figura 3-14. Na Figura 3-14 à esquerda, regiões mais claras representam altos valores de teor de argila e vice-versa. Diferente dos métodos determinísticos (ver Figura 3-4), o uso da krigagem ordinária como método de interpolação espacial permitiu capturar e, portanto, representar com mais qualidade, a variabilidade espacial inerente à propriedade em estudo. Além disso, conforme ilustra a Figura 3-14 à direita, a krigagem ordinária fornece a variância da estimativa (denominada variância de krigagem). Tal informação pode ser útil para identificar regiões onde a amostragem pode ser melhorada.

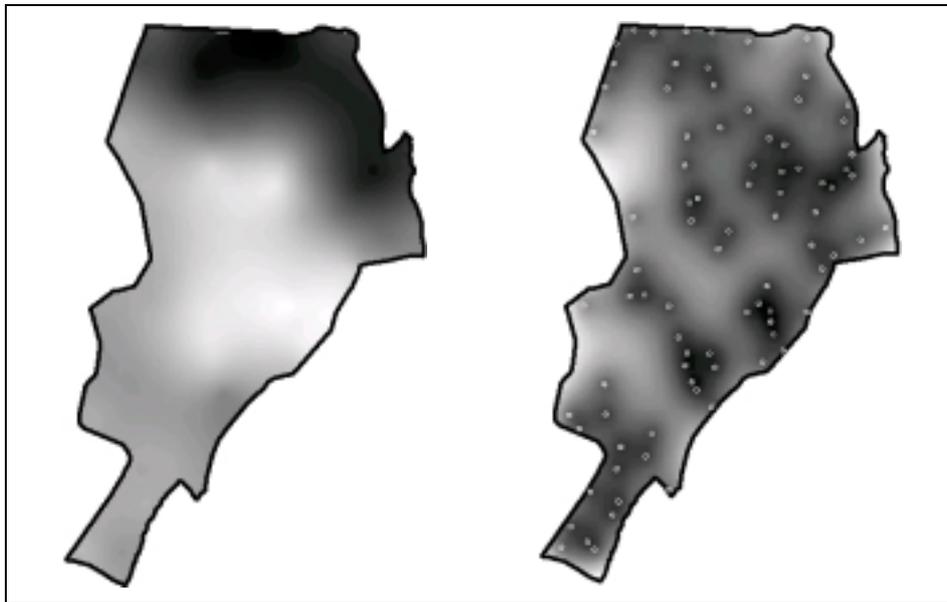


Figura 3-14 – À esquerda a superfície do teor de argila e à direita a variância de krigagem.

Com algumas ressalvas, o método da *média ponderada pelo inverso do quadrado da distância*, produz resultado que se assemelha ao resultado da krigagem ordinária. O ponto crítico, porém, ocorre em regiões onde há agrupamento (“clusters”) de amostras. A krigagem ordinária, por utilizar intrinsecamente uma estrutura de covariância, consegue tratar redundâncias (“clusters”), isto é, atribuir pesos adequados para os agrupamentos de amostras. Fato este não considerado nos procedimentos determinísticos. Além disso, na krigagem ordinária, a área de influência na interpolação é indicada pelo alcance; já nos procedimentos determinísticos, como o método da *média ponderada pelo inverso do quadrado da distância*, o raio de busca é arbitrário.

Os resultados produzidos pelos métodos *média simples* e *vizinho mais próximo*, são menos expressivos com relação aos demais. O método da *média simples* produz resultado que apresenta imbricação, principalmente na região central da área de estudo. Já o método de inferência relativo ao *vizinho mais próximo*, embora sendo o que pior expressa a variabilidade espacial do fenômeno estudado, revela a área de influência de cada ponto de observação. Tal informação é de grande valia, como, por exemplo, numa análise preliminar para detecção de valores amostrais suspeitos.

Um outro fato que merece atenção, é que os resultados apresentados na Figura 3-14 são oriundos de um modelo isotrópico. A suposição de isotropia, que é rara em fenômenos naturais, simplifica a modelagem por procedimentos geoestatísticos. Se a anisotropia existe, deve ser detectada e modelada, afim de representar com mais qualidade, a variabilidade espacial inerente à propriedade em estudo. No

apêndice ao Capítulo, são apresentados alguns tópicos sobre anisotropia e uma técnica para a modelagem da mesma.

3.7 CONCLUSÕES

Conclui-se que é possível melhorar a distribuição espacial das variáveis ambientais significativamente quando procedimentos geoestatísticos são aplicados. Ficou constatado que o teor de argila varia mais intensamente numa direção do que em outra. Tal fato refere-se à anisotropia da variável em estudo. Muitos aspectos particulares dos dados ficariam ocultos sem o uso de semivariogramas e da modelagem da anisotropia, mostrando, por exemplo, a tendência da distribuição espacial nos dados de teor de argila. Informações como estas não são apresentadas quando se usam apenas parâmetros estatísticos clássicos como médias e variâncias ou então, procedimentos determinísticos.

3.8 REFERÊNCIAS BIBLIOGRÁFICAS

A estrutura teórica da geoestatística está apresentadas na *Teoria das Variáveis Regionalizadas*, desenvolvida por Matheron (1971) e um artigo detalhado e teórico sobre geoestatística é escrito por Journel (1988). A 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). 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. Referências básicas sobre métodos de interpolação são descritas por Burrough (1987). O exemplo de superfícies de tendência está baseado no trabalho de Bönisch (2001).

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. (1987). *Principles of geographical information systems for land resources assessment*. Oxford, Clarendon Press.

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.

- Deutsch, C. e A. Journel (1992). *GSLIB: Geostatistical Software Library and user's guide*. New York, Oxford University Press.
- Issaks, M. e E. Srivastava (1989). *An Introduction to Applied Geostatistics*. New York, Oxford University Press, 1989.
- Journel, A. (1988). *Fundamentals of geostatistics in five lessons*. California, Stanford Center for Reservoir Forecasting Applied Earth Sciences Department.
- Matheron (1963, 1971). *The theory of regionalized variables and its applications*. Paris, Les Cahiers du Centre de Morphologie Mathematique de Fontainebleu, 1971. 211p.

APÊNDICE

MODELAGEM DA ANISOTROPIA

A anisotropia é uma característica muito freqüente nos elementos da natureza, isto é, a variabilidade ou distribuição espacial de tais elementos ocorre mais intensamente numa direção e menos intensamente em outra direção. Tome como exemplo o mapeamento do teor de zinco, dentro de uma região de interesse, é pouco provável que tal propriedade se espalhe igualmente em todas as direções.

Para lidar com a anisotropia, é importante que o modelo proposto represente bem a variabilidade espacial da propriedade em estudo. Procedimentos determinísticos para este fim são limitados, porque não consideram a estrutura de autocorrelação espacial bem como a anisotropia presente. Modelos mais adequados para este objetivo vem sendo propostos e a geoestatística engloba esses modelos.

TIPOS DE ANISOTROPIA

Antes de apresentar os tipos de anisotropia, é necessário mostrar as convenções direcionais usadas na geoestatística. Isto é resumido conforme ilustra a Figura 3-15.

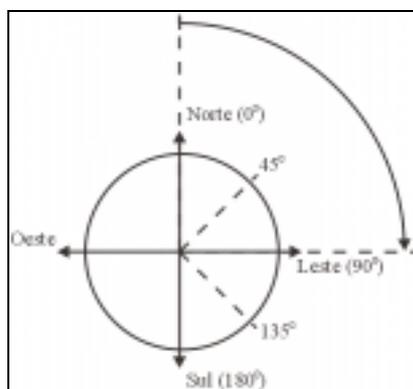


Figura 3-15 - Convenções direcionais usadas na geoestatística.

Quando os semivariogramas experimentais direcionais apresentam diferenças acentuadas, a distribuição é denominada anisotrópica. Se a anisotropia é observada e é refletida pelo mesmo Patamar (C) com diferentes Alcances (a) do mesmo modelo, então ela é denominada Geométrica, conforme ilustra a Figura 3-16. Existe ainda um outro tipo de anisotropia em que os semivariogramas experimentais direcionais apresentam os mesmos Alcances (a) e diferentes Patamares (C). Neste caso, a anisotropia é denominada *zonal*. Como a isotropia, a anisotropia *zonal* também é pouco presente nas variáveis ambientais. O mais comum é encontrar combinações da anisotropia Zonal e Geométrica, denominada anisotropia

Combinada, conforme Figura 3-16. Na Figura 3-16, a_1 e a_2 estão relacionados às direções de menor e maior continuidade espacial da variável, respectivamente.

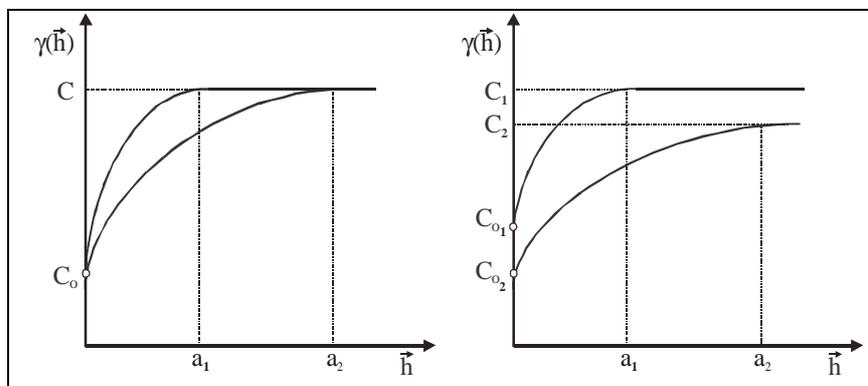


Figura 3-16 – À esquerda Anisotropia Geométrica e à direita Anisotropia Combinada.

DETEÇÃO DA ANISOTROPIA

Existem várias formas de detectar a anisotropia, por exemplo calculando-se os semivariogramas experimentais direcionais em várias direções, desenhando todos num único gráfico, e visualmente avaliando suas similaridades. Outra forma, é através do esboço gráfico de uma elipse (conhecido também como diagrama da rosa), calculada através dos alcances obtidos em direções distintas.

A forma mais eficiente e direta de detectar a anisotropia é através do mapa de semivariograma, conhecido também como semivariograma de superfície, que é um gráfico, 2D, no qual obtém-se uma visão geral da variabilidade espacial da variável em estudo. Além disso, sobre o mapa de semivariograma é possível detectar rapidamente os eixos de anisotropia, isto é, as direções de maior e menor variabilidade espacial da variável em análise. A Figura 3-17 ilustra o mapa de semivariograma aplicado aos dados da EMBRAPA – Solos, obtidos na Fazenda Canchim, em São Carlos - SP., conforme descritos na Seção 3.1. Os eixos maior e menor, da elipse, correspondem às direções de maior e menor variabilidade espacial do teor de argila respectivamente. O ângulo de anisotropia é tomado da direção norte, em sentido horário, até o eixo maior; neste caso igual a 17° . Conseqüentemente a direção de menor variabilidade é $17^\circ + 90^\circ = 107^\circ$. Obviamente que a exigência de ortogonalidade entre os eixos, pode não corresponder à realidade, mas é necessário para modelagem dos semivariogramas como será visto mais adiante.

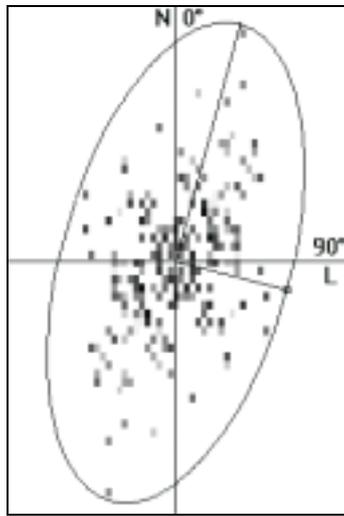


Figura 3-17 – Mapa de Semivariograma do teor de argila.

MODELAGEM DA ANISOTROPIA

O princípio fundamental na modelagem de anisotropia (geométrica, zonal ou combinada), consiste em usar todas as estruturas presentes em todas as direções, atribuindo um alcance infinito às inexistentes. Inicialmente identificam-se os eixos de anisotropia, isto é, os eixos de maior e de menor variabilidade espacial da variável em estudo. Isto é realizado com auxílio do mapa de semivariograma conforme descrito na seção anterior. Identificados os eixos de anisotropia, calculam-se os dois semivariogramas experimentais direcionais, relativos às direções de maior e menor variabilidade espacial da variável em estudo, e procede-se o ajuste dos mesmos. Estabelecidos os dois modelos, o passo seguinte é combiná-los num modelo único e consistente para todas as direções.

MODELAGEM DA ANISOTROPIA GEOMÉTRICA

Como dito anteriormente, se a anisotropia é observada e é refletida pelo mesmo Patamar (C) com diferentes Alcances (a) do mesmo tipo de modelo, então ela é denominada geométrica. Considere o exemplo da Figura 3-18, as direções de menor e maior variabilidade espacial são 0° e 90° respectivamente e os modelos de ajustes são esféricos em ambas direções.

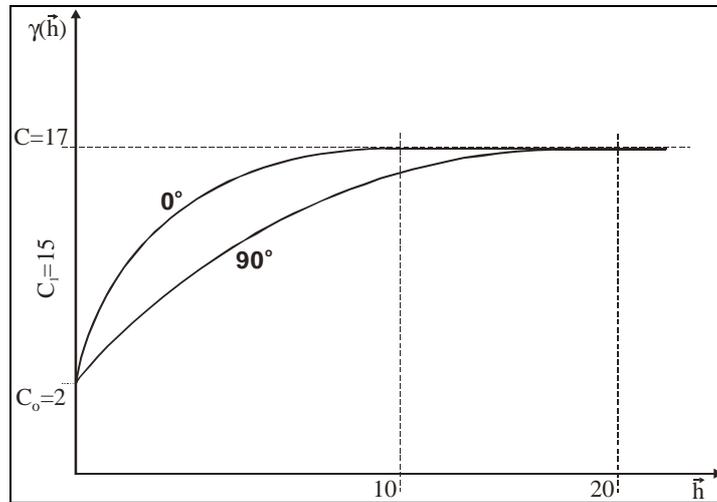


Figura 3-18 – Exemplo de anisotropia geométrica.

O modelo de semivariograma relativo à direção 0° é:

$$\gamma_{0^\circ}(\mathbf{h}) = C_0 + C_1 [\text{Sph}(\mathbf{h})] \quad (3.36)$$

O termo Sph(h) é apenas uma notação representativa do modelo teórico esférico normalizado, conforme apresentado na Seção 0. Lembre-se que \mathbf{h} é um vetor, portanto seu módulo pode ser decomposto; isto é:

$$|\mathbf{h}| = \sqrt{(h_{0^\circ})^2 + (h_{90^\circ})^2} \quad (3.37)$$

A Figura 3-19 ilustra uma decomposição genérica para o vetor \mathbf{h} .

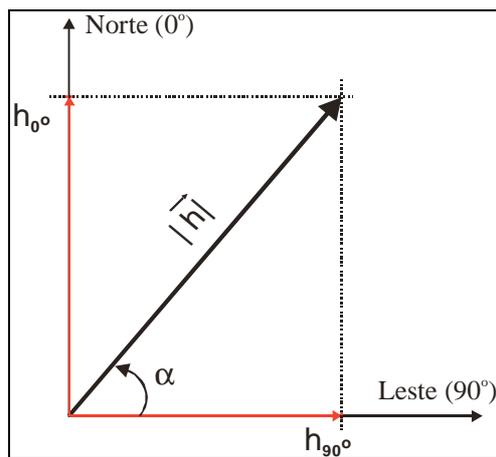


Figura 3-19 – Decomposição genérica do vetor \mathbf{h} .

Para direção de análise em questão, 0° , o vetor \mathbf{h} está sobre o eixo Norte, portanto não possui componente na direção 90° ; isto é, para $0^\circ \Rightarrow \alpha=90^\circ$ (ver α na Figura 3-19), $h_{0^\circ} = |\mathbf{h}| \cdot \text{sen}(90^\circ) = |\mathbf{h}|$ e $h_{90^\circ} = |\mathbf{h}| \cdot \text{cos}(90^\circ) = 0$.

Normalizando 3.39 em relação ao alcance (a), tem-se:

$$\left| \frac{\mathbf{h}}{a} \right| = \sqrt{\left(\frac{h_{0^\circ}}{a} \right)^2 + \left(\frac{h_{90^\circ}}{a} \right)^2} \quad (3.38)$$

Neste caso, como a componente $\frac{h_{90^\circ}}{a}$ é sempre nula, podemos atribuir um alcance infinito à direção 90° . Assim, a equação 3.21 é escrita da forma:

$$\left| \frac{\mathbf{h}}{a} \right| = \sqrt{\left(\frac{h_{0^\circ}}{a} \right)^2 + \left(\frac{h_{90^\circ}}{\infty} \right)^2} \quad (3.39)$$

O modelo normalizado do semivariograma relativo à direção 0° é definido como:

$$\gamma_{0^\circ}(\mathbf{h}) = C_0 + C_1 \left[\text{Sph} \left(\sqrt{\left(\frac{h_{0^\circ}}{a} \right)^2 + \left(\frac{h_{90^\circ}}{\infty} \right)^2} \right) \right] \quad (3.40)$$

Substituindo os valores de C_0 , C_1 e a , conforme Figura 3-18, tem-se:

$$\gamma_{0^\circ}(\mathbf{h}) = 2 + 15 \left[\text{Sph} \left(\sqrt{\left(\frac{h_{0^\circ}}{10} \right)^2 + \left(\frac{h_{90^\circ}}{\infty} \right)^2} \right) \right] \quad (3.41)$$

De maneira análoga, o modelo de semivariograma relativo à direção 90° é:

$$\gamma_{90^\circ}(\mathbf{h}) = 2 + 15 \left[\text{Sph} \left(\sqrt{\left(\frac{h_{0^\circ}}{\infty} \right)^2 + \left(\frac{h_{90^\circ}}{20} \right)^2} \right) \right] \quad (3.42)$$

Uma vez definidos os modelos relativos às direções de 0° e 90° , determina-se o modelo único e consistente para qualquer distância e direção do vetor \mathbf{h} . Das Equações 3.44 e 3.45, obtém-se o modelo único que é expresso através da seguinte equação:

$$\gamma(\mathbf{h}) = 2 + 15 \left[\text{Sph} \left(\sqrt{\left(\frac{h_{0^\circ}}{10} \right)^2 + \left(\frac{h_{90^\circ}}{20} \right)^2} \right) \right] \quad (3.43)$$

A consistência desse modelo é verificada primeiro determinando-se os valores das componentes h_0° e h_{90° para um determinado vetor h . Em seguida, calcula-se o valor de $\gamma(h)$. Por exemplo, deseja-se saber o valor de $\gamma(h)$ na direção 0° quando $|h| = \text{alcance}$; isto é, $|h| = 10$. Neste caso, as componentes h_0° e h_{90° valem:

$$h_0^\circ = |h| \cdot \sin(\alpha) = a \cdot \sin(\alpha) = 10 \cdot \sin(90^\circ) = 10.$$

$$h_{90^\circ} = |h| \cdot \cos(\alpha) = a \cdot \cos(\alpha) = 10 \cdot \cos(90^\circ) = 0.$$

$$\text{Sph}(|h|) = 1,5 \left[\sqrt{\left(\frac{10}{10}\right)^2 + \left(\frac{0}{20}\right)^2} \right] - 0,5 \left[\sqrt{\left(\frac{10}{10}\right)^2 + \left(\frac{0}{20}\right)^2} \right] = 1$$

Seguindo, determina-se $\gamma(h)$:

$$\gamma(h) = 2 + 15 \left[\text{Sph} \left(\sqrt{\left(\frac{h_0^\circ}{10}\right)^2 + \left(\frac{h_{90^\circ}}{20}\right)^2} \right) \right] = 2 + 15 \cdot [1] = 17$$

De maneira análoga, na direção 90° quando $|h| = 20$, tem-se que $\gamma(h) = 17$. E assim por diante, para uma direção θ qualquer quando $|h| \rightarrow 0$, tem-se que $\gamma(h) = 2$, que é o Efeito Pepita.

MODELAGEM DA ANISOTROPIA COMBINADA

Neste caso, a anisotropia é observada e é refletida com diferentes Patamares (C) e Alcances (a) do mesmo tipo de modelo, podendo ainda apresentar dois valores distintos de Efeito Pepita (C_0). O exemplo da Figura 3-20, referem-se aos semivariogramas nas direções de maior e menor variabilidade espacial do teor de argila, detectadas na Seção 0. Ambos semivariogramas foram ajustados com modelos esféricos.

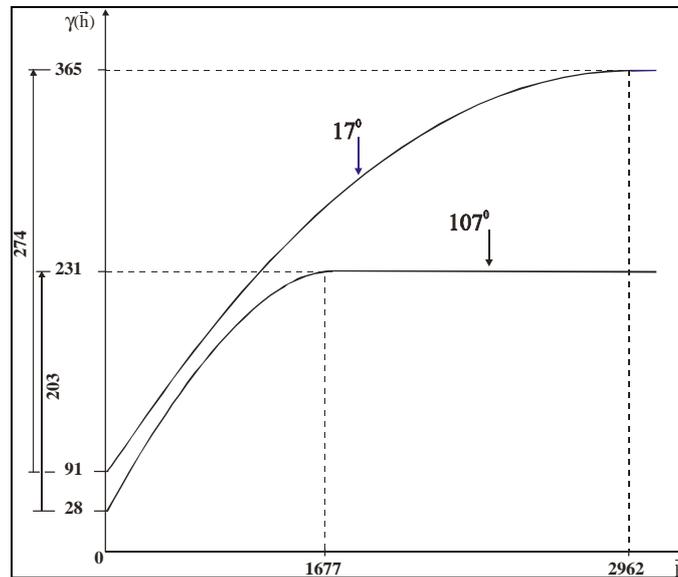


Figura 3-20 – Anisotropia combinada referente ao teor de argila.

O modelo de semivariograma relativo à direção 17° é:

$$\gamma_{17^{\circ}}(h) = 91 + 274 \left[\text{Sph} \left(\sqrt{\left(\frac{h_{17^{\circ}}}{2962} \right)^2 + \left(\frac{h_{107^{\circ}}}{\infty} \right)^2} \right) \right] \quad (3.44)$$

O modelo de semivariograma relativo à direção 107° é:

$$\gamma_{107^{\circ}}(h) = 28 + 203 \left[\text{Sph} \left(\sqrt{\left(\frac{h_{17^{\circ}}}{\infty} \right)^2 + \left(\frac{h_{107^{\circ}}}{1677} \right)^2} \right) \right] \quad (3.45)$$

Uma vez estabelecidos os modelos relativos às direções de mínima e máxima continuidade espacial do fenômeno, procede-se à modelagem da anisotropia combinada. A modelagem da anisotropia combinada é um caso mais complexo que a modelagem da anisotropia geométrica. A idéia básica é dividir em faixas convenientes o gráfico de semivariograma, conforme ilustra a Figura 3-21, de maneira que, em cada faixa reste somente a anisotropia geométrica. Evidentemente que esta técnica exige o conhecimento e prática com semivariogramas e modelagem da anisotropia.

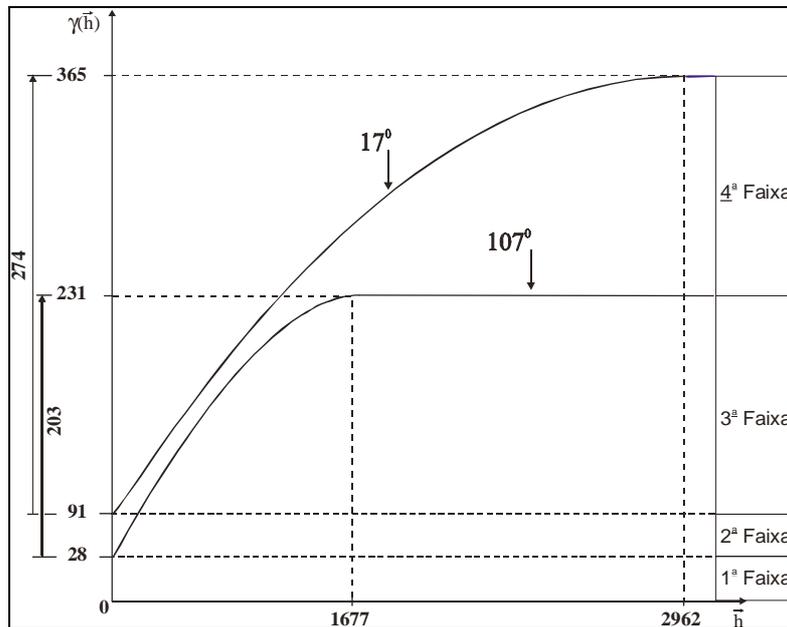


Figura 3-21 – Definição das faixas para modelagem da anisotropia combinada.

Uma vez estabelecido de forma conveniente as faixas, a anisotropia combinada é decomposta graficamente, conforme ilustra a Figura 3-22, de modo que, cada parcela represente somente a anisotropia geométrica.

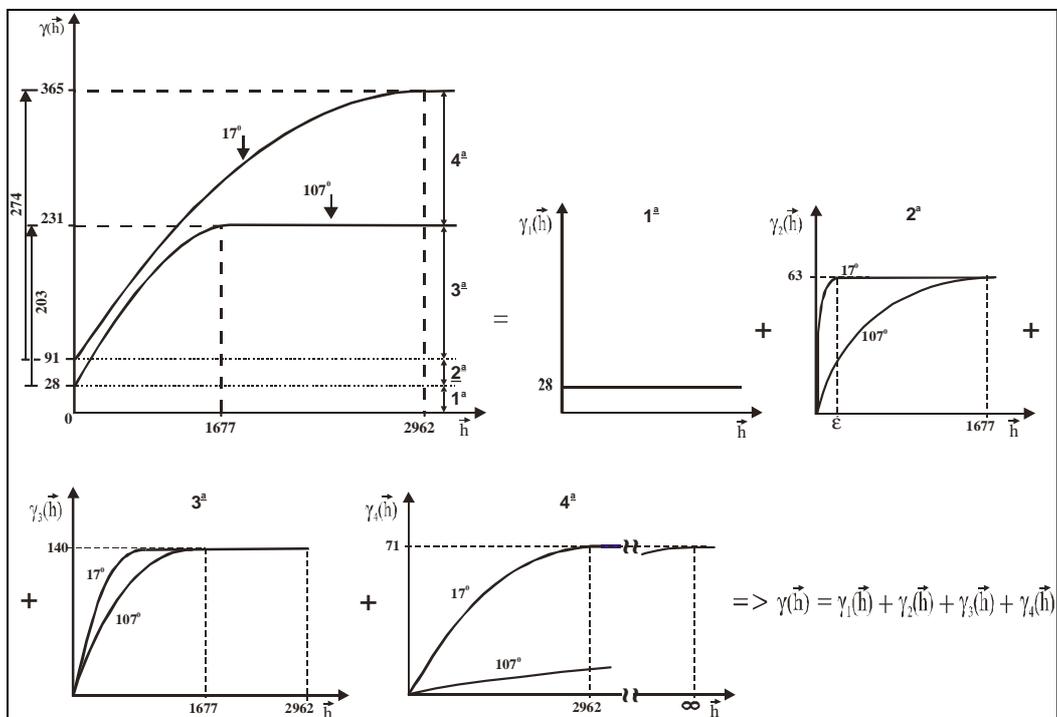


Figura 3-22 – Decomposição da anisotropia combinada.

A anisotropia combinada apresentada na Figura 3-22 é decomposta da seguinte forma:

A 1ª parcela refere-se a um valor constante, o Efeito Pepita ($C_0 = 28$). O modelo relativo a 1ª parcela é:

$$\gamma_1(\mathbf{h}) = C_0 \quad (3.46)$$

Para estabelecer a anisotropia geométrica na 2ª parcela, é necessário empregar um artifício. Este consiste em utilizar um modelo esférico com alcance muito pequeno (ϵ). Isto é necessário para modelar o segundo efeito pepita (91) relativo à direção de 17°. Com relação a outra direção, 107°, observa-se que parte do modelo esférico participa com uma pequena contribuição. Desta forma, a anisotropia geométrica é caracterizada da seguinte forma: em ambas direções modelos esféricos com contribuição 63 (91 - 28), alcance (ϵ) para a direção 17° e alcance 1677m para a direção 107°. O modelo único e consistente de semivariograma relativo à 2ª parcela é:

$$\gamma_2(\mathbf{h}) = 63 \left[\text{Sph} \left(\sqrt{\left(\frac{h_{17^\circ}}{\epsilon} \right)^2 + \left(\frac{h_{107^\circ}}{1677} \right)^2} \right) \right] \quad (3.47)$$

Na 3ª parcela, a anisotropia geométrica é obtida de forma direta. Isto é, parte de ambos modelos contribuem para a caracterização da mesma. Conforme pode ser visto na Figura 3-22, esta é composta de uma estrutura esférica com alcance de 1677m na direção 107°, uma estrutura esférica com alcance de 2962m na direção 17° e ambas com contribuição de 140 (231 - 91). O modelo único e consistente de semivariograma relativo à 3ª parcela é:

$$\gamma_3(\mathbf{h}) = 140 \left[\text{Sph} \left(\sqrt{\left(\frac{h_{17^\circ}}{2962} \right)^2 + \left(\frac{h_{107^\circ}}{1677} \right)^2} \right) \right] \quad (3.48)$$

Para estabelecer uma anisotropia geométrica à 4ª parcela é necessário empregar um outro artifício. Observando a Figura 3-22, nota-se que não existe um modelo associado à direção 107°. O segredo então é, atribuir um alcance muito grande, ∞ , a esta direção. Tal artifício é utilizado apenas para estabelecer a anisotropia geométrica. Isto não influencia em nada no modelo final a ser determinado. O resultado disto é uma estrutura esférica com alcance na direção 17° de 2962m, uma

estrutura esférica com alcance na direção 107° muito grande (∞) e ambas estruturas com contribuição de 71 (274 – 203). O modelo único e consistente de semivariograma relativo à 4ª parcela é:

$$\gamma_4(\mathbf{h}) = 90 \left[\text{Sph} \left(\sqrt{\left(\frac{h_{17^\circ}}{2962} \right)^2 + \left(\frac{h_{107^\circ}}{\infty} \right)^2} \right) \right] \quad (3.49)$$

Finalmente, o modelo completo, $\gamma(\mathbf{h})$, e consistente para qualquer distância e direção do vetor \mathbf{h} , resume-se na soma das estruturas $\gamma_1(\mathbf{h})$, $\gamma_2(\mathbf{h})$, $\gamma_3(\mathbf{h})$ e $\gamma_4(\mathbf{h})$. Então,

$$\begin{aligned} \gamma(\mathbf{h}) &= \gamma_1(\mathbf{h}) + \gamma_2(\mathbf{h}) + \gamma_3(\mathbf{h}) + \gamma_4(\mathbf{h}) \quad (3.50) \\ \gamma(\mathbf{h}) &= 28 + 63 \left[\text{Sph} \left(\sqrt{\left(\frac{h_{17^\circ}}{\varepsilon} \right)^2 + \left(\frac{h_{107^\circ}}{1677} \right)^2} \right) \right] + 140 \left[\text{Sph} \left(\sqrt{\left(\frac{h_{17^\circ}}{2962} \right)^2 + \left(\frac{h_{107^\circ}}{1677} \right)^2} \right) \right] + \\ &+ 71 \left[\text{Sph} \left(\sqrt{\left(\frac{h_{17^\circ}}{2962} \right)^2 + \left(\frac{h_{107^\circ}}{\infty} \right)^2} \right) \right] \quad (3.51) \end{aligned}$$

A Tabela 3.3 resume os parâmetros estruturais que compõem o modelo expresso na Equação (3.54), e sua consistência é verificada de maneira análoga ao caso de anisotropia geométrica, conforme descrita anteriormente.

Tabela 3.3 – Sumarização dos Parâmetros Estruturais.

Número de Estruturas		3	
Efeito Pepita		28	
Primeira Estrutura – Tipo: Esférica			
Contribuição	63	Ângulo de anisotropia	17°
Menor Alcance	ε	Maior Alcance	1677
Segunda Estrutura – Tipo: Esférica			
Contribuição	140	Ângulo de anisotropia	17°
Menor Alcance	1677	Maior Alcance	2962
Terceira Estrutura – Tipo: Esférica			
Contribuição	71	Ângulo de anisotropia	17°
Menor Alcance	2962	Maior Alcance	∞

A etapa seguinte refere-se à estimação de krigagem ordinária. Como resultado, têm-se uma grade de valores estimados e uma outra que refere-se à variância de krigagem. Ambas são convertidas em superfícies e apresentadas na Figura 3-23.

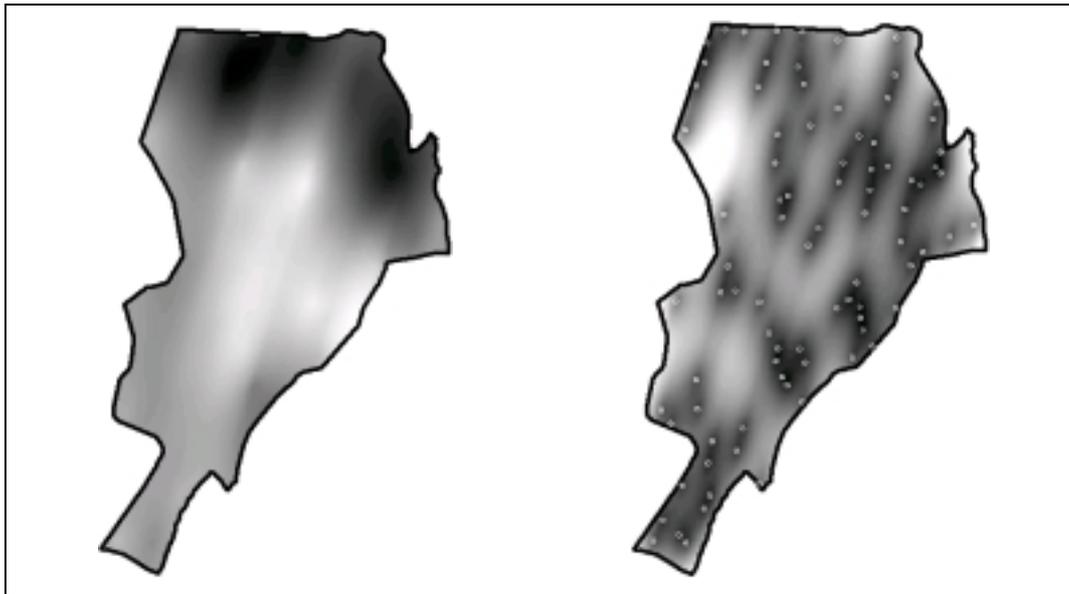


Figura 3-23 – À esquerda superfície anisotrópica do teor de argila e à direita a variância de krigagem.

Analisando os resultados apresentados nas Figura 3-14 e Figura 3-23, observa-se que as diferenças na distribuição espacial do teor de argila são acentuadas. O resultado oriundo do modelo anisotrópico, Figura 3-23, mostra que a variável em estudo possui uma tendência maior de espalhamento na direção de aproximadamente 17° (ângulo de anisotropia) e uma menor tendência na direção ortogonal (107°). Este fato, mostra a importância da modelagem da anisotropia na reconstrução da distribuição espacial do teor de argila, proporcionando resultados e análises mais representativas.

ANÁLISE ESPACIAL DE SUPERFÍCIES: O ENFOQUE DA GEOESTATÍSTICA POR INDICAÇÃO

*Carlos Alberto Felgueiras
Suzana Druck
Antônio Miguel Vieira Monteiro*

4.1 Introdução

Os procedimentos de krigagem ordinária apresentados no capítulo anterior (vide Seção 3.4) buscavam previsões ótimas da variável em estudo, em locais não observados, minimizando a variância do erro associado a essa estimativa. Neste capítulo, o foco será na análise de modelos de incerteza, ou seja, na inferência das distribuições de probabilidade para cada posição do espaço considerado, representadas pelos vetores x . Os novos procedimentos vão permitir a definição de estimadores obtidos segundo a minimização de outras funções de erro inferencial, e não, como efetuado pela krigagem linear (vide Seção 3.5), um estimador baseado apenas na minimização da variância do erro. Situações em que a análise da incerteza é relevante podem ser ilustradas na aplicação da krigagem nos estudos de reposição de nutrientes nos solos. Neste caso, o que se deseja é determinar a quantidade de nutrientes que deve ser reposta nos solos de uma região de maneira a *maximizar* a produção e tornar *mínimo* os custos. O processo inferencial tem como objetivo evidenciar os locais em que um determinado fator dos solos, $Z(x)$, é deficiente, ou seja, os locais em que o valor estimado, $\hat{Z}(x)$, seja igual ou abaixo de um valor crítico, z_{lim} , isto é, quando $\hat{Z}(x) \leq z_{lim}$. Assim, o que interessa não é inferir exatamente um determinado valor, mas definir áreas com maior probabilidade que o evento ocorra, ou seja, áreas onde a probabilidade do valor estimado $\hat{Z}(x)$ ser menor ou igual a um limite z_{lim} , definida por $Prob\{\hat{Z}(x) \leq z_{lim}\}$, tem um valor determinado.

Por outro lado, os erros inferenciais, que são a subestimação (estimar um valor menor do que seria o valor real) ou, a sobre-estimação (estimar um valor maior do que seria o valor real) vão produzir efeitos diferentes no processo produtivo. Enquanto a subestimação pode levar a repor nutrientes onde não é necessário, e contaminar os solos, a sobre-estimação pode conduzir a não repor nutrientes onde é necessário e prejudicar a produtividade. Dessa forma, esses erros inferenciais não podem ser tratados como se tivessem o mesmo impacto, e a minimização de um, ou de outro, ou de ambos, vai depender dos objetivos impostos pelo trabalho a ser executado. Neste contexto, o estimador de krigagem linear obtido pela

minimização da variância (vide Seção 3.5), que considera equivalentes e simétricos os impactos de subestimar ou sobre-estimar, seria insuficiente para apoiar as decisões necessárias a melhor solução do problema.

Este capítulo apresenta um conjunto de técnicas que procura construir o modelo de incerteza associado a uma determinada posição do espaço, representada pelo vetor x . O modelo a ser produzido é condicionado a um conjunto de dados geográficos, coletados previamente a partir de suportes amostrais pontuais. Os exemplos, utilizados para ilustrar os conceitos deste capítulo, referem-se a conjuntos amostrais obtidos no levantamento de solos executado na região de Canchim (vide Seção 3.4, Figura 4-1 e Tabela 4-1). No que segue, admite-se que o leitor esteja familiarizado com os conceitos de krigagem apresentados no capítulo 3 (Seção 3.4 a Seção 3.7).

4.2 Incertezas locais

A geoestatística considera os valores de um atributo para cada posição $x \in A$ (uma região da superfície terrestre) como uma realização de uma variável aleatória (VA), descrita como $Z(x)$. Isto significa que, na posição x , $Z(x)$ pode assumir diferentes valores para o atributo considerado, cada valor com uma probabilidade de ocorrência associada a ele. Uma VA $Z(x)$ ordenada, contínua ou discreta, é caracterizada pela sua *função de distribuição de probabilidade acumulada, fdpa, univariada*, $F(x, z)$, definida como:

$$F(x; z) = Prob\{Z(x) \leq z\} \quad (4.1)$$

Os procedimentos por indicação (também conhecidos por funções indicatriz) estão interessados na modelagem da *função de distribuição univariada acumulada condicionada (fdpac)*, isto é, a *função de distribuição* que pode ser construída condicionada aos n dados amostrados, $F(x; z/(n))$, que é dada por:

$$F(x; z/(n)) = Prob\{Z(x) \leq z/(n)\} \quad (4.2)$$

A $F(x; z/(n))$ modela a incerteza da V.A. Z no local x , e uma vez estimada essa função de distribuição de probabilidade ela pode ser utilizada para:

- produzir uma estimativa de valores do atributo em posições não conhecidas;
- modelar a incerteza dos valores para o atributo nas posições estimadas;

O enfoque tradicional, oferecido pela krigagem linear, para modelar a incerteza em locais não amostrados, consiste em computar estimativas do valor desconhecido $\hat{z}(x)$ e de sua respectiva variância $\hat{\sigma}^2(x)$, e construir um intervalo de confiança do tipo gaussiano, centrado em $\hat{z}(x)$,

$$Prob\{Z(\mathbf{x}) \in [\hat{z}(\mathbf{x}) - 2\hat{\sigma}(\mathbf{x}), \hat{z}(\mathbf{x}) + 2\hat{\sigma}(\mathbf{x})]\} \quad (4.3)$$

A construção deste tipo de intervalo de confiança fundamenta-se nas hipóteses:

- os erros locais de estimação têm distribuição gaussiana;
- o intervalo de confiança pode ser construído através da variância destes erros.

Essas hipóteses são fortemente restritivas, uma vez que a distribuição local dos erros pode apresentar severas assimetrias, principalmente quando o histograma das amostras apresenta-se assimétrico, não se adequando a hipótese gaussiana sendo implicitamente considerada. Por outro lado, a variância obtida através da krigagem linear depende unicamente da configuração geométrica dos dados, e não do valor de seu atributo naquela posição, e uma variância com essas características pode não ser adequada para representar as incertezas na estimativa de valor para o atributo, principalmente em regiões onde amostras próximas apresentam valores para seu atributo, medido ou observado, muito discrepantes.

Um outro enfoque possível é considerar que primeiro é necessário modelar a incerteza, ou seja inferir as distribuições de probabilidades locais, as distribuições para cada ponto do espaço a ser estudado, representado pelo vetor \mathbf{x} . Uma vez estabelecidas as funções, $F(\mathbf{x}; z/(n))$, e só então deduzir as estimativas ótimas para cada ponto. Observe que o procedimento tradicional primeiro calcula a estimativa, os valores estimados para os pontos não observados, e depois acrescenta o intervalo de confiança, com base na variância dos erros produzidos pelo estimador. A modelagem da incerteza, sendo construída diretamente através da *fdpac*, $F(\mathbf{x}; z/(n))$, condiciona, por construção, essa *fdpac* aos dados amostrais, e produz então um modelo que é independente de uma particular estimativa $\hat{z}(\mathbf{x})$, obtida com base em um particular estimador, no nosso caso o estimador por krigagem linear. Ficamos agora com o problema da inferência desta função de distribuição de probabilidade acumulada condicionada para cada ponto do espaço, da $F(\mathbf{x}; z/(n))$.

Vamos abordar dois enfoques, mais presentes na literatura :

- O *multigaussiano*, que estabelece o modelo de distribuição a ser considerado à priori;
- O *enfoque por indicação*, que não estabelece nenhum modelo de distribuição para os dados. A *fdpac* é modelada de forma aproximada pela sua discretização numa série de K cortes $z_k, k = 1, \dots, k$.

O primeiro enfoque, o *multigaussiano*, é o mais fácil de ser utilizado, mas apresenta algumas restrições importantes:

1. estabelece a hipótese multigaussiana para a distribuição multivariada que não pode ser inteiramente verificada;
2. é inadequada para fenômenos que apresentam uma expressiva correlação em valores extremos da distribuição.

O *enfoque por indicação* pode ser considerado mais geral. Não restringe o fenômeno em estudo a ser representado por uma distribuição específica. Deve ser utilizado quando os dados não se ajustam a uma distribuição multigaussiana, ou quando os valores extremos da distribuição das amostras apresentam significativa conectividade. Esse capítulo, por essas razões, focaliza esse procedimento.

4.3 O Enfoque por Indicação

O enfoque por indicação está fundamentado na interpretação da probabilidade condicional $Prob\{Z(\mathbf{x}) \in [\hat{z}(\mathbf{x}) - 2\hat{\sigma}(\mathbf{x}), \hat{z}(\mathbf{x}) + 2\hat{\sigma}(\mathbf{x})]\}$ como uma esperança condicional de uma variável aleatória por indicação, $I(\mathbf{x}, z_k | (n))$, considerada a informação disponível nas (n) amostras, isto é:

$$F(\mathbf{x}; z_k | (n)) = E\{I(\mathbf{x}, z_k | (n))\} \quad k = 1, \dots, K \quad (4.4)$$

onde $I(\mathbf{x}, z_k | (n)) = 1$ se $Z(\mathbf{x}) \leq z_k$ e $I(\mathbf{x}, z_k | (n)) = 0$ se $Z(\mathbf{x}) > z_k$

A estimativa de krigagem de uma variável por indicação, $I(\mathbf{x}, z_k | (n))$, é também uma estimativa de sua esperança condicional. Portanto, as estimativas de $\hat{F}(\mathbf{x}, z_k | (n))$, para $k = 1, \dots, K$, podem ser calculadas estimando-se o valor $\hat{i}(\mathbf{x}, z_k | (n))$, que utiliza para sua estimativa os dados transformados para dados por indicação, com valores 1 e 0.

Dessa forma, os procedimentos por indicação iniciam-se por uma transformação não linear, chamada de *codificação por indicação*, que transforma cada valor do conjunto amostral, $z(\mathbf{x})$, em valores por indicação, $i(\mathbf{x}, z_k)$.

A codificação por indicação dos dados amostrais

Na distribuição de um conjunto de dados amostrais, um determinado número de cortes K e seus respectivos valores de cortes $z_k, k = 1, \dots, k$, são definidos. A codificação por indicação, se processa para cada valor de corte z_k , e gera um *conjunto amostral por indicação* $i(\mathbf{x}, z_k)$ do tipo:

$$i(\mathbf{x}; z_k) = \begin{cases} 1, & \text{se } z(\mathbf{x}) \leq z_k \\ 0, & \text{se } z(\mathbf{x}) > z_k \end{cases} \quad (4.5)$$

A codificação por indicação é aplicada sobre todo conjunto amostral criando, para cada valor de corte, um conjunto cujos valores são 0 ou 1. Os K valores de corte, são definidos em função do número de amostras e devem ser escolhidos de tal forma que os $K + 1$ cortes contenham aproximadamente as mesmas frequências. Entretanto, existem algumas critérios para a escolha de K :

1. Os valores de k , devem ser representativos de toda a gama de valores apresentados pelos dados.
2. Os valores de k devem destacar os pontos importantes da distribuição.
3. O número de cortes K não deve ser muito grande, o que demandaria grande esforço computacional, mas principalmente não deve ser muito pequeno, pois pode resumir aspectos relevantes da distribuição. Uma regra razoável é considerar que o valor de K não deve ser menor que cinco (5), nem maior que quinze (15).

Se para um determinado conjunto de dados cujos valores variam no intervalo [5, 43] podemos definir $z_k = 20, 30, 39$ correspondentes respectivamente a três quantis de sua distribuição ($p = 0.25, 0.50, 0.75$). A codificação associará a cada valor amostral um vetor com 3 dados por indicação com valores 0 ou 1. Por exemplo, se o valor amostral $z(u_j) = 25.2$, então o valor por indicação $i(u_j, 20) = 0$ e representa a probabilidade de $Z(u_j)$ assumir valores menores ou iguais a 20, dado que o valor de $z(u_j) = 25.2$, $Prob\{Z(u_j) \leq 20 | z(u_j) = 25.2\}$. Considerando os três valores de z_k , o vetor por indicação seria representado como abaixo descrito:

$$\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} i(\mathbf{x}_j, 20) \\ i(\mathbf{x}_j, 30) \\ i(\mathbf{x}_j, 39) \end{bmatrix} \rightarrow \begin{bmatrix} Prob\{Z(\mathbf{x}_j) \leq 20 | z(\mathbf{x}_j) = 25.2\} \\ Prob\{Z(\mathbf{x}_j) \leq 30 | z(\mathbf{x}_j) = 25.2\} \\ Prob\{Z(\mathbf{x}_j) \leq 39 | z(\mathbf{x}_j) = 25.2\} \end{bmatrix}$$

4.3.2 A variografia por indicação

A análise de variografia se processa de forma semelhante a realizada na krigeagem linear (vide Seção 3.5), considerando-se separadamente o conjunto de valores por indicação para cada valor de corte, z_k . Dessa forma, para cada valor de corte z_k um modelo de variograma deve ser estabelecido, o que corresponderia, no exemplo anterior, ao ajuste de 3 modelos de semivariogramas a partir de 3 variogramas experimentais computados como:

$$\gamma_I(\mathbf{h}, z_k) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [i(\mathbf{h}_\alpha; z_k) - i(\mathbf{h}_\alpha + \mathbf{h}; z_k)]^2 \quad (4.6)$$

Como os valores das variáveis por indicação são 0 e 1, o variograma por indicação é, usualmente, bem comportado e resistente a valores extremos ("outliers"). Também as amostras de $i(u, z_k)$ para cada z_k são considerados como amostras de uma distribuição Bernouilli cuja variância máxima é 0.25. Dessa forma o efeito pepita somado ao patamar, que são aproximadamente iguais ao valor da variância, terá como valor máximo 0.25. Calcular os variogramas é relativamente simples, sendo a única dificuldade prática o número de variogramas a ser modelados.

4.3.3 A estimação dos valores por indicação

Como mencionado anteriormente para cada valor de corte $z_k, k = 1, \dots, k$, a $\hat{F}(\mathbf{x}, z_k | (n))$ pode ser estimada através da combinação linear dos dados por indicação $i(\mathbf{x}, z_k)$. O estimador linear é expresso em termos de VAs por indicação.

$$\hat{F}(\mathbf{x}; z_k | (n)) = \sum_{\alpha=1}^{n(u)} \lambda_\alpha(\mathbf{x}; z_k) I(\mathbf{x}_\alpha; z_k) + \left[1 - \sum_{\alpha=1}^{n(u)} \lambda_\alpha(\mathbf{x}; z_k) \right] I(\mathbf{x}; z_k) \quad (4.7)$$

onde $\lambda_\alpha(\mathbf{x}; z_k)$ é o peso assinalado a cada dado convertido interpretado como uma realização de uma variável aleatória por indicação. Se a média por indicação, $E[I(\mathbf{x}; z_k)]$, é considerada constante dentro da área em estudo dois procedimentos podem ser considerados, descritos a seguir.

Krigeagem por Indicação Simples

Neste caso a média por indicação é conhecida e constante, isto é:

$$E\{I(\mathbf{x}; z_k)\} = F(z_k) \quad (4.8)$$

e o preditor linear (4.6) é então rescrito,

$$\hat{F}_{KS}(\mathbf{x}; z_k | (n)) = \sum_{\alpha=1}^{n(u)} \lambda_\alpha^{KS}(\mathbf{x}; z_k) I(\mathbf{x}_\alpha; z_k) + \left[1 - \sum_{\alpha=1}^{n(x)} \lambda_\alpha^{KS}(\mathbf{x}; z_k) \right] F(z_k) \quad (4.9)$$

onde os pesos $\lambda_\alpha^{KS}(\mathbf{x}, z_k)$ são determinados através do sistema de krigeagem simples.

$$\sum_{\beta=1}^{n(u)} \lambda_\beta^{KS}(\mathbf{x}; z_k) C_I(\mathbf{h}_{\alpha\beta}; z_k) = C_I(\mathbf{h}_\alpha; z_k) \quad \forall \alpha = 1, 2, \dots, n(\mathbf{x}) \quad (4.10)$$

onde $\mathbf{h}_{\alpha\beta}$ é o vetor de separação definido pelas posições \mathbf{x}_α e \mathbf{x}_β , \mathbf{h}_α é o vetor definido entre as posições \mathbf{x}_α , e a posição a ser estimada \mathbf{x}_0 . $C_I(\mathbf{h}_{\alpha\beta}; z_k)$ é a

autocovariância definida por $\mathbf{h}_{\alpha\beta}$ e $C_I(\mathbf{h}_\alpha; z_k)$ é a autocovariância definida por \mathbf{h}_α em $z = z_k$. As autocovariâncias são determinadas pelo modelo de variografia teórico definido pelo conjunto I para $z = z_k$.

Krigeagem por Indicação Ordinária

A krigeagem por indicação ordinária permite considerar flutuações locais da média limitando seu domínio de estacionariedade a vizinhança local $W(\mathbf{x})$

$$E\{I(\mathbf{x}; z_k)\} = \text{constante mas desconhecida para } \forall \mathbf{x} \in W(\mathbf{x})$$

$$E\{I(\mathbf{x}; z_k)\} = \hat{F}(\mathbf{x}; z_k) \text{ estimado no domínio } W(\mathbf{x})$$

O estimador de *krigeagem por indicação ordinária* tem a seguinte expressão:

$$\hat{F}_{KS}(\mathbf{x}; z_k | (n)) = \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_{\alpha}^{KS}(\mathbf{x}; z_k) I(\mathbf{x}_{\alpha}; z_k) + \left[1 - \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_{\alpha}^{KS}(\mathbf{x}; z_k) \right] \hat{F}(\mathbf{x}; z_k) \quad (4.11)$$

sendo que os pesos $\lambda_{\alpha}^{KS}(\mathbf{x}; z_k)$ são obtidos pela solução do seguinte sistema de equações de krigeagem por indicação ordinária:

$$\begin{cases} \sum_{\beta=1}^{n(\mathbf{x})} \lambda_{\beta}^{KO}(\mathbf{x}; z_k) C_I(\mathbf{h}_{\alpha\beta}; z_k) + \phi(\mathbf{x}; z_k) = C_I(\mathbf{h}_{\alpha}; z_k) & \forall \alpha = 1, 2, \dots, n(\xi) \\ \sum_{\beta=1}^{n(\mathbf{x})} \lambda_{\beta}^{KO}(\mathbf{x}; z_k) = 1 \end{cases} \quad (4.12)$$

onde $\phi(\mathbf{x}; z_k)$ é o multiplicador de Lagrange.

A krigeagem por indicação, simples ou ordinária, fornece, para cada valor de corte z_k , a melhor estimativa da esperança condicional da VA $I(\mathbf{x}, z_k)$, $\hat{I}(\mathbf{x}, z_k)$. Utilizando esta propriedade, e o teorema que estabelece que $\hat{I}(\mathbf{x}, z_k) = \hat{F}(\mathbf{x}, z_k)$ pode-se calcular estimativas dos valores da *fdpac* de $Z(\mathbf{x})$ para vários valores de $z = z_k$, pertencentes ao domínio de $Z(\mathbf{x})$. O conjunto dos valores das *fdpac*'s, estimados nos valores de corte, é considerado uma *aproximação discretizada* da *fdpac* real de $Z(\mathbf{x})$. Quanto maior a quantidade de valores de corte, melhor é a aproximação. A aproximação é complementada pela definição de *uma função de ajuste para a distribuição*, que deve ser utilizada para se inferir a *fdpac* para valores diferentes dos valores de corte. Um ajuste linear é o mais simples de se definir, porém funções de maior grau podem ser usadas.

4.3.4 Correção dos Desvios de Ordem

A aproximação da função de distribuição apresenta alguns problemas, conhecidos como *desvios de relação de ordem*, que devem ser corrigidos automaticamente pelo procedimento. Os valores de probabilidades acumuladas condicionadas, para cada valor de corte, são inferidos independentemente. Para que esses valores de probabilidade constituam uma distribuição legítima, eles devem verificar as seguintes relações de ordem:

1. Os valores inferidos de $\hat{F}(\mathbf{x}, z_k | (n))$ devem satisfazer a seguinte relação $0 \leq F^*(\mathbf{x}; z_k / (n)) \leq 1 \quad \forall z_k, k = 1, \dots, K$
2. O valor estimado de $\hat{F}(\mathbf{x}, z_k | (n))$ não deve ser maior do que a $\hat{F}(\mathbf{x}; z_{k+1} | (n))$ quando $z_k \leq z_{k+1}$, ou seja $\hat{F}(\mathbf{x}; z_k / (n)) \leq \hat{F}(\mathbf{x}; z_{k+1} / (n))$ se $z_k \leq z_{k+1}$

A primeira condição pode ser garantida quando todos os pesos do estimador são positivos e somam 1. A krigeagem não garante que os pesos sejam todos positivos. Por isso é possível a inferência de valores da *fdpac* fora do intervalo [0,1]. A solução para este problema é ajustar os valores estimados para as bordas, ou seja, valores negativos são mapeados para 0 e valores maiores que 1, para 1. A segunda condição é garantida com o uso de ponderadores positivos que somam 1, e com a utilização dos mesmos pesos de estimação para todos os valores de corte, o que não pode ser garantido pela krigeagem por indicação. Portanto, estas inconsistências podem ocorrer e devem ser corrigidas. Um procedimento simples de correção é verificar pares de *fdac's* estimadas, em valores sucessivos de cortes, e ajustá-los para o valor médio das duas, sempre que a relação de ordem, $\hat{F}(\mathbf{x}; z_k / (n)) \leq \hat{F}(\mathbf{x}; z_{k+1} / (n))$ se $z_k \leq z_{k+1}$, não for satisfeita. A Figura 4-2 ilustra os problemas e as soluções das 2 condições acima descritas.

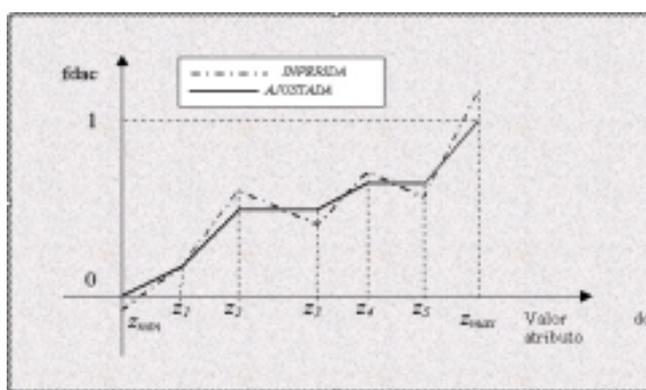


Figura 4-2 Correção dos desvios de relação de ordem

A Figura 4-3 e a Figura 4-4 que seguem buscam ilustrar as etapas descritas para a obtenção do modelo de incerteza para um conjunto amostral tomado conceitualmente como variáveis aleatórias.

Primeira Etapa do Processo de Krigagem por Indicação para atributos contínuos

- ① - Codificação Probabilística dos Dados => Função de Transformação por Indicação
- ② - Análise Estrutural => Função Semivariograma por Indicação e Modelagem

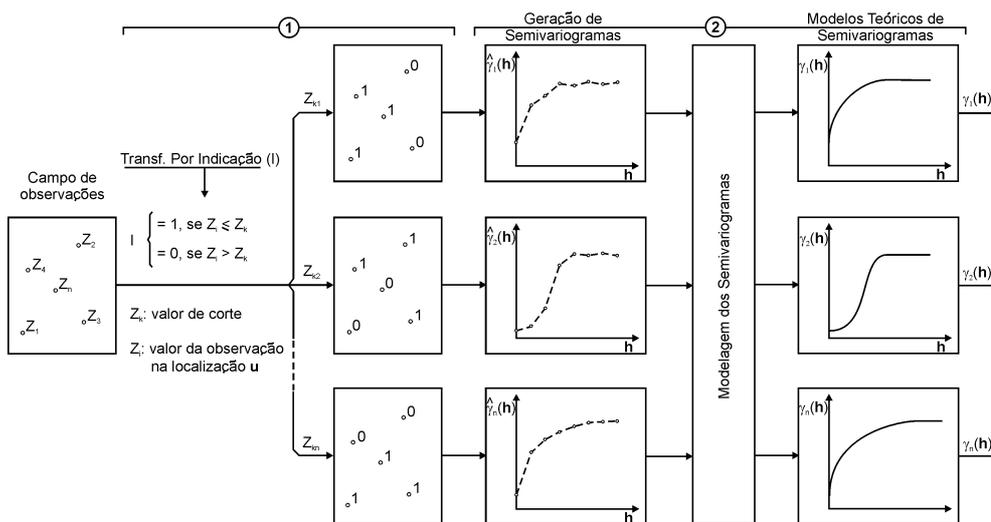


Figura 4-3 Primeira etapa do processo de krigagem por Indicação

Segunda Etapa do Processo de Krigagem por Indicação para atributos contínuos

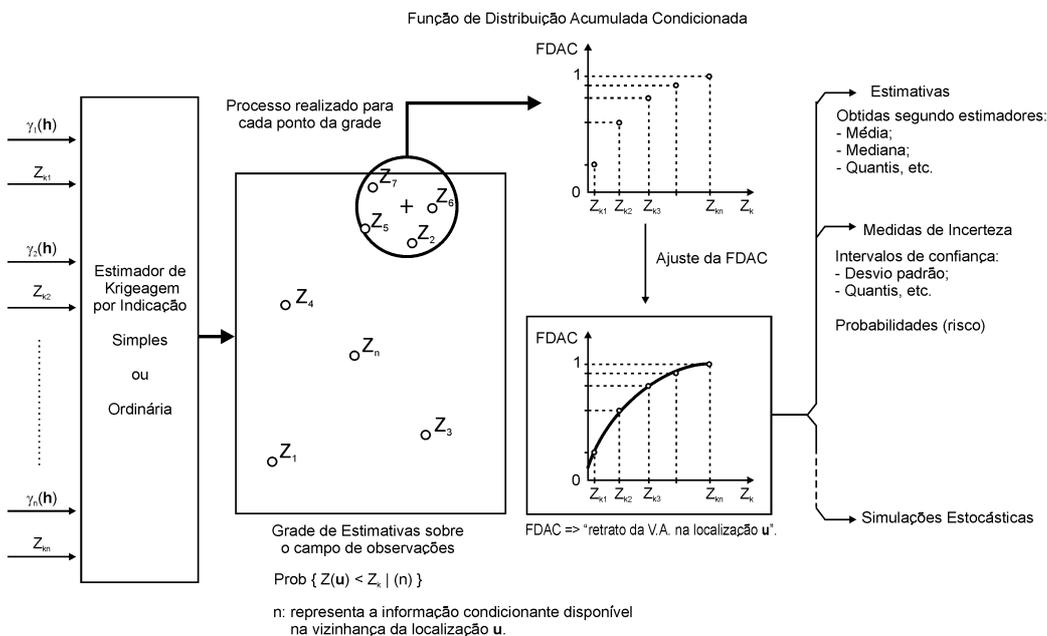


Figura 4-4 Segunda etapa do processo de krigagem por Indicação

4.4 Estimativa de incertezas locais

O conhecimento da *fdpac*, $F(\mathbf{x}; z_k | (n))$, em uma localização \mathbf{x} , possibilita a estimativa direta da incerteza, sobre o valor não conhecido $z_k(\mathbf{x})$, independente da escolha de um estimador para $z_k(\mathbf{x})$. Vamos ver agora como a incerteza pode ser estimada quando adotamos o enfoque por indicação aqui apresentado.

Intervalos de probabilidade

A incerteza pode ser estimada através de intervalos de valores do atributo. A probabilidade de um valor $z_k(\mathbf{x})$ estar dentro de um intervalo $(a, b]$ qualquer, chamado *intervalo de probabilidade*, é computado como a diferença entre os valores da *fdpac* para os limiares b e a , ou seja:

$$Prob\{Z(\mathbf{x}) \in (a, b] | (n)\} = F(\mathbf{x}; b | (n)) - F(\mathbf{x}; a | (n)) \quad (4.13)$$

Um intervalo de probabilidade dado por $Prob\{Z(\mathbf{x}) \in (a, b] | (n)\} = 0.7$, significa que $z(\mathbf{x})$ tem 70% de chance de estar dentro e, portanto, 30% de chance de estar fora do intervalo $(a, b]$. Quando $b = \infty$ obtêm-se a probabilidade de se exceder um limiar a , ou seja:

$$Prob\{Z(\mathbf{x}) \in (a, +\infty] | (n)\} = Prob\{Z(\mathbf{x}) > a | (n)\} = 1 - F(\mathbf{x}; a | (n)) \quad (4.14)$$

Esta probabilidade é particularmente importante em aplicações ambientais focadas em medir os riscos de se exceder limites regulatórios. Para exemplificar a utilização dessas medidas de incerteza, numa situação real, considere o conjunto amostral de altimetria de Canchim, apresentado na Figura 4-5. Esse conjunto amostral foi utilizado como entrada para produção do mapa temático de altimetria e do mapa de incertezas apresentados na Figura 4-6 (a) e (b), respectivamente.

A classificação apresentada no mapa da Figura 4-6(a) foi obtida a partir dos modelos de distribuição probabilística inferidos pelo procedimento de krigeagem por indicação condicionado às amostras de altimetria. Neste caso, foram definidas 3 faixas distintas de valores de altimetria, 3 classes, e para cada ponto desse mapa, as probabilidades de pertinência a cada um dos intervalos de valores, definidos para as classes, foram calculadas pela formulação apresentada na equação 4.13. Para classificação de cada ponto do mapa temático de altimetria, utilizou-se o critério de máxima probabilidade, ou seja, atribuiu-se, a cada ponto do mapa, a classe de maior probabilidade de ocorrência nesse local. Os valores de incerteza apresentados na Figura 4-6(b), mapa da direita, foram calculados a partir do valor da probabilidade da classe que foi associada a cada ponto do mapa temático de altimetria gerado. Assim, calculou-se a incerteza como:

$$Inc(\mathbf{x}) = 1 - Prob\{z(\mathbf{x}) \in s_k(\mathbf{x}), k = 1, 2 \text{ ou } 3\} \quad (4.15)$$

onde $s_k(\mathbf{x})$ é a classe atribuída a localização (\mathbf{x}) .

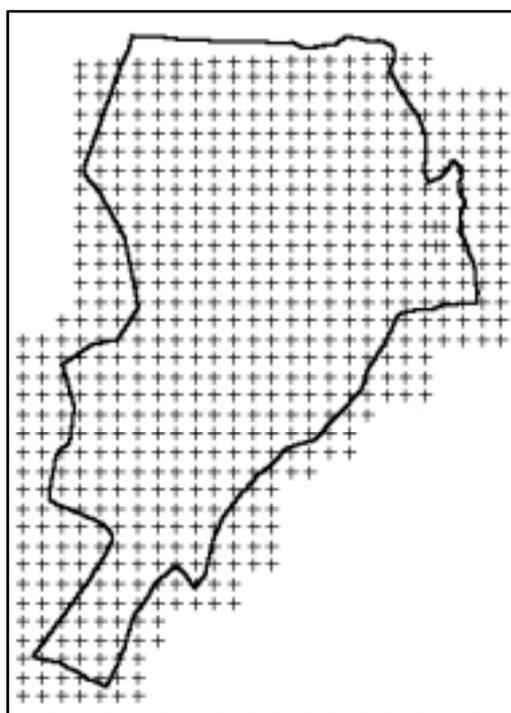


Figura 4-5 Distribuição espacial das amostras de altimetria na região de Canchim

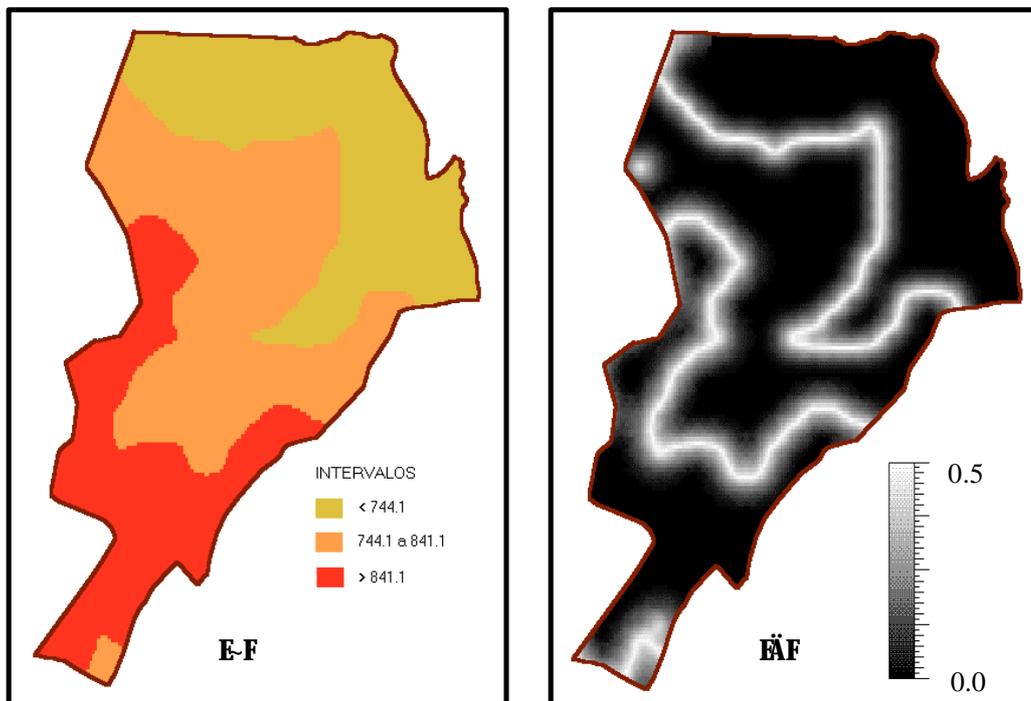


Figura 4-6 Mapa temático de altimetria (a) e respectivas medidas de incerteza (b)

Distância interquantil

Uma medida mais robusta de espalhamento é um intervalo interquantil. Por exemplo, o intervalo interquartil, $q_R(\mathbf{x})$ é definido por:

$$q_R(\mathbf{x}) = q_{0.75}(\mathbf{x}) - q_{0.25}(\mathbf{x}) = F^{-1}(\mathbf{x}; 0.75 / (n)) - F^{-1}(\mathbf{x}; 0.25 / (n)) \quad (4.16)$$

Para distribuições altamente assimétricas, uma medida mais robusta é o intervalo interquantil, que é definido como a diferença entre dois quantis, simétricos em relação a mediana. A partir da função de distribuição acumulada condicionada inferida, $\hat{F}(\mathbf{x}; z|(n))$, pode-se derivar vários intervalos de probabilidade tais como o intervalo 95%, $[q_{0.025}; q_{0.975}]$, tal que:

$$Prob\{Z(\mathbf{x}) \in [q_{0.025}; q_{0.975}] | (n)\} = 0.95 \quad (4.17)$$

com $q_{0.025}$ e $q_{0.975}$ sendo os quantis relativos aos valores de probabilidade da *fdpac* 2,5% e 97.5%, ou seja, $F^*(\mathbf{x}; q_{0.025} | (n)) = 0.025$, e $F^*(\mathbf{x}; q_{0.975} | (n)) = 0.975$. Os valores do atributo, referentes aos quantis, são estimados a partir da função de ajuste e dos valores de corte usados na krigagem por indicação. Um mapa de incertezas obtido pelos valores de uma grade de intervalos interquartis, diferença entre o primeiro e o terceiro quartil de altimetria, e estimados segundo a equação 4.16, está apresentado na Figura 4-7.

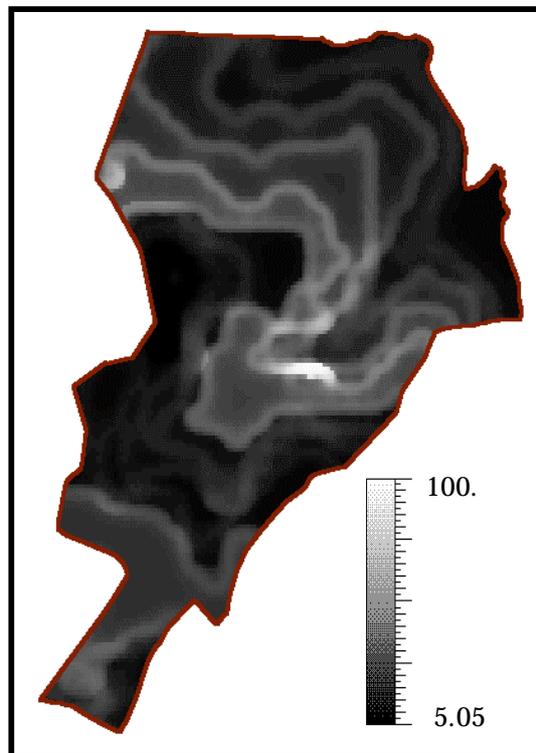


Figura 4-7 Mapa de incertezas locais obtido a partir dos quartis, primeiro e terceiro, dos modelos de distribuição probabilística locais inferidos pela krigagem por indicação

Variância condicional

Uma medida importante de espalhamento de uma distribuição é a variância condicional que mede os desvios da *fdpac* em torno da média da distribuição, $\bar{z}_{z_k}(u)$. Diferente das medidas de incerteza anteriormente descritas, esta necessita da estimação da média da distribuição, isto é, da definição desse estimador. É possível obter-se uma estimativa da variância da distribuição condicionada, $\hat{\sigma}^2(\mathbf{x})$, pela seguinte formulação:

$$\begin{aligned}(\hat{\sigma}^2)(\mathbf{x}) &= \int_{-\infty}^{\infty} [z - \bar{z}_{z_k}(\mathbf{x})]^2 dF(\mathbf{x}; z / (n)) \\ &\approx \sum_{k=1}^{K+1} [z'_k - \bar{z}_{z_k}(\mathbf{x})]^2 [\hat{F}(\mathbf{x}; z_k / (n)) - \hat{F}(\mathbf{x}; z_{k-1} / (n))]\end{aligned}\quad (4.18)$$

onde \bar{z}_{z_k} é o valor da média da classe $(z_{k-1}, z_k]$.

A Figura 4-8 apresenta um mapa de variâncias para os valores de altimetria, da região de Canchim, obtidas pela equação 4.18.

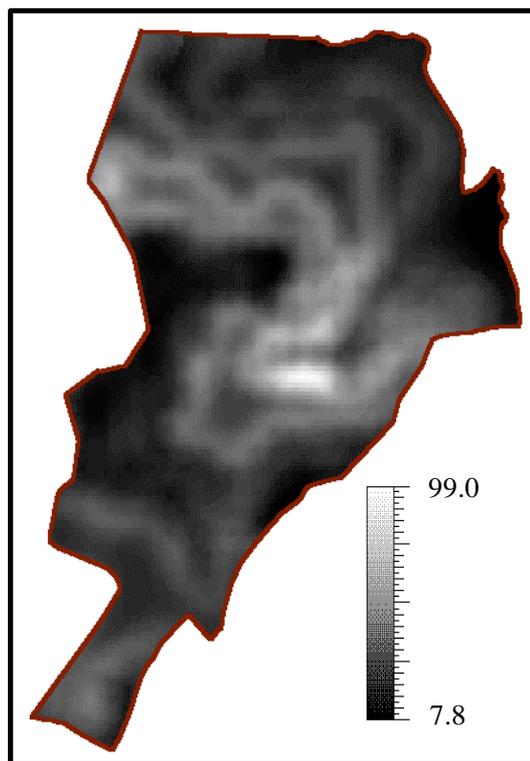


Figura 4-8 Mapa de incertezas locais obtido a partir das variâncias inferidas dos modelos de distribuição probabilística construídos pela krigeagem por indicação.

Entropia de Shannon

Uma medida de incerteza local, não relacionada a qualquer intervalo $(a, b]$, é dada pela medida de entropia da função de densidade de probabilidade local. Essa medida é definida como:

$$H(\mathbf{x}) = \int_{-\infty}^{\infty} -[\ln f(\mathbf{x}; z|(n))] \cdot f(\mathbf{x}; z|(n)) dz \quad (4.19)$$

onde $f(\mathbf{x}; z|(n)) = \partial F(\mathbf{x}; z|(n)) / \partial z$ é a função de distribuição de probabilidade. Na prática a amplitude de variação de z é discretizada em K classes, que não se interceptam, $(z_{k-1}, z_k]$, computando-se a probabilidade desses K intervalos como:

$$p_k(\mathbf{x}) = [\hat{F}(\mathbf{x}; z_k | (n)) - \hat{F}(\mathbf{x}; z_{k-1} | (n))] \quad (4.20)$$

A entropia para a distribuição condicional em \mathbf{x} é computada como:

$$H(\mathbf{x}) \cong - \sum_{k=1}^K [\ln(p_k(\mathbf{x}))] \cdot p_k(\mathbf{x}) \geq 0, \quad \forall p_k \neq 0 \quad (4.21)$$

4.5 Estimadores Ótimos para as Superfícies Interpoladas

O processo inferencial visa calcular uma estimativa do valor de $z(\mathbf{x})$ através de um estimador que é caracterizado por uma determinada função dos dados. Esse estimador, no que concerne aos objetivos do processo inferencial, deve minimizar algum tipo de erro que se deseja evitar, maximizando os acertos de interesse. Por essa razão, um estimador é dito ótimo quando minimiza perdas, isto é, uma particular função dos erros inferenciais, $L(\varepsilon)$, onde $\varepsilon = z(\mathbf{x}) - \hat{z}(\mathbf{x})$. Entretanto, minimizar $L(\varepsilon)$ significa conhecer $z(\mathbf{x})$, que é desconhecido. Portanto, a idéia é utilizar o modelo de incerteza definido para determinar a perda esperada, $E[L(\varepsilon)]$.

$$\begin{aligned} E[L(\varepsilon)] &= E\{L(\varepsilon(\mathbf{x}))|(n)\} \\ &= \int_{-\infty}^{+\infty} L(\varepsilon(\mathbf{x})) dF(\mathbf{x}, z|(n)) \end{aligned} \quad (4.22)$$

Na prática, a seguinte aproximação é utilizada

$$E[L(\varepsilon)] \cong \sum_{k=1}^{K+1} L(\hat{z}(\mathbf{x}) - \bar{z}_k) [\hat{F}(\mathbf{x}, z_k | (n)) - \hat{F}(\mathbf{x}, z_{k-1} | (n))] \quad (4.23)$$

Assim sendo a determinação de estimativas ótimas se processa em duas etapas:

1. A incerteza sobre o valor desconhecido $z(\mathbf{x})$ é inicialmente modelada pela sua *fdpac* $\hat{F}(\mathbf{x}, z_k | (n))$;

2. Desse modelo uma estimativa de $\hat{z}(\mathbf{x})$ é obtida tal que minimiza $E[L(\varepsilon)]$.

Estimativa do valor esperado

A estimativa do valor esperado para cada valor espacial da distribuição é realizada a partir do de mínimos quadrados onde $L[\varepsilon(u)] = [\varepsilon(u)]^2$. Mostra-se que essa função é minimizada quando z é o valor esperado, $\hat{z}(\mathbf{x}) = z_E(\mathbf{x})$. A estimativa do valor esperado, $\hat{z}_E(\mathbf{x}) = E\{Z(\mathbf{x})\}$ onde:

$$E[Z(\mathbf{x})] = \int_{-\infty}^{\infty} z \cdot f(\mathbf{x}; z/(n)) dz = \int_{-\infty}^{\infty} z \cdot dF(\mathbf{x}; z/(n)) \quad (4.24)$$

é obtida pela função de densidade de probabilidade condicionada as n amostras, $f(\mathbf{x}, z_k | (n))$, e a partir dos K valores de corte, z_k , pela aproximação:

$$E[Z(\mathbf{x})] = \int_{-\infty}^{\infty} z \cdot dF(\mathbf{x}; z/(n)) \approx \sum_{k=1}^{K+1} \bar{z}_k [\hat{F}(\mathbf{x}; z_k | (n)) - \hat{F}(\mathbf{x}; z_{k-1} | (n))] \quad (4.25)$$

A estimativa do valor esperado como definida em (4.25) e aquela obtida por krigagem linear são ambas ótimas no sentido de minimizar variâncias inferenciais, entretanto produzem resultados diferentes. São diferentes porque, no caso do enfoque aqui adotado, derivam de uma *fdpac* que depende dos valores dos dados.

Estimativa da mediana

O estimador de mínimos quadrados não é a única função de otimização de erros possível. Uma outra função $L(\varepsilon(\mathbf{x}))$ pode também ser considerada. Podemos tomá-la como sendo dada pelo valor absoluto dos erros estimados $L(\varepsilon(\mathbf{x})) = |\varepsilon(\mathbf{x})|$. Mostra-se que o valor de z que minimiza $E[L(\varepsilon(\mathbf{x}))]$, quando $L(\varepsilon(\mathbf{x}))$ é o modulo de $\varepsilon(\mathbf{x})$, é a mediana da distribuição $q_{0.5}(\mathbf{x})$, definida como:

$$q_{0.5}(\mathbf{x}) = F^{-1}(\mathbf{x}; 0.5/(n)) \quad (4.26)$$

A mediana é inferida aplicando-se a função de ajuste da distribuição sobre os valores de corte com probabilidades acumuladas vizinhas ao valor 0.5. Para distribuições com alto grau de assimetria, a mediana é um estimador mais robusto do que a média. Os mapas de média e mediana, dos dados de altimetria de Canchim, estão mostrados na Figura 4-9.

Estimativa de quantis

A função de perda considerada nos dois estimadores anteriormente definidos não discriminava as diferenças de impacto dos erros de sub-estimação ou sobre-estimação. Entretanto, existem situações, como a descrita no início desse capítulo (vide Seção 4.1), em que cada um desses erros produz diferentes impactos, e essas diferenças devem ser também consideradas no processo inferencial. Assim, funções de perdas assimétricas devem ser utilizadas

$$L[\varepsilon(\mathbf{x})] = \begin{cases} w_1 \cdot \varepsilon(\mathbf{x}) & \text{se } \varepsilon(\mathbf{x}) \geq 0 \quad \text{sobrestimado} \\ w_2 \cdot \varepsilon(\mathbf{x}) & \text{se } \varepsilon(\mathbf{x}) < 0 \quad \text{subestimado} \end{cases} \quad (4.27)$$

onde w_1 e w_2 são parâmetros não negativos, e medem o relativo impacto de sub ou sobre estimar. O estimador que minimiza essa função $L(\varepsilon(\mathbf{x}))$ é chamado de p-quantil, e definido como:

$$\hat{z}_q = F^{-1}(\mathbf{x}; p|(n)) = q_p(\mathbf{x}) \quad (4.28)$$

onde $p = \frac{w_2}{w_1 + w_2}$

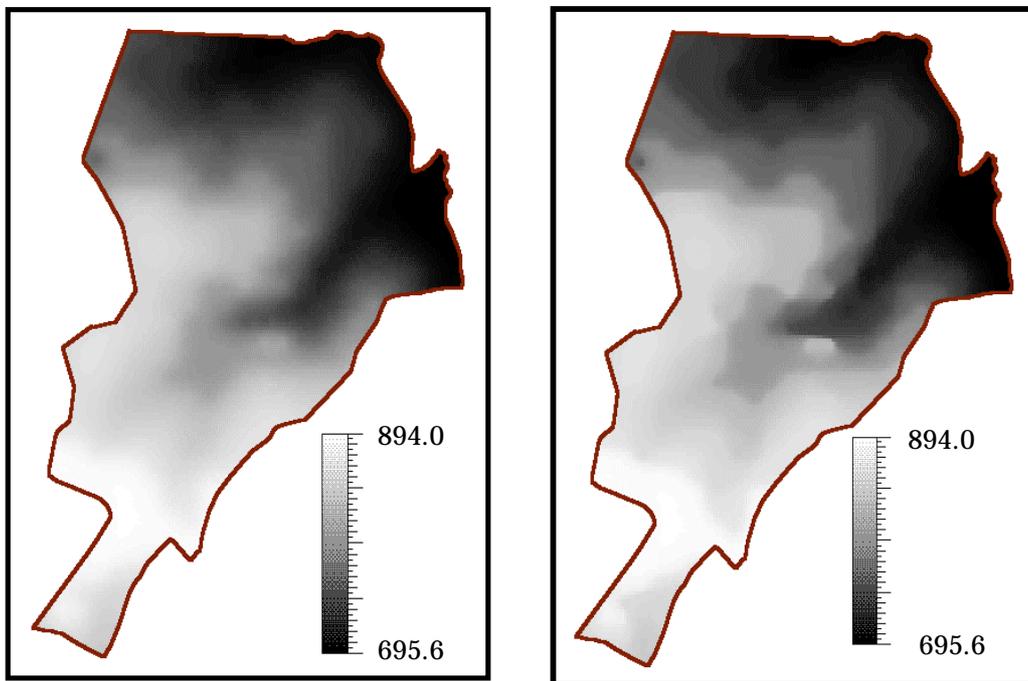


Figura 4-9 Mapas de média (a) e mediana (b) inferidos pelo procedimento por indicação, para os dados de altimetria da região de Canchim.

Considerando o exemplo de aplicação apresentado na introdução desse capítulo, seja w_1 o impacto de sobre-estimar um determinado nutriente no solo, e w_2 o impacto de subestimar este mesmo nutriente. Vamos supor que se deseja estimar $z(\mathbf{x})$ de forma a reduzir o risco de comprometimento da produção, que é motivado pelos erros de sobre-estimação. Dessa forma, $w_1 > w_2$ e $p < 0.5$, ou seja, um estimador ótimo seria um quantil menor do que a mediana, onde $p = 0.5$. Ou ainda, se $w_1 = 0.9$ e $w_2 = 0.1$, $p = 0.1$. A estimativa ótima seria considerando o quantil de 10%.

4.6 Incertezas locais para atributos Categóricos

O enfoque por indicação, semelhante àquele aplicado aos dados com atributos numéricos, pode ser também aplicado a dados com atributos categóricos, também chamados dados temáticos. O dado categórico é aqui considerado como o dado cujo atributo é discreto e sem ordenação, para o qual não é possível um cálculo de distribuições acumuladas, a menos que se defina uma ordenação para os mesmos. Um exemplo típico de dados categóricos é o atributo textura do solo, cujas classes são derivadas de atributos granulométricos do solo. Outros exemplos podem ser: tipos de rochas, classes de solo, etc. A metodologia geoestatística, aqui apresentada, utilizada para espacialização de dados categóricos, baseia-se na krigeagem por indicação e, equivale a um processo de classificação de dados categóricos a partir de amostras individuais. Os principais conceitos abordados aqui são exemplificados a partir do mesmo conjunto de dados coletados na região de Canchim (vide Seção 3.4, Figura 4-10 e Tabela 4-2).

O Enfoque por Indicação para Atributos Categóricos

Considere-se um dado espacial cujo atributo é categórico, podendo assumir K classes, ou estados diferentes, $s_k, k = 1, \dots, K$. Para cada posição (\mathbf{x}) do espaço, o dado categórico pode ser representado por uma variável aleatória $S(\mathbf{x})$ que pode assumir s_k estados, cada um associado a uma probabilidade de ocorrência. Os procedimentos por indicação para atributos categóricos baseiam-se na modelagem da *função de distribuição de probabilidade condicionada*, (fdpc), isto é, a modelagem da distribuição condicionada aos n dados amostrados, $p(\mathbf{x}; s_k / (n))$, que é definida como:

$$p(\mathbf{x}; s_k / (n)) = Prob\{S(\mathbf{x}) = s_k / (n)\} \quad (4.29)$$

A $p(\mathbf{x}; s_k / (n))$ modela a incerteza da variável aleatória S no ponto (\mathbf{x}) e, uma vez estimada, essa função de distribuição de probabilidade pode ser utilizada para:

- classificar o atributo em posições não conhecidas;

- modelar a incerteza das classificações efetuadas.

Pela metodologia por indicação, a definição da *fdpc* depende, inicialmente, da definição de um conjunto de valores de cortes para a variável em questão. Para um conjunto de amostras de uma variável aleatória categórica qualquer, o número de cortes K é definido pela quantidade de classes que essa variável pode assumir no seu domínio. Neste caso, a codificação por indicação, se processa em valor de cortes s_k , e gera um *conjunto amostral por indicação* $i(\mathbf{x}; s_k)$ do tipo:

$$i(\mathbf{x}; s_k) = \begin{cases} 1, & \text{se } s(\mathbf{x}) = s_k \\ 0, & \text{se } s(\mathbf{x}) \neq s_k \end{cases} \quad (4.30)$$

A codificação por indicação é aplicada sobre todo conjunto amostral criando, para cada corte s_k , um conjunto amostral por indicação, $I(\mathbf{x}; s_k / (n))$, cujos valores são 0 ou 1. Cada probabilidade condicional $p(\mathbf{x}; s_k / (n))$ é, também, a esperança condicional da variável aleatória por indicação $I(\mathbf{x}; s_k / (n))$, a saber:

$$p(\mathbf{x}; s_k / (n)) = E\{I(\mathbf{x}; s_k | (n))\} \quad (4.31)$$

onde $I(\mathbf{x}; s_k) = 1$ se $S(\mathbf{x}) = s_k$, e 0 (zero) caso contrário .

Assim, a *fdpc* da variável categórica $S(\mathbf{x})$ pode ser modelada usando-se um enfoque por indicação, semelhante àquele aplicado às variáveis de natureza contínua. Para cada um dos K conjuntos $I(\mathbf{x}; s_k / (n))$, define-se um variograma experimental, ajustado *a posteriori* por um modelo teórico, que busca representar a variabilidade espacial do conjunto de dados codificados por indicação sendo considerados. Cada modelo de variograma teórico, em conjunto com as amostras, codificadas por indicação, é usado para se estimar o valor da probabilidade condicional $[p(\mathbf{x}; s_k / (n))]^*$. O conjunto dessas probabilidades estimadas, considerando-se os K valores de corte, determina uma aproximação discreta da *fdpc* de $S(\mathbf{x})$. Essa *fdpc* deve, ainda, sofrer uma correção dos desvios de relação de ordem para se garantir as relações:

$$[p(\mathbf{x}; s_k / (n))]^* \in [0,1] \quad k = 1, \dots, K \quad (4.32)$$

$$\sum_{k=1}^K [p(\mathbf{x}; s_k / (n))]^* = 1 \quad (4.33)$$

ou seja, cada valor deve estar no intervalo $[0,1]$ e a soma total desses valores deve ser igual a 1.

4.7 Classificadores para Atributos Categóricos

No enfoque por indicação, os classificadores locais para atributos categóricos são definidos a partir da distribuição de probabilidade inferida para cada uma das s_k classes de $S(\mathbf{x})$. Em geral, esse classificador é implementado segundo um *estimador de moda*, que determina o valor de $S(\mathbf{x})$ como sendo a classe com a maior probabilidade inferida em (\mathbf{x}) , ou seja:

$$S^*(\mathbf{x}) = s_{k_{max}}(\mathbf{x}) = s_k(\mathbf{x}) \text{ sse } [p(\mathbf{x}; s_k / (n))]^* > [p(\mathbf{x}; s_i / (n))]^* \forall i = 1, \dots, K \text{ e } k \neq i \quad (4.34)$$

Uma variante do classificador de moda considera também a reprodução das proporções globais definidas a priori. O mapa da Figura 4-11 mostra o resultado de uma classificação, pelo estimador de moda, a partir de um conjunto de amostras do atributo textura do solo.

4.8 Medidas de incerteza para atributos Categóricos

Apresentam-se, a seguir, dois procedimentos de medida de incertezas para atributos categóricos, a incerteza do classificador de moda e a incerteza por entropia de Shannon.

A Incerteza do classificador de moda

A incerteza local $Inc(\mathbf{x})$ pode ser definida como 1(um) menos a maior probabilidade condicional, estimada em \mathbf{x} para as diversas classes de corte s_k :

$$Inc(\mathbf{x}) = 1 - [p(\mathbf{x}; s_{k_{max}}(\mathbf{x}) / (n))]^* \quad (4.35)$$

A Figura 4-12 mostra o mapa de incertezas locais do classificador de moda usado na geração do mapa da Figura 4-11. Analisando-se a classificação apresentada na Figura 4-11 e o mapa de incertezas da Figura 4-12, observa-se que este último mostra um campo com variação proporcional ao comportamento do atributo na região. Nas regiões de transição entre as classes, os valores de incerteza por moda aumentam, com os valores mais baixos longe das transições, como ocorre naturalmente com muitas propriedades naturais nas proximidades de zonas de fronteira.

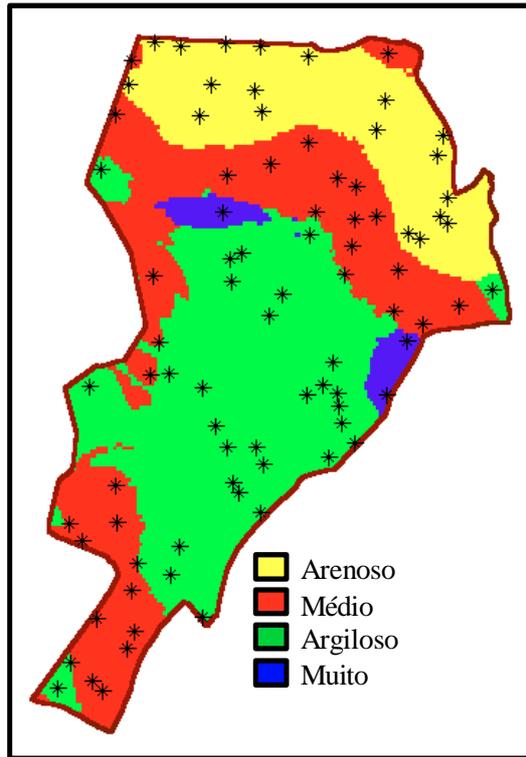


Figura 4-11 Mapa de valores de textura do solo inferidos, pelo valor de moda, a partir do procedimento de krigagem por indicação

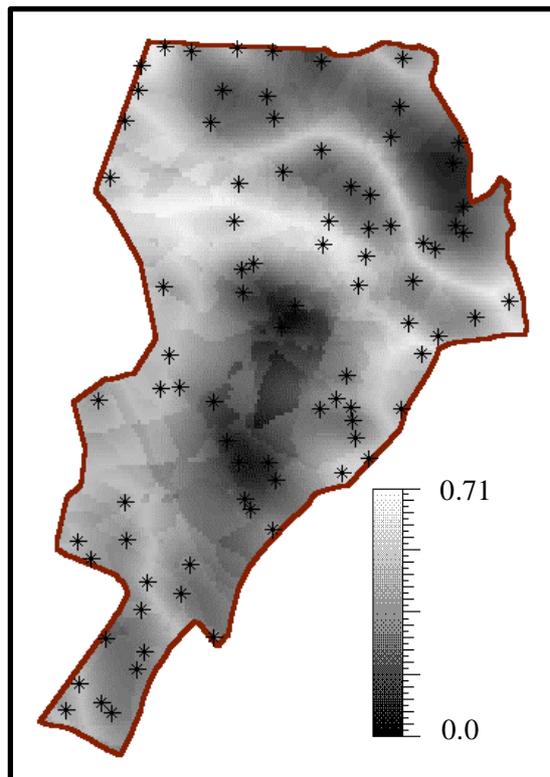


Figura 4-12 Mapa de incerteza por moda estimado a partir do procedimento de krigagem por indicação usado para inferir o mapa da Figura 4-11

Incerteza por entropia de Shannon

Outra medida da incerteza local $Inc(x)$ é a entropia de Shannon das probabilidades condicionais das diversas classes de corte s_k , definida como:

$$Inc(x) = H(x) \cong -\sum_{k=1}^K \ln[p(x; s_k / (n))]^* [p(x; s_k / (n))]^* \geq 0 \quad (4.36)$$

A entropia de Shannon é maximizada para distribuições uniformes, ou seja, quando as probabilidades de ocorrência das classes se igualam. Assim, os valores de incerteza por entropia de Shannon são maiores onde existe uma confusão maior entre as classes consideradas. Isto pode ser observado no mapa de incertezas da Figura 4-13.

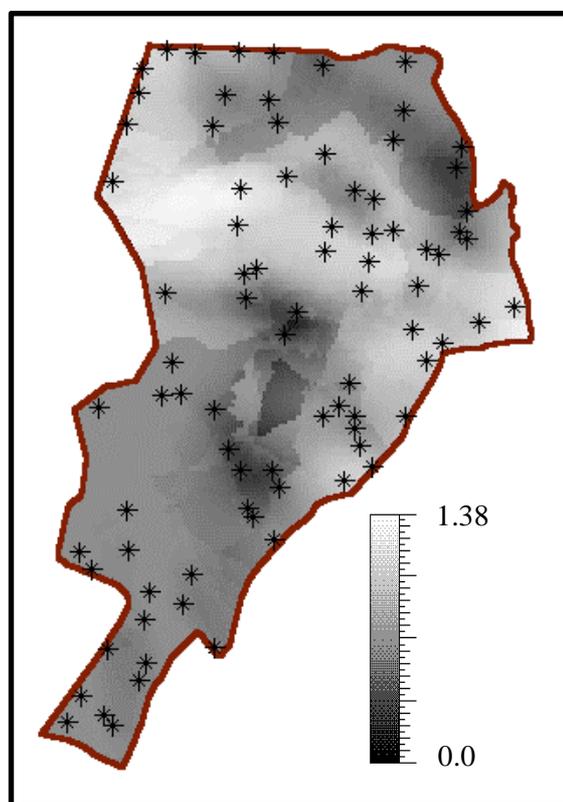


Figura 4-13 Mapa de incerteza por entropia de Shannon estimado a partir do procedimento de krigeagem por indicação usado para inferir o mapa da Figura 4-11

Comparando-se os mapas das Figura 4-12 e Figura 4-13, pode-se analisar as diferenças existentes entre o mapa de incertezas por moda e o mapa de incertezas por entropia. As diferenças são mais aparentes nas regiões onde várias classes se confundem. Este é um resultado esperado, uma vez que, nestas regiões a

distribuição de probabilidade das variáveis aleatórias está mais próxima de uma distribuição uniforme, quando então a incerteza medida pela entropia tem seus valores maximizados. A incerteza por moda mostra um crescimento a partir da parte central de uma classe em direção as zonas de transição. Os valores máximos de incerteza por moda aparecem nas bordas entre as classes e, não têm influência do número de classes próximos as bordas. Dependendo da aplicação, o especialista é responsável por decidir sobre qual medida de incerteza estará trabalhando. Quando a confusão entre classes é importante deve-se optar pela incerteza por entropia. Caso o interesse seja somente nas transições entre as classes, a incerteza por moda deve ser priorizada.

Conclusões

Apresentamos neste capítulo a formalização do procedimentos geoestatísticos da krigagem por indicação. Estes procedimentos servem não apenas para produzir uma predição de valores sobre uma superfície, mas essencialmente como uma poderosa ferramenta para produzir modelos de incertezas locais para dados geográficos que compartilham uma base de informações. Estes dados são sempre usados em conjunto com outros para produzir novas informações, através de operações e transformações. Os procedimentos da geoestatística, em seu enfoque por indicação, nos permitem produzir informações espaciais qualificadas por uma métrica de “confiança” nas informações representadas naqueles suportes, os mapas. Temos a possibilidade concreta de produzir e operar com os mapas e suas “barras de erro”. Podemos ainda ressaltar as seguintes características, específicas do procedimento de krigagem por indicação:

- a krigagem por indicação é não paramétrica. Não considera nenhum tipo de distribuição de probabilidade a priori para a variável aleatória. Ao invés disso, ela possibilita a construção de uma aproximação discretizada da *fdpac*. Os valores de probabilidades discretizados podem ser usados diretamente para se estimar valores característicos da distribuição, tais como: quantis, valor esperado e variância. Portanto, ela não se restringe a modelagem de atributos com distribuições simétricas como, por exemplo, a gaussiana;
- a krigagem por indicação fornece uma metodologia única para espacialização, com estimativa de incertezas, para atributos espaciais tanto de natureza temática quanto numérica;
- diferentemente da krigagem linear, que estima a variância do erro de estimação em função do estimador e da distribuição geométrica das amostras, a krigagem por indicação possibilita a estimativa de incertezas, utilizando a função de distribuição acumulada condicionada da VA que representa o atributo, independentemente do estimador;

- a krigagem por indicação pode ser usada para modelar atributos com alta variabilidade espacial sem a necessidade de se filtrar amostras cujos valores estão muito distantes de uma tendência (“outliers”);
- a krigagem por indicação permite melhorar a qualidade de estimação com o uso de amostras indiretas, retiradas de fontes auxiliares, que são acrescentadas ao conjunto amostral do atributo, as amostras diretas.

No entanto, os procedimentos de krigagem por indicação apresentam também alguns problemas, além das probabilidades negativas e funções acumuladas inválidas já mencionados. Este procedimento requer, do especialista, um alto grau de interatividade para a definição da quantidade e dos valores de corte a serem utilizados. Também, exige que seja definido um variograma para cada valor de corte considerado.

A ferramenta geoestatística de krigagem é utilizada para inferir valores de atributos, em posições não observadas, e também incertezas associadas aos valores inferidos. Mostrou-se que a krigagem por indicação tem aplicação mais geral, principalmente porque não supõe nenhum tipo de distribuição de probabilidade a priori e pode ser usado com atributos numéricos e temáticos. Por exemplo, a krigagem por indicação permite a inferência de valores temáticos e, portanto, pode ser considerada um classificador estocástico, que fornece estimativas de incertezas associadas aos valores das classes atribuídos a cada ponto do espaço. Apresentou-se, ainda, alternativas para estimativas de incertezas que devem ser escolhidas de acordo com a natureza do atributo, que está sendo modelado, e também de acordo com os objetivos de uma aplicação.

Salienta-se que os procedimentos geoestatísticos por indicação incluem também os simuladores estocásticos, que não foram abordados neste capítulo. Também não foi abordado o uso de informação indireta para a melhora das inferências. Estes tópicos são de extrema relevância para o contexto do uso efetivo da geoestatística em análise de dados geográficos e deverão ser considerados em futuras edições. Mesmo no método por indicação algumas limitações da krigagem permanecem – uso dos dados para estimar o variograma e prever a incerteza, deficiência na extrapolação, ou seja, avaliar a incerteza fora dos dados. Novas generalizações começam a surgir, tomando como base a teoria dos campos aleatórios espaço-temporais.

REFERÊNCIAS BIBLIOGRÁFICAS

A estrutura teórica da geoestatística em seu enfoque por indicação está bem apresentada em Goovaerts (1997) e em Isaaks e Srivastava (1989). Algoritmos implementados e explicações didáticas sobre como operar a Krigagem por indicação pode ser encontrada no livro de Deutsch e Journel (1992). Com relação à integração entre geoestatística e SIGs e modelagem e tratamento de incertezas em SIG, o leitor deve referir-se a Felgueiras C. A. (1999), Felgueiras et al (1999) e Heuvelink (1998). As questões sobre medidas de entropia podem ser apreciadas no clássico Shannon, and Weaver (1949). Para uma discussão sobre diferentes medidas de incerteza no enfoque por indicação veja Soares(1992). Referente a modelagem espaço-temporal, deve-se consultar o artigo de Kyriakidis e Journel (1999) e o livro do George Christakos (2000). Referências básicas sobre os dados da Fazenda Canchim podem ser encontrados em Calderano Filho et al. (1996). Estes dados também estão disponíveis no site do livro (www.dpi.inpe.br/gilberto/livro/analise).

Calderano Filho, B.; Fonseca, O. O. M.; Santos, H. G. e Lemos A. L.. *Levantamento Semidetalhado dos Solos da Fazenda Canchim São Carlos - SP*. Rio de Janeiro, EMBRAPA- CNPS, 1996. 261p.

Christakos, G. *Modern Spatiotemporal Geostatistics*; IAMG Studies no. 6, Oxford University Press, 2000

.Deutsch e Journel (1992). *GSLIB: Geostatistical Software Library and user's guide*. New York, Oxford University Press, 1992. 339p.

Felgueiras C. A. *Modelagem Ambiental com Tratamento de Incertezas em Sistemas de Informação Geográfica: O Paradigma Geoestatístico por Indicação*. Tese (Doutorado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, Publicado em <http://www.dpi.inpe.br/teses/carlos/>, 1999.

Felgueiras C. A., Monteiro A. M. V., Fuks S. D. and E. C. G. Camargo. *Inferências e Estimativas de Incertezas Utilizando Técnicas de Krigagem Não Linear* [CD-ROM]. In: *V Congresso e Feira para Usuários de Geoprocessamento da América Latina*, 7, Salvador, 1999. Anais. Bahia, GisBrasil'99. Seção de Palestras Técnico-Científicas.

Goovaerts, P. *Geostatistics for Natural Resources Evaluation*. New York, Oxford University Press, 1997. 481p.;

Isaaks E. H. and Srivastava R. M. *An Introduction to Applied Geostatistics*, Oxford University Press, 1989. 560p.

Kyriakidis, P. C. e Journel, A. G. Geostatistical Space-Time Models: A Review. *Mathematical Geology*, Vol. 31, No. 6, 1999

Heuvelink G. B. M. *Error Propagation in Environmental Modeling with GIS*, Bristol, Taylor and Francis Inc, 1998.

Shannon, C. E. e Weaver, W. *The Mathematical Theory of Communication*. Urbana, The University of Illinois Press, 1949. 117p.

Soares, A. *Geostatistical Estimation of Multi-Phase Structures*. *Mathematical Geology*, 24(2):140-160, 1992.

5 ANÁLISE ESPACIAL DE ÁREAS

Gilberto Câmara
Marília Sá Carvalho
Oswaldo Gonçalves Cruz
Virginia Correa

5.1 INTRODUÇÃO

Este capítulo discute métodos de análise de dados espaciais cuja localização está associada a áreas delimitadas por polígonos. Este caso ocorre com muita frequência quando lidamos com eventos agregados por municípios, bairros ou setores censitários, onde não se dispõe da localização exata dos eventos, mas de um valor por área. Alguns desses indicadores são contagens, como é o caso da maior parte das variáveis coletadas no censo: por exemplo, o IBGE fornece, para cada setor censitário, o número de chefes de família em cada uma das faixas de renda consideradas. Diversos indicadores de saúde também são deste tipo: o Ministério e Secretarias de Saúde organizam e disponibilizam dados de óbitos, partos, doenças transmissíveis por município. Utilizando duas contagens – óbitos e população, por ex. – taxas de densidade de ocorrência, como taxas de mortalidade ou incidência são estimados. Outros indicadores bastante úteis são: (a) proporções, como percentual de adultos analfabetos; (b) médias, como renda média do chefe da família por setor censitário; e (c) medianas, como mediana etária em homens.

A forma usual de apresentação de dados agregados por áreas é o uso de mapas coloridos com o padrão espacial do fenômeno. A Figura 5-1 mostra a distribuição espacial do índice de exclusão social¹ para os 96 distritos da cidade de São Paulo, para os dados do censo de 1991. Verifica-se que 2/3 dos 96 distritos de São Paulo estavam abaixo dos índices mínimos de inclusão social em 1991. Uma forte polarização centro-periferia é claramente perceptível no mapa, que apresenta duas grandes regiões de exclusão social, as zonas Sul e Leste da cidade. Na zona Leste, nota-se um gradiente do índice de exclusão/inclusão social, que piora à medida que nos afastamos do centro. Na zona Sul, a descontinuidade do índice é mais

¹ O índice de exclusão/inclusão social é uma medida agregada das disparidades socioeconômicas, que varia de -1 a +1, onde o valor 0 (zero) indica o um nível básico de inclusão social.

abrupta, e verificamos a existência de distritos com altos índices de exclusão/inclusão social próximos a áreas excluídas.

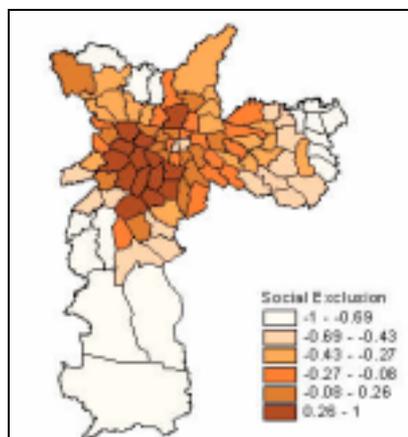


Figura 5-1– Índice de Exclusão/Inclusão Social dos Bairros da Cidade de São Paulo para os dados de 1991, com 96 distritos agrupados por sextis.

Grande parte dos usuários limita seu uso de SIG a essas operações de visualização, tirando conclusões intuitivas. Mas é possível ir muito além. Quando visualizamos um padrão espacial, é muito útil traduzi-lo em considerações objetivas: o padrão que observamos é aleatório ou apresenta uma agregação definida? Esta distribuição pode ser associada a causas mensuráveis? Os valores observados são suficientes para analisar o fenômeno espacial a ser estudado? Existem agrupamentos de áreas com padrões diferenciados dentro da região de estudo?

Para abordar estas questões, este capítulo apresenta um conjunto de técnicas de análise espacial de dados agregados por áreas. O primeiro passo é escolher o modelo inferencial a ser utilizado. A hipótese mais comum é supor que as áreas são diferenciadas, e que cada uma delas possui uma “identidade” própria. Do ponto de vista estatístico, isto implica em que cada área apresenta uma distribuição de probabilidade distinta das demais, o chamado *modelo espacial discreto*. A alternativa é supor que o fenômeno estudado apresenta continuidade espacial, formando uma superfície, o chamado *modelo espacial contínuo* estudado no capítulo anterior. Neste caso, as áreas são consideradas apenas um suporte para coleta de dados, e o modelo inferencial desconsidera os limites de cada área. A produção de superfícies a partir de dados de área será discutida no final deste capítulo.

A questão de agregação de contagens em áreas levanta ainda problemas conceituais importantes: Pode-se estimar comportamentos individuais a partir de dados agregados? Em que medida a comportamento dos

agregados reflete mais do que a soma dos indivíduos? Qual o erro cometido ao estimar indicadores onde as contagens são muito pequenas? Neste capítulo, após a apresentação dos modelos adequados à análise de dados agregados por áreas serão abordados os conceitos básicos da análise espacial, para dados agregados por área.

5.2 MODELOS DE DISTRIBUIÇÃO DE DADOS EM ÁREAS

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

De forma semelhante ao modelo de eventos pontuais discutido no capítulo 2, considere-se Z_i como a variável aleatória que descreve a contagem, indicador ou taxa associada à área A_i . Dispomos de um valor observado z_i , correspondente à contagem na i -ésima área. A hipótese mais comum é supor que a variável aleatória Z_i , que descreve o número de ocorrências em cada área pode ser associada a uma distribuição de probabilidade de Poisson. Tal hipótese justifica-se por ser esta a distribuição estatística mais adequada a fenômenos que envolvem contagens de eventos, como é o caso na maioria dos dados agregados por áreas. Evidentemente outras distribuições podem ser mais adequadas, dependendo da variável a ser analisada. Taxas podem ser modeladas utilizando a distribuição normal, pois ainda que esta admita valores negativos, evidentemente impossíveis neste tipo de indicador, as propriedades da distribuição normal podem ser adequadas.

A alternativa à hipótese de *variação espacial discreta* é supor que os dados apresentam *variação espacial contínua*. Considera-se um processo estocástico $\{Z(x), x \in A, A \subset \mathcal{R}^2\}$, cujos valores podem ser conhecidos em todos os pontos da área de estudo. Neste caso, as contagens agregadas devem ser transformadas em taxas ou indicadores, pois o que varia continuamente no espaço são as taxas e não as contagens. A estimação deste processo estocástico pode ser feita como descrito nos capítulos 3 e 4 deste livro. O uso de modelos espaciais contínuos será discutido na seção 5.8.

5.3 PROBLEMAS DE ESCALA E A RELAÇÃO ÁREA-INDIVÍDUO

Um dos problemas básicos com dados agregados por área é que, para uma mesma população estudada, a definição espacial das fronteiras das áreas afeta os resultados obtidos. As estimativas obtidas dentro de um sistema de unidades de área são função das diversas maneiras que estas unidades podem ser agrupadas; pode-se obter resultados diferentes simplesmente alterando as fronteiras destas zonas. Este problema é conhecido como “problema da unidade de área modificável”.

Em muitos dos estudos envolvendo dados de área, o dado agregado é a única fonte disponível, porém o objeto de estudo diz respeito a características e relacionamentos individuais. Alguns destes estudos procuram estabelecer relações de causa-efeito entre diferentes medidas, como o uso de modelos de regressão; um exemplo clássico é correlacionar anos de estudo do chefe de família e sua renda, que usualmente apresenta forte correlação. Note-se, no entanto, que devido aos efeitos de escala e de agregação de áreas, os coeficientes de correlação podem ser inteiramente diferentes no indivíduo e nas áreas. Este fenômeno, nas ciências sociais e na epidemiologia, é chamado de “falácia ecológica”.

Considere um conjunto de indivíduos onde são medidas duas características de cada um dos indivíduos, conforme estimado na Figura 5-2. Uma regressão considerando todos os indivíduos (linha negra do quadro à esquerda) resulta em coeficiente positivo de 0,1469. Esses indivíduos pertencem a grupos distintos, separando cada grupo conforme o atributo cor, obtém-se correlação negativa, variando entre $-0,5$ e $-0,8$. Utilizando as médias de cada grupo (linha negra do quadro à direita), o coeficiente vai a 0,99. É importante observar que cada modelo mede um aspecto diferente e que não há modelo correto. No primeiro caso, pode-se dizer que sem informações que permitam separar os indivíduos nos grupos coloridos, as variáveis se relacionam positivamente. No último exemplo, o interesse do estudo é o efeito da variação na média de uma variável sobre a média da outra, nos grupos. São perguntas diferentes, e modelos diferentes.

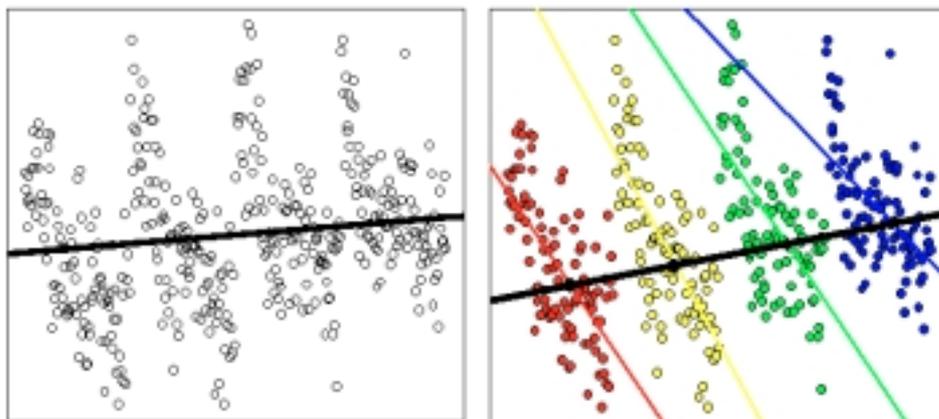


Figura 5-2 – Modelos de regressão: indivíduos, indivíduos em estratos diferentes e grupos.

Para ilustrar o problema das mudanças de unidade de análise, estudou-se os dados de censo de Belo Horizonte para o ano de 1991, em duas escalas: os setores censitários e as unidades de planejamento (UP), mostradas na Figura 5-2. Os setores censitários foram utilizados pelo IBGE para o censo de 1991, e as unidades de planejamento correspondem a agregamentos de áreas utilizados pela prefeitura de Belo Horizonte.

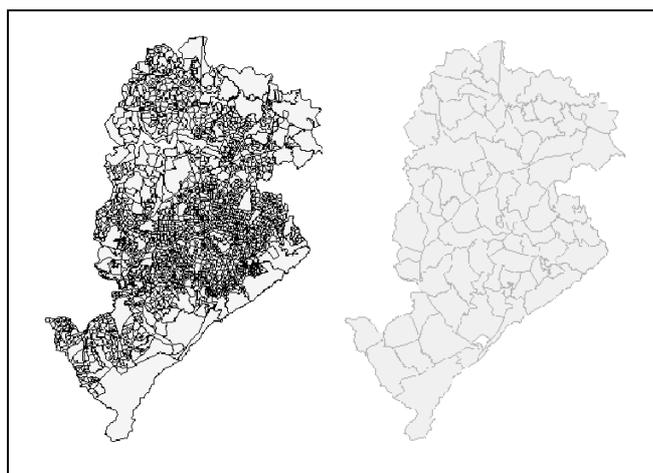


Figura 5-3. Setores censitários (à esquerda) e Unidades de Planejamento (à direita) para o município de Belo Horizonte.

A partir das variáveis do censo, foram computadas 1000 correlações entre pares de variáveis, tanto por setor censitário como por UP. Por exemplo, tomou-se as variáveis “número de chefes de família com rendimento entre 0,5 e 1 salário mínimo” e “número de chefes de família com 1 a 3 anos de estudo” e computou-se a correlação para o caso de setores censitários (0,79) e para o caso de UP (0,96). Os resultados, mostrados na Tabela 5-1, indicam que as correlações nos setores censitários são significativamente menores que as correlações por unidades de planejamento. Nada menos que 773 correlações são menores para os

setores censitários que para as UPs. Apenas 40 (4%) tem o comportamento oposto. Em algumas situações, ocorre inclusive mudança de sinal, isto é, variáveis correlacionadas negativamente no nível dos setores censitários passam a ser correlacionadas positivamente. Verifica-se que a redução de escala (áreas maiores) tende a homogeneizar os dados, reduzir a flutuação aleatória e reforçar correlações que, assim, aparentam ser mais fortes que em áreas menores.

Os resultados acima indicam que não se pode afirmar que qualquer escala seja a “certa”, mas apenas qual dos modelos melhor serve ao que se deseja esclarecer: correlações mais fracas e maior flutuação aleatória, porém com mais homogeneidade interna, ou mais fortes com o viés ocasionado por desconsiderar a dispersão e a heterogeneidade em torno da média nas grande áreas. Como regra geral, quanto mais desagregado o dado, maior a flexibilidade na escolha de modelos; pois agregar em regiões maiores é fácil, mas desagregar impossível.

Tabela 5-1

CORRELAÇÕES ENTRE PARES DE VARIÁVEIS SEGUNDO
DIFERENTES UNIDADES DE ÁREAS – SETOR CENSITÁRIO E UNIDADE DE
PLANEJAMENTO - PARA O CENSO DE 1991 EM BELO HORIZONTE

Correlações por Unidade de Planejamento

	-0,4/-0,2	-0,2/0,0	0,0/0,2	0,2/0,4	0,4/0,6	0,6/0,8	0,8/1,0	Pares
-0,8/-0,6	0	0	1	1	1	0	2	5
-0,6/-0,4	2	11	7	4	2	7	0	33
-0,4/-0,2	3	23	14	11	10	3	6	70
-0,2/0,0	3	5	9	27	34	13	21	112
0,0/0,2	0	1	2	42	75	32	55	207
0,2/0,4	0	2	0	17	44	50	68	181
0,4/0,6	0	2	3	1	10	42	110	168
0,6/0,8	0	0	2	7	8	9	75	101
0,8/1,0	0	0	0	4	4	3	112	123
Totais	8	45	38	114	187	159	449	1000

Na prática, por razões de confidencialidade, os dados individuais muito raramente estão disponíveis. O que fazer então? Uma possibilidade é trabalhar com os dados na maior escala

espacial possível, usualmente denominadas micro-áreas, por exemplo, setores censitários. E utilizar técnicas de agregação ou de otimização combinatória para obter regiões mais agregadas, mas que preservem o fenômeno estudado da melhor forma possível. Deste modo, deve-se reconhecer que o problema da escala é um efeito inerente aos dados agregados por áreas. Ele não pode ser removido e não pode ser ignorado. Para minimizar seu impacto com relação a esses estudos, deve-se procurar utilizar a melhor escala de levantamento de dados disponível e utilizar técnicas que permitam tratar a flutuação aleatória, sempre buscando critérios de agregação dos dados que sejam consistentes com os objetivos do estudo.

5.4 ANÁLISE EXPLORATÓRIA

As técnicas de análise exploratória aplicadas a dados espaciais são essenciais ao desenvolvimento das etapas da modelagem estatística espacial, em geral sensível ao tipo de distribuição, à presença de valores extremos e à ausência de estacionariedade. As técnicas empregadas são, em geral, adaptações das ferramentas usuais. Assim, se na investigação de valores extremos se utiliza ferramentas gráficas como histogramas ou *boxplots*, na análise espacial é importante também investigar *outliers* não só no conjunto dos dados mas também em relação aos vizinhos. Além disso, a não-estacionariedade do processo espacial na região de estudo também deve ser investigada, nos seus vários aspectos: variação na média (primeira ordem), na variância e na covariância espacial.

Visualização de Dados

A forma mais simples e intuitiva de análise exploratória é a visualização de valores extremos nos mapas. Vale ressaltar que o uso de diferentes pontos de corte da variável induz a visualização de diferentes aspectos. Os SIGs dispõem usualmente de três métodos de corte de variável: intervalos iguais, percentis e desvios padrões. No caso de *intervalos iguais*, em que os valores máximo e mínimo são divididos pelo número de classes. Se a variável tem uma distribuição muito concentrada de um lado, este corte deixa apenas um número muito pequeno de áreas nas classes da perna mais longa da distribuição; como resultado, a maior parte das áreas será alocada a uma ou duas cores. O uso de *percentis* para definição de classes obriga a alocação dos polígonos em quantidades iguais pelas cores; isto pode mascarar diferenças significativas em valores extremos e dificultar a identificação de áreas críticas. Finalmente, o uso de desvios padrões, no qual a distribuição da variável é apresentada em gradações de cores diferentes para valores acima e abaixo da média, faz a suposição da normalidade da distribuição da variável; esta hipótese é pouco realista no caso de variáveis censitárias em países de

grande desigualdade social com o Brasil.” Em resumo, é parte importante da análise exploratória experimentar diferentes pontos de corte da variável na visualização dos mapas.

As diferentes técnicas de visualização estão ilustradas no exemplo a seguir, em que mostramos a distribuição espacial do indicador que mede a proporção de recém-natos que nasce em boas condições de saúde (Índice de APGAR) para os bairros do Rio de Janeiro, no ano de 1994. Foram geradas duas visualizações, ambas com 5 pontos de corte e 5 cores. Na Figura 5-4, utilizou-se quintis; na Figura 5-5, cinco classes de igual tamanho. Como a distribuição da variável não é simétrica, quando se divide em classes de amplitudes iguais as de valores mais baixos (ou piores), assinaladas em vermelho ficam reduzidas a poucas áreas, enquanto que na divisão em quintis, por definição, um quinto das áreas ficará em cada classe. A pergunta então é: o que se deseja mostrar? Certamente o responsável pela assistência peri-natal da região não ficará satisfeito visualizando um quinto dos bairros como sendo de “alto” risco. Por outro lado, como as áreas onde o índice é mais baixo têm população pequena, a confiabilidade dos valores encontrados pode ser efeito apenas da flutuação aleatória descrita anteriormente. Vale a pena então olhar mapas? Claro que sim, da mesma forma como olhamos histogramas e box-plots, e procurando sempre ver a distribuição utilizando diferentes pontos de corte. Os SIGs em geral tem uma forma padrão, mas dezenas de possibilidades podem e devem ser exploradas.

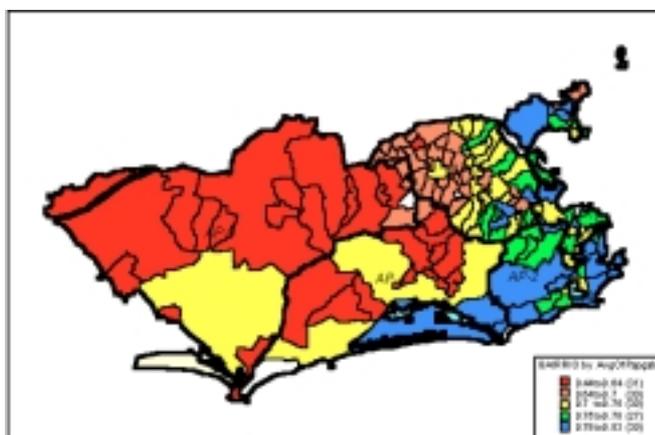


Figura 5-4– Distribuição do índice de APGAR, agrupada em quintis.

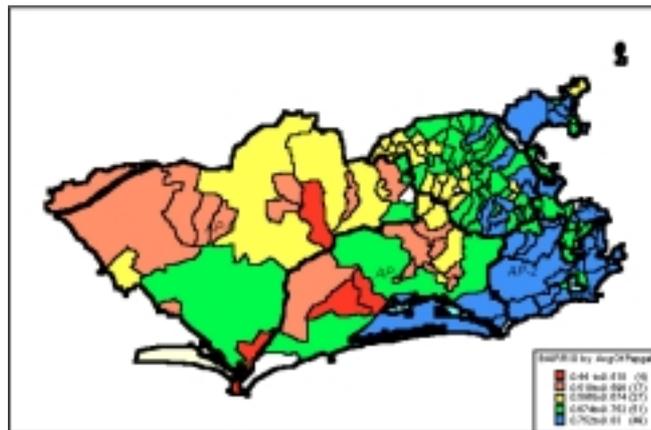


Figura 5-5 - Distribuição do índice de APGAR, agrupada em classes de igual amplitude.

Outra questão interessante é a comparação de mapas. Supondo a distribuição espacial de um indicador em diferentes anos: como visualizar a evolução temporal? Certamente os pontos de corte da variável nos diferentes períodos devem ser os mesmos. Observe na Figura 5-4 a evolução temporal da mortalidade por homicídios para os triênios 79-81 e 90-92, no Estado do Rio de Janeiro. A apresentação dos quintis da *distribuição conjunta* dos indicadores permite visualizar bem o espalhamento desta “doença”.

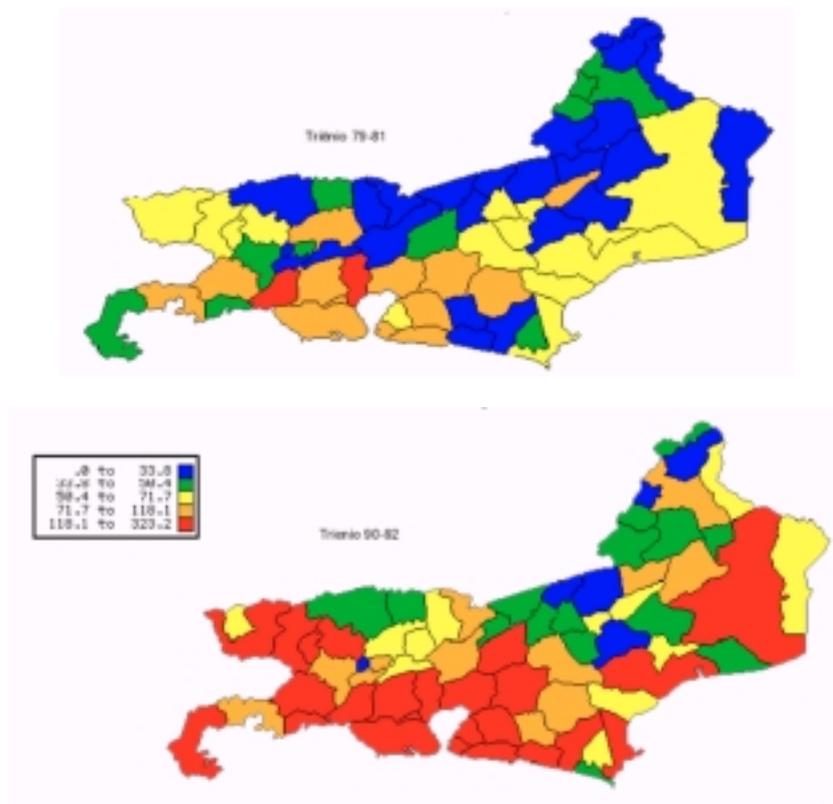


Figura 5-6– Mortalidade por homicídios no Rio de Janeiro, para os triênios 79-81 e 80-92.

Gráficos de Médias e Medianas

Os gráficos de médias e medianas segundo linhas e colunas permitem explorar simultaneamente a presença de tendência (não-estacionariedade de primeira ordem), e não-estacionariedade de segunda ordem, onde a variância e a covariância entre vizinhos não se mantém constante. Para construir estes gráficos, utiliza-se as coordenadas dos centróides das áreas, aproximando-as para um espaçamento regular de forma a montar uma matriz. Calcula-se então as médias e as medianas do indicador ao longo das linhas (eixo Leste-Oeste) e colunas (eixo Norte-Sul) desta matriz. Esta técnica permite identificar a flutuação das medidas ao longo de duas direções, sugerindo a presença de valores discrepantes quando a diferença entre estas é grande, e a tendência ao longo de uma direção quando os valores variam suavemente.

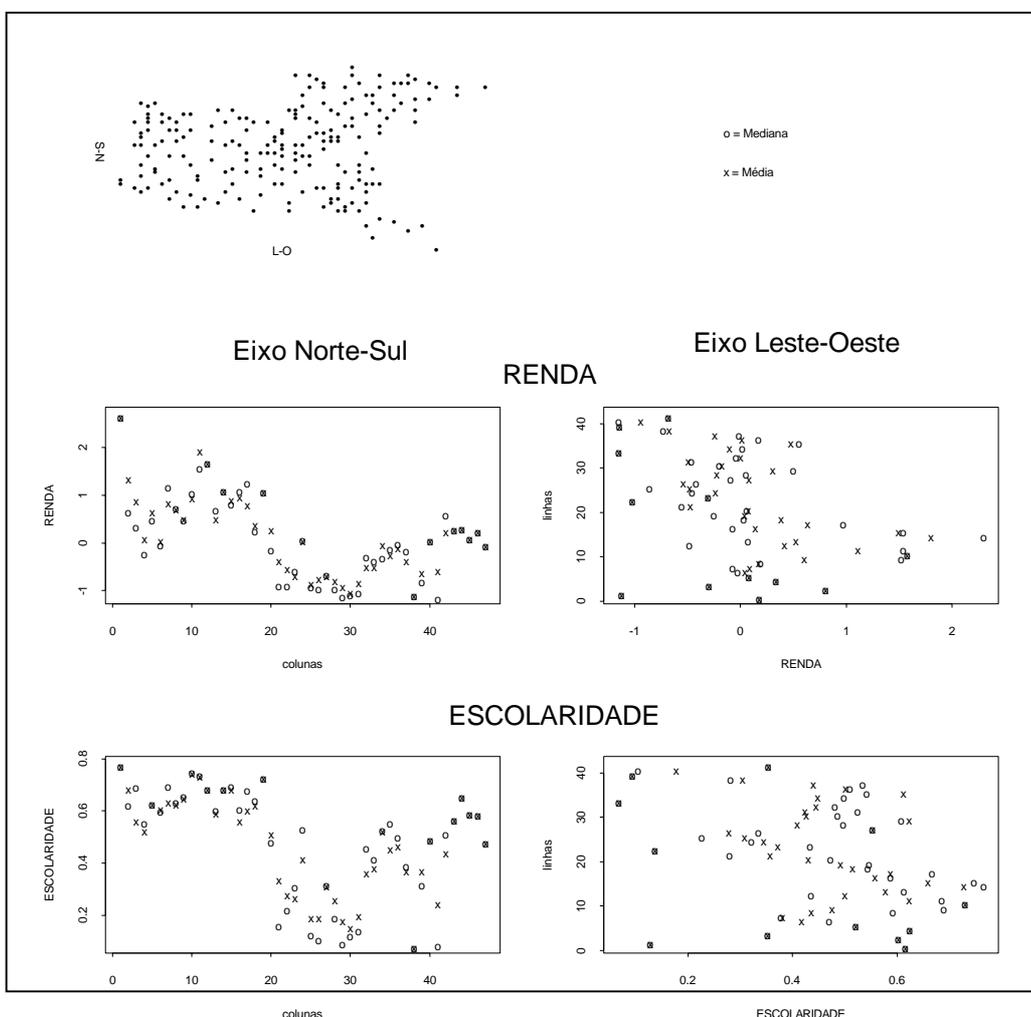


Figura 5-7 – Médias e medianas para escolaridade e renda na Ilha do Governador.

Na Figura 5-7, apresenta-se o resultado desta técnica aplicada a dois indicadores socioeconômicos do censo 1991 – renda média do chefe da família e proporção de chefes de família com escolaridade igual ou superior ao segundo grau – para setores censitários da Ilha do Governador, no Rio de Janeiro. Esta é composta por 225 setores censitários, cujos centróides estão assinalados no primeiro quadro da figura: observe que nas extremidades do “mapa” a quantidade de pontos é muito pequena, e, conseqüentemente, qualquer medida nesta área será pouco robusta.

No eixo Norte-Sul (colunas) pode-se observar que a renda média do chefe da família apresenta tendência variável, bem menor no centro da região. A mesma coisa acontece para escolaridade, embora com maior flutuação. No eixo Leste-Oeste (linhas), também parece haver algum deslocamento para valores mais altos no sentido leste, mas o descolamento de médias (\bar{x}) e medianas (o) sugere a presença de valores extremos dos indicadores. A variação na média dos indicadores na região está, aparentemente, dividida entre as duas direções analisadas, e pode-se explorar melhor a tendência através da rotação dos eixos de referência.

Análise de Autocorrelação Espacial

Outra etapa da análise exploratória visa identificar a estrutura de correlação espacial que melhor descreva os dados. A idéia básica é estimar a magnitude da autocorrelação espacial entre as áreas. Neste caso, as ferramentas utilizadas são o índice global de Moran, o índice de Geary e o variograma. Quando se dispõe de grande número de áreas, resultantes por exemplo de escalas espaciais detalhadas, a natureza dos processos envolvidos é tal que é muito provável a existência de diferentes regimes de correlação espacial em diferentes sub-regiões. Para evidenciar estes regimes espaciais, pode-se utilizar os indicadores locais de autocorrelação espacial e o mapa de espalhamento de Moran, descritos também nesta seção. Todas estas estatísticas dependem da definição de vizinhança adotada, discutida a seguir.

Matrizes de Proximidade Espacial

Para estimar a variabilidade espacial de dados de área, uma ferramenta básica é a *matriz de proximidade espacial*, também chamada matriz de vizinhança. Dado um conjunto de n áreas $\{A_1, \dots, A_n\}$, construímos a matriz $\mathbf{W}^{(1)}$ ($n \times n$), onde cada um dos elementos w_{ij} representa uma medida de proximidade entre A_i e A_j . Esta medida de proximidade pode ser calculada a partir de um dos seguintes critérios:

- $w_{ij} = 1$, se o centróide de A_i está a uma determinada distância de A_j ; caso contrário $w_{ij} = 0$

- $w_{ij} = 1$, se A_i compartilha um lado comum com A_j , caso contrário $w_{ij} = 0$
- $w_{ij} = l_{ij}/l_i$, onde l_{ij} é o comprimento da fronteira entre A_i e A_j e l_i é o perímetro de A_i

Como a matriz de proximidade é utilizada em cálculos de indicadores na fase de análise exploratória, é muito útil normalizar suas linhas, para que a soma dos pesos de cada linha seja igual a 1. Isto simplifica muito vários cálculos de índices de autocorrelação espacial, como se verá a seguir. A Figura 5-8 ilustra um exemplo simples de matriz de proximidade espacial, em que os valores dos elementos da matriz refletem o critério de adjacência e foram normalizados.

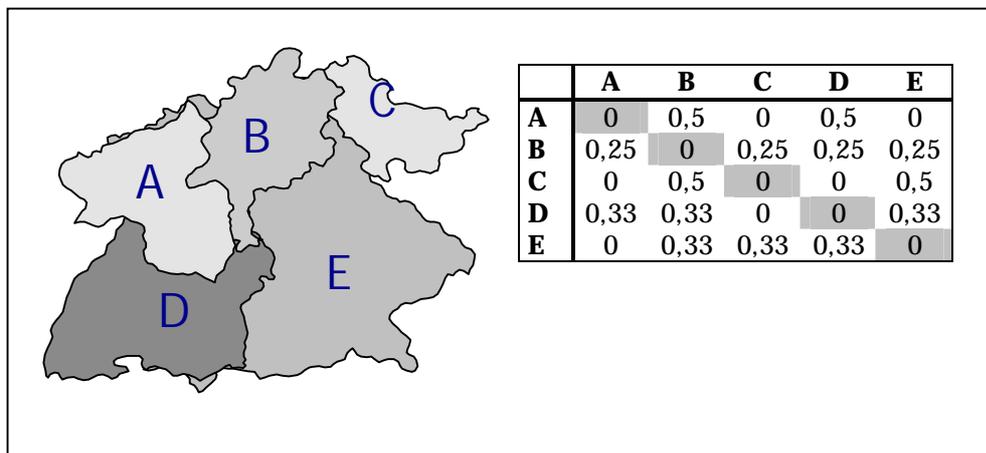


Figura 5-8- Matriz de proximidade espacial de primeira ordem, normalizada pelas linhas.

A idéia da matriz de proximidade espacial pode ser generalizada para vizinhos de maior ordem (vizinhos dos vizinhos). Com critério análogo ao adotado para a matriz de vizinhança de primeira ordem, pode-se construir as matrizes $\mathbf{W}^{(2)}$, ..., $\mathbf{W}^{(n)}$. Por exemplo, na Figura 5-6, as áreas A e C são vizinhas na matriz de proximidade espacial de ordem 2. No que segue, por simplicidade, os coeficientes da matriz de primeira ordem são designados simplesmente por w_{ij} , e os das matrizes de ordem k por $w_{ij}^{(k)}$ e que essas matrizes estão normalizadas por linhas.

Média Móvel Espacial

Uma forma simples de explorar a variação da tendência espacial dos dados é calcular a média dos valores dos vizinhos. Isto reduz a variabilidade espacial, pois a operação tende a produzir uma superfície com menor flutuação que os dados originais. A média móvel $\hat{\mu}_i$ associada ao atributo z_i relativo à i-ésima área, pode ser calculada a partir dos elementos w_{ij} da matriz normalizada de proximidade espacial $\mathbf{W}^{(1)}$, tomando-se simplesmente a média dos vizinhos:

$$\hat{\mu}_i = \sum_{j=1}^n w_{ij} z_i \quad (5.1.)$$

A Figura 5-9 ilustra o uso do estimador de média móvel para o percentual de idosos (mais de 70 anos) para os 96 distritos da cidade de São Paulo. Estes dados são indicadores da grande disparidade social da cidade, com uma grande variação entre o centro (onde a proporção de idosos chega a 8%) com a periferia (onde há várias regiões com menos de 1%). O valor máximo do percentual de idosos é de 8,2% e o mínimo de 0,8%, com um desvio padrão de aproximadamente 2%. Com a média local, há um alisamento: o valor mínimo é de 1% e o máximo é reduzido a 6,8%. Pode-se notar, ao comparar os dois mapas da Figura 5-9, que a média móvel local fornece uma visão das grandes *tendências* do fenômeno em estudo e no caso do percentual de idosos, mostra um forte gradiente centro-periferia.

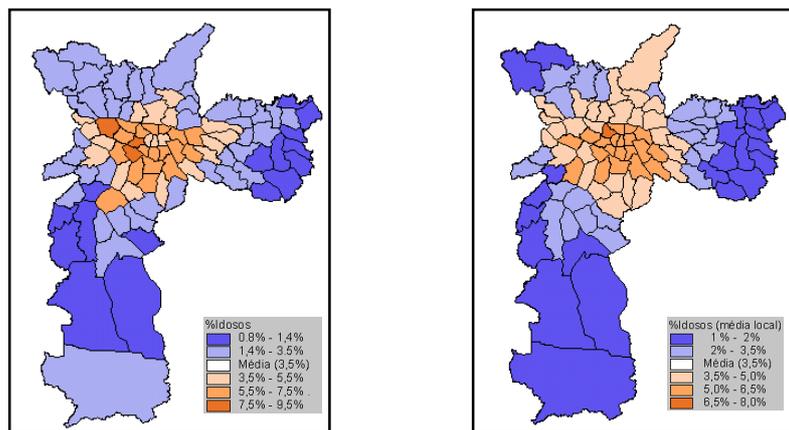


Figura 5-9- Distribuição dos idosos na cidade de São Paulo (censo de 1991). À esquerda, apresentação dos valores por distribuição estatística. À direita, média móvel local.

Indicadores Globais de Autocorrelação Espacial: Índices de Moran e Geary

Um aspecto fundamental da análise exploratória espacial é a caracterização da dependência espacial, mostrando como os valores estão correlacionados no espaço. Neste contexto, as funções utilizadas para estimar quanto o valor observado de um atributo numa região é dependente dos valores desta mesma variável nas localizações vizinhas são a *autocorrelação espacial* e o *variograma*. O índice global de Moran I , é a expressão da autocorrelação considerando apenas o primeiro vizinho:

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (z_i - \bar{z})(z_j - \bar{z})}{\sum_{i=1}^n (z_i - \bar{z})^2} \quad (5.2.)$$

Na equação acima, n é o número de áreas, z_i o valor do atributo considerado na área i , \bar{z} é o valor médio do atributo na região de estudo e w_{ij} os elementos da matriz normalizada de proximidade espacial. Neste caso a correlação será computada apenas para os vizinhos de primeira ordem no espaço, conforme estabelecido pelos pesos w_{ij} . O mesmo cálculo feito para matrizes de proximidade de maior ordem permite estimar a função de autocorrelação para cada ordem de vizinhança (ou “lag”).

$$I^{(k)} = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij}^{(k)} (z_i - \bar{z})(z_j - \bar{z})}{\sum_{i=1}^n (z_i - \bar{z})^2} \quad (5.3.)$$

De uma forma geral, o índice de Moran presta-se a um teste cuja hipótese nula é de independência espacial; neste caso, seu valor seria zero. Valores positivos (entre 0 e +1) indicam para correlação direta e negativos, (entre 0 e -1) correlação inversa. Uma vez calculado, é importante estabelecer sua validade estatística. Em outras palavras, será que os valores medidos representam correlação espacial significativa? Para estimar a significância do índice, será preciso associar a este uma distribuição estatística, sendo mais usual relacionar a estatística de teste à distribuição normal. Outra possibilidade, sem pressupostos em relação à distribuição, e abordagem mais comum é um *teste de pseudo-significância*. Neste caso, são geradas diferentes permutações dos valores de atributos associados às regiões; cada permutação produz um novo arranjo espacial, onde os valores estão redistribuídos entre as áreas. Como apenas um dos arranjos corresponde à situação observada, pode-se construir uma distribuição empírica de I , como mostrado na Figura 5-10. Se o valor do índice I medido originalmente corresponder a um “extremo” da distribuição simulada, então trata-se de valor com significância estatística.

No caso do índice exclusão/inclusão social em São Paulo, apresentado na Figura 5-1, o índice global de Moran medido é 0,642. Uma pseudo-distribuição com 100 valores está mostrada na Figura 5-10. Neste caso, o valor de significância associado é de 5,23, o que nos leva a rejeitar a hipótese nula (não correlação entre as regiões), com significância de 99,5%. Pode-se dizer então que a exclusão social em São Paulo apresenta forte estrutura espacial,

parte variação ampla, ou tendência, parte dependência espacial entre vizinhos.

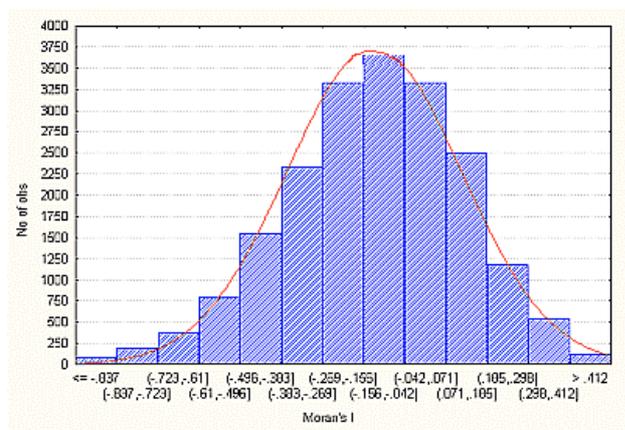


Figura 5-10– Exemplo de distribuição simulada para o índice de Moran.

A hipótese implícita do cálculo do índice de Moran é a estacionariedade de primeira e segunda ordem, e o índice perde sua validade ao ser calculado para dados não estacionários. Quando existir não-estacionariedade de primeira ordem (tendência), os vizinhos tenderão a ter valores mais parecidos que áreas distantes, pois cada valor é comparado à média global, inflacionando o índice. Da mesma forma, se a variância não é constante, nos locais de maior variância o índice será mais baixo, e vice-versa. Quando o dado é não-estacionário, a função de autocorrelação continua decaindo mesmo após ultrapassar a distância onde há influências locais. Algumas variações deste modelo são o teste *C* de Geary e o teste *Ipop*. O primeiro (*C* de Geary) difere do teste *I* de Moran por utilizar a diferença entre os pares, enquanto que Moran utiliza a diferença entre cada ponto e a média global. Assim, o indicador *C* de Geary assemelha-se ao variograma, e o *I* de Moran ao correlograma.

$$C = \frac{(n-1) \sum_{i=1}^n \sum_{j=1}^n w_{ij} (z_i - z_j)^2}{\sum_{i=1}^n \sum_{j=1}^n w_{ij} \sum_{i=1}^n z_i^2} \quad (5.4.)$$

O teste *Ipop* também é utilizado para detectar desvios de uma distribuição espacial aleatória, porém incorpora a variação da população nas áreas. Assim, é sensível à ocorrência de aglomerado intra-área – ou seja, a ocorrência de elevado número de casos numa pequena população de um único município – além dos aglomerados entre áreas, onde municípios com muitos casos são adjacentes. Portanto o índice *Ipop* pode ser decomposto em

um componente intra-áreas e outro inter-áreas, que podem ser apresentados sob forma percentual nos resultados. A hipótese nula (H_0) assume que a variação geográfica do número de casos segue a variação geográfica do tamanho da população, sendo particularmente útil quando a população das áreas não é homogênea.

$$I_{pop} = \frac{N^2 \sum_{i=1}^m \sum_{j=1}^m w_{ij}(e_i - d_i)(e_j - d_j) - N(1 - 2\bar{b}) \sum_{i=1}^m w_{ij}e_i - N\bar{b} \sum_{i=1}^m w_{ii}d_i}{\left(X^2 \sum_{i=1}^m \sum_{j=1}^m d_i d_j w_{ij} - X \sum_{i=1}^m d_i w_{ii} \right) \bar{b}(1 - \bar{b})} \quad (5.5.)$$

- onde:
- \tilde{a} → Número de áreas
 - k → Número total de casos em todas as áreas.
 - $\hat{a}_{\tilde{a}}$ → Número de casos na área \tilde{a}
 - $\hat{E}_{\tilde{a}}$ → Proporção de casos na área \tilde{a} $\hat{E}_{\tilde{a}} = \frac{\hat{a}_{\tilde{a}}}{k}$
 - u → População total em todas as áreas
 - $\hat{n}_{\tilde{a}}$ → Tamanho da população na área \tilde{a}
 - $\hat{C}_{\tilde{a}}$ → Proporção de população na área \tilde{a} $\hat{C}_{\tilde{a}} = \frac{\hat{n}_{\tilde{a}}}{u}$
 - $w_{\tilde{a}}$ → Diferença entre a taxa $u_{\tilde{a}}$ e a média de u
 - $\hat{i}_{\tilde{a}}$ → Pesos atribuídos conforme a conexão entre as áreas $\hat{i}_{\tilde{a}} = \frac{\hat{a}_{\tilde{a}}}{k}$
 - \hat{A} → Prevalência média $\hat{A} = \frac{k}{u}$

A tabela 5.2 apresenta os resultados dos testes de aglomerado espacial para a mortalidade por homicídios no Estado do Rio. Observe que o grau de significância do teste I_{pop} é maior que o Moran, e que aproximadamente metade da agregação deve-se a fatores intra-municipais. Ou seja, além de municípios próximos apresentarem padrões semelhantes, existe um excesso de casos dentro dos municípios violentos, que ultrapassa o esperado em função da população.

TABELA 5.2

RESULTADOS DOS TESTES DE AGLOMERADOS ESPACIAIS:
HOMICÍDIOS NO RIO DE JANEIRO, 90-92

	Moran I	Ipop
Indicador	0,5861	0,00015
p-valor	7,5091	88,9238
% entre áreas	-	54,3
% intra áreas	-	45,7

Variograma

De maneira análoga ao apresentado no capítulo 3, podemos utilizar o variograma como indicador da dependência espacial. Para tanto, associamos o valor único do atributo de cada área a um ponto, usualmente o centro geométrico ou populacional do polígono. Com base nestas localizações, calcula-se a função variograma. Note-se quando o dado é não-estacionário, também o variograma não se estabiliza, mas continua crescendo sempre com a distância. Como exemplo de uso do variograma para dados de área, a Figura 5-11 ilustra o Índice de Desenvolvimento Humano – IDH – para o estado de São Paulo, calculado pelo IPEA, com base no censo de 1991. A Figura 5-12 apresenta o variograma do IDH, computado a partir dos centróide de cada município.

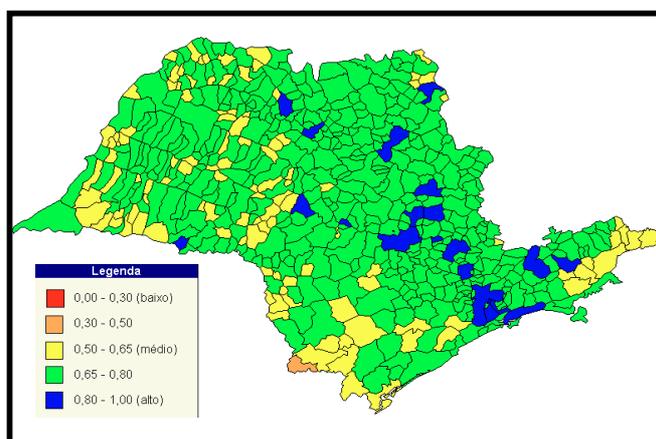


Figura 5-11- IDH para São Paulo (censo de 1991)

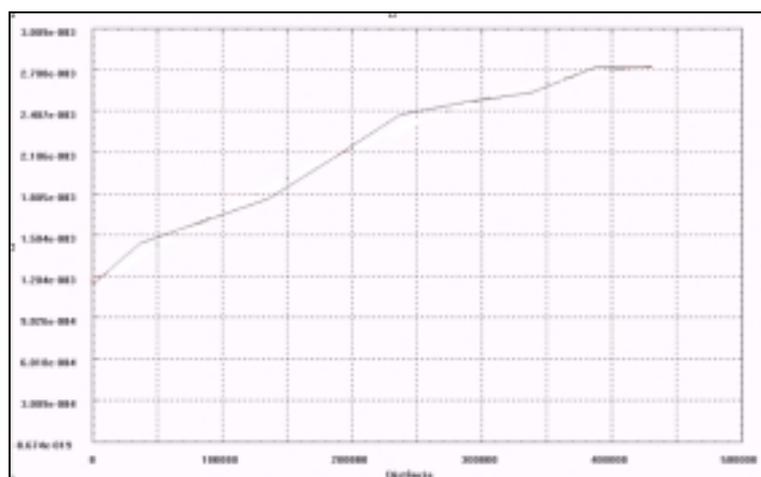


Figura 5-12 Variograma experimental do IDH para São Paulo (censo de 1991). Passo de amostragem: 40 km (tolerância : 20 km).

O que mostra o variograma da Figura 5-10? No eixo dos X, apresentam-se as distâncias entre os municípios, e no eixo Y, a média do quadrado das diferenças do IDH, para municípios separados por faixas de distância, com intervalos de 40 km e tolerância de 20 km. Assim, o primeiro ponto calcula a diferença de IDH entre os municípios cuja distância entre os centros seja de 20 a 60 Km, e assim por diante, até a distância de 400 km. O gráfico evidencia uma forte dependência espacial entre os indicadores de qualidade de vida dos municípios de São Paulo. Trata-se de um resultado dos processos de ocupação do estado, que seguiram perspectivas regionais. A partir da lógica de expansão do café do século XIX, observa-se hoje uma região de forte produção agrícola situada ao longo do eixo da rodovia Anhanguera, a predominância da pecuária na região do Oeste Paulista, e uma forte concentração industrial na região metropolitana de São Paulo, no ABC e no médio Vale do Paraíba. Assim, todos os processos históricos apontam para uma dependência espacial no desenvolvimento econômico no estado.

Para considerar um exemplo adicional, considere-se o estudo sobre mortalidade por homicídios na região Sudeste, que são a causa de mais de 20% dos óbitos dos homens entre 15 e 45 anos no Brasil. A Figura 5-13 ilustra a distribuição espacial da mortalidade por homicídios, usando como indicador o logaritmo do coeficiente de mortalidade específico, por 100.000 residentes do mesmo grupo etário. Entendendo o processo da violência como o de uma "epidemia" da modernidade, que se "propaga" no espaço, uma simples observação visual permite identificar uma elevada ocorrência de mortes violentas no RJ, com uma tendência espacial capital-interior. No caso de ES e SP, há uma concentração próxima da capital e grandes cidades. No entanto, em MG, as áreas mais violentas situam-se longe das regiões metropolitanas, o que indica um padrão espacial distinto. Adicionalmente, há uma marcada transição na fronteira entre MG e RJ, indicando uma mudança nas condições de disseminação da "epidemia da violência". Cabe lembrar que foi utilizado o logaritmo do indicador, dado ser a distribuição do mesmo bastante concentrada em torno de valores muito baixos, com uma grande cauda a direita.

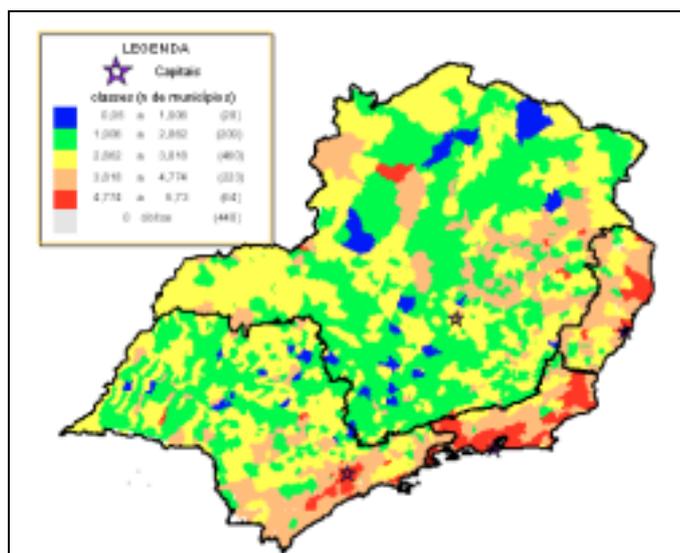


Figura 5-13 - Mortalidade por homicídios, região Sudeste do Brasil

O correlograma da Figura 5-14 apresenta a autocorrelação espacial entre os municípios de cada estado, expressa através da função definida pela equação 5.3. O gráfico indica a existência de uma forte tendência espacial no RJ, pois a função de autocorrelação não se estabiliza com a distância, mas continua decrescente, ao contrário de MG, que não apresenta dependência espacial marcante. Em outras palavras, no RJ, se o município vizinho ao seu é violento, é altamente provável que a sua cidade também o seja; todo o estado apresenta uma estrutura de violência regionalizada, e a violência decai no interior do estado. Em MG, este padrão não é observado: a violência parece flutuar aleatoriamente.

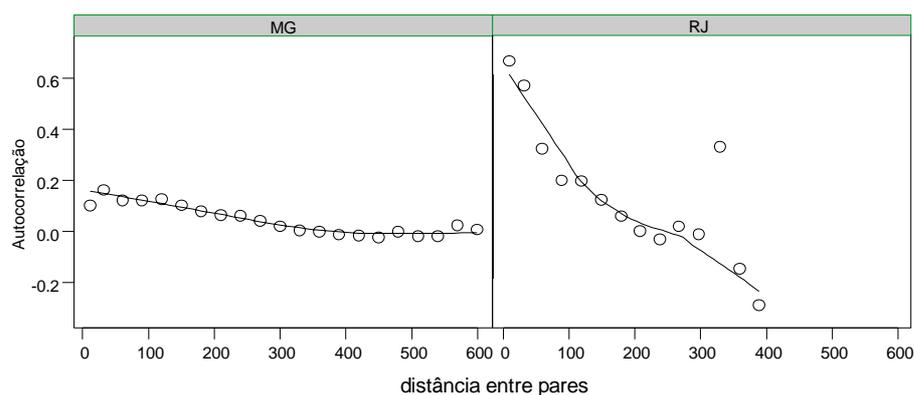


Figura 5-14. Correlograma da mortalidade por homicídios nos estados do Sudeste.

Diagrama de Espalhamento de Moran

O diagrama de espalhamento de Moran é uma maneira adicional de visualizar a dependência espacial. Construído com base nos valores

normalizados (valores de atributos subtraídos de sua média e divididos pelo desvio padrão), permite analisar o comportamento da variabilidade espacial. A idéia é comparar os valores normalizados do atributo numa área com a média dos seus vizinhos, construindo um gráfico bidimensional de z (valores normalizados) por wz (média dos vizinhos), que é dividido em quatro quadrantes, como mostrado na Figura 5-15 para o índice de exclusão/inclusão social de São Paulo, censo de 1991. Os quadrantes podem ser interpretados como:

- Q1 (valores positivos, médias positivas) e Q2 (valores negativos, médias negativas): indicam pontos de associação espacial positiva, no sentido que uma localização possui vizinhos com valores semelhantes.
- Q3 (valores positivos, médias negativas) e Q4 (valores negativos, médias positivas): indicam pontos de associação espacial negativa, no sentido que uma localização possui vizinhos com valores distintos.

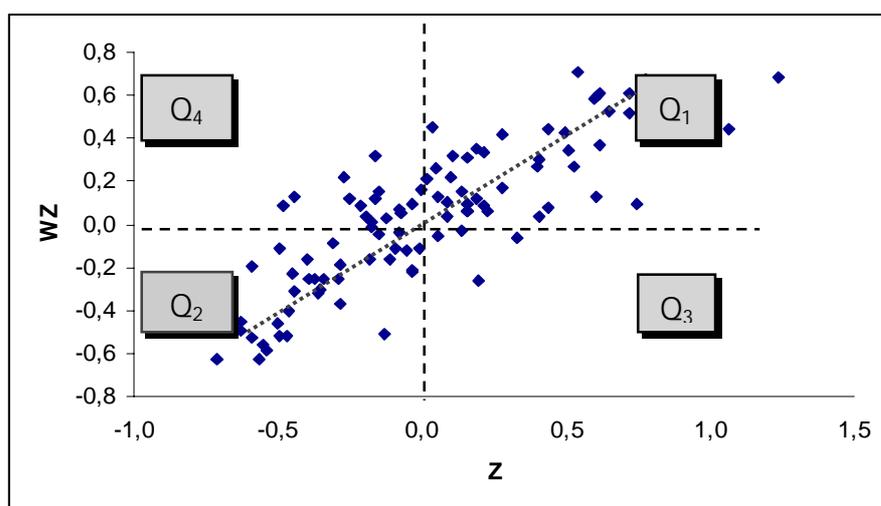
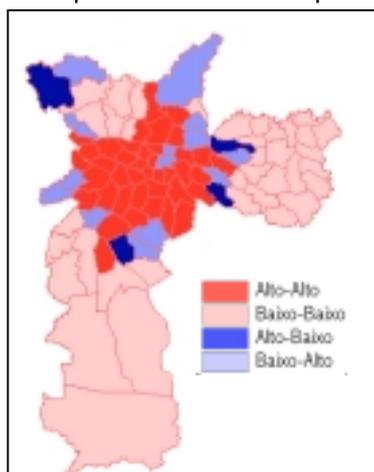


Figura 5-15 Diagrama de Espalhamento de Moran para o índice de exclusão/inclusão social de São Paulo, censo de 1991.

O diagrama de espalhamento de Moran corrobora os resultados apresentados, onde indicamos que o índice global de Moran para o indicador de exclusão/inclusão social para os distritos de São Paulo apresentava valor estatisticamente significativo. Como mostrado na Figura 5-15, a maior parte dos distritos de São Paulo está localizado nos quadrantes Q1 e Q2, que apresentam associação espacial positiva. Os pontos localizados nos quadrantes Q3 e Q4 podem ser vistos como regiões que não seguem o mesmo processo de dependência espacial das demais observações. Evidentemente, o diagrama reflete a estrutura espacial nas duas escalas de análise: vizinhança e tendência.

O índice de Moran I é equivalente ao coeficiente de regressão linear que indica a inclinação da reta de regressão (α) de wz em z . Para o caso dos dados apresentados na Figura 5-15, este coeficiente é igual a 0,642, o mesmo valor calculado aplicando-se a fórmula da equação 5.3. O diagrama de espalhamento de Moran também pode ser apresentado na forma de um mapa temático bidimensional, no qual cada polígono é apresentado indicando-se seu quadrante no diagrama de espalhamento, como ilustra a Figura 5-16, em que mostramos o mapa do espalhamento do índice de Moran para o índice de exclusão/inclusão social da cidade de São Paulo em 1991. Nesta figura, “Alto-Alto”, “Baixo-Baixo”, “Alto-Baixo” e “Baixo-Alto” indicam, respectivamente, os quadrantes Q1, Q2, Q3 e Q4, mostrados na Figura 5-15. Nota-se uma forte polarização centro-periferia e observa-se que os distritos localizados nos quadrantes Q3 e Q4 (indicados pela cor azul) podem ser entendidos como regiões de transição entre o centro da cidade (que tende a apresentar valores positivos do índice de exclusão/inclusão social) e as duas grandes periferias de São Paulo (zona Sul e zona Leste).

Figura 5-16 Mapa de Espalhamento de Moran para o índice de exclusão/inclusão



social da cidade de São Paulo, censo 1991

Indicadores Locais de Associação Espacial

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

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

$$I_i = \frac{z_i \sum_{j=1}^n w_{ij} z_j}{\sum_{j=1}^n z_j^2} \quad (5.6.)$$

A significância estatística do uso do índice de Moran local é computada de forma similar ao caso do índice global. Para cada área, calcula-se o índice local, e depois permuta-se aleatoriamente o valor das demais áreas, até obter uma pseudo-distribuição para a qual possamos computar os parâmetros de significância. Uma vez determinada a significância estatística do índice local de Moran, é útil gerar um mapa indicando as regiões que apresentam correlação local significativamente diferente do resto do dados. Estas regiões podem ser vistas como "bolsões" de não-estacionariedade, pois são áreas com dinâmica espacial própria e que merecem análise detalhada. Para o caso do índice de exclusão/inclusão social da cidade de São Paulo (censo de 1991), esse mapa (Figura 5-17) mostra claramente os agregados de pobreza e de riqueza na cidade. Na zona Leste e na zona Sul de São Paulo há regiões críticas, onde o agravamento das condições sociais resulta numa degradação significativa das condições de vida.

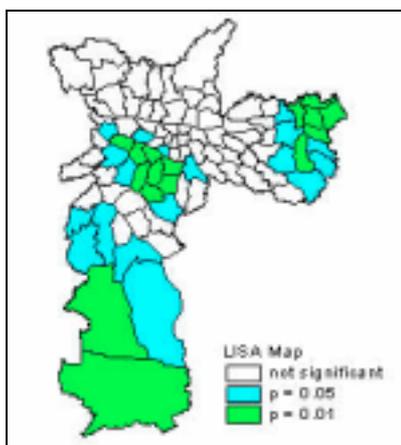


Figura 5-17 – Indicador de autocorrelação espacial para o índice de exclusão/inclusão social de São Paulo (censo de 1991). Apenas os valores com significância maior que 95% estão mostrados.

5.5 ESTIMAÇÃO DE INDICADORES:

A seção 5.3 apresentou o problema de agregação de contagens em áreas, com a recomendação final de utilizar a melhor resolução espacial disponível. Na prática, o uso desta estratégia requer um tratamento adicional nos dados, principalmente nos casos de pequenas áreas em que calculamos taxas sobre um universo populacional reduzido. Para entender melhor o problema, considere-se a Figura 5-18 que apresenta um mapa temático com a mortalidade infantil dos bairros do Rio de Janeiro, em 1994. Neste mapa, o Rio está dividido em 148 bairros, e a taxa de mortalidade infantil anual para cada bairro, expressa o número de óbitos de menores de 1 ano, por mil nascidos vivos.

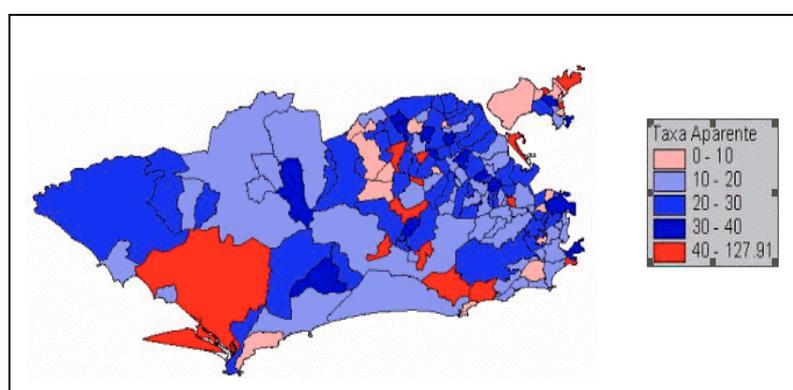


Figura 5-18 - Taxa total de mortalidade infantil por mil nascidos vivos no Rio de Janeiro, em 1994.

Numa primeira leitura, este mapa choca pelas altas taxas de mortalidade de vários bairros, com 15 bairros apresentando uma taxa maior que 40 óbitos por mil nascidos, e 2 casos com taxas acima de 100 por mil nascidos. Um observador desatento poderia concluir que todos estes bairros apresentam um grave problema social. Na realidade, muitos destes valores extremos ocorrem nos bairros com pequenas populações, pois a divisão da cidade utilizada esconde enormes diferenças na população em risco, variando de 15 até 7500 crianças por bairro. Por exemplo, considere uma região com 15 crianças nascidas e nenhuma morte, o que aparentemente indicaria uma situação ideal. Se apenas uma criança morre neste ano, a taxa passa de 0 por mil para 66 por mil !

Tais problemas são típicos de recobrimentos espaciais sobre divisões político-administrativas, onde se analisam áreas com valores muito distintos de população em risco. Vários estudos têm mostrado que em divisões políticas como bairros e municípios apresentam relações inversas de área e população, isto é, os maiores bairros em população tendem a ter menores

áreas, e vice-versa. Por isso mesmo, freqüentemente o que mais chama a atenção num mapa temático de taxas, que são os valores extremos, muitas vezes são resultado de um número reduzidíssimo de observações, sendo portanto menos confiável, ou seja, apenas flutuação aleatória .

Para suavizar a flutuação aleatória, considera-se que a taxa estimada pela divisão simples entre contagem de óbitos e de população – taxa observada – é apenas **uma** realização de um processo não observado, e que é tanto menos confiável quanto menor a população. Assim, propõe-se re-estimar uma taxa mais próxima do risco real ao qual a população está exposta. A primeira providência é fazer um gráfico que expresse a taxa em função da população em risco, como mostrado na Figura 5-19.

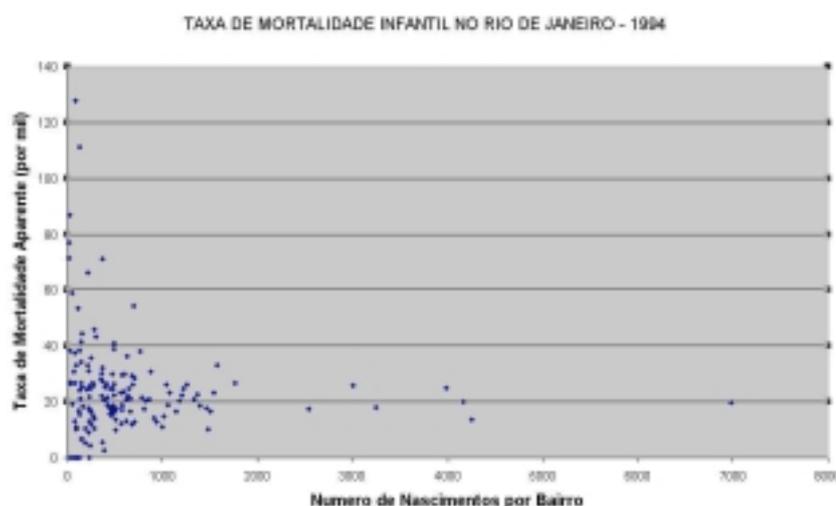


Figura 5-19 Taxa de mortalidade infantil no Rio de Janeiro em 1994 em função do número de nascimentos por bairro.

No caso do Rio, a taxa média de mortalidade infantil da cidade, em 1994, foi de 21 óbitos por mil nascidos. Neste gráfico, observa-se que os bairros com maior população apresentam taxas próximas da média da cidade. Conforme diminui a população em risco, aumenta muito a flutuação da taxa medida, formando o que já foi denominado de “efeito funil”. Nos bairros de menor população, esta variação oscilou de 0 a quase 130 por mil.

É razoável supor que as taxas das diferentes regiões estão autocorrelacionadas, e levar em conta o comportamento dos vizinhos para estimar uma taxa mais realista para as regiões de menor população. Esta formulação sugere o uso de técnicas de estimação bayesiana. Nesse contexto, considera-se que a taxa “real” θ_i associada a cada área não é conhecida, e dispomos de uma taxa observada $t_i = z_i/n_i$, onde n_i é o número de pessoas observadas, z_i é o número de eventos na i -ésima área.

A idéia do estimador bayesiano é supor que a taxa θ_i é uma variável aleatória, que possui uma média μ_i e uma variância σ_i^2 . Pode ser demonstrado que o melhor estimador bayesiano é dado por uma combinação linear entre a taxa observada e a média μ_i :

$$\hat{\theta}_i = w_i t_i + (1 - w_i) \mu_i, \quad (5.7.)$$

O fator w_i é dado por:

$$w_i = \frac{\sigma_i^2}{\sigma_i^2 + \mu_i/n_i} \quad (5.8.)$$

O peso w_i é tanto menor quanto menor for a população em estudo da i -ésima área e reflete o grau de confiança a respeito de cada taxa. Para o caso de populações reduzidas, a confiança na taxa observada diminui e a estimativa da taxa se aproxima de nosso modelo a priori (ou seja, se aproxima de μ). Regiões com populações muito baixas terão uma correção maior, e regiões populosas terão pouca alteração em suas taxas. Logo θ_i será estimado, quando n for pequeno, com maior peso da média da vizinhança.

Neste ponto, deve-se observar que a formulação bayesiana requer as médias e variâncias μ_i e σ_i^2 para cada uma das áreas. A abordagem mais simples para tratar a estimação destes parâmetros é o chamado *estimador bayesiano empírico*. Este estimador parte da hipótese que a distribuição da variável aleatória θ_i é a mesma para todas as áreas; isto implica que todas as médias e variâncias são iguais. Pode-se então estimar μ_i e σ_i^2 diretamente a partir dos dados. Neste caso, calcula-se μ_i a partir das taxas observadas:

$$\hat{\mu} = \frac{\sum y_i}{\sum n_i} \quad (5.9.)$$

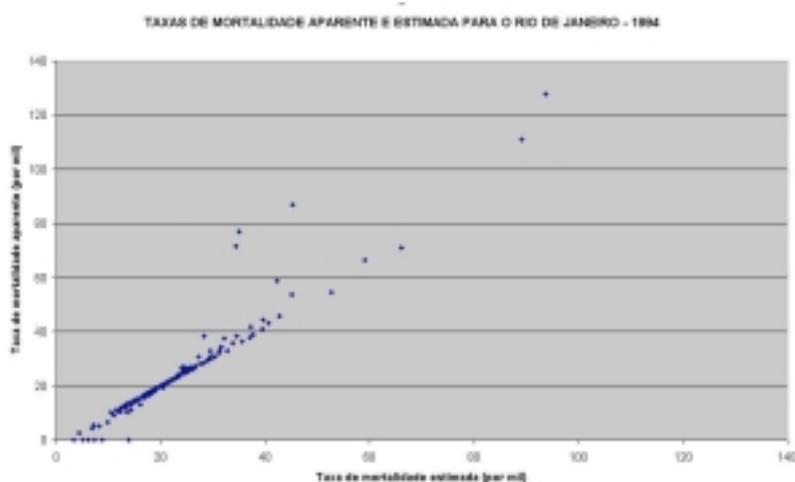
E estima-se a variância σ_i^2 a partir da variância das taxas observadas com relação à média estimada:

$$\sigma^2 = \frac{\sum n_i (t_i - \hat{\mu})^2}{\sum n_i} - \frac{\hat{\mu}}{\bar{n}} \quad (5.10.)$$

As regiões terão suas taxas re-estimadas aplicando-se uma média ponderada entre o valor medido e a taxa média global, em que o peso da média será inversamente proporcional à população da região. Ao aplicarmos esta correção às taxas de mortalidade infantil do Rio de Janeiro, observamos que há uma redução significativa nos valores extremos. Por exemplo, a Cidade Universitária (Ilha do Fundão), onde nasceram 13 crianças em 1994, apresentou uma taxa aparente de 76 por mil nascidos vivos e uma taxa

corrigida de 36 por mil. Bairros com pouca população no grupo de risco apresentaram reduções semelhantes, enquanto que bairros mais populosos mantiveram as taxas originalmente medidas. A comparação entre a taxa primária e o valor estimado está apresentada na Figura 5-18. Em resumo, é preciso extremo cuidado ao produzir mapas temáticos, especialmente em casos onde apresentamos taxas medidas sobre populações com valores reduzidos.

Figura 5-18. Comparação entre a taxa de mortalidade infantil observada e a taxa estimada



pele método bayesiano empírico.

O estimador bayesiano empírico pode ser generalizado para incluir efeitos espaciais. Neste caso, a idéia é fazer a estimativa bayesiana localmente, convergindo em direção a uma média local e não a uma média global. Basta aplicar o método anterior em cada área considerando como “região” a sua vizinhança. Isto é equivalente a supor que as taxas da vizinhança da área i possuem média μ_i e variância σ_i^2 comuns. Neste caso, pode-se falar em *estimativa bayesiana empírica local*. A seguir, apresenta-se a detecção de hanseníase em Recife (Figura 5-20) onde foi utilizado esse método local para estimar a taxa da doença nos bairros da cidade. Através do mapa “corrigido” foi possível indicar bairros prioritários para a atuação da vigilância epidemiológica por apresentarem valores altos mesmo após suavização do indicador.

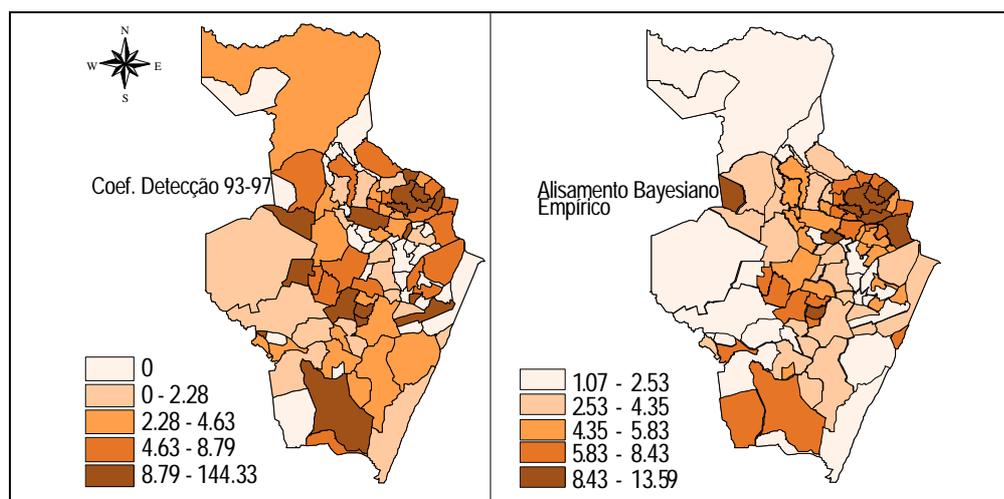


Figura 5-20 - Taxas de detecção média de hanseníase em menores de 15 anos, período 1993-1997, por bairro do Recife, e taxas estimadas através de alisamento bayesiano.

Como apresentado acima, o *estimador bayesiano empírico* parte da hipótese que a distribuição da variável aleatória θ_i é a mesma para todas as áreas e que as médias e variâncias μ_i e σ_i^2 para cada uma das áreas são iguais. Deve-se lembrar que esta hipótese nem sempre é realista, pois em estatísticas socioeconômicas (como no caso dos dados de saúde discutidos) as características das populações estudadas são muito heterogêneas. Deste modo, em muitos casos é desejável fazer a hipótese de que cada área tem seu próprio padrão (e os μ_i e σ_i^2 são distintos); isto implica em estimar a distribuição conjunta $Z = \{Z_1, \dots, Z_n\}$ das variáveis aleatórias.

À primeira vista, a estimativa da distribuição conjunta pode parecer impossível, dado que está disponível para análise apenas uma amostra de cada uma das variáveis aleatórias, ou seja, sabe-se apenas o valor coletado em cada unidade de área. Entretanto, os *estimadores bayesianos completos (full Bayes)* tornaram possível resolver o problema, através da utilização de técnicas de simulação baseadas em MCMC – *Markov Chain Monte Carlo* – para a inferência dos parâmetros de interesse. Em função da complexidade de formulação, este livro não aborda os estimadores bayesianos baseados em MCMC. O leitor deve referir-se à bibliografia no final do capítulo para maiores detalhes.

5.6 MODELOS DE REGRESSÃO

Um dos tipos de estudos mais comuns com dados de área é o uso de modelos de regressão. Um modelo de regressão é uma ferramenta estatística que utiliza o relacionamento existente entre duas ou mais variáveis de maneira que uma delas possa ser descrita ou o seu valor estimado a partir das demais. Na situação dos dados espaciais, quando está presente a autocorrelação espacial, as estimativas do modelo devem incorporar esta estrutura espacial, uma vez que a dependência entre as observações altera o poder explicativo do modelo. A significância dos parâmetros é usualmente superestimada, e a existência de variações em larga escala pode até mesmo induzir a presença de associações espúrias.

Neste livro, não será feita uma descrição detalhada dos modelos tradicionais de regressão, disponível em diversos livros consagrados, mas apenas será apresentado um breve resumo, necessário ao entendimento dos modelos de regressão espacial. O objetivo geral de uma análise de regressão linear é quantificar a relação linear entre uma variável dependente e um conjunto de variáveis explicativas, conforme expresso na equação matricial:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(0, \sigma^2) \quad \text{ou} \quad (5.11.)$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & \dots & X_{1k-1} \\ 1 & X_{21} & \dots & X_{2k-1} \\ \dots & \dots & \dots & \dots \\ 1 & X_{n1} & \dots & X_{nk-1} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_{k-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{bmatrix} \quad (5.12.)$$

onde \mathbf{Y} é a variável dependente, composta de um vetor ($n \times 1$) de observações tomadas em cada um das n áreas, \mathbf{X} é uma matriz ($n \times k$) com $k-1$ variáveis explicativas também tomadas nas n áreas, $\boldsymbol{\beta}$ é vetor ($k \times 1$) com os coeficientes de regressão, e $\boldsymbol{\varepsilon}$ é um vetor ($n \times 1$) de erros aleatórios, ou resíduos.

Tipicamente, quando se faz uma análise de regressão, procura-se alcançar dois objetivos: (a) encontrar um bom ajuste entre os valores preditos pelo modelo e os valores observados da variável dependente; (b) descobrir quais das variáveis explicativas contribuem de forma significativa para este relacionamento linear. Para tanto, a hipótese padrão é que as observações não são correlacionadas, e, conseqüentemente, que os resíduos ε_i do modelo também são independentes e não-correlacionados com a variável dependente, tem variância constante, e apresentam distribuição normal com média zero.

No entanto, no caso de dados espaciais, onde está presente a dependência espacial, é muito pouco provável que a hipótese padrão de observações não correlacionadas seja verdadeira. No caso mais comum os resíduos continuam apresentando a autocorrelação espacial presente nos dados, que pode se manifestar por diferenças regionais sistemáticas nas relações do modelo, ou ainda por uma tendência espacial contínua.

A investigação dos resíduos da regressão em busca de sinais de estrutura espacial é o primeiro passo em uma regressão espacial. As ferramentas usuais de análise gráfica e o mapeamento de resíduos, podem dar as primeiras indicações de que os valores observados estão mais correlacionados do que seria esperado sob uma condição de independência. Neste caso, utilizar os testes de autocorrelação espacial – Moran e Geary – nos resíduos da regressão informa sobre sua presença. Em caso de existir autocorrelação, deve-se especificar um modelo que considere a interferência causada pela mesma.

No restante desta seção, apresentamos vários tipos de modelos de regressão que permitem incorporar efeitos espaciais, desde aqueles que tratam a estrutura espacial de forma global (como um único parâmetro) até modelos em que os parâmetros variam continuamente no espaço.

Modelos com Efeitos Espaciais Globais

A inclusão explícita de efeitos espaciais em modelos de regressão pode ser feita de diferentes formas. A classe de modelos de regressão espacial mais simples, chamados de *modelos com efeitos espaciais globais*, supõe que é possível capturar a estrutura de correlação espacial num único parâmetro, que é adicionado ao modelo de regressão tradicional. Neste caso, tem-se duas alternativas para tratar a autocorrelação global em um modelo de regressão. Na primeira, a autocorrelação espacial ignorada é atribuída à variável dependente Y . Esta abordagem é denominada como *modelo espacial autoregressivo misto* (“*Spatial AutoRegressive– SAR*” ou ainda como “*spatial lag model*”), dado que se considera a dependência espacial através da adição ao modelo de regressão de um novo termo na forma de uma relação espacial para a variável dependente. Formalmente isto é expresso como:

$$Y = \rho WY + X\beta + \varepsilon, \quad (5.13.)$$

onde W é a matriz de proximidade espacial, e o produto WY expressa a dependência espacial em Y e ρ é o *coeficiente espacial autoregressivo*. A hipótese nula para a não existência de autocorrelação é que $\rho = 0$. A idéia básica neste modelo é incorporar a autocorrelação espacial como componente do modelo. Em termos de componentes individuais, este modelo pode ser expresso como

$$y_i = \rho \left(\sum_j w_{ij} y_j \right) + \sum_{i=1} x_i \beta_i + \varepsilon_i \quad (5.14.)$$

O segundo tipo de modelo de regressão espacial com parâmetros globais considera que os efeitos espaciais são um ruído, ou perturbação, ou seja, fator que precisa ser removido. Neste caso, os efeitos da autocorrelação espacial são associados ao termo de erro ε e o modelo pode ser expresso por:

$$Y = X\beta + \varepsilon, \quad \varepsilon = \lambda W + \xi, \quad (5.15.)$$

onde $W\varepsilon$ é a componente do erro com efeitos espaciais, λ é o coeficiente autoregressivo e ξ é a componente do erro com variância constante e não correlacionada. A hipótese nula para a não existência de autocorrelação é que $\lambda = 0$, ou seja, o termo de erro não é espacialmente correlacionado. Este modelo é também chamado de modelo do *erro espacial* (“*spatial error model*” ou ainda “*Conditional AutoRegressive*” - CAR).

A partir da equação 5.15, pode-se mostrar que o modelo de erro espacial pode também ser expresso como:

$$Y - \lambda WY = X\beta - \lambda WX\beta + \xi \quad (5.16.)$$

ou ainda como

$$(I - \lambda W)Y = (I - \lambda W)X\beta + \xi \quad (5.17.)$$

o que pode ser visto como uma regressão não-espacial nas variáveis “filtradas”

$$Y^* = (I - \lambda W)Y, \quad X^* = (I - \lambda W)X \quad (5.18.)$$

Na prática, a distinção entre os dois tipos de modelos de regressão espacial com parâmetros globais é difícil pois, apesar da diferença nas suas motivação, eles são muito próximos em termos formais. Estes modelos estão incluídos em ambientes de estatística espacial avançados, como nos softwares SpaceSat™, S-Plus™ e R, esse de domínio público. Nas referências no final do capítulo, o leitor poderá encontrar indicações sobre como tais modelos podem ser estimados e sobre testes de hipóteses sobre seu comportamento.

Os modelos de regressão espacial com efeitos globais partem do princípio de que o processo espacial subjacente aos dados analisados é estacionário. Isto implica que os padrões de autocorrelação espacial existentes nos dados podem ser capturados num único parâmetro. Na prática, para conjuntos de dados censitários de médio e grande porte, a natureza dos processos espaciais é tal que diversos padrões de associação espacial podem estar presentes. Esta hipótese, que pode ser verificada, por

exemplo, pelos indicadores locais de autocorrelação espacial, está na origem aos modelos cujos parâmetros variam no espaço, discutidos a seguir.

Modelos de Regressão com Efeitos Espaciais Locais

(a) Caso Discreto – Modelos de Regressão com Regimes Espaciais

Quando o processo espacial é não-estacionário, os coeficientes de regressão precisam refletir a heterogeneidade espacial. Para tanto, há duas grandes alternativas: (a) modelar a tendência espacial de forma contínua, com parâmetros variantes no espaço; (b) modelar a variação espacial de forma discreta, ao dividir o espaço em sub-regiões estacionárias, chamadas de *regimes espaciais*.

A idéia de *regimes espaciais* é dividir a região de estudo em sub-regiões, cada uma com seu padrão espacial próprio, e realizar regressões em separado, uma para cada região. As observações são classificadas em dois ou mais subconjuntos, a partir de uma variável por indicação, a saber:

$$Y_1 = X_1\beta_1 + \varepsilon_1, \text{ ind} = 1 \quad (5.19.)$$

$$Y_2 = X_2\beta_2 + \varepsilon_2, \text{ ind} = 2 \quad (5.20.)$$

Apesar de cada regime possuir os seus próprios valores de coeficientes, estes valores são estimados conjuntamente, ou seja, todo o conjunto de observações disponível é utilizado na regressão. Para a determinação dos regimes espaciais, as técnicas de análise exploratória apresentadas no início do capítulo são muito úteis, especialmente o mapa de espalhamento de Moran e os indicadores locais de autocorrelação espacial.

Na prática, para os dados sócio-econômicos típicos de cidades brasileiras, o modelo de regimes espaciais tende a apresentar resultados melhores que os modelos de regressão simples ou de regressão espacial com efeitos globais. Isto ocorre em função das fortes desigualdades sociais no Brasil, que ocasionam descontinuidades abruptas nos fenômenos estudados, como no caso do recorte entre favelas e áreas ricas, como é freqüente nas em nossas grandes cidades.

Modelos de Regressão com Efeitos Espaciais Locais

(b) Modelos de Regressão com Efeitos espaciais contínuos

Esta classe de modelos procura modelar fenômenos não-estacionários. Diferentemente do modelo por regimes espaciais, os efeitos espaciais são modelados de forma contínua, com duas hipóteses: (a) a existência de uma variação suave em larga escala, sem efeitos locais significativos ou (b) a existência de variações locais contínuas, sem uma forte tendência global. O primeiro caso corresponde às *superfícies de tendência*, descritas no capítulo 3 deste livro, resumidas no que segue para conveniência de leitura. O modelo

de *superfícies de tendência* considera um processo espacial onde o valor da variável é uma função polinomial de sua posição no espaço. O modelo de regressão múltipla utilizando notação vetorial é:

$$Y(s) = X(s)\beta + \varepsilon(s) \quad (5.21.)$$

onde, $Y(s)$ → variável aleatória representando o processo no ponto s ,

$X(s)\beta$ → tendência (ou seja, o valor médio $\mu(s)$),

$\varepsilon(s)$ → erro aleatório com média zero e variância σ^2

O vetor $x(s)$ consiste em p funções das coordenadas espaciais (s_1, s_2) , do ponto amostrado s . Para uma superfície de tendência linear é apenas $(1, s_1, s_2)$, para quadrática é $(1, s_1, s_2, s_1^2, s_2^2, s_1 \cdot s_2)$, e assim sucessivamente. β é o vetor $(p+1)$ de parâmetros a ser ajustado. O pressuposto básico deste modelo supõe que os erros têm variância constante e são independentes em cada local, conseqüentemente, a covariância é zero: não há efeitos de segunda ordem presentes no processo. Neste contexto, é feito o ajuste do modelo por mínimos quadrados ordinários. O modelo de *superfícies de tendência* é útil sobretudo como uma primeira aproximação do fenômeno, pois na prática, são limitados os casos em que a variação espacial pode ser expressa desta forma. No entanto, os resíduos destes modelos são muito informativos sobre a natureza das variações locais.

No caso de *modelos de variações locais contínuas*, é idéia é ajustar um modelo de regressão a cada ponto observado, ponderando todas as demais observações como função da distância a este ponto. Desta forma, serão feitos tantos ajustes quantas observações existirem e o resultado será um conjunto de parâmetros, sendo que cada ponto considerado terá seus próprios coeficientes de ajuste. Estes parâmetros podem ser apresentados visualmente para identificar como se comportam espacialmente os relacionamentos entre variáveis. Esta técnica é denominada *geographically weighted regression* (GWR ou regressão ponderada espacialmente). Para aplicar o modelo GWR, o modelo padrão de regressão é reescrito na forma:

$$Y(s) = \beta(s)X + \varepsilon, \quad (5.22.)$$

onde, $Y(s)$ é a variável aleatória representando o processo no ponto s , e $\beta(s)$ indica que os parâmetros são estimados no ponto s . Para estimar os parâmetros deste modelo, a solução padrão por mínimos quadrados para o caso não-espacial, dada por

$$\beta = (X^T X)^{-1} X^T Y \quad (5.23.)$$

é generalizada usando um método de ajuste local:

$$\beta(s) = (X^T W(s) X)^{-1} X^T W(s) Y \quad (5.24.)$$

O ajuste local é feito de forma a garantir uma influência maior dos pontos mais próximos, de forma semelhante aos estimadores de densidade por *kernel*, discutidos no capítulo 2 do livro. Um exemplo é o uso de uma função gaussiana, do tipo

$$w_{ij}(s, \tau) = \frac{1}{2\pi\tau} \exp\left(-\frac{d_{ij}^2}{2\tau^2}\right) \quad (5.25.)$$

onde τ representa o raio de influência considerado, e d_{ij} a distância entre a localização considerada e o j -ésimo ponto. Pode-se fazer testes de hipóteses para verificar se as variações espaciais têm significado estatístico ou são aleatórias. Para maiores detalhes sobre o modelo GWR, o leitor deve referir-se à bibliografia no final do capítulo.

Diagnóstico de Modelos com Efeitos Espaciais

A análise gráfica dos resíduos é o primeiro passo para avaliar a qualidade do ajuste da regressão. Mapear os resíduos é uma etapa importante no diagnóstico do modelo, buscando indícios de ruptura dos pressupostos de independência. Uma alta concentração de resíduos positivos (ou negativos) numa parte do mapa é um bom indicador da presença de autocorrelação espacial. Para um teste quantitativo, o mais comum é utilizar o índice I de Moran sobre os resíduos.

Como os estimadores e os diagnósticos tradicionais de regressão não levam em conta os efeitos espaciais, as inferências, como por exemplo as indicações de qualidade de ajuste baseadas em R^2 (coeficiente de determinação), serão incorretas. Estas consequências são similares às que acontecem quando uma variável explicativa significativa é omitida do modelo de regressão. Quando se quer comparar um ajuste obtido por um modelo de regressão padrão, com um ajuste obtido por um dos modelos cuja especificação considera a autocorrelação espacial, uma medida como o R^2 não é mais confiável.

O método mais usual de seleção de modelos de regressão baseia-se nos valores de *máxima verossimilhança* dos diferentes modelos, ponderando pela diferença no número de parâmetros estimados. Nos modelos com estrutura de dependência – espacial ou temporal – utilizam-se os *critérios de informação* onde a avaliação do ajuste é penalizada por uma função do número de parâmetros. Cabe observar que é necessário ainda levar em conta o número de parâmetros independentes ao se incluir funções espaciais nos modelos. Para cada nova variável em modelo de regressão, acrescenta-se um parâmetro.

Usualmente a comparação de modelos é feita utilizando o logaritmo da máxima verossimilhança, que é o que possui melhor ajuste para os dados observados. O critério de informação de Akaike (AIC) é expresso por:

$$AIC = -2 * LIK + 2k \quad (5. 26.)$$

onde LIK é o log de verossimilhança maximizado e k é o número de coeficientes de regressão. Segundo este critério, o melhor modelo é o que possui menor valor de AIC. Diversos outros critérios de informação estão disponíveis, a maior parte dos quais são variações do AIC, com mudanças na forma de penalização de parâmetros ou observações.

Exemplo Ilustrativo

Como exemplo ilustrativo das técnicas de regressão espacial, estudou-se o relacionamento entre renda e longevidade na cidade de São Paulo, para os dados do Censo de 1991. Tratam-se de duas das três variáveis utilizadas para compor o IDH (índice de desenvolvimento humano) da ONU. A variável dependente a ser explicada é denotada por PERIDOSO (percentual de pessoas com mais de 70 anos por distrito de São Paulo) e a variável independente é indicada por PERREN20 (percentual de chefes de família com renda de mais de 20 salários mínimos mensais). A distribuição espacial destas variáveis está mostrados na Figura 5-21.

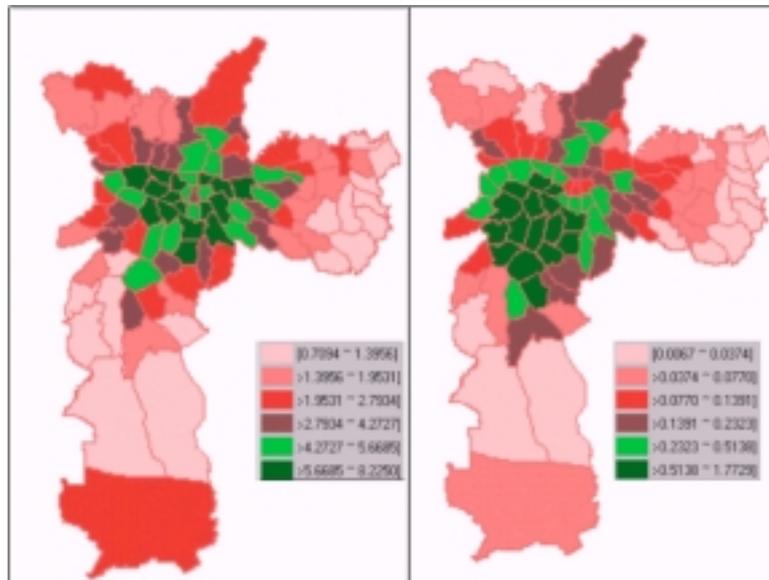


Figura 5-21. Percentual de idosos (à esquerda) e de chefes de família com renda maior que 20 SM mensais (à direita) para os distritos de São Paulo (1991).

Foram comparados três modelos de regressão: o modelo padrão não-espacial, o modelo autoregressivo (*spatial lag*) e o modelo em regimes espaciais. No caso dos regimes espaciais foram consideradas três regiões da

cidade (centro, periferia e a transição centro-periferia). O modelo padrão é expresso como:

$$\text{PERIDOSO} = \beta_0 + \beta_1 \text{PERREN20} + \varepsilon \quad (5. 27.)$$

Utilizando-se a matriz de vizinhança W dos distritos, o modelo “spatial lag” pode ser expresso como:

$$\text{PERIDOSO} = \beta_0 + \beta_1 \text{PERREN20} + \rho W(\text{PERIDOSO}) + \varepsilon \quad (5. 28.)$$

Considerando-se três regiões da cidade, o modelo de regimes espaciais pode ser expresso como

$$\text{PERIDOSO_1} = \beta_0^1 + \beta_1^1 \text{PERREN20_1}, \text{ reg}= 1 \quad (5. 29.)$$

$$\text{PERIDOSO_2} = \beta_0^2 + \beta_1^2 \text{PERREN20_2}, \text{ reg}= 2 \quad (5. 30.)$$

$$\text{PERIDOSO_3} = \beta_0^3 + \beta_1^3 \text{PERREN20_3}, \text{ reg}= 3 \quad (5. 31.)$$

Os resultados destes modelos de regressão são apresentados na Tabela 5-3. No modelo de regressão tradicional, a relação entre renda e longevidade em São Paulo é muito reduzida, o que dá suporte a idéia do IDH de que tratam-se de dimensões complementares da desenvolvimento humano. No entanto, quando os efeitos espaciais são levados em conta, verifica-se que a existência de real dependência entre os dois fatores. Na Figura 5-22, apresenta-se a distribuição espacial dos resíduos da regressão para os modelos de mínimos quadrados e *spatial lag*. Uma análise visual dos resíduos da regressão tradicional indica uma prevalência de resíduos positivos no centro da cidade e resíduos negativos na periferia, principalmente nas Zonas Leste e Sul. Os resultados numéricos confirmam esta análise, pois o índice de Moran dos resíduos é altamente significativo. Com relação ao desempenho global, as medidas R^2 são indicadores limitados e devem ser encaradas com cuidados, e deve-se preferir as medidas baseadas em verossimilhança (LIK, AIC). Neste caso, o modelo *spatial lag* teve um desempenho muito superior ao modelo padrão. Este efeito é esperado, pela existência de um índice de Moran significativo nos resíduos, que é capturado no coeficiente de efeito espacial (ρ).

Os regimes espaciais escolhidos para São Paulo são mostrados na Figura 5-23, bem como os resíduos da regressão considerando estes regimes. Da análise visual dos resíduos, verifica-se a não-existência de forte tendência espacial, o que é evidenciado pelo baixo índice de Moran dos mesmos, indicado na Tabela 5-3. No geral, o modelo de regimes espaciais apresentou o melhor desempenho, por qualquer dos critérios (R^2 , LIK e AIC). O resultado reflete a forte polarização centro-periferia da cidade de São Paulo, e é compatível com estudos que mostram os resultados da violência urbana nas taxas de mortalidade, especialmente de homens dos 15 aos 25 anos.

Tabela 5-3

Resultados da Regressão para Longevidade e Renda em São Paulo, 1991

	Regressão MMQ	Spatial Lag	Regimes Espaciais
R ² ajustado	0,280	0,586	0,80
Log verossimilhança	-187,92	-150,02	-124,04
AIC (Critério de Inf. Akaike)	379,84	306,51	260,09
Índice de Moran dos resíduos	0,620	-	0,020

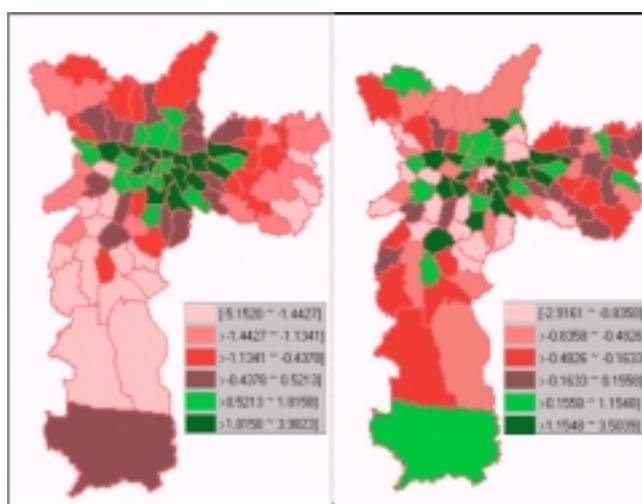


Figura 5-22- Resíduos da regressão por mínimos quadrados (à esquerda) e resíduos da regressão com o modelo *spatial lag* (à direita).

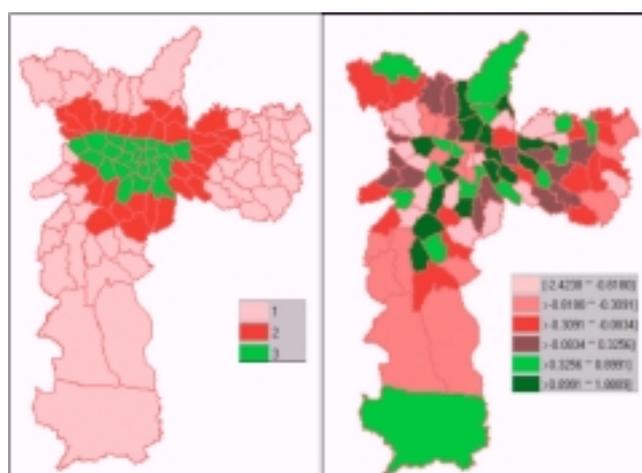


Figura 5-23 Regimes espaciais para os distritos de São Paulo (à esquerda) e resíduos da regressão por regimes espaciais (à esquerda).

5.7 ESTIMAÇÃO DE MODELOS CONTÍNUOS A PARTIR DE DADOS DE ÁREA

As seções anteriores apresentaram técnicas de análise espacial de dados de área tomando por base o modelo de *variação espacial discreta*, onde cada área é modelada respeitando seus limites, adjacências e vizinhança. Nesta seção, considera-se o modelo de *variação espacial contínua*, que supõe um processo estocástico $\{Z(x), x \in A, A \subset \mathfrak{R}^2\}$, cujos valores podem ser conhecidos em todos os pontos da área de estudo. A idéia de modelos contínuos para dados socioeconômicos decorre do fato que os levantamentos censitários muitas vezes impõem limites de áreas a partir de critérios puramente operacionais, que não têm relação direta com o fenômeno modelado. Este fato leva à idéia de dissolver os limites das áreas em superfícies contínuas, de forma a modelar melhor a real continuidade de, por exemplo, setores censitários em regiões urbanas densamente povoadas.

No caso de estimadores de superfícies, as principais alternativas são o uso de técnicas não-paramétricas e o uso de interpoladores geoestatísticos, descritos nos capítulos 3 deste livro e que são brevemente resumidos no que segue.

Estimador de Intensidade Não-Paramétrico

De forma similar como no caso de superfícies, podemos utilizar o estimador de intensidade (*kernel estimator*) para nos fornecer uma primeira aproximação da distribuição espacial do fenômeno ou variável. Neste caso, quando os valores observados representam uma medida “média” como taxa de mortalidade ou renda per capita, podemos utilizar um estimador que nos permitiria calcular o valor do atributo por unidade de área. Para toda posição $(\mathbf{x}; \mathbf{y})$ cujo valor queremos estimar, o estimador de intensidade será computado a partir dos valores $\{z_1, \dots, z_n\}$ contidos num raio de tamanho τ , a partir da equação

$$\hat{z}_i = \frac{\sum_{j=1}^n k\left(\frac{d_{ij}}{\tau}\right) z_j}{\sum_{j=1}^n k\left(\frac{d_{ij}}{\tau}\right)}, \quad d_{ij} \leq \tau \quad (5. 32.)$$

Na equação acima, a função $k()$ é um interpolador não-paramétrico, que pode ser, por exemplo, um *kernel* gaussiano, como apresentado nos capítulos 2 e 3 deste livro, onde o leitor poderá encontrar uma discussão mais aprofundada sobre os estimadores de intensidade não-paramétricos. Um exemplo do estimador de intensidade para taxas pode ser visto na Figura 5-22, onde são apresentados os dados de mortalidade por homicídios para o Estado do Rio de Janeiro, para o triênio 90-92 interpolados pelo estimador

de intensidade, que nos dá uma idéia da distribuição espacial da variável estudada. Na Figura 5-24(a) é apresentado um mapa com os valores de indicadores de taxa de mortalidade, agregados por município. Na Figura 5-24(b), apresentamos o resultado do estimador de intensidade, que nos dá uma idéia melhor da distribuição espacial da variável estudada.

Quando as observações nas áreas representam contagens, como as obtidas pelo censo, o estimador de kernel apresentado acima não é apropriado. Um valor “médio” de um atributo como “número de domicílios precários” não faria sentido, e deve-se pensar em termos de “número de domicílios precários por unidade de área”. Neste caso, pode-se utilizar o numerador da equação (5.32), dividido pela área do círculo definido pelo raio de busca:

$$\hat{z}_i = \frac{1}{\pi\tau^2} \sum_{j=1}^n k\left(\frac{d_{ij}}{\tau}\right) z_j, \quad d_{ij} \leq \tau \quad (5.33.)$$

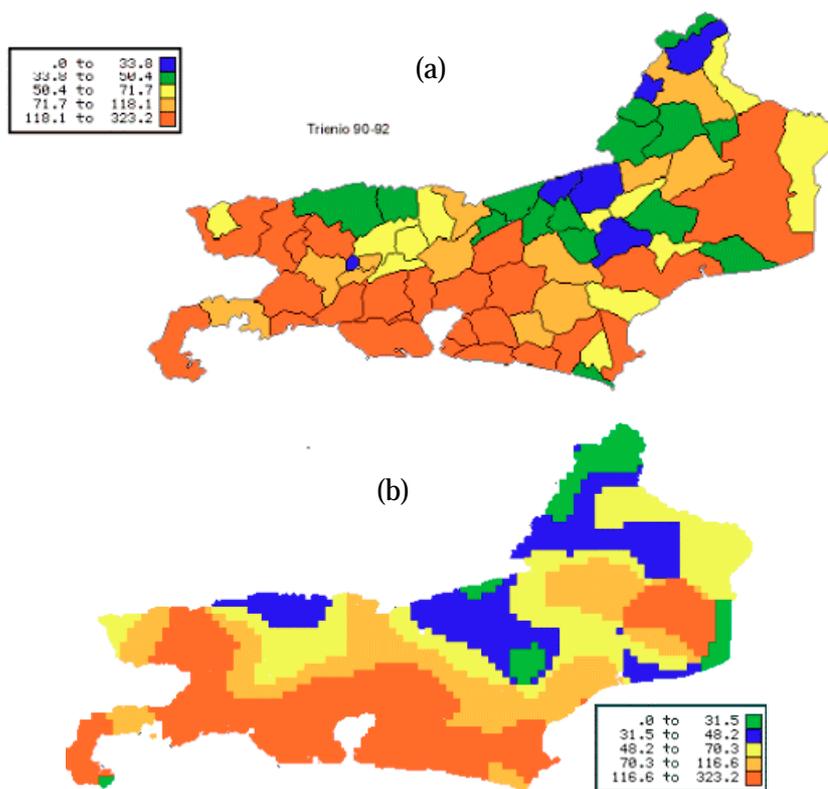


Figura 5-24 (a) Mortalidade por homicídios no RJ (1990-1992). Mapa temático com valores por município. (b) Superfície obtida por estimador de intensidade não-paramétrico

Uso de Interpoladores Geoestatísticos

No capítulo 3, apresenta-se a teoria básica da geoestatística, cuja motivação tradicional está associada a dados do meio físico como medidas de teor mineral ou de poluição. No caso da krigagem ordinária, a hipótese subjacente é que os dados apresentam distribuição gaussiana, e neste caso as propriedades ótimas dos estimadores (como a mínima variância do resultado) são garantidas. Para o caso de dados socioeconômicos ou de saúde coletiva, a hipótese da normalidade dos dados muito raramente é realista, sendo mais comum supor uma distribuição de Poisson, por se tratar de contagens de eventos. No entanto, as propriedades ótimas do estimador de krigagem e sua ampla disponibilidade em diferentes sistemas de informação geográfica fazem com que seja importante investigar seu uso para dados socioeconômicos. Neste caso, a primeira providência é investigar quão aproximados da distribuição normal se apresentam os dados; se for necessário, pode-se aplicar transformações apropriadas (com a transformação logarítmica) para “simetrizar” a distribuição empírica e assim aproximar-se da distribuição normal. Para considerar uma situação concreta, Figura 5-25 apresenta a distribuição da taxa de homicídios por 100 mil habitantes, para os 96 distritos de São Paulo em 1996, acompanhada do gráfico de probabilidade normal, que indica o quanto estes dados se aproximam de uma distribuição gaussiana. Da análise dos dois dados, e considerando-se ainda que a média (43,6) é suficientemente próxima da mediana (39,3), e como o teste de normalidade de Shapiro-Wilk indica um valor de 0,9653 (p-valor de 0,012), a hipótese de normalidade não pode ser rejeitada e permite aplicar uma interpolador de krigagem.

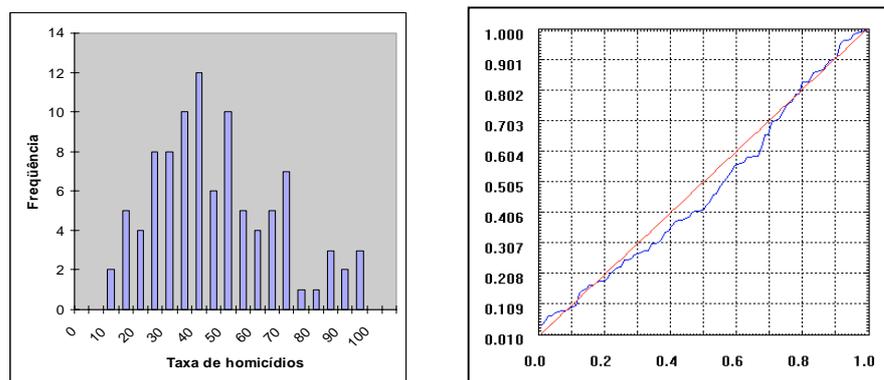


Figura 5-25. Distribuição da taxa de homicídios por 100 mil habitantes para São Paulo em 1996. À direita: frequência relativa; à esquerda: gráfico de probabilidade normal.

Com base nestas hipóteses, e com o objetivo de entender os padrões espaço-temporais em São Paulo, utilizou-se a krigagem ordinária para produzir superfícies das taxas de homicídio para os 96 distritos de São Paulo para os anos de 1996 e 1999 (a distribuição de taxas de 1999 apresentou padrões semelhantes que a de 1996). Para tal, o conjunto de pontos obtido pela associação do valor do parâmetro de cada área, ao seu centróide, foi tomado como uma amostra, usada para computar um variograma que modelou a estrutura de correlação espacial. A superfície obtida está apresentada na Figura 5-26 e mostra uma queda significativa nas áreas com as menores taxas de homicídios (menos que 30 mortes por 100,000 pessoas) em 1999 com relação a 1996. Como as áreas de menor taxa de homicídio correspondem às áreas mais ricas da cidade (compare com as figuras 5.1), o resultado mostra um espalhamento espacial do crime, com a violência ocupando progressivamente toda a cidade.



Figura 5-26. Superfícies estimadas para as taxas de homicídio em São Paulo em 1996 (esquerda) e 1999 (direita).

5.8 COMENTÁRIOS FINAIS

Este capítulo mostrou que as técnicas de análise espacial podem ampliar consideravelmente a capacidade de compreender os padrões espaciais associados a dados de área, especialmente quando se trata de indicadores sociais, que apresentam autocorrelação espacial global e local. Técnicas exploratórias como os indicadores de Moran e os mapas de espalhamento de Moran são muito úteis para mostrar as agregações espaciais e indicar áreas prioritárias em termos de política pública. Métodos de estimação bayesiana

para taxas permitem a correção de efeitos associados a pequenas populações. Modelos de regressão espacial permitem estabelecer as relações entre as variáveis, levando em conta os efeitos espaciais; neste caso, o poder explicativo dos modelos pode ter ganhos significativos. A geração de superfícies é um maneira eficiente de apreensão visual dos padrões espaciais. Em resumo, estudiosos de dados sócio-econômicos podem se beneficiar substancialmente das técnicas deste capítulo.

5.9 REFERÊNCIAS

A referência básica para a maior parte das técnicas apresentadas neste capítulo é o livro de Trevor Bailey, “*Spatial Data Analysis by Example*” (Bailey and Gattrel, 1995) e uma discussão geral sobre os modelos de distribuição para dados espaciais é apresentada em Diggle (2001). A homepage de Peter Diggle (www.maths.lancs.ac.uk/~diggle) contém material relevante sobre estatística espacial.

No caso dos modelos de regressão espacial, o software SpaceStat de Luc Anselin, e a documentação associada (Anselin, 1992) apresenta em detalhe os modelos de regressão com efeitos globais (*spatial lag e spatial error*), e o modelo de regimes espaciais. O SpaceStat foi utilizado para computar os modelos no exemplo apresentado no capítulo. Os trabalhos de Luc Anselin no campo de indicadores locais de autocorrelação espacial (Anselin, 1995; Anselin, 1996) também são referências importantes. O sítio do SpaceStat é www.spacestat.com.

O modelo de regressão GWR (*geographically weighted regression*) foi idealizado por A.Stewart Fotheringham, e está descrito em seu livro *Quantitative Geography* (Fotheringham et al., 2000) e outros trabalhos (Fotheringham et al., 1996) (Brunsdon et al., 1996). Maiores informações podem ser encontradas no sítio <http://www.ncl.ac.uk/~ngeog/GWR/>.

A discussão sobre o problema dos efeitos de escala e a chamada “falácia ecológica” deve muito aos trabalhos de Stan Openshaw; como exemplo, veja-se Openshaw (1997). Seus trabalho sobre o uso de técnicas de otimização combinatória para obter regiões mais agregadas, também são muito importantes (Openshaw and Albanides, 1999).

A questão da geração de superfícies a partir de dados socioeconomicos deve muito aos trabalhos de David Martin, em seu livro “*Geographic Information Systems: Socioeconomic Applications*” (Martin, 1995) e seus trabalhos sobre os dados censitários no Reino Unido (Martin, 1996; Martin, 1998).

Os estimadores bayesianos empíricos foram inicialmente propostos em (Marshall, 1991). Uma discussão geral sobre o assunto, incluindo uma discussão sobre os estimadores bayesianos completos, pode ser encontrada no excelente trabalho de Renato Assunção (Assunção, 2001) ou na revisão abrangente de Trevor Bailey, publicada nos Cadernos de Saúde Pública (Bailey, 2001).

Os dados de São Paulo do censo de 1991 foram extraídos do trabalho "Mapa de Exclusão/Inclusão Social na Cidade de São Paulo", coordenado pela prof. Aldaíza Sposati, da PUC/SP (Sposati, 1996). As taxas de homicídio para os distritos de São Paulo em 1996 e 1999 foram produzidas pela Fundação SEADE e a geração de superfícies por krigeagem foi feita por José Luiz Rodriguez Yi.

Os dados do censo de Belo Horizonte para o ano de 1991 foram cedidos pela PRODABEL, e o estudo do problema das mudanças de unidade de análise foi realizado por Taciana Dias e Maria Piedade Oliveira.

Os dados de mortalidade infantil para a cidade do Rio de Janeiro foram organizados pela FIOCRUZ e estão apresentados no trabalho de Eleonora D'Orsi e Marília Carvalho (D'Órsi & Carvalho, 1998). Os dados do estudo sobre mortalidade por homicídios na Região Sudeste também foram publicados pela equipe da FIOCRUZ, e podem ser acessados nas páginas pessoais dos autores: <http://www.procc.fiocruz.br/~marilia/> e www.procc.fiocruz.br/~oswaldo/.

O número especial dos Cadernos de Saúde Pública sobre o tema de estatísticas espaciais em saúde (volume 17(5), outubro-novembro 2001), disponível na Internet (www.scielo.br) representa um bom ponto de partida sobre o tema, com vários estudos relevantes.

1. ANSELIN, L. **SpaceStat tutorial: a workbook for using SpaceStat in the analysis of spatial data**. Santa Barbara, NCGIA (National Center for Geographic Information and Analysis), 1992.
 2. ANSELIN, L. Local indicators of spatial association - LISA. **Geographical Analysis** v.27, p.91-115, 1995.
 3. ANSELIN, L. The Moran scatterplot as ESDA tool to assess local instability in spatial association. In: M. Fisher, H. J. Scholten and D. Unwin (ed). **Spatial Analytical Perspectives on GIS**. London, Taylor & Francis, 1996. v., p.111-126.
 4. ASSUNÇÃO, R. **Estatística Espacial com Aplicações em Epidemiologia, Economia e Sociologia**. São Carlos, SP, UFScar, 2001. Disponível na homepage www.est.ufmg.br/~assuncao.
-

5. BAILEY, T. Spatial Statistics Methods in Health. **Cadernos de Saúde Pública** v.17, n.5., 2001.
 6. BAILEY, T. and A. GATTREL. **Spatial Data Analysis by Example**. London, Longman, 1995.
 7. BRUNSDON, C. A.S. FOTHERINGHAM AND M.E. CHARLTON, Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. **Geographical Analysis**, 28(4), 281-298, 1996.
 8. CRUZ, O. C. **Homicídios no Estado do Rio de Janeiro: análise da distribuição espacial e sua evolução**. Dissertação de mestrado/Faculdade de saúde Pública-USP, 1996.
<http://malaria.procc.fiocruz.br/~oswaldo/publi/ogc-diss.pdf>
 9. DIGGLE, P. **Spatial statistics in the biomedical science: future directions**. Lancaster, Lancaster University, 2001.
 10. D'ÓRSI, E. and M. S. CARVALHO. Perfil de Nascimentos no Município do Rio de Janeiro - Uma Análise Espacial. **Cadernos de Saúde Pública** v.14, n.1, p.367-379, 1998.
 11. FOTHERINGHAM, A.S., C. BRUNSDON AND M.E. CHARLTON, 2000, **Quantitative Geography**, London: Sage
 12. FOTHERINGHAM, A.S., M.E. CHARLTON AND C. BRUNSDON, The Geography of Parameter Space: An Investigation into Spatial Non-Stationarity. **International Journal of Geographic Information Systems**, 10: 605-627, 1996.
 13. GELMAN, A., CARLIN, J.B., STERN, H.S., RUBIN, D.B. (1995) **Bayesian Data Analysis** Chapman & Hall/CRC.
 14. GILKS, W.R., RICHARDSON, S., SPIEGELHALTER, D.J. (orgs) (1998), *Markov Chain Monte Carlo in Practice*, Chapman & Hall.
 15. MARSHALL, R. Mapping disease and mortality rates using empirical Bayes estimators. **Applied Statistics** v.40, p.283-294, 1991.
 16. MARTIN, D. **Geographic Information Systems: Socioeconomic Applications**. London, Routledge, 1995.
 17. MARTIN, D. An assessment of surface and zonal models of population. **International Journal of Geographical Information Systems** v.10, p.973-989, 1996.
 18. MARTIN, D. Optimizing census geography: the separation of collection and output geographies. **International Journal of Geographical Information Science** v.12, p.673-685, 1998.
-

19. OPENSHAW, S. Developing GIS-relevant zone-based spatial analysis methods. In: P. Longley and M. Batty (ed). **Spatial Analysis: Modelling in a GIS Environment**. New York, John Wiley, 1997. v., p.55-73.
 20. OPENSHAW, S. and S. ALVANIDES. Applying Geocomputation to the analysis of spatial distributions. In: P. A. Longley, Goodchild, M. F., Maguire, D. J. and Rhind, D. W (ed). **Geographical Information Systems: Principles, Techniques, Management and Applications**. Chichester, Wiley, 1999. v., p.267-282.
 21. SPOSATI, A. **Mapa de Exclusão/Inclusão Social de São Paulo**. São Paulo, EDUC, 1996.
-

MODELAGEM DINÂMICA E GEOPROCESSAMENTO

Bianca Maria Pedrosa

Gilberto Câmara

8.1 INTRODUÇÃO

Historicamente, a tecnologia de Geoprocessamento enfatizou a representação de fenômenos espaciais no computador de forma estática. Isto se deve ao fato de que a principal abstração utilizada em Sistemas de Informação Geográfica (GIS) é o mapa. No entanto, um significativo conjunto de fenômenos espaciais, tais como escoamento de água da chuva, planejamento urbano e dispersão de sementes, entre outros, são inerentemente dinâmicos e as representações estáticas comumente utilizadas em GIS não os capturam de forma adequada. Deste modo, um dos grandes desafios da Ciência da Informação Espacial é o desenvolvimento de técnicas e abstrações que sejam capazes de representar adequadamente fenômenos espaço-temporais dinâmicos.

O uso de modelos temporais em GIS vem sendo investigado com afinco na literatura recente (Worboys 1995). Neste trabalho, estaremos dando ênfase a abordagem de representação da dinâmica espaço-temporal por autômatos celulares. Nesta abordagem o espaço é representado através de um *array* de células em que cada célula pode assumir diferentes estados ao longo do tempo. O tempo varia em intervalos discretos e o estado de todas as células muda simultaneamente em função do seu próprio estado, do estado das demais células em sua vizinhança e de acordo com um conjunto específico de regras de transição (Engelen 1995).

Neste trabalho, será abordado os requisitos necessários as abordagens utilizadas para desenvolver sistemas espaciais dinâmicos, os conceitos computacionais e geográficos envolvidos e algumas das aplicações de Modelagem dinâmica em Geoprocessamento.

No capítulo 2 serão apresentados os princípios básicos para representar os principais componentes de um modelo espacial dinâmico. No capítulo 3 serão abordados os conceitos básicos e os principais aspectos computacionais envolvidos na modelagem dinâmica de processos físicos. Será apresentado também um exemplo de modelagem utilizando o aplicativo PCRaster. No capítulo 4 serão apresentados dois exemplos de aplicações computacionais para modelagem dinâmica de processos urbanos, o Citylife e o modelo Multi-escala.

8.2 PRINCÍPIOS BÁSICOS

A atual geração de GIS configura uma tecnologia estabelecida para armazenar, organizar, recuperar e modificar informações sobre a distribuição espacial de recursos naturais, dados geo-demográficos, redes de utilidade pública e muitos outros tipos de dados localizáveis na superfície da terra. Nesta área, um dos principais desafios para os próximos anos é transformar estes sistemas, essencialmente estáticos, em ferramentas capazes de prover representações realistas de processos espaço-temporais. A modelagem de grande quantidade de processos físicos, em aplicações como Geomorfologia, Estudos Climáticos, Dinâmica Populacional e Impacto Ambiental, requer que os GIS tenham capacidade de representar os tipos de processos dinâmicos encontrados em estudos de sistemas físicos e sócio-econômicos.

Neste contexto, a Modelagem Dinâmica (Burrough 1998) procura transcender as limitações atuais da tecnologia de Geoprocessamento, fortemente baseada numa visão estática, bidimensional do mundo. O objetivo dos modelos dinâmicos em GIS é realizar a simulação numérica de processos dependentes do tempo, como nos modelos hidrológicos, que simulam o fluxo e transporte de água. Na definição de Burrough, “um modelo espacial dinâmico é uma representação matemática de um processo do mundo real em que uma localização na superfície terrestre muda em resposta a variações nas forças dirigidas”.

Tipicamente, GIS são desenvolvidos a partir de suposições pré-estabelecidas quanto a homogeneidade, uniformidade e universalidade das propriedades de seus principais componentes, que incluem o espaço e as relações espaciais, o tempo e o modelo matemático que descreve o fenômeno. Entretanto, para modelar processos dinâmicos em GIS com o nível necessário de realismo, estas suposições rígidas têm que ser flexibilizadas de tal forma que o sistema seja capaz de representar (Couclelis 1997):

O espaço como uma entidade não homogênea tanto nas suas propriedades quanto na sua estrutura.

As vizinhanças como relações não estacionárias

As regras de transição como regras não universais.

A variação do tempo como um processo regular ou irregular.

O sistema como um ambiente aberto a influências externas.

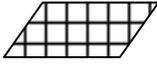
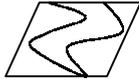
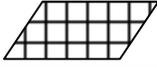
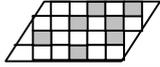
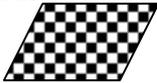
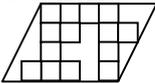
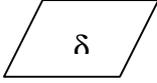
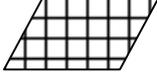
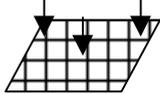
Espaço	Estrutura	regular 	irregular 
	Propriedades	uniforme 	não uniforme 
Vizinhança	estacionária 	não estacionária 	
Função de Transição	universal 	não universal 	
Tempo	regularidade 	irregularidade 	
Sistema	fechado 	aberto 	

Figura 8.1 – Requisitos para modelagem dinâmica em GIS (fonte: Couclelis, 1997)

Na figura 8.1 estão representados os requisitos mencionados acima. A regularidade do espaço diz respeito a forma como ele é distribuído e pode ser regular, isto é, dividido em partes iguais, ou irregular, distribuído de forma diferenciada. As vizinhanças, que geralmente são concebidas como tendo a mesma configuração para todo ponto no espaço, deve superar esta estacionaridade e poder ser representada com diferentes configurações em diferentes pontos do espaço. Por exemplo, em determinado ponto uma célula pode ter vizinhança 4 e em outro vizinhança 8. O sistema deve permitir que mais de uma função de transição possa ser aplicada, permitir que o tempo seja representado em intervalos variáveis (meses, anos) e suportar a inclusão de variáveis externas.

Para implementar sistemas espaciais dinâmicos com as características mencionadas acima, alguns princípios básicos relativos aos principais elementos destes sistemas devem ser considerados. Entre estes elementos destacam-se a questão da representação do espaço e do tempo, o modelo dinâmico a ser utilizado para a representação do fenômeno espacial e a abordagem computacional para implementar estes princípios de forma integrada e consistente. Nas seções seguintes, discutiremos cada um destes elementos.

8.2.1 O Espaço

O espaço é o conceito chave na geografia e, por extensão, na Ciência da Informação Espacial. Tradicionalmente, os geógrafos fazem uma distinção entre os conceitos de espaço absoluto e espaço relativo.

“Espaço absoluto, também chamado Cartesiano ou Newtoniano, é um container de coisas e eventos, uma estrutura para localizar pontos, trajetórias e objetos. Espaço relativo, ou Leibnitziano, é o espaço constituído pelas relações espaciais entre coisas e eventos” (Couclelis 1997).

Santos (1996) refere-se a distinção entre espaço absoluto e espaço relativo como o “espaço dos fixos” e o “espaço dos fluxos”. Em termos de representações computacionais pode-se, de forma aproximada, traduzir estes conceitos como a distinção entre as representações associadas a recobrimentos planares (mapas de polígonos e matrizes) e representações associadas a conectividade (grafos). Um caso típico de medida realizada no espaço absoluto são os índices de auto-correlação espacial. Neste caso, um dos instrumentos básicos é a matriz de proximidade espacial, cujo cálculo usualmente é feito em função de distância euclidiana entre objetos ou da existência de uma fronteira entre eles. Na Figura 8.2 está representado um mapa temático e sua respectiva matriz de proximidade, definida com base nas fronteiras existentes entre os objetos.

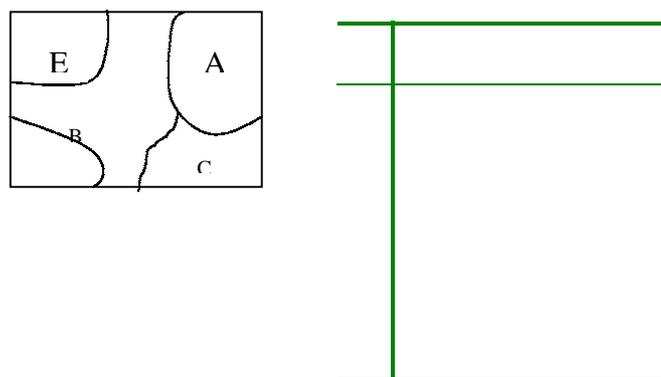


Figura 8.2 – Um mapa poligonal e sua matriz de proximidade

Em muitos fenômenos geográficos, os objetos estabelecem relações entre si que independem das relações espaciais típicas como as relações topológicas, direcionais e de distância. Estes fenômenos geralmente incluem relações como fluxo de pessoas ou materiais, conexões funcionais de influência, comunicação e acessibilidade, entre outras (Couclelis 1999). Um exemplo de fenômeno em que a dimensão espacial requer o

conceito de espaço relativo é o caso de fluxo de pessoas pela rede de transporte metroviário de uma cidade. O fluxo de pessoas a partir de uma mesma origem tem diferentes destinos, Figura 8.3, e a relação entre a origem e destino é estabelecida com base em relações de conectividade e acessibilidade.

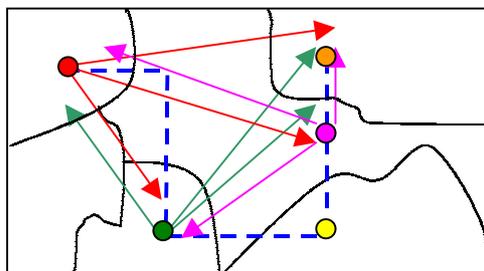
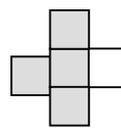
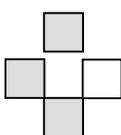
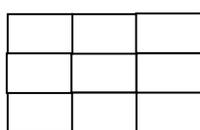


Figura 8.3 – Mapa do Fluxo de Pessoas em uma rede de transporte

Couclelis (1997) propõe a idéia de espaço próximo, como uma extensão dos conceitos de espaço absoluto e relativo. No espaço próximo o conceito chave é a vizinhança associada à noção de proximidade, que conduzem também ao conceito de proximidade funcional ou influência. O conceito de vizinhança é facilmente visualizado em representações matriciais do espaço. Algumas operações espaciais disponíveis em GIS como filtros espaciais, por exemplo, utilizam a noção de espaço próximo de forma limitada. No filtro espacial, o estado de uma célula (um pixel de uma imagem) é modificado com base nos estados das demais células em sua vizinhança, definida através de uma máscara. A seguir é apresentado um exemplo de filtro espacial e o estado de uma célula qualquer antes (Figura 8.4b) e após (Figura 8.4c) a aplicação do filtro espacial.



a) Máscara

b) estado da célula central antes do filtro

c) estado da célula central após o filtro

Figura 8.4 – Exemplo de Filtro espacial

A abstração fundamental na maior parte dos GIS atuais é o conceito de mapa, fortemente relacionado com noções cartográficas e portanto, do espaço absoluto. Em processos dinâmicos a noção de espaço relativo e próximo são fundamentais para estabelecer e representar fluxos e conexões entre entidades do sistema.

8.2.2 O Tempo

Conceitualmente, pode-se representar o tempo através de diferentes estruturas, definidas, principalmente, com base em três aspectos da representação temporal: granularidade, variação e ordem no tempo (Figura 5).

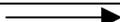
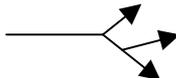
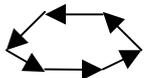
Ordem no tempo	Variação Temporal	Granularidade
linear 	discreto 	instante 
ramificado 	contínuo 	intervalo 
ciclício 		período 

Figura 5 - Estruturas temporais (fonte: Worboys, 1998)

A ordem temporal refere-se ao modo como o tempo flui. Neste caso, pode-se assumir que o tempo flui de forma linear, ramificada ou cíclica. No tempo linear considera-se que o tempo flui seqüencialmente, ou seja, existe uma ordem de precedência entre os pontos no tempo, de tal forma que cada ponto tenha apenas um sucessor e um antecessor. No tempo ramificado múltiplos pontos podem ser os sucessores ou antecessores imediatos de um mesmo ponto. O tempo cíclico é utilizado para modelar eventos e processos recorrentes (Edelweiss and Oliveira 1994).

Com relação à variação temporal duas possibilidades podem ser consideradas: tempo contínuo e discreto. Uma variável temporal contínua é usada em processos que demandam medidas de tempo com níveis arbitrários de precisão. Por exemplo, a expansão da área de desmatamento de uma floresta entre dois instantes de tempo medidos pode ser interpolada. A Figura 6 apresenta mapas de uma área desmatada em dois instantes t e t' . Se necessário, pode-se gerar um novo mapa para representar a área desmatada entre os instantes t e t' porque este processo é contínuo no tempo.

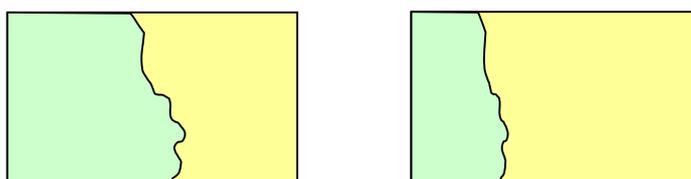


Figura 6 – Área Desmatada da floresta em dois instantes t e t'

Uma variável temporal discreta é usada quando o tempo é medido em certos pontos ou intervalos e a variação é descontínua entre estes pontos. Uma delimitação de lotes de um cadastro imobiliário pode ocupar uma posição num instante t e outra num instante t' , mas não faz sentido dizer que a delimitação ocupou alguma posição intermediária entre t e t' . Na figura 7, no instante t existem 4 lotes em uma área residencial, no instante t' os lotes 2 e 3 são unidos formando um novo lote 5.

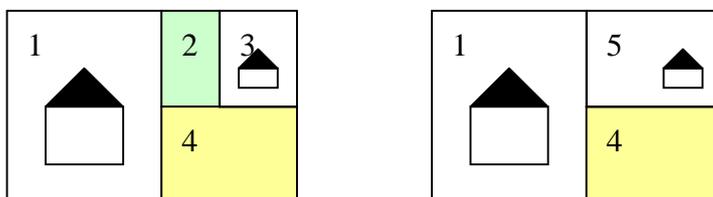


Figura 7 – Delimitação de lotes de um cadastro imobiliário em dois instantes t e t'

Associado ao conceito de variação temporal discreta, existe o conceito de *Chronos*. Um *chronon* é a menor duração de tempo suportada por um sistema e pode variar em diferentes aplicações (Edelweiss and Oliveira 1994).

A granularidade temporal de um sistema está diretamente relacionada com a duração de um *chronon*. As diferentes granularidades de um sistema temporal conduzem à definição de instante e intervalo de tempo. Um instante de tempo representa um ponto particular no tempo, um intervalo é o tempo decorrido entre dois instantes e um período consiste de uma seqüência de intervalos de tempo .

Em sistemas computacionais, representa-se o tempo em pelo menos duas dimensões:

- tempo válido (*valid time*) - corresponde ao tempo em que um evento ocorre no domínio da aplicação.
- tempo de transação (*transaction time*) – corresponde ao tempo em que transações acontecem dentro do sistema de informação (Worboys 1995).

Adicionalmente, existe o conceito de “tempo definido pelo usuário”, consistindo de propriedades definidas explicitamente pelos usuários em um domínio temporal e manipuladas pelo programa de aplicação (Edelweiss and Oliveira 1994).

A incorporação da dimensão temporal em um sistema de informação não se restringe apenas à questão da representação do tempo, mas inclui também questões relativas a sua recuperação. Um GIS temporal deve ser capaz de recuperar informações através de consultas definidas sobre critérios temporais, como por exemplo:

Quais rodovias do Brasil foram recuperadas **a partir de 1980** e **agora** permitem uma velocidade superior a 100km/h ?

Qual rio teve a maior taxa de poluição **entre 1970 e 1985**?

Quais as cidades em que a cobertura vegetal aumentou em pelo menos 5% **durante** os últimos 5 anos?

Para resolver consultas como as relacionadas acima, um GIS tem que prover um conjunto de operadores e funções que permitam a avaliação de relacionamentos como os de precedência, sobreposição, igualdade e pertinência entre dois intervalos de tempo (Figura 8).

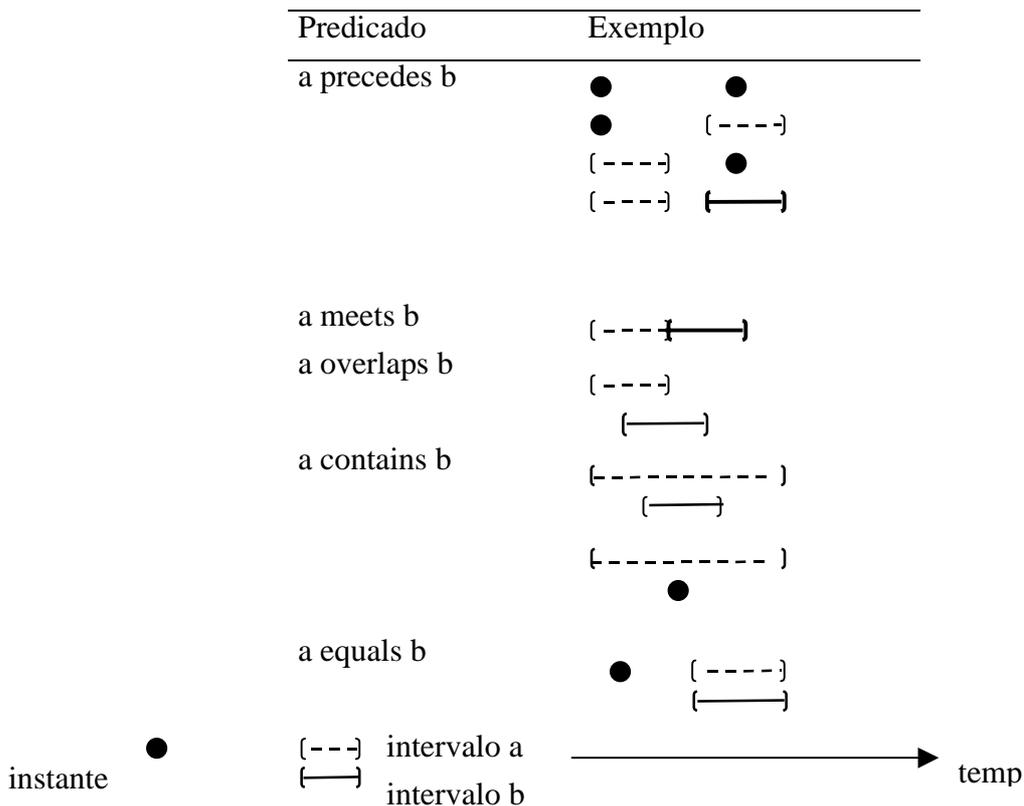


Figura 8 – Predicados temporais (fonte: Voigtmann, 1996)

Para exemplificar consultas envolvendo predicados como os apresentados na Figura 8, utilizaremos uma linguagem de consulta temporal especialmente desenvolvida para aplicações em geoprocessamento, chamada T/OOGQL (Voigtmann 1996). Esta linguagem é uma extensão ao SQL (Structured Query Language) com suporte para tipos de dados espaciais e temporais.

Quanto aos operadores espaciais, a linguagem T/OOGQL oferece os clássicos *cross*, *overlap*, *touch* e *in*, entre outros. Quanto aos aspectos temporais, a linguagem T/OOGQL trabalha com os conceitos de timestamp (definição explícita de tempo associada a uma informação), tempo válido e tempo transacional, e oferece as seguintes funções:

Função	Descrição
First(), Last()	retorna o primeiro e último timestamp associado a um atributo, objeto ou relacionamento
FirstValue(), LastValue()	retorna o primeiro e último valor associado a um atributo, objeto ou relacionamento
Begin(), End()	retorna o início e o fim de um timestamp
Period(b,f), Period(d)	retorna um período tendo início b e final f ou um período de duração d
Day(), Month(), Year()	construtores de timestamps para descrever um dia, mês ou ano
date(), time(), datetime()	construtores de timestamps para descrever uma data, hora ou data e hora
years(n), months(n), days(n)	retorna um intervalo de tempo com a duração de n dias, n meses ou n anos

Figura 9 – Funções Temporais do T/OOGQL

A seguir, demonstraremos o uso da linguagem T/OOGQL em alguns exemplos de consultas que envolvem tanto operadores temporais quanto espaciais:

1. Quais rodovias do Brasil foram recuperadas a partir de 1980 e agora permitem uma velocidade ≥ 100 km/h ?

```
select snapshot r
from railroad r, state s
where s.name="Brasil" and (r cross s or r in s) and
      r.max_speed>=100 and
      Begin( Year(1980) ) vt_precedes r.max_speed
```

2. Qual rio teve a maior taxa de poluição entre 1970 e 1985?

```
query_time:Period (Begin (Year(1970)), End(Year(1985)))
select r.name
from river r
where exists rp in r.pollution:
      rp.max_pollution(query_time) >=max (select
      max_pollution (query_time) from river_pollution)
```

3. Quais as cidades em que a cobertura vegetal aumentou em pelo menos 5% durante os últimos 5 anos?

```
select c.name
from   city c
where  c.vegetation.coverage("Wood",c.geometry,now)
      >= min(c.vegetation.coverage("Wood",c.geometry,
      Period(now-years(5)),now))+5
```

Nas consultas apresentadas acima, as cláusulas *select*, *from* e *where* são similares às de qualquer linguagem baseada em SQL. Os prefixos *vt* e *tt* são abreviações para tempo válido e tempo de transação, respectivamente. As palavra reservada *Snapshot*, presente na cláusula *select* da primeira consulta, tem o mesmo significado da Linguagem TSQL2 (uma extensão temporal para a linguagem SQL2), ou seja, gera um resultado de consulta instântaneo, sem timestamps associados (Voigtmann 1996).

8.3 MODELOS

Modelos espaciais dinâmicos descrevem a evolução de padrões espaciais de um sistema ao longo do tempo. Segundo Lambin(1994) um modelo deve responder as seguintes questões:

- Quais variáveis ambientais e culturais contribuem para explicar o fenômeno e quais são os processos ecológicos e sócio-econômicos existentes por trás do fenômeno?
- Como o processo evolui?
- Onde ocorrem os fenômenos?

Estas questões chaves podem ser identificadas como as clássicas “Porque”, “Quando” e “Onde”. Um modelo que responde a estas questões é capaz de descrever quantitativamente um fenômeno e prever sua evolução, integrando suas escalas temporal e espacial.

8.3.1 Tipos de modelos

Um modelo é constituído de pelo menos três elementos: variáveis, relacionamentos e processos. Ao se construir um modelo, dependendo do objetivo, pode-se dar ênfase a um ou outro destes elementos. Nesta visão, os modelos podem ser classificados em empíricos e de sistemas. Modelos empíricos focalizam os relacionamentos entre as variáveis do modelo, a partir da suposição de que os relacionamentos observados no passado continuarão no futuro. Modelos de sistemas são descrições matemáticas de processos complexos que interagem entre si, enfatizando as interações entre todos os componentes de um sistema (Lambin 1994).

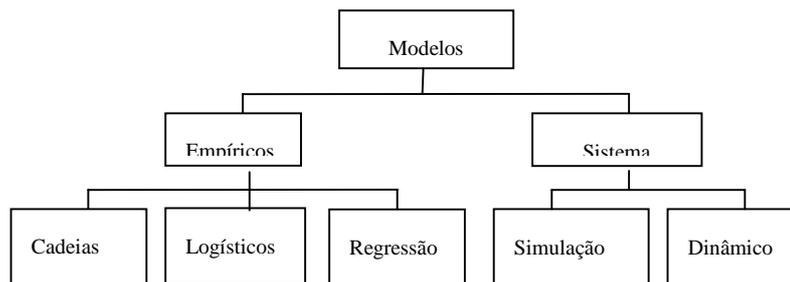


Figura 8.10 – Tipos de modelos

8.3.1.1 Modelos Empíricos

Os modelos empíricos, em sua dimensão procedural, possuem três componentes chaves: uma configuração inicial, uma função de mudança e uma configuração de saída. A configuração inicial de um modelo dinâmico pode ser obtida através de dados históricos do fenômeno em estudo, chamados de séries temporais. Neste caso, equações diferenciais (totais ou parciais) que incluem pelo menos um termo derivado no tempo podem ser utilizadas para representar o modelo e o processo é classificado como determinístico. Quando variáveis aleatórias são utilizadas para explicar um sistema o processo é classificado como estocástico-probabilístico.

Modelos empíricos são caracterizados pela simplicidade dos modelos matemáticos empregados e pelo número reduzido de variáveis envolvidas. Estes modelos são eficientes em fazer previsões, embora apresentem limitações em abordar a evolução espacial e identificar os aspectos causais do sistema. A seguir, serão apresentados três modelos empíricos: cadeias de markov, modelos logísticos de difusão e modelos de regressão.

Cadeias de Markov

Cadeias de Markov são modelos matemáticos para descrever processos estocásticos e podem denotadas por:

$$\Pi_{(t+1)} = P^n \cdot \Pi_{(t)}$$

onde $\Pi_{(t)}$ é o estado do sistema no tempo t , $\Pi_{(t+1)}$ é o estado do sistema após o intervalo $t+1$ e P^n são os estados passíveis de acontecer, que são representados em matrizes de possibilidades de transição. Essas matrizes de transição representam a possibilidade de um determinado estado i permanecer o mesmo ou mudar para o estado j durante o intervalo de tempo $t \rightarrow t+1$. As probabilidades de transição são usualmente derivadas de amostras relativas a um certo intervalo de tempo. Cadeias de Markov de 1ª ordem assumem que o estado futuro do sistema depende apenas do seu estado presente e das possibilidades de transição, sendo independente da trajetória que o levou aquele estado (estados em um tempo $t-1$). Este modelo não ignora o passado, mas assume que toda a informação do passado está concentrada no presente estado do sistema. Desta forma, as interações são instantâneas, sendo irrelevante o tempo de permanência das variáveis em cada estado (Soares Filho 1998).

Outra característica das cadeias de Markov é que as probabilidades de transição não mudam com o tempo, o que o caracteriza como um processo estacionário.

As principais vantagens das cadeias de Markov são a simplicidade operacional e matemática do modelo aliadas a facilidade com que podem ser aplicadas a dados provenientes de sensoriamento remoto e implementadas em GIS. Outra grande vantagem é o fato de não necessitar de grande quantidade de dados antigos para prever o futuro.

As principais limitações das cadeias de markov incluem o fato do modelo não explicar o fenômeno (Porque) e ser limitado na resposta espacial (Onde), entretanto o modelo pode fazer previsões (Quando) desde que os processos sejam estacionários. Além disto, o modelo não suporta de imediato a inclusão de variáveis exógenas como variáveis sócio-econômicas ou outras forças dirigidas, embora esta limitação possa ser superada. Em (Lambin 1994) são apresentadas várias abordagens para superar as principais limitações de cadeias de Markov em modelagem dinâmica.

Modelos logísticos de Difusão

Modelos logísticos são utilizados para descrever matematicamente fenômenos em que as variáveis inicialmente apresentam variações em um ritmo lento, depois o ritmo de variações se intensifica, voltando a reduzir-se até que o nível de saturação seja atingido. Este modelo leva em conta as interações temporais entre as variáveis do sistema, podendo ser expresso por:

$$dP/dt = r P [(U - P) / U]$$

onde P é a variável de um fenômeno de crescimento ao longo do tempo t, como aumento da população, por exemplo; r é a taxa de crescimento e U uma função de crescimento (Lambin 1994). Dentre os modelos baseados em funções logísticas destacam-se os modelos de difusão. Tais modelos enfatizam a velocidade do processo e permitem a inclusão de variáveis relacionadas às causas do fenômeno.

Os principais elementos de um modelo espacial de difusão são (Soares Filho 1998):

- meio ambiente (isotrópico ou heterogêneo)
- tempo (contínuo ou discretizado)
- item a ser difundido (material, pessoas, informação, doença)
- locais de origem
- locais de destino
- caminhos a serem percorridos

Estes elementos interagem entre si através de um mecanismo em que pode-se identificar quatro estágios:

- Estágio inicial – neste estágio tem início o processo de difusão.
- Estágio de difusão – tem início o processo de espalhamento
- Estágio de condensação – diminui o ritmo do espalhamento.
- Estágio de saturação – ocorre a desaceleração ou encerramento do processo de difusão.

O processo de espalhamento em modelos de difusão pode se dar por expansão ou realocação. Nos modelos de difusão por expansão a informação ou material se espalha de uma região para outra, permanecendo na região original. Nos modelos de difusão por realocação os objetos se movem para novas regiões, abandonando as áreas originais (Soares Filho 1998). Modelos de difusão não explicam as causas de um fenômeno, embora possam integrar variáveis ecológicas e sócio-econômicas. Sua maior contribuição está na predição do comportamento futuro do fenômeno. Quanto à dimensão espacial, o modelo em si não a incorpora, mas ela pode ser introduzida através da integração deste modelo com um GIS (Lambin 1994).

Modelos de Regressão

O objetivo dos modelos de regressão é estabelecer relações estatísticas entre um fenômeno em estudo e as variáveis independentes, chamadas forças dirigidas, que exercem influência sobre ele. Sendo assim, o modelo suporta a inclusão de variáveis exógenas como as sócio-econômicas. Isto contribui para o entendimento do fenômeno em estudo, mas é insuficiente para explicá-lo, pois a identificação de um relacionamento estatístico entre duas variáveis por si só não estabelece um relacionamento causal entre elas. Por exemplo, pode-se identificar através de um modelo de regressão que o crescimento populacional tem relação com o crescimento do desmatamento de uma determinada região, entretanto, o modelo de regressão não explica os mecanismos que ligam estas variáveis (Lambin 1994).

Matematicamente, o modelo estabelece um relacionamento linear entre as variáveis dependentes e independentes através da expressão:

$$y = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_i x_i + E$$

onde:

y = mudança ocorrida em um determinado tempo

x_i = variáveis independentes (forças dirigidas)

A_i = Coeficientes de regressão dos relacionamentos

E = Componente de erro

Em modelos de regressão a dimensão temporal é considerada, mas a distribuição espacial do fenômeno não é abordada, limitação esta que pode ser superada se o modelo for combinado com GIS. Outra limitação deste modelo é que ele se aplica apenas a processos estacionários (Lambin 1994).

Um exemplo de modelo de regressão é o implementado por Reis e Margulis (1991) para modelar o desmatamento da Amazônia em função da densidade espacial das atividades econômicas da região. Neste modelo, num primeiro estágio, áreas desmatadas são relacionadas com a densidade populacional, áreas cultivadas, distância de centros urbanos e proximidade de rodovias, entre outras variáveis. Num segundo estágio, o modelo relaciona o crescimento de determinadas atividades (colonização, cultivo, pecuária) entre 1980 e 1985 com a densidade destas atividades em 1980, obtendo assim o padrão de crescimento espacial de cada atividade. Então, partindo da suposição de que este padrão espacial de crescimento irá se manter no futuro, o modelo faz projeções sobre a tendência de desmatamento para o período de 1985-2000 (Lambin 1994).

8.3.2 Modelos de Sistemas

Modelos de sistemas procuram descrever o sistema como um todo, isto é, tentam representar as interações entre todos os seus componentes. Uma característica chave destes modelos é a eficiência com que abordam a dimensão espacial, implementando conceitos como as relações de vizinhança e suportando o uso combinado de múltiplas escalas. A seguir, descreveremos as características de gerais de duas classes de modelos de sistema: os modelos de simulação de ecossistemas e os modelos de simulação dinâmica espacial.

Modelos de Simulação de Ecossistemas

Modelos de Ecossistemas são projetados para imitar o comportamento de um sistema, enfatizando as interações entre todos os seus componentes. Estes modelos são baseados na composição de ecossistemas complexos em um número de equações diferenciais (Lambin 1994). A construção de um modelo de simulação requer que os principais aspectos que afetam o fenômeno estejam bem integrados, que seus relacionamentos funcionais estejam bem representados e que o modelo possa prever os impactos ecológicos e econômicos das mudanças ao longo do tempo.

Estes modelos são adequados para representar processos não estacionários, mas apresentam limitações quanto ao aspecto espacial, pois tratam o espaço como uma entidade homogênea (Lambin 1994).

Modelos de Simulação Dinâmica Espacial

Modelos de Simulação Dinâmica Espacial baseiam-se em modelos de ecossistemas com extensões para acomodar a heterogeneidade espacial e processos humanos de tomada de decisão.

Uma abordagem para desenvolver modelos de simulação dinâmica espacial é representar o espaço como uma matriz de células e aplicar as equações matemáticas a cada uma das células da matriz, simultaneamente. Cada célula do modelo está conectada com suas células vizinhas, de tal forma que é possível estabelecer um fluxo entre células adjacentes. Isto simplifica sobremaneira o mecanismo de predições do sistema porque por exemplo, se uma célula tem três vizinhos com estado x , é altamente provável que o estado desta célula venha a ser x também. Entretanto, este raciocínio simplista pode ser aperfeiçoado em regras de transição. Outro aperfeiçoamento desse modelo é a possibilidade de incorporar processos de tomada de decisões. Modelos que incorporam este mecanismo são chamados modelos baseados em regras. As regras de tomada de decisão são representadas através de abstrações muito semelhantes àquelas que ocorrem na mente humana.

Um exemplo de modelo com as funcionalidades mencionadas acima é o DELTA (Dynamic Ecological Land Tenure Analysis), um sistema desenvolvido para integrar

aspectos sócio-econômicos da colonização amazônica e aspectos ecológicos do desmatamento e da liberação de carbono no Estado de Rondônia.

O DELTA consiste em três submodelos integrados que simulam, respectivamente, a difusão da colonização, mudança do uso do solo e liberação de carbono. Os submodelos são examinados em diferentes escalas, o que caracteriza o modelo como multi-escala. Além disso, o modelo é considerado “a playing game tool”, pois não se restringe a fazer previsões, mas sim a servir como instrumento para responder “what if questions” (Lambin 1994).

Para finalizar, um resumo das características-chaves de cada tipo de modelo, segundo o potencial de cada um deles para responder as perguntas porque, quando e onde é apresentado a seguir:

Modelo	Porquê	Quando	Onde
Cadeias de Markov	não pode explicar a razão de um fenômeno por ser processo estocástico e não suportar a inclusão de variáveis exógenas	pode prever a evolução de processos estacionários	pode prever distribuições espaciais de elementos do modelo se for combinado com GIS
Logístico de Difusão	permite a inclusão de poucas variáveis exógenas, entretanto é um modelo descritivo, não suportando investigações exploratórias	suporta a dimensão temporal, podendo prever a evolução de processos não estacionários	pode prever distribuições espaciais de elementos do modelo se for combinado com GIS
Regressão	contribui para identificar forças direcionadoras, entretanto são modelos descritivos, não sendo capaz de estabelecer relações causais entre as variáveis	pode prever a evolução de processos estacionários	não são modelos espaciais, entretanto podem ser combinados com GIS
Simulação de Ecossistemas	modelo exploratório que requer descrições funcionais dos sistemas ecológicos	pode formular cenários de mudanças futuras no uso do solo, baseado nos parâmetros do modelo	apresenta dificuldades na representação espacial
Simulação Espacial Dinâmica	requer modelos funcionais espacialmente definidos	pode prever mudanças temporais no uso do solo, baseado nos parâmetros do modelo	pode prever evolução de padrões espaciais em processos determinísticos

8.3.3 Autômatos Celulares

Em 1982 John Conway apresentou o Jogo da Vida (*The Game of Life*), demonstrando que regras muito simples quando aplicadas repetidamente sobre estados aleatórios, produzem resultados semelhantes à forma como certos sistemas evoluem no mundo real. No Jogo da Vida o espaço é representado como uma grade de células, algumas das quais estão vivas e outras mortas. Dado um estado inicial aleatório, a cada geração, novas células nascem e algumas morrem. O que determina o estado de uma célula é sua vizinhança que, neste caso, é definida por quatro células adjacentes. Uma célula viva morre se tiver duas ou três células vizinhas vivas. Por outro lado, uma célula morta renasce, se tiver três células vizinhas vivas. Este sistema deu grande popularidade aos conceitos de autômatos celulares, que foram inicialmente apresentados por John Von Neumann (Roy, 1996).

Nos últimos anos, os conceitos de autômatos celulares tem sido utilizados para modelar fenômenos físicos e urbanos (Batty 1999, Burrough, 1998; Roy, 1996; Engelen, 1995; Câmara, 1996). Nesta abordagem o espaço é representado por um mosaico de células, geralmente de tamanhos e formatos idênticos (*regular tessellations*). Algumas das formas mais simples utilizadas para representar células em autômatos celulares são apresentadas na figura 8.12.

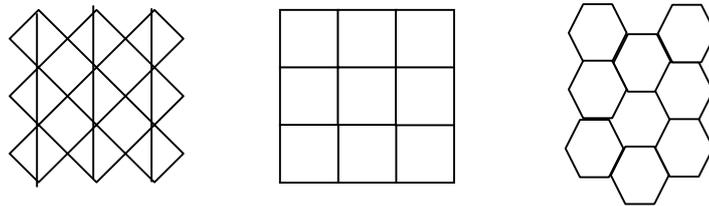


Figura 8.12 – Representações de células em autômatos celulares (fonte: Câmara, 1996)

Sobre cada célula de um autômato celular são aplicadas regras de transição. Regras de transição determinam quando e porque o estado de uma célula se altera e podem ser qualitativas ou quantitativas.

Para ilustrar como se dá o mecanismo de aplicação das regras de transição, apresentaremos um exemplo simples baseado em (Câmara 1996). Neste exemplo, uma célula pode assumir dois estados (branco e preto) e sua vizinhança é definida sobre duas células adjacentes. As regras de transição especificam que o estado de uma célula num instante $t+1$ é igual ao dos seus vizinhos no instante t , se estes vizinhos tiverem os estados iguais; caso contrário, o estado da célula permanece o mesmo. Para entender o exemplo é necessário identificar os componentes básicos do autômato celular clássico, que são:

- espaço euclidiano, dividido em um *array* de células
- uma vizinhança de tamanho e formato definidos (Figura 13a)
- um conjunto de estados discretos (Figura 13b)
- um conjunto de regras de transição (Figura 13c)
- um conjunto de intervalos de tempo, com atualização simultânea das células (Figura 13d)

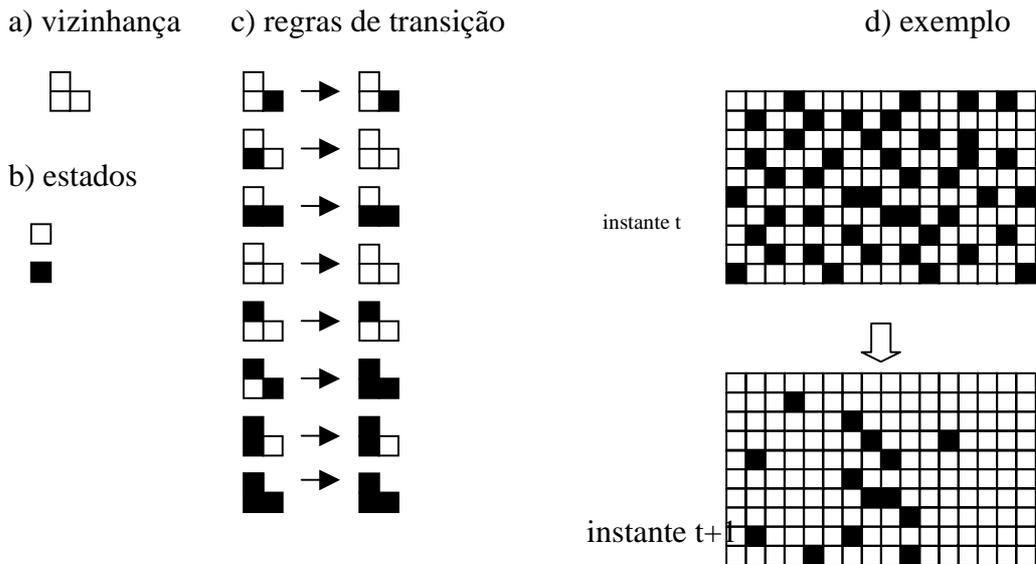


Figura 8.13 – Exemplo de autômato celular (Fonte: Câmara,1966)

A dinâmica de aplicação das regras de transição em um autômato celular é semelhante a de um filtro espacial. Desta forma, todas as células são avaliadas e, quando for o caso, modificadas para um novo estado. Na figura 13 d, a primeira célula da segunda linha do autômato tem, no instante t, o estado branco e suas vizinhas possuem estados diferentes (uma é branca e outra preta). Neste caso o estado da célula permanece o mesmo (1ª regra de transição). Seguindo o mesmo mecanismo, a segunda célula da segunda linha, tem no instante t o estado preto e suas vizinhas tem ambas o estado branco, logo o estado desta célula sofre uma transição para branco (2ª regra de transição). O processo segue este mecanismo para as demais células até que todas tenham sido avaliadas.

No exemplo acima, pode-se observar que as mudanças geradas por autômatos celulares são estritamente locais, isto é, baseadas nas vizinhanças de cada célula. Nesta perspectiva, pode-se dizer que sua aplicação é eficiente em processos em que a ordem global emerge de ações locais e descentralizadas (Batty 2000).

8.4 MODELAGEM DINÂMICA DE PROCESSOS FÍSICOS

Na seção anterior foram apresentados os princípios básicos relativos aos principais componentes de um modelo espacial dinâmico. Estes modelos são utilizados para explicar a ocorrência de um fenômeno, seu padrão espacial e sua evolução ao longo do tempo, respondendo assim às questões chaves porque, onde e quando, respectivamente. Dentre as diversas aplicações da modelagem espaço-temporal podemos identificar dois grandes grupos de processos: os físicos e os de planejamento urbano. Estes grupos possuem variáveis e comportamentos diferenciados que exigem diferentes abordagens de implementação. Nesta seção focalizaremos alguns aspectos computacionais presentes em modelagem dinâmica de processos físicos e apresentaremos um exemplo de modelagem dinâmica utilizando o aplicativo PCRaster.

8.4.1 Aspectos Computacionais

Fenômenos físicos tais como o escoamento da água da chuva e a difusão de plantas, encontrados na hidrologia e ecologia, respectivamente, são exemplos de fenômenos com alto índice de variação do estado da superfície ao longo do tempo. A complexidade dos modelos dinâmicos depende da dimensão em que tais modelos operam, 2 ou 3D, e dos equacionamentos matemáticos que utiliza. O mais simples dos modelos dinâmicos é chamado **modelo pontual sem memória**. Neste modelo, o estado de uma célula é modificado apenas pela variável fornecida como entrada para esta célula em um determinado instante t (Figura 14a). As demais células, bem como o estado desta célula em instantes anteriores não afetam o estado da célula naquele momento. Sendo assim, o estado de uma célula num processo pontual sem memória é uma função matemática operando na variável de entrada da célula no instante t (Figura 14b).

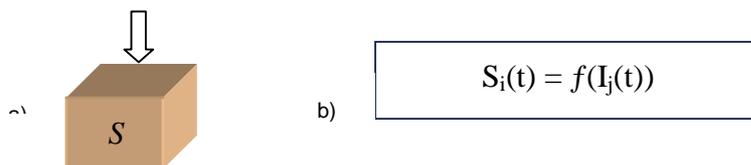


Figura 14 – Célula cujo estado depende apenas da variável de entrada (I) (fonte: Burrough, 1998)

Processos pontuais com memória referem-se a processos em que o estado de uma célula no instante $t+1$ retêm informações sobre seu estado no instante t . A memória da célula é determinada por uma função g operando no estado inicial da célula (Figura 15b). Um exemplo clássico para ilustrar o caso de processos pontuais com memória é o caso da água da chuva no solo. Quando o solo não consegue mais absorver água deve ocorrer algum transporte de material. Neste caso esta distribuição de material pode ser feita verticalmente, isto é, das células superiores para as inferiores (fluxo gravitacional – *Darcy's law*), como mostra a Figura 15a.

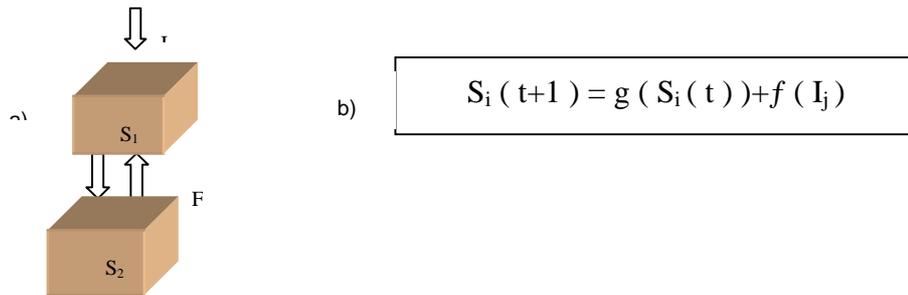


Figura 15 – Célula com adjacência vertical (fonte: Burrough, 1998)

Outra forma possível de transporte de material, considerando ainda o exemplo da água da chuva no solo, é através da adjacência lateral (processo de dispersão), Figura 16a. Neste caso existem mais variáveis envolvidas, o estado de uma célula depende do seu estado anterior, do fluxo de material e das entradas naquele intervalo de tempo (Figura 16b).

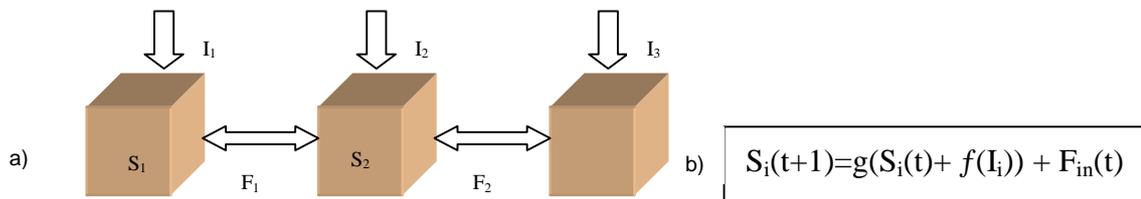


Figura 16 – Célula com adjacência horizontal (fonte: Burrough, 1998)

As células podem também ser conectadas por ligações topológicas (Figura 17). Estas ligações geralmente são baseadas em aspectos físicos do transporte de material. No caso da água de chuva, uma boa orientação para modelar o fluxo de material de célula para célula é conhecer a topologia do terreno.

Para que o fluxo de material de célula para célula possa ser computado a partir da topologia do terreno é necessário modelar esta topologia em uma forma computacionalmente apropriada. As redes *Local Drain Direction* (LDD), Figura 18, apresentam uma estrutura de dados computacional que viabiliza a interação entre as células. Nesta estrutura, que corresponde a um *array* de células (*cellarray*), cada célula

possui um atributo que indica a direção de fluxo. Esta direção pode ser para um dos seus oito vizinhos, considerando o espaço 2D.

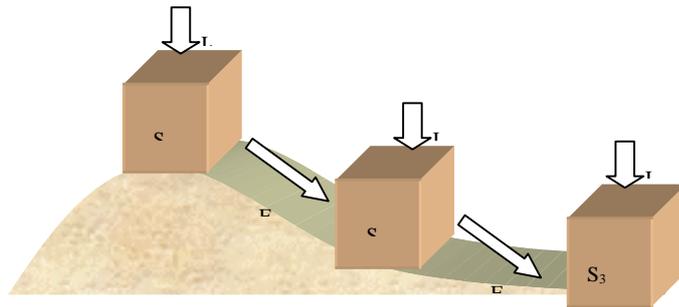


Figura 17 – Células com adjacência direcionada pela topologia (fonte: Burrough, 1998)

Para gerar a rede LDD existem vários algoritmos dentre os quais o D8 (*Deterministic algorithm*) destaca-se pela sua simplicidade (Burrough and McDonnell 1998). Neste algoritmo, a direção do fluxo é determinada pela direção mais inclinada dentro de uma janela 3x3 de células. Numa rede LDD existem dois tipos de células *upstream* e *target*. A célula *target* é a célula para onde todo o fluxo é direcionado. As demais células são chamadas *upstream*.

C_u ↓	C_u ↓	C_u ↓	C_u ↙	C_u ↙
C_u ↓	C_u ↓	C_u ↙	C_u ↙	C_u ↙
C_u ↘	C_u ↓	C_u ↙	C_u ↙	C_u ↙
C_u ↘	C_u ↓	C_u ←	C_u ←	C_u ←
C_u →	C_g ●	C_u ←	C_u ←	C_u ←

Figura 18 – Local Drain Direction (fonte: Burrough, 1998)

O material pode fluir por uma rede LDD a partir de diferentes funções, tais como:

- fluxo acumulado - calcula o novo estado dos atributos de uma célula, somando o valor original da célula mais a soma acumulada de todas as células *upstream*, cujo fluxo passa por esta célula;
- capacidade de transporte de uma célula - limita o fluxo de célula para célula a um atributo de capacidade de transporte fornecido em valores absolutos;
- fração de transporte - limita o fluxo sobre a rede a um parâmetro que controla a proporção de material que pode fluir por cada célula.

- valor limite – modifica o acúmulo de fluxo sobre a rede limitando o transporte de valores superiores a um determinado limite mínimo por célula.
- valor de disparo – permite o fluxo de material apenas se um valor de disparo for excedido.

8.4.2 PCRaster

Para demonstrar a aplicabilidade dos conceitos vistos nas seções anteriores, apresentaremos um exemplo de fenômeno físico que requer modelagem dinâmica para ser representado. Para modelar este fenômeno utilizaremos um aplicativo chamado PCRaster.

PCRaster é um *toolkit* para modelagem dinâmica que opera no modo matricial (*Raster*) e oferece um conjunto de ferramentas para análise espacial e temporal, funções para dispersão espacial e transporte sobre redes topológicas e um conjunto de métodos geoestatísticos para interpolação e simulação espacial. No PCRaster os resultados podem ser exibidos de forma dinâmica em 2 ou 3D. Para ilustrar o uso do PCRaster utilizaremos como exemplo um caso de escoamento de água da chuva em uma Bacia.

Para modelar este processo é necessário fornecer como entradas para o sistema o Modelo Numérico do Terreno (MNT) e as séries temporais com os dados de precipitação pluviométrica. A partir do MNT (Figura 20) é gerada a rede LDD, que é a rede de drenagem por onde a água excedente flui. A água excedente é toda a água que não foi infiltrada, por já ter excedido a capacidade de infiltração da célula. Para determinar o padrão espacial do processo de infiltração, um mapa de solos da área em estudo tem que ser fornecido. A partir destes dados, o programa é executado e gera um conjunto de mapas resultantes (Figura 19).

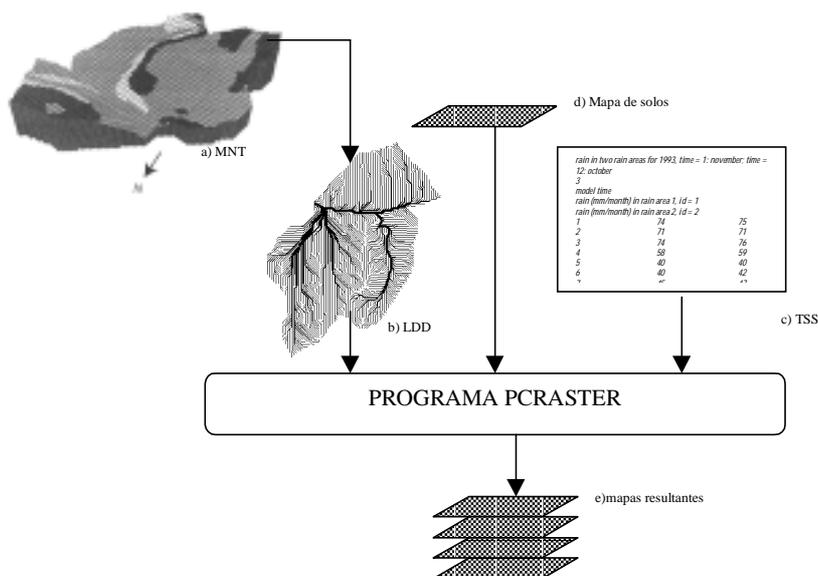


Figura 19 – Esquema simplificado das entradas e saídas do PCRaster

As séries temporais são arquivos `ascii`, contendo dados armazenados de forma tabular (Figura 21). Os dados fornecidos na séries temporais são utilizados para calcular o novo estado das células. Uma das funções de fluxo de material, comentadas na seção anterior, é selecionada pelo usuário para calcular o fluxo de água de célula para célula.

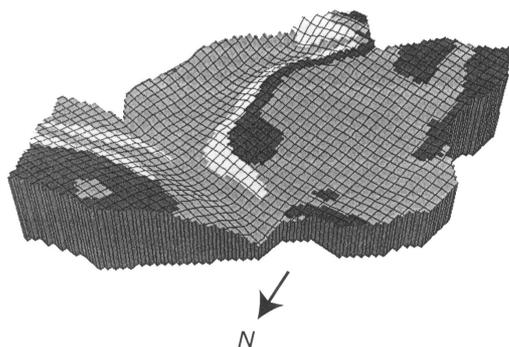


Figura 20 - MNT da área de estudo

```
rain in two rain areas for 1993, time =
1: november; time = 12: october
3
model time
rain (mm/month) in rain area 1, id = 1
rain (mm/month) in rain area 2, id = 2
1      74      75
2      71      71
3      74      76
4      58      59
5      40      40
6      40      42
7      45      47
8      62      64
9      80      78
10     80      85
11     75      79
12     67      69
```

Figura 21 – Série temporal dos índices pluviométricos

Um programa `PCRaster` (arquivos `.mod`) é organizado em cinco seções: *binding*, *areamap*, *timer*, *initial* e *dynamic*. A seção *binding* é onde são definidas as ligações entre as variáveis do programa e os arquivos. Estas ligações tem dupla direção, tanto podem determinar que as variáveis serão gravadas nos arquivos especificados (caso em que é executado um comando `report` na seção *dynamic*), quanto podem apenas indicar que as variáveis receberão valores provenientes dos arquivos especificados. Depois, na seção *areamap*, deve ser definido o formato geral dos mapas do modelo. Todos os mapas utilizados em um modelo devem ter o mesmo tamanho, localização geográfica e resolução. Na seção *timer*, o domínio de tempo do modelo é definido através de uma declaração que fornece os tempos inicial e final da execução do modelo, bem como o intervalo ou passo em que este tempo deve variar ao longo da execução do modelo. A seção *initial* é utilizada para inicializar as variáveis do programa. Esta seção é executada antes da primeira execução da seção *dynamic*. A seção *dynamic* é a parte principal de um programa `PCRaster`. Descreve as mudanças temporais das variáveis ou mapas do modelo. A principal característica desta seção é ser iterativa, isto é, é repetida, do início ao final, para todo o intervalo de tempo definido na seção *timer*. A seguir é apresentado um exemplo de programa `PCRaster` para calcular a precipitação da bacia, de nosso exemplo.

```

# model for simulation of rainfall
# one timeslice represents one month

binding
  RainTimeSeries=rain12.tss;# timeseries with rainfall (mm) per month
                                # for two rain areas
  Precip=rain;                  # reported maps with precipitation,
                                # rain is suffix of filenames
  RainAreas=rainarea.map;      # map with two rain areas

areamap
  clone.map;

timer
  1 12 1;

initial
  # this section is empty

dynamic
  # precipitation
  report Precip=timeinputscalar(RainTimeSeries,RainAreas);

```

Figura 22 – Programa PCRaster para calcular a precipitação

No programa acima, na seção `binding` as variáveis dos programas foram associadas a arquivos do Banco de dados. A declaração `RainTimeSeries=rain12.tss;` especifica que será utilizada a série temporal armazenada no arquivo `rain12.tss` (Figura 21). Depois, na declaração `Precip=rain;` é definido que o nome dos mapas de precipitação resultantes serão gravados em arquivos nomeados por `rain0000.xxx` (onde `xxx` varia de 001 a 012, porque a precipitação será calculada para 12 meses, conforme definido na seção `timer`). Na seção `timer` é especificado `1 12 1`, que significa que o programa deve executar a seção `dynamic` 12 vezes, a variável que controla estas repetições começa com o valor 1 e é incrementada no passo 1. Na seção `dynamic`, a precipitação é calculada através da expressão `report Precip = timeinputscalar (RainTimeSeries, RainAreas);` onde: `Timeinputscalar` é uma função que requer dois parâmetros: a série temporal e o mapa sobre o qual deve ser calculada a precipitação.

Depois de calculada a precipitação, pode-se facilmente estender o programa anterior para calcular a precipitação total em m^3/s . Para isto, basta incluir na seção `dynamic` a seguinte expressão:

```
report VolumePrecip=maptotal(Precip)*(cellarea()/2628);
```

onde $1/2628$ é o fator de conversão da área celular (Km^2) e `Precip`(mm/month) para m^3/s .

Nesta instrução a operação `maptotal` calcula a soma dos valores das células em `Precip`. Esta soma é multiplicada pela área de uma célula do mapa (`cellArea()`) e dividida pelo fator de conversão para m^3/s .

Para calcular a precipitação efetiva, a evapo-transpiração deve ser incluída no modelo. Assumindo que as condições do solo não influenciam na taxa de evapo-transpiração, pode-se calcular a evaporação para um mês (`Evap`, mm/mês) através da expressão:

$$\text{Evap} = K * \text{EvapRef}$$

onde:

- `EvapRef` é uma referência, um padrão da superfície do solo durante o mês em questão. Existe um valor diferente para cada mês e para cada classe de uso do solo.
- `K` é um coeficiente constante no tempo para uma classe de uso do solo.

Depois, calcula-se a precipitação excedente, através da expressão:

$$\text{PrecipSurplus} = \text{Precip} - \text{Evap};$$

Se a precipitação excedente for positiva em um mês, será adicionada ao solo. Se a quantidade máxima de água no solo for atingida, a parte restante do excedente não será mais adicionada no solo. Esta quantidade é chamada água excedente no solo e será escoada para o subsolo. Quando a precipitação excedente for negativa, a quantidade de água no solo será subtraída, em valores absolutos, pela precipitação excedente naquele mês.

Depois de calculada o balanço de água no solo, pode-se estender o modelo para modelar o escoamento de água na área em estudo. Isto é feito com o mapa de direção de drenagem local (`local drain direction map – ldd`).

Uma função de transporte de material tem que ser selecionada. Neste exemplo vamos utilizar a função de fluxo acumulado, que no PCRaster é implementada com o nome de `accuflux` e tem a seguinte sintaxe:

$$\text{Resultfluxmap} = \text{accuflux}(\text{lddmap}, \text{materialmap});$$

onde: `lddmap` é a rede `ldd` (Figura 23), `materialmap` é um mapa do material a ser transportado e `Resultfluxmap` é o mapa resultante. Na Figura 8.24 é apresentada a seção `dynamic` do programa PCRaster que implementa o modelo descrito.

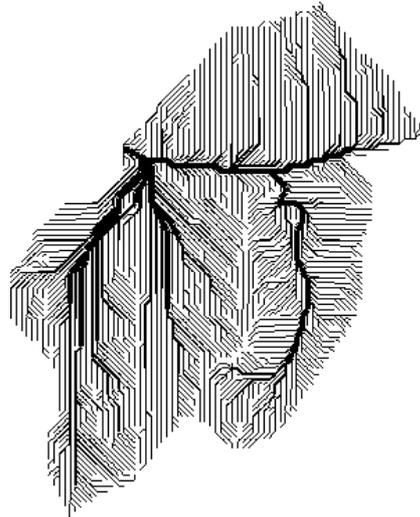


Figura 23 - LDD

```

dynamic
# precipitation
report Precip=timeinputscalar(RainTimeSeries,RainAreas);
# total volume precipitation over study area, in cubic metres per second
report VolumePrecip=maptotal(Precip)*(cellarea()/2628);
# reference evapotranspiration
EvapRef=timeinputscalar(EvapRefTimeSeries,1);
# evapotranspiration
report Evap=K*EvapRef;
# precipitation surplus
report PrecipSurplus=Precip-Evap;
# intermediate soilwater content: soilwater plus precipitation surplus
Soilwater=Soilwater+PrecipSurplus;
# soil water surplus (mm/month)
report SoilwaterSurplus=max(Soilwater-MaxSoilwater,0);
# soilwater content, no saturation
report Soilwater=min(Soilwater,MaxSoilwater);
# discharge in mm/month
DischargeMM=accuflux(Ldd,SoilwaterSurplus);
# discharge in metres3/second
report Discharge=DischargeMM*(cellarea()/2628);

```

Figura 24 – Seção dynamic de programa PCRaster para escoamento da água da chuva

8.5 MODELAGEM DINÂMICA DE PROCESSOS URBANOS

Na modelagem dinâmica de processos urbanos, os autômatos celulares são usualmente utilizados para modelar o uso do solo. Tradicionalmente, autômatos celulares são implementados segundo critérios estritamente locais, isto é, a dinâmica de aplicação das regras de transição baseiam-se principalmente na vizinhança de uma célula. Entretanto, em muitos casos de processos urbanos, a função de transição deve levar em conta diferentes fatores, incluindo: os efeitos da vizinhança, a qualidade do solo (fator ambiental), as taxas demográficas da região (fator social), a demanda por uma determinada atividade econômica e o comportamento dos agentes econômicos.

Nesta seção, para representar a modelagem dinâmica de processos urbanos serão apresentadas duas aplicações diferentes. A primeira consiste numa aplicação baseada nos princípios básicos de autômatos celulares, proposta por (Roy and Snickars 1996). A segunda aplicação, chamada Modelo Multi-Escala Integrado (Engelen 1995), apresenta uma estrutura sofisticada, capaz de integrar as variáveis sócio-econômicas e ambientais de sistemas urbanos.

8.5.1 *Citylife*

Numa tentativa de estudar a aplicabilidade de autômatos celulares na dinâmica urbana, (Roy and Snickars 1996) implementou o *Citylife*, baseado no *The Game of life*. No *Citylife* o espaço é representado como uma grade regular de células em que cada célula representa uma unidade do espaço ocupada por alguma atividade urbana típica, como por exemplo: área verde, residencial e trabalho. A partir de um estado inicial e um conjunto de regras de transição, o sistema cresce e evolui espacialmente.

Cada célula no sistema tem uma atratividade para cada tipo de atividade urbana definida pela função:

$$A_i(k) = \sum_l b(k,l) * a_i(l)$$

onde: $b(k,l)$ é um coeficiente que indica a probabilidade de uma célula do tipo k se transformar em uma célula tipo l (Tabela 1).

$a_i(l)$ “acessibilidade” da célula i para células contendo uma atividade do tipo l . Definida pela função:

$$a_i(k) = \sum_j \exp(-\mu(k) * d_{ij}(k)) * x_j(k) / N(k)$$

onde: $x_j(k) = 1$ se a célula j é utilizada para a atividade k , 0 caso contrário;

$d_{ij}(k)$ = distância da célula i para a célula j para uma atividade do tipo k

$\mu(k)$ = coeficiente de “acessibilidade” para uma atividade do tipo k

$N(k)$ = número de células contendo uma atividade do tipo k , onde \sum_j

$x_j(k)=N(k)$

Dado um estado inicial, o sistema aloca novas células para cada uma das atividades disponíveis. O número de células a ser alocado para cada atividade depende do número de células do estado inicial. Assim, por exemplo se no estado inicial são alocadas duas células para áreas verdes, duas células para área de trabalho e quatro células para áreas residenciais (Figura 25 a), então a cada geração o sistema alocará mais duas células para áreas verdes, mais duas células para áreas de trabalho e quatro células adicionais para áreas residenciais. O critério para seleção de uma célula é a sua atratividade para a atividade, será selecionada a célula com maior atratividade ($A_i(k)$).

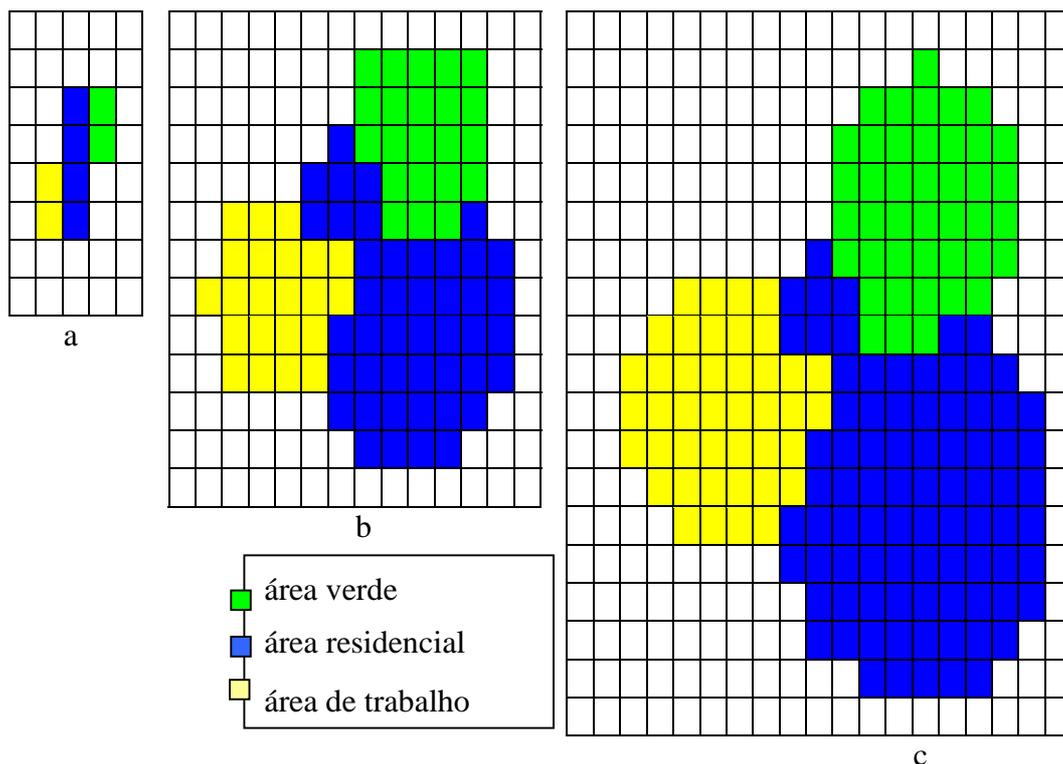


Figura 25 – Citylife a) estado inicial b) após 10 gerações c) após 20 gerações (fonte: Roy and Snickars 1996).

O mecanismo de expansão do Citylife é considerado evolucionário (estado futuro do sistema depende da trajetória seguida) e baseado na competição entre as células (Roy and Snickars 1996).

Para o exemplo apresentado na Figura 25 adotou-se os seguintes coeficientes de interação entre as atividades:

	Verde	Residencial	Trabalho
Verde	1	0	0
Residencial	0	1	0
Trabalho	0	0	1

Tabela 1 – Coeficientes de probabilidade

Uma restrição do *citylife* é considerar que uma vez que uma célula for ocupada por uma atividade ela permanecerá nesta atividade. Desta forma, este sistema adota um modelo dinâmico espacial de difusão por expansão e não de realocação.

8.5.2 Modelo Multi-Escala Integrado

Como já mencionado, fatores sócio-econômicos são inerentes ao planejamento urbano. Na literatura recente, verifica-se uma tendência de propostas de extensões ao modelo de autômato celular clássico, visando integrar fatores ambientais e sócio-econômicos, para representar a dinâmica espacial de fenômenos urbanos.

Entre estas propostas destaca-se a de Engelen (1995), que apresenta uma estrutura de modelagem dinâmica e de suporte a decisão capaz de operar em uma variedade de escalas. Esta estrutura é constituída de dois níveis denominados macro e micro escalas. Na macro escala estão representadas as variáveis ecológicas e sócio-econômicas que afetam o sistema como um todo. A micro escala representa a dimensão espacial do modelo. Estas escalas interagem intensivamente entre si e com um Banco de Dados Geográfico, a partir do qual obtêm os dados necessários para as simulações (Figura 26).

A macro escala possui três componentes representando os subsistemas natural, econômico e social. Estes sub-modelos estão conectados através de uma rede de influência mútua e recíproca. O subsistema natural representa condições ambientais tais como temperatura, precipitação e poluição. O subsistema social inclui dados demográficos como nascimentos, morte e migração. O subsistema econômico é fortemente determinado pelas mudanças ocorridas no subsistema natural e pelas demandas sociais. Neste sentido, ele pode gerar demandas como, por exemplo, a necessidade por mais células residenciais quando a população aumenta.

A micro escala consiste em um autômato celular sobre o qual são aplicadas regras de transição para calcular as mudanças no uso do solo.

Para ilustrar o uso deste modelo, consideraremos dados de um estudo para analisar os impactos de mudanças climáticas em uma ilha do Caribe. Estes dados estão disponíveis na homepage do RIKS (*Research Institute for Knowledge Systems* – www.riks.nl).

Neste exemplo, a macro escala inclui no subsistema natural apenas mudanças climáticas, no subsistema social inclui dados relativos a população, nascimentos e mortes e no subsistema econômico as demandas geradas a partir da interação deste subsistema com os demais (Figura 27).

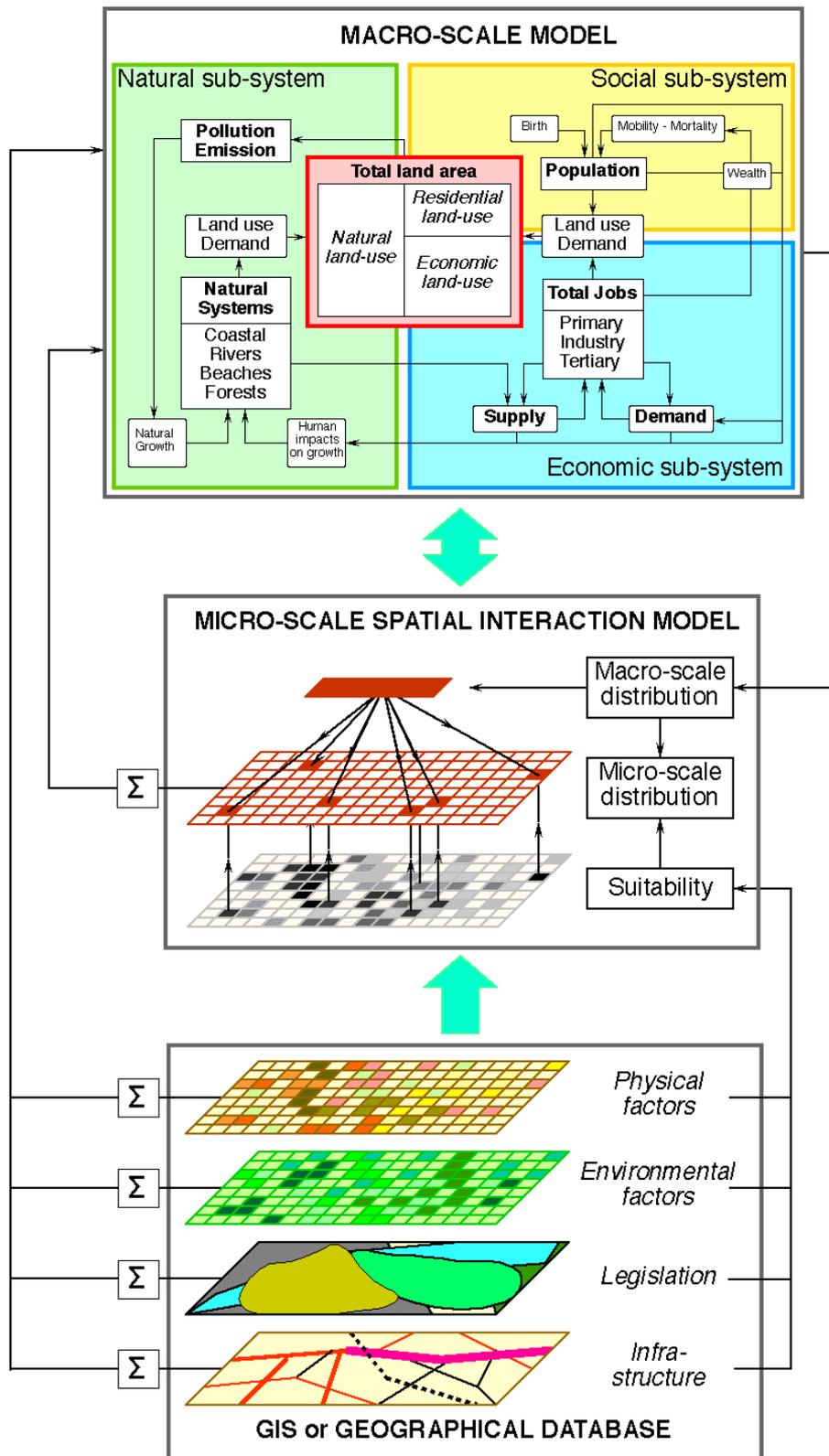


Figura 26 Integração entre o Modelo Multi-Escala e GIS fonte:(Engelen 1995) .

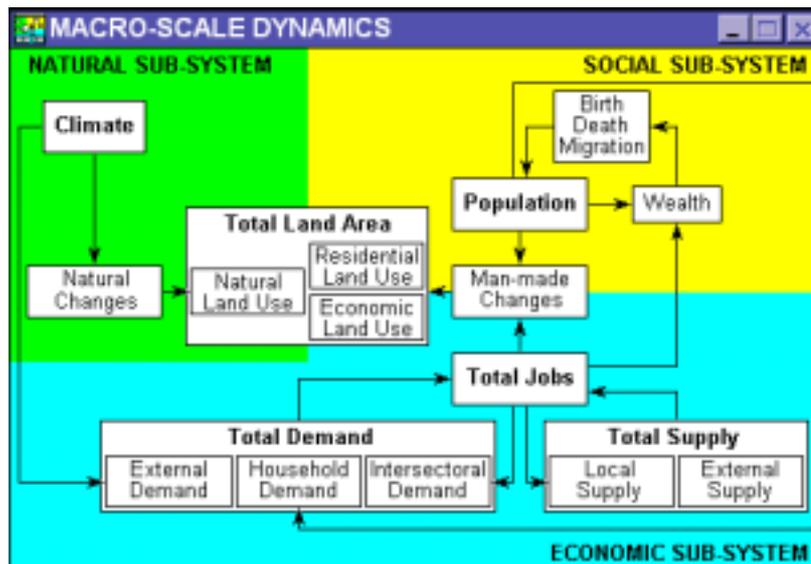


Figura 27 – A macro escala (fonte: Engelen 1995)

As condições climáticas (*climate*) do modelo são definidas a partir de variáveis ambientais tais como temperatura, precipitação e nível do mar, e das relações de influência existentes entre elas.

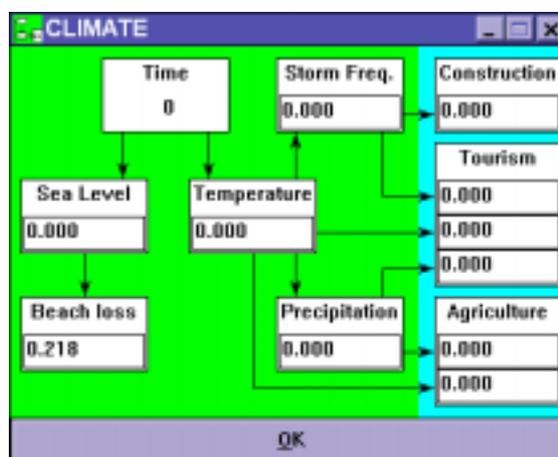


Figura 28 – O Subsistema Natural

A Figura 28 mostra que variações na temperatura e no nível do mar afetam as demais variáveis. Estas relações de influência são também expressas de forma explícita, através de gráficos e tabelas e podem ser manipulados pelo usuário de forma independente e interativa. Esta funcionalidade caracteriza este modelo como um

modelo exploratório, pois permite que o usuário avalie um fenômeno a partir de diferentes cenários (*what if questions*).

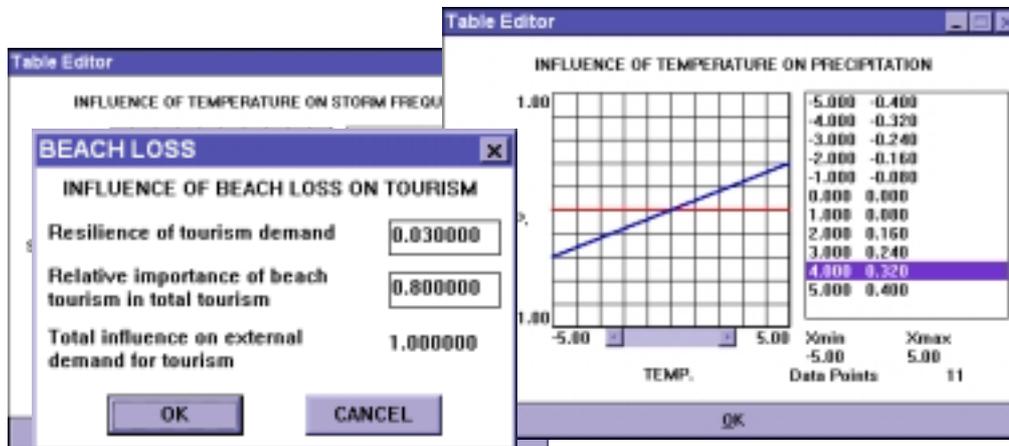


Figura 29 – Os componentes do clima

As variáveis sociais, de forma análoga às naturais, podem ser manipuladas de forma interativa e independente. Entretanto, as variáveis econômicas são geradas a partir do comportamento dos subsistemas natural e social. Para o cálculo destas variáveis são utilizados coeficientes para medir o crescimento populacional e da oferta de empregos, por exemplo, e determinar o espaço necessário (demanda do solo) para acomodar as atividades econômicas (turismo, indústrias) afetadas por estes coeficientes.

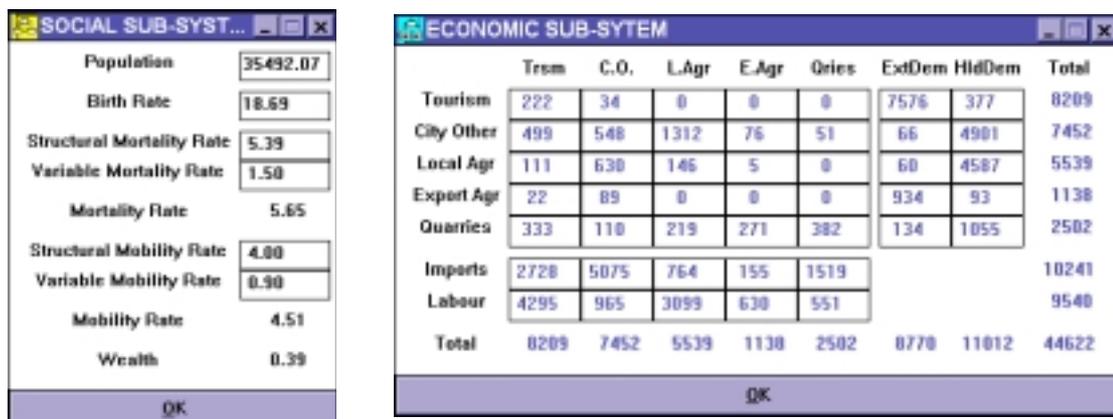


Figura 30 – Subsistemas Social e Econômico

Na micro escala, cada estado de célula representa uma categoria de uso do solo (Figura 31). Os estados são divididos em duas categorias: funções e feições. Funções são usos do solo ativos, tais como residencial, floresta, comercial. Em princípio, uma célula função pode mudar para qualquer um dos estados possíveis. Feições são usos do solo fixos, tais como rios, parques e aeroportos. Embora as feições não estejam sujeitas às mudanças geradas pelas regras de transição do autômato celular, eventualmente elas podem ser convertidas através de um processo especial ou uma intervenção exógena. Feições aparecem como argumentos das regras e podem afetar a transição de células vizinhas. Assim, por exemplo, a existência de um parque pode influenciar a transição de células vizinhas em células residenciais (White and Engelen 1997).

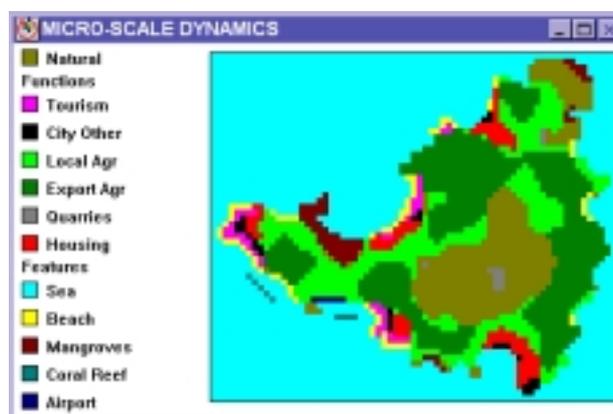


Figura 31 – A micro escala

A demanda do solo para as várias atividades é fornecida pela macro escala, de acordo com um mecanismo baseado em três classes de prioridades. As regras de prioridade um são intervenções do usuário como, por exemplo, a inclusão de um aeroporto. As regras de prioridade dois são regidas pelo subsistema natural e geram certas transições diretamente, sem interferência do autômato celular. Por exemplo, se o nível do mar sobe, células com baixa elevação são convertidas em praias ou mangues (White and Engelen 1997). As regras de prioridade três se aplicam às células ativas (funções). Para cada célula ativa é calculado um vetor de potencialidades, em que cada potencialidade representa o grau de atração de uma célula para um determinado estado (z). O potencial (P_z) de uma célula depende de três fatores:

- a adequabilidade da célula para a atividade z (S_z)
- efeito agregado das células na vizinhança (N_z)
- uma perturbação estocástica (ϵ_z)

Estes fatores se relacionam da seguinte forma:

$$P_z = S_z N_z + \epsilon_z$$

O efeito agregado da vizinhança (N_z) leva em consideração principalmente a localização das células. A vizinhança de uma célula é uma região circular com um número variável de células, organizadas em zonas de distâncias. Assim, a fórmula para calcular o efeito agregado de vizinhança consiste em:

$$N_z = \sum L_{d,i} W_{z,y,d}$$

onde:

- $W_{z,y,d}$ parâmetro de peso aplicado a células no estado y na distância d
- i índice das células na zona de distância d
- $L_{d,i}$ 1 se a célula i na distância d está no estado y ; 0, caso contrário.

A regra de transição estabelece que cada célula ativa é convertida para o estado para o qual seu potencial é maior, mas até que a demanda por células deste estado seja atendida. Depois deste ponto, nenhuma outra célula é convertida para este estado. Os potenciais para tal estado são ignorados nas conversões subsequentes.

Os resultados das simulações são apresentados de forma dinâmica na tela do computador, isto é, o usuário acompanha todas as transições. Na Figura 32, são apresentados os resultados de uma simulação para 40 anos em que se trabalhou com dois cenários. No primeiro cenário, Figura 32 a, assumiu-se que não haveria mudanças climáticas, ou seja, a temperatura e o nível do mar se manteriam ao longo do tempo. No segundo cenário, Figura 32 b, considerou-se que a temperatura aumentaria em 2C e o nível do mar em +20cm. Em ambos os casos partiu-se da configuração inicial apresentada na Figura 31 e considerou-se que a população apresentaria um crescimento de 2% ao ano e um número total de 11000 vagas de empregos, distribuídas em diferentes atividades.

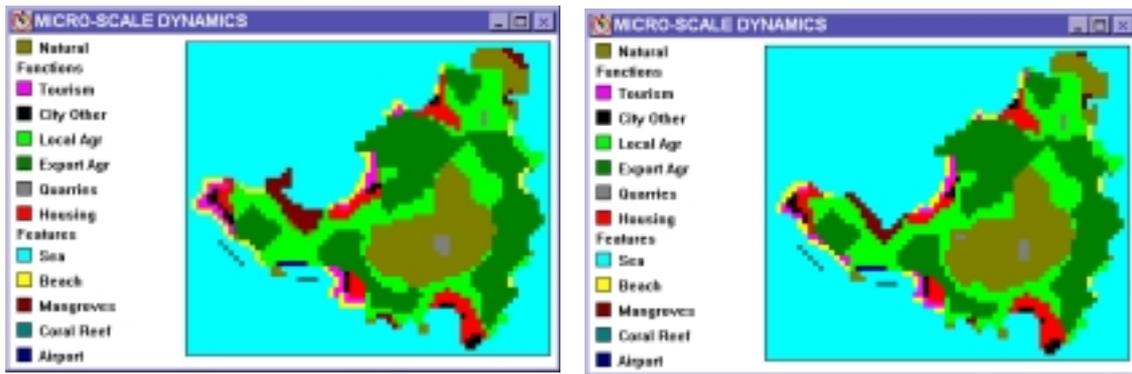


Figura 32 – Simulações a) sem mudanças climáticas b) com mudanças climáticas

Na Figura 32 b, simulada com o cenário de mudanças climáticas, pode-se observar uma redução da área das praias e mangue, como consequência do aumento do nível do mar.

8.6 CONSIDERAÇÕES FINAIS

Este capítulo teve por objetivo apresentar os principais conceitos e aspectos computacionais envolvidos em sistemas de modelagem espacial dinâmica. Na dimensão conceitual, verificou-se que cada um dos elementos chaves de um modelo dinâmico, tais como espaço, tempo e modelo matemático permitem diferentes representações computacionais. A escolha de uma forma de representação para um destes elementos afeta os demais, uma vez que as escalas de todos os elementos devem ser integradas.

No contexto computacional, foi explorado a solução baseada em autômatos celulares. No estudo desta abordagem de implementação, verificou-se que processos físicos e urbanos possuem mecanismos distintos para aplicação de regras de transição. Enquanto os processos físicos podem ser descritos por modelos determinísticos, os processos urbanos são caracterizados como processos estocásticos e são altamente influenciados por variáveis exógenas. Nos processos físicos, pode-se considerar a topologia do terreno, o que, apesar das críticas ao modo como as redes de drenagem LDD são geradas (discretização do fluxo em 45°, introdução de artefatos, (Burrough and McDonnel 1998)), produz padrões espaciais coerentes com os que acabam se desenvolvendo naturalmente no mundo real. Quanto à modelagem de processos urbanos, sistemas como os desenvolvidos pelo RIKS (www.riks.nl) apresentam grande flexibilidade para a inclusão de variáveis que aumentam a precisão das predições,

entretanto, são sistemas de estrutura complexa, de difícil entendimento e implementação.

Não existe uma solução única para modelar fenômenos espaciais dinâmicos. A solução ideal para cada caso deve ser buscada tentando responder o porque, onde e quando de cada fenômeno, através da integração das escalas temporal e espacial articulada com o modelo matemático definido para descrever o fenômeno e prever sua evolução.

Cadeias de Markov, modelos logísticos de Difusão e Regressão são eficientes em modelar processos estacionários mas são desprovidos de funcionalidades específicas para a representação espacial. Entretanto, estes modelos utilizam equações matemáticas simples e requerem poucos dados, além de serem compatíveis com o formato de dados oriundos de fontes de sensoriamento remoto e, como consequência, facilmente implementados em GIS.

Modelos de Sistemas são classificados como modelos exploratórios, porque fornecem condições para que várias simulações possam ser investigadas a partir de diferentes cenários. Entretanto, estes modelos requerem um profundo nível de conhecimento do fenômeno em estudo e acabam por se tornar sistemas altamente especializados, não podendo ser aplicados a outras classes de fenômenos.

Diferentes modelos servem a diferentes propósitos, logo eles não são excludentes, mas sim complementares. Nesta perspectiva, Lambin (1994) sugere que ao se construir um modelo deve-se fazê-lo de forma gradual, começando por Cadeias de Markov, que são os mais simples, e ir incorporando novos elementos (variáveis exógenas) e funções (determinísticas) ao projeto.

A dimensão espacial deve ser também introduzida de forma gradual, começando com as relações espaciais mais elementares como as de vizinhança, refinando continuamente, de forma a contemplar a noção de espaço relativo e suporte a representações em múltiplas escalas.

Modelos espaciais dinâmicos construídos com esta visão de projeto devem ser capazes de representar de forma realista os fenômenos dinâmicos encontrados na natureza, superando as limitações dos modelos atuais, baseados em concepções limitadas quanto às representações do espaço, do tempo e dos processos.

Referências Bibliográficas

- Batty, M. (1999). Modeling urban dynamics through GIS-based cellular automata. Computers, Environment and Urban Systems **23**: 205-233.
- Batty, M. (2000). GeoComputation Using Cellular Automata. GeoComputation. S. Openshaw and R. J. Abraham, Taylor&Francis: 95-126.
- Burrough, P. (1998). Dynamic Modelling And Geocomputation. Geocomputation: A Primer. P. Longley, M. Batty and R. McDonnel. London, John Wiley & Sons.
- Burrough, P. and R. McDonnel (1998). Principles of Geographical Information Systems. Oxford, OUP.
- Câmara, A. S. (1996). Spatial Simulation Modelling. Spatial Analytical Perspectives on GIS. M. Fisher. London, Taylor & Francis: 213-218.
- Couclelis, H. (1997). "From Cellular Automata to Urban Models: New Principles for Model Development and Implementation." Environment and Planning B: Planning and Design **24**: 165-174.
- Couclelis, H. (1999). Space, Time, Geography. Geographical Information Systems. P. Longley, M. Goodchild, D. Maguire and D. Rhind. New York, John Wiley: 29-38.
- Edelweiss, N. and J. P. M. Oliveira (1994). Modelagem de Aspectos Temporais de Sistemas de Informação. Recife, UFPE-DI.
- Engelen, G. (1995). "Using Cellular Automata for Integrated Modelling of Socio-environmental Systems." Environmental Monitoring and Assessment **34**: 203-214.
- Lambin, E. F. (1994). Modeling Deforestation Processes - A Review, Trees series B: Research Report . European Commission, Luxembourg.
- Reis, E. J. and S. Margulis (1991). Options for Slowing Amazon Jungle Clearing. Global warming: economic policy responses. R. Dornbusch and J. M. Poterba. Cambridge, The MIT Press: 335-375.
- Roy, G. G. and F. Snickars (1996). Citylife: A study of cellular automata in urban dynamics. Spatial Analytical Perspectives on GIS. M. Fisher. London, Taylor & Francis: 213-218.
- Santos, M. (1996). A Natureza do Espaço: Técnica e Tempo, Razão e Emoção. São Paulo, HUCITEC.
- Soares Filho, B. S. (1998). Modelagem dinâmica de paisagem de uma região de fronteira de colonização amazônica. Escola Politécnica. São Paulo, Universidade de São Paulo.
- Voigtmann, A. B., L.; Hinrichs K.H. (1996). Temporal Extensions for an Object-Oriented Geo-Data Model. Spatial Data Handling - 1996 (SDH'96).

White, R. and G. Engelen (1997). "Cellular Automata as the Basis of Integrated Dynamic Regional Modelling." Environment and Planning B: Planning and Design **24**: 165-174.

Worboys, M. F. (1995). GIS - A Computing Perspective. Bristol, PA, Taylor & Francis Inc.