CAPÍTULO 2
MODELAGEM AMBIENTAL EM SISTEMAS DE INFORMAÇÃO GEOGRÁFICA COM TRATAMENTO DE INCERTEZAS

2.1 INTRODUÇÃO

Cada vez mais, os computadores vêm sendo utilizados como ferramentas de apoio a procedimentos de estudos, de análises e de simulações em vários campos do conhecimento humano. Sistemas complexos para análises e para modelagens foram desenvolvidos para se trabalhar dados relacionados com áreas específicas como finanças, transportes, geologia, solos, etc.. Nessa mesma tendência, sistemas de armazenamento, manipulação e apresentação de dados espaciais, conhecidos como SIG, foram criados e estão sendo utilizados no campo das ciências ambientais. Modelos matemáticos, aritméticos e lógicos, buscando representar propriedades e processos do meio físico natural, têm sido implementados, nos SIG, com o objetivo de facilitar o seu estudo e compreensão para que se possa atuar sobre o meio ambiente de forma responsável e cooperativa. Para que isto ocorra, este trabalho defende a idéia de que é imperativo que se considere o tratamento das incertezas das representações dos dados espaciais envolvidos em um modelo matemático ambiental apoiado em SIG. As incertezas destas representações, uma vez propagadas para os resultados finais das modelagens, qualificam os produtos gerados em SIG para o apoio efetivo aos processos de tomadas de decisão baseados nesses produtos.

Considerando-se esse contexto os objetivos deste capítulo são:

- apresentar conceitos importantes relacionados com a representação dos dados espaciais e com seu uso em modelagem ambiental produzida utilizando-se da tecnologia de SIG;

- tratar os temas de estimativa de incertezas para dados espaciais e de propagação destas incertezas em modelagens computacionais desenvolvidas em SIG e;

- apresentar, analisar e propor o uso do paradigma da modelagem geoestatística por indicação para modelagem computacional em SIG envolvendo dados espaciais com atributos temáticos e numéricos.

O presente capítulo tem a seguinte organização:

- na seção 2.2 são apresentados os aspectos gerais da natureza da modelagem ambiental, tratados neste trabalho, os conceitos gerais sobre SIG, os detalhes da modelagem de dados espaciais, com e sem inclusão de incertezas, e as características importantes da modelagem computacional no ambiente de um SIG;

- na seção 2.3 discute-se o problema da propagação de incertezas em modelagens computacionais, com a descrição de alguns modelos de propagação envolvendo atributos de natureza numérica.

- a seção 2.4 apresenta um histórico evolutivo da modelagem computacional enfatizando-se a necessidade de se adotar o paradigma da modelagem geoestatística por indicação para o efetivo uso de modelagem computacional e;

- a seção 2.5 encerra este capítulo tecendo algumas conclusões.

2.2 MODELAGEM AMBIENTAL E SISTEMAS DE INFORMAÇÃO GEOGRÁFICA

A potencialidade principal de um SIG está na sua capacidade de realizar análises complexas a partir da integração, em uma base de dados única, de representações de dados espaciais. Como já definido, na seção 1.2, um dado espacial é caracterizado por sua posição espaço-temporal, (x, y, z, t), e por atributos a ele associados. Um dado de temperatura, por exemplo, pode ser observado em várias posições do espaço tridimensional (x, y, z) e em diferentes períodos de tempo t. As medidas de temperatura são os valores do atributo observados em posições desse sistema espaço-temporal.

Os procedimentos de análise espacial, desenvolvidos no ambiente de um SIG, possibilitam, no estágio tecnológico atual, a análise de processos, alguns simples e outros mais complexos, do mundo real. Para isto é necessário a criação de modelos ambientais, que representem adequadamente o fenômeno[1] natural em estudo. Assim, a *modelagem ambiental*, no contexto deste trabalho, consiste na criação de modelos matemáticos, determinísticos ou estocásticos, que relacionam atributos ambientais na tentativa de representar o comportamento de um processo ocorrendo na natureza. Os modelos ambientais são, então, transformados em *modelos computacionais* para serem executados no ambiente de um SIG.

2.2.1 A natureza da modelagem ambiental em meio digital

2.2.1.1 A dimensão espaço-temporal e a natureza contínua dos processos da natureza

As pessoas observam o mundo, diretamente ou com auxílio de instrumentos, e percebem fenômenos da natureza que se modificam com maior ou menor intensidade em função do tempo e do espaço (Burrough e McDonnell, 1998), ou seja, apresentam uma *variação contínua espaço-temporal*. Segundo Steyaert, 1993, os processos ambientais no mundo real são, tipicamente, tridimensionais, dependentes do tempo e complexos. Essa complexidade pode incluir comportamento não linear, componentes estocásticos e realimentações em múltiplas escalas de tempo e de espaço. Assim sendo, a natureza espaço temporal dos fenômenos ambientais, devido a sua importância para compreensão do fenômeno em estudo, coloca algumas questões à modelagem ambiental. Estas questões se referem a criação de modelos matemáticos e

[1] Os termos fenômeno e processo são usados como sinônimos neste texto com o significado de sequência de estados de um sistema que se transforma.

computacionais com representações, uma vez que o fenômeno deve ser, de alguma forma, representado.

A modelagem deve considerar que os processos da natureza resultam de interações espaço-temporais complexas entre os diversos elementos que os compõem, ou seja, as propriedades ambientais*.* No modelo matemático de um processo, as propriedades ambientais são tratadas como variáveis do modelo enquanto que suas inter-relações são representadas por operações aritméticas ou lógicas.

A representação digital envolve procedimentos de discretizações e quantizações, nos domínios do espaço-tempo e do atributo, para se obter estruturas computacionalmente manipuláveis. Essas estruturas representam digitalmente o comportamento do atributo no espaço-tempo e são usadas diretamente pelos modelos computacionais.

2.2.1.2 O mundo contínuo e o mundo discreto dos computadores digitais

Como já citado, os dados espaciais, assim como os processos ambientais, apresentam a característica de continuidade no espaço e no tempo. Por outro lado, atualmente, a tecnologia de computadores digitais vem sendo intensivamente utilizada para representar dados espaciais e simular processos ambientais com o objetivo de melhor compreendê-los para, então, atuar sobre o meio ambiente de forma cooperativa. Por exemplo, modelos de simulação ambiental fornecem diagnósticos e saídas preditivas que podem ser combinadas com dados sócio-econômicos para avaliação de riscos ambientais regionais e locais ou decisões relacionadas com o gerenciamento de recursos naturais (Steyaert, 1993).

Entretanto, os computadores digitais trabalham com um número fixo de campos, ou bytes, para armazenar qualquer valor numérico. Isto vale para os valores das posições espaciais e também para os valores dos atributos relacionados aos dados espaciais. Assim, uma base de dados computacional contém aproximações, ou representações simplificadas, dos dados ou fenômenos espaciais. Goodchild e Guoging, 1992, argumentam que: “por que uma base de dados computacional é um armazenador discreto e finito, é necessário amostrar, abstrair, generalizar ou, até mesmo, comprimir informações do mundo real”.

2.2.1.3 Representações de continuidade a partir de dados amostrais

As observações de propriedades e processos da natureza contam, atualmente, com um conjunto cada vez mais numeroso de tecnologias. As imagens de sensoriamento remoto orbital, por exemplo, possibilitam acompanhar diversos processos ambientais, tais como, as modificações na paisagem por ações humanas de engenharia e agricultura, os aumentos de áreas desertificadas, erodidas ou reflorestadas, o tamanho e a direção de crescimento de áreas urbanas, a extensão de áreas inundadas ou queimadas, etc.. Equipamentos de medidas com grande grau de acurácia, tais como, os sistemas de posicionamento globais, GPS, as plataformas de coletas de dados meteorológicos e outros, são hoje, cada vez com mais frequência, utilizados para se obter amostras da distribuição de um atributo espacial dentro de uma área de interesse.

O computador digital não pode armazenar um atributo de forma contínua e, também, por razões de custo de aquisição, os atributos da natureza são representados digitalmente por *conjuntos de amostras* observadas, direta ou indiretamente, do atributo. A quantidade e a distribuição das amostras, que são coletadas, devem representar a variação do atributo dentro de uma região espacial de interesse. Valores de máximo, de mínimo e de inflexão relacionados aos atributos de um dado espacial são informações obrigatórias de um conjunto amostral.

Os dados amostrais podem ser pontuais, quando o valor de cada dado amostrado está relacionado exclusivamente a uma posição observada. Por exemplo, o valor da elevação, em relação ao nível do mar, em uma posição do terreno, para levantamento topográfico de uma região, só é válido para essa localização. Por outro lado, um dado amostral por área tem um valor do atributo que é válido para uma área predeterminada. Um pixel de uma imagem de sensoriamento remoto representa uma amostra de uma região do espaço definida pela resolução espacial do sensor. Este trabalho supõe a existência de amostragens pontuais para os atributos a serem modelados por procedimentos geoestatísticos de krigeagem e simulação.

2.2.2 Modelagem de dados espaciais em SIG

Existe uma grande quantidade de definições para SIG, cada uma delas baseada no tipo de usuário e no domínio da aplicação (Maguirre et al., 1991). A metodologia de banco de dados define o SIG como *um banco de dados geográficos não convencional que possibilita gerenciamento de dados espaciais*. A visão orientada a processos considera o SIG como *uma coleção de subsistemas integrados, onde dados passam por uma sequência de procedimentos de transformação*. A definição de aplicação ou utilização conceitua o SIG de acordo com *o tipo de problema a ser solucionado e o tipo de dado manipulado* (De Oliveira et al., 1997). Burrough e McDonnell, 1998, definem o SIG como "*um conjunto poderoso de ferramentas para coleta, armazenamento, recuperação, transformação e apresentação de dados espaciais do mundo real para um conjunto particular de propósitos*". Esta definição considera o SIG como uma caixa de ferramentas computacionais para se trabalhar com dados espaciais. No contexto deste trabalho, é esta a definição que será enfocada. Segundo Câmara e Medeiros, 1998, como instrumentos computacionais para geoprocessamento, os SIG permitem a realização de análises complexas ao integrar dados de diversas fontes e ao criar banco de dados georreferenciados.

Um dado espacial, para ser trabalhado dentro de um SIG, deve ser representado por um modelo computacional. Um modelo é uma abstração de fatos ou de entidades do mundo real. A modelagem de dados geográficos é o processo de discretização que converte a realidade geográfica complexa em um número finito de registros ou objetos (Goodchild, 1993a).

Num nível de abstração mais alto, ou *nível conceitual*, o mundo real é representado segundo duas visões complementares: *o modelo de campos* e o *modelo de objetos* (Goodchild, 1993a, Câmara e Medeiros, 1998). É este o modelo que será adotado neste trabalho.

O modelo de campos representa os dados espaciais cujos atributos têm distribuição contínua no espaço. Um campo é formalizado como uma função matemática cujo domínio é uma região geográfica e cujo contradomínio é o conjunto de valores que o campo pode tomar (Câmara et al., 1996b). Os campos podem ser especializados em *campos numéricos*, quando o contradomínio pode conter um número infinito de valores numéricos e *campos temáticos*, quando o conjunto de valores do contradomínio é um conjunto limitado de classes, ou temas. Dados de elevação e temperatura são exemplos de campos numéricos, enquanto que classes de solos e classes de vegetação são exemplos típicos de campos temáticos.

O modelo de objetos é utilizado, tipicamente, para representar facilidades construídas pelo homem. Este modelo considera que o mundo está povoado por um conjunto de objetos, ou entidades. Os objetos têm identificações únicas e são caracterizados por suas propriedades geométricas e topológicas e por valores de atributos não espaciais. Neste modelo, uma posição espacial pode, até mesmo, ser ocupada por um ou mais objetos. Tubos de água ou de esgoto, prédios, ginásios esportivos e postes de eletricidade são entidades do mundo real modeladas como objetos.

Este trabalho vai lidar com propriedades ambientais cujas variáveis podem ser representadas no modelo de campos. A informação inicial se constitui de dados de propriedades ambientais coletados em forma de amostras pontuais georeferenciadas no espaço.

Amostras de dados espaciais, modelados como campos, podem ser espacializadas utilizando-se algoritmos de inferência ou interpolação. O *procedimento de espacialização* consiste em inferir valores do atributo, em posições não amostradas, a partir dos valores observados. Dessa forma, pode-se obter o valor do atributo em qualquer posição do espaço, ainda que a representação por campo seja discreta. É muito comum, no ambiente de um SIG, a criação de estruturas de representação por *grades regulares retangulares* onde os valores do atributo nos vértices da grade são obtidos por procedimentos de interpolação local a partir do conjunto de amostras.

Os algoritmos para inferência de atributos de dados espaciais, representados por amostras pontuais, podem ser classificados em dois (2) tipos básicos: *interpoladores determinísticos* e *interpoladores estocásticos*.

Os algoritmos de interpolação determinísticos mais utilizados na prática são os de média móvel ponderada (Burrough, 1987, Felgueiras, 1987, McCullagh, 1988), que têm a seguinte formulação geral:

$$z^*(\mathbf{u}) = \frac{\sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u})\, z(\mathbf{u}_\alpha)}{\sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u})} \quad \boldsymbol{e} \quad \lambda_\alpha(\mathbf{u}) = 1 / d_\alpha(\mathbf{u}) \tag{2.1}$$

onde $z^*(\mathbf{u})$ é o valor do atributo inferido numa posição **u** do espaço, $z(\mathbf{u}_\alpha)$ o valor do atributo na amostra vizinha α, *onde* α=1,..,n(**u**) com n(**u**) igual ao número total de amostras vizinhas de **u,** e $\lambda_\alpha(\mathbf{u})$ é o valor de ponderação relativo à amostra α e à posição **u**. Em geral, o valor de ponderação é definido igual ao inverso, simples ou ao quadrado, da distância euclidiana, $d_\alpha(\mathbf{u})$, da amostra α à posição **u**. Estes interpoladores são denominados locais pois utilizam um subconjunto das amostras na vizinhança de cada ponto a ser interpolado. Variações deste algoritmo também são usadas, tais como: interpolação por vizinho mais próximo, quando o número de vizinhos é igual a

1; interpolação por média simples, quando os pesos de interpolação são todos iguais a 1; interpolação por quadrantes, quando as amostras vizinhas são escolhidas por quadrantes no sentido de resolver problemas causados por amostras aglomeradas e; interpolação por cotas, quando as amostras estão representadas por curvas de nível e se deseja evitar tendenciosidade em relação à uma das curvas.

Os valores inferidos pelos interpoladores determinísticos são tratados como dados sem erros, ou seja, os valores obtidos não estão contaminados por erros nos dados de entrada ou pelo algoritmo de interpolação. Esta é a grande desvantagem dos interpoladores determinísticos em relação aos estocásticos.

Os interpoladores estocásticos utilizam as ferramentas da geoestatística para inferências de valores de atributos com estimativas de incertezas. A geoestatística considera que a distribuição espacial do atributo define uma função aleatória, FA, dentro de uma região de interesse A (Delfiner e Delhomme, 1975, Isaaks e Srivastava, 1989, Journel, 1988, Cressie, 1991, Goovaerts, 1997, Deutsch e Journel, 1998). O valor *z* do atributo em estudo, em uma localização qualquer de A, é considerado uma realização de uma VA *Z*, cuja distribuição de probabilidade modela a incerteza local a respeito de *z*.

Os procedimentos inferenciais da geoestatística, conhecidos como métodos de *krigeagem*, baseiam-se na análise e na modelagem da variabilidade espacial do atributo a partir de um conjunto amostral pontual desse atributo. Supõem, ainda, a hipótese de estacionariedade de segunda ordem para a propriedade que está sendo modelada, ou seja, a média é constante, em todas as posições do campo, e a covariância só depende da distância entre as amostras. *A krigeagem ordinária* é o estimador estocástico mais utilizado na prática e tem a seguinte formulação:

$$\boxed{z_O^*(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{O\alpha}(\mathbf{u}) \cdot z(\mathbf{u}_\alpha)} \qquad (2.2)$$

onde $z_0^*(\mathbf{u})$ é o valor do atributo inferido, por krigeagem ordinária, numa posição **u** do espaço, $z(\mathbf{u}_\alpha)$ o valor do atributo na amostra vizinha α, *onde* α=1,..,$n(\mathbf{u})$ e $n(\mathbf{u})$ é o número total de amostras vizinhas de **u,** e $\lambda_{0\alpha}(\mathbf{u})$ é o valor de ponderação da krigeagem ordinária relativo à amostra α e à posição **u**. A soma dos valores de ponderação deve ser igual a 1, para evitar tendenciosidade do estimador. Os pesos da krigeagem, diferentemente dos pesos do interpolador por média móvel, são obtidos a partir da hipótese de minimização da variância do erro de estimação e a partir do estudo de variabilidade do atributo, ou, mais formalmente, a partir de modelos de covariância inferidos sobre o conjunto amostral do atributo. Além disso, os procedimentos de krigeagem possibilitam a estimativa de incertezas relacionadas aos valores inferidos para o atributo.

Um valor observado, ou sorteado, é conhecido como uma *realização* de uma variável aleatória. A *simulação estocástica*, outro procedimento geoestatístico, se utiliza da krigeagem para obtenção de realizações de variáveis e campos aleatórios. Os valores obtidos, em várias realizações de uma mesma VA, possibilitam inferências de valores e estimativas de incertezas para o atributo.

O capítulo 3, deste trabalho, é dedicado a formalização dos procedimentos de krigeagem linear e de krigeagem por indicação, enquanto que o capítulo 4 formaliza os procedimentos de simulação estocástica condicionada e descreve as principais diferenças destes com os procedimentos de krigeagem. Esses capítulos mostram, ainda, que os procedimentos geoestatísticos, baseados em krigeagem por indicação, possibilitam inferências para atributos temáticos, além de atributos numéricos, o que não pode ser realizado com estimadores determinísticos ou com a krigeagem linear.

2.2.3 Modelagem computacional para modelos ambientais em SIG

Modelos ambientais, aqui também chamados modelos matemáticos, são representações matemáticas criadas para representar fenômenos ou processos do mundo real. Estes modelos são simplificações da realidade, de onde se abstraem os elementos mais importantes para uma aplicação, e são construídos a partir da observação dos dados espaciais e seus relacionamentos. Os modelos ambientais são usados para aumentar o conhecimento sobre um processo, predizer valores ou comportamentos em áreas não observadas e comprovar, ou não, hipóteses feitas sobre processos. Estes modelos variam de equações empíricas simples, tais como, equações de regressão linear, até conjuntos de equações diferenciais complexas derivadas dos fundamentos da física (Moore et al., 1993)

Modelagem computacional em SIG é a implementação de um modelo matemático, que representa um fenômeno natural, no contexto de um Sistema de Informação Geográfica. Segundo Heuvelink, 1998, os modelos espaciais podem ser classificados em 3 tipos, lógicos, empíricos e conceituais.

Os *modelos lógicos* computam um atributo de saída, resultado do modelo, pela aplicação de regras lógicas simples sobre os atributos de entrada. Por exemplo, um mapa de risco de erosão pode ser obtido pelo cruzamento dos dados de declividade, cobertura vegetal e tipos de solo.

Os *modelos empíricos* baseiam-se em experiências ou em conhecimentos obtidos por percepção dos fenômenos ambientais. Em geral são formulados por regressões cujos coeficientes são definidos experimentalmente e, via de regra, se aplicam somente às áreas usadas para derivá-los. Burrough e McDonnell, 1998, apresenta o relacionamento empírico de dois modelos de perdas de solo por erosão: o modelo USLE – "Universal Soil Loss Equation" – e o modelo SLEMSA – "Soil Loss Estimation Model for Southern Africa". O modelo USLE, por exemplo, depende de alguns atributos, tais como, o comprimento de rampa L e a declividade S que, por sua vez, são, também, derivados de regressões empíricas.

Os *modelos conceituais*, também conhecidos como *modelos físicos,* são concebidos a partir do entendimento dos processos físicos do fenômeno que está sendo modelado. Têm aplicações mais gerais e seus coeficientes referem-se às propriedades físicas, já comprovadas ou aceitas, do mundo real. A maioria dos modelos atmosféricos e hidrológicos são construídos como modelos conceituais. Lee et al., 1993, descreve modelos atmosféricos baseados nas leis da conservação da física. Maidment, 1993, apresenta aspectos importantes relacionados com a modelagem hidrológica em ambiente de SIG.

Na prática, muitos dos modelos usados nas ciências ambientais contém componentes empíricas e conceituais (Heuvelink, 1998), como exemplos típicos podem ser citados os modelos de erosão de solo e os modelos de aptidão agrícola. Além disso, Burrough et al., 1996, argumenta que muitos modelos físicos e empíricos são lineares porque modelos lineares são fáceis para manipular computacionalmente e eles têm um comportamento previsível quando possuem realimentações.

Numa formulação geral, a modelagem computacional estima o valor de um atributo de saída *Y*, para uma determinada posição, ou região, a partir dos valores de *N* atributos, ou variáveis, de entradas, Z_1, Z_2,...,Z_N, ou seja:

$$Y = \boldsymbol{g}(\boldsymbol{Z}_1, \boldsymbol{Z}_2 ..., \boldsymbol{Z}_N) \tag{2.3}$$

Algumas considerações podem ser feitas em relação à esta abordagem:

- Os valores dos atributos fontes (z_1,...,z_N) não são exatos, ou seja, $z_i = z_{im} + \varepsilon$ onde z_{im} é o valor médio do atributo de entrada e ε é uma componente de erro aleatório com média zero;
- Os valores dos atributos z_i's podem ser independentes ou apresentar dependência entre eles;
- A função ***g*** é uma relação aritmética quando a natureza dos atributos dos dados dos mapas fontes for quantitativa, dados numéricos;
- A função ***g*** é uma relação lógica quando a natureza dos atributos dos mapas fontes for qualitativa, dados temáticos;

- A função ***g*** pode ser uma composição de relações aritméticas e lógicas para aplicações complexas envolvendo atributos numéricos e temáticos.

Atualmente, modelos computacionais simples são executados diretamente nos SIG através de operações básicas, lógicas e aritméticas, contidas nos seus módulos de análise ou de álgebra de dados espaciais. Modelos complexos são, muitas vezes, executados fora do ambiente do SIG, por sistemas de modelagem específicos. Nestes casos, os SIG são usados como base de armazenamento de dados espaciais e também como ferramentas de visualização para os dados de entrada e de saída dos modelos. Aqui, coloca-se um dilema entre se dotar o SIG de todas as potencialidades de análises ou criar interfaces inteligentes entre o SIG e os sistemas específicos de modelagem.

Outra questão importante se refere à inclusão de informação de incertezas nos modelos computacionais. Como já apresentado neste capítulo, a espacialização de um atributo pode ser feita por algoritmos de inferência determinísticos e estocásticos. Os modelos computacionais estocásticos, diferentemente dos determinísticos, consideram as incertezas associadas aos dados e aos métodos envolvidos no modelo. Assim, quando um modelo computacional integra dados espaciais, modelados segundo o paradigma estocástico, as incertezas, atribuídas aos dados de entrada, se propagam para o dado de saída. A estimativa e a propagação de incertezas são temas ainda pouco explorados e usados, porém são de extrema relevância, uma vez que objetivam adicionar informação de qualidade aos dados e produtos disponibilizados pela tecnologia dos SIG.

2.3 PROPAGAÇÃO DE INCERTEZAS EM MODELAGEM AMBIENTAL COM SIG

A propagação de incertezas nos atributos espaciais, em modelagem ambiental com SIG, tem sido abordada com frequência na literatura. Trabalhos como os de Arbia, 1993, Burrough, 1992, Goodchild e Guoging, 1992 e Goodchild, 1993b, Heuvelink e Stein, 1989, Heuvelink et al., 1993 e Heuvelink, 1998, são exemplos reais dessa preocupação. Entretanto, segundo Heuvelink, 1996, desenvolvimentos teóricos determinam que é relativamente direto avaliar como incertezas de atributos são propagadas através de modelagem espacial quantitativa em SIG. Mais recentemente, Heuvelink, 1998, apresentou os fundamentos teóricos necessários para a solução dos problemas de propagação de incerteza em modelagens espaciais considerando que:

- os dados espaciais estão representados por campos com atributos de natureza contínua, ou seja, modelados como campos numéricos, e cada campo, de entrada ou saída, é modelado como uma função aleatória e;
- a modelagem espacial envolve classes de operações locais, pontuais ou de vizinhança, conforme ilustrado na Figura 2.1.

Segundo esta abordagem, os valores de saída $y(\mathbf{u})$, para cada posição $\mathbf{u}$, são derivados dos N atributos de entrada, segundo uma relação do tipo:

$$Y(\mathbf{u}) = \boldsymbol{g}(Z_1(\mathbf{u}), Z_2(\mathbf{u}) ..., Z_N(\mathbf{u})) \quad (2.4)$$

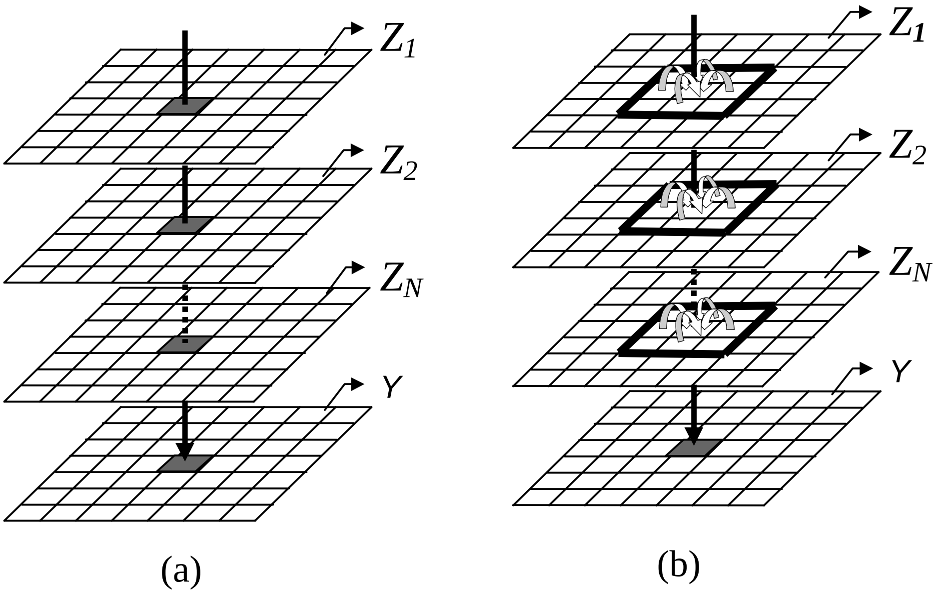

Figura 2.1: Operações locais: (a) pontuais e (b) de vizinhança em modelagem computacional (baseado em Heuvelink, 1998).

Supondo-se conhecidos os campos de incerteza de cada uma das representações das entrada, pode-se, também, obter um campo de incertezas do campo resultante *Y*. Isto é feito por procedimentos de propagação de incerteza, que consideram os tipos de dados, temáticos ou numéricos, e os tipos de operações, lógicas ou aritméticas, e seus parâmetros, envolvidos nas relações definidas sobre os dados de entrada. Heuvelink, 1998, apresenta o formalismo de quatro métodos de propagação de incertezas para campos numéricos. O autor assume que cada campo de incerteza, dos atributos de entrada, é modelado por intervalos de confiança, (Isaaks e Srivastava, 1989), definidos pelas variâncias do erro de estimação obtidas por procedimentos de krigeagem linear. As principais características dos métodos de propagação de incerteza formalizados por Heuvelink, 1998, são apresentadas e analisadas, segundo restrições de aplicação e vantagens, nas quatro seções que seguem.

2.3.1 Método de Taylor de primeira ordem

A expansão da série de Taylor de primeira ordem, em torno do vetor de médias, $\vec{\mu}_Z = \left(\mu_{Z_1}, \ldots, \mu_{Z_n}\right)$, de *N* variáveis aleatórias Z_i, $i = 1,\ldots,N$, é dada por:

$$\boxed{Y = g\left(\vec{\mu}_Z\right) + \sum_{i=1}^{N}\left\{\left(Z_i - \mu_{Z_i}\right)\cdot\left(\frac{\partial g}{\partial z_i}\left(\vec{\mu}_Z\right)\right)\right\} + \text{resíduo}} \qquad (2.5)$$

A partir desta expansão, e desconsiderando-se o resíduo, obtêm-se as seguintes aproximações para média μ_Y e variância σ_Y^2 da VA *Y*:

$$\boxed{\mu_Y = E(Y) \approx g\left(\vec{\mu}_Z\right)} \qquad (2.6)$$

$$\boxed{\sigma_Y^2 \approx \sum_{i=1}^{N}\sum_{j=1}^{N}\left[\frac{\partial g}{\partial z_i}\left(\vec{\mu}_Z\right)\cdot\frac{\partial g}{\partial z_j}\left(\vec{\mu}_Z\right)\cdot\sigma_{Z_i}\cdot\sigma_{Z_j}\cdot\rho_{ij}\right]} \qquad (2.7)$$

onde σ_{Z_k} é o desvio padrão da *k*-ésima VA de entrada, $k = 1,\ldots,N$, e ρ_{ij} é o coeficiente de correlação entre as VA Z_i e Z_j.

As seguintes considerações, sobre essa metodologia, podem ser ressaltadas:

- o método aplica-se somente para operações aritméticas, Z_i's numéricos;
- o valor médio de Y depende apenas dos valores médios das entradas;
- a variância da saída depende dos desvios padrão das variáveis de entrada, dos coeficientes de correlação entre elas e, também, das derivadas parciais de $\boldsymbol{g}$ em relação às entradas. Isto implica que este método só é aplicável para funções $\boldsymbol{g}$ continuamente diferenciáveis e;
- quando as variáveis de entrada são independentes, o coeficiente de correlação, ρ_{ij}, é igual a 0 para i≠j, e é igual a 1 para i=j. Neste caso, e somente neste caso, a formulação da variância é simplificada para:

$$\boldsymbol{\sigma}_Y^2 = \sum_{i=1}^{N} \left[\frac{\partial \boldsymbol{g}}{\partial z_i} \left(\vec{\boldsymbol{\mu}}_Z \right) \right]^2 \cdot \boldsymbol{\sigma}_{Z_i}^2 \tag{2.8}$$

2.3.2 Método de Taylor de segunda ordem

O método de Taylor de segunda ordem é uma extensão do método de Taylor de primeira ordem, incluindo-se o termo de segunda ordem na expansão, ou seja:

$$\begin{aligned} Y = \boldsymbol{g}\left(\vec{\mu}_Z\right) + \sum_{\boldsymbol{i}=1}^{\boldsymbol{N}} \left\{ \left(Z_i - \mu_{Z_i} \right) \cdot \left(\frac{\partial \boldsymbol{g}}{\partial z_i} \left(\vec{\mu}_Z\right) \right) \right\} \\ + \frac{1}{2} \sum_{\mathrm{i}=1}^{N} \sum_{\mathrm{j}=1}^{N} \left\{ \left(Z_i - \mu_{Z_i} \right) \cdot \left[Z_j - \mu_{Z_j} \right] \cdot \left(\frac{\partial^2 \boldsymbol{g}}{\partial z_i \, \delta z_{\mathrm{j}}} \left(\vec{\mu}_Z\right) \right) \right\} + \text{resíduo} \end{aligned} \tag{2.9}$$

Neste caso, a aproximação da média $\boldsymbol{\mu}_Y$ é dada por:

$$\mu_Y = E(Y) \approx \boldsymbol{g}\left(\vec{\mu}_Z\right) + \frac{1}{2} \sum_{\boldsymbol{i}=1}^{\boldsymbol{N}} \sum_{\boldsymbol{j}=1}^{\boldsymbol{N}} \left\{ \rho_{i,j} \boldsymbol{\sigma}_{z_i} \boldsymbol{\sigma}_{z_j} \frac{\partial^2 g}{\partial_{z_i} \partial_{z_j}} \left(\vec{\mu}_Z\right) \right\} \tag{2.10}$$

Contrário ao que se obteve com o método de Taylor de 1ª ordem, a média $\boldsymbol{\mu}_Y$, da equação 2.10, pode diferir do valor de ***g*** aplicado às médias das entradas. Além disso, a estimativa da variância, para o método de Taylor de segunda ordem, tem uma formulação mais complexa, pois requer o cálculo do 1º, 2º, 3º e 4º momentos das entradas e, também, derivadas parciais de primeira e segunda ordens de ***g*** em relação às entradas (Heuvelink, 1998). Assim, o método só é aplicável para ***g*** continuamente diferenciáveis até segunda ordem.

2.3.3 Método de Rosenblueth

O método de Rosenblueth é equivalente ao de Taylor de 1ª ordem, podendo ser usado também nos casos em que a função ***g*** não é continuamente diferenciável. Como no método de Taylor de 1ª ordem, os valores de média e a variância da VA *Y* de saída são estimados a partir da aplicação da função ***g*** e do 1º e 2º momentos das VA Z_i de entrada, (Heuvelink, 1998).

2.3.4 Método de Monte Carlo

A idéia principal do método de simulação de Monte Carlo é obter realizações da VA de saída *Y* a partir de realizações, das VA de entrada Z_i's, obtidas a partir da distribuição conjunta destas variáveis. Os valores realizados de *Y* são calculados pela aplicação da função ***g*** sobre os valores realizados das entradas. As estatísticas da saída, como média e variância, são obtidas a partir dos valores realizados de *Y*. Considerando-se operações pontuais com *N* entradas, o método de Monte Carlo consiste dos seguintes passos:

Início
  Para cada localização $\mathbf{u} \in A$
  Início
    Repita ***k*** vezes:
    Início
      Obtenha um conjunto de realizações $z_i(\mathbf{u})$, i=1,...,*N*.
      Calcule e armazene realizações de Y, $y(\mathbf{u})=\boldsymbol{g}(z_1(\mathbf{u}),..z_N(\mathbf{u}))$
    Fim
    Calcule estatísticas da distribuição das ***k*** saídas y_i de *Y* em **u**
  Fim
Fim

As estatísticas de média e variância de *Y* são calculadas, em **u**, por:

$$ {}_{Y}(\mathbf{u}) = \frac{1}{\boldsymbol{N}} \sum_{i=1}^{N} y_i(\mathbf{u}) \quad \boldsymbol{e} \quad {}^{2}_{Y}(\mathbf{u}) = \frac{1}{\boldsymbol{N}-1} \sum_{\boldsymbol{i}=1}^{N} \left(y_i(\mathbf{u}) - {}_{Y}(\mathbf{u}) \right)^2 \qquad (2.11) $$

A principal vantagem do método de Monte Carlo é que ele é de aplicação geral. Pode ser aplicado a modelos lógicos, conceituais ou empíricos. O método considera a função ***g*** como uma caixa preta cujas perturbações nos dados de entrada são estudadas a partir do resultado da saída.

Por outro lado, como os resultados não são obtidos numa forma analítica, análises de sensibilidade do modelo são mais difíceis, uma vez que a simulação deve ser executada várias vezes. Isto requer recursos computacionais com processamento rápido e possibilidade de armazenamento de grandes volumes de dados. O número de vezes que o algoritmo é repetido é, usualmente, maior que uma centena mas pode chegar a centenas de milhões.

2.4 O PARADIGMA DA MODELAGEM ESTOCÁSTICA POR INDICAÇÃO

Em função de um melhor entendimento do contexto deste trabalho, a modelagem computacional em ambiente de SIG será classificada segundo 3 (três) etapas de evolução distintas, a modelagem determinística, a modelagem estocástica[2] linear e a modelagem estocástica, não linear, por indicação.

A *modelagem ambiental determinística* se caracteriza pela modelagem dos dados espaciais segundo o paradigma determinístico. Essa modelagem é praticada desde os primeiros modelos computacionais implementados em um SIG. Neste tipo de modelagem, os modelos de dados espaciais não contém componente aleatória e são, portanto, considerados exatos. Consequentemente, os produtos da modelagem computacional não contém informação de qualidade. A qualidade dos resultados é neglicenciada ou é

[2] O termo *modelagem estocástica* é utilizado, neste texto, como referência exclusiva à modelagem que se utiliza de procedimentos geoestatísticos para representação de atributos espaciais.

inferida subjetivamente pelos usuários do sistema. Este tipo de modelagem ainda é praticado nos tempos atuais, mas obviamente é pobre e tende a ser substituído por modelos mais representativos.

A *modelagem estocástica linear* é mais recente e se caracteriza pela modelagem dos atributos espaciais segundo o paradigma geoestatístico linear, na qual os atributos espaciais são considerados variáveis aleatórias e são tratados pela teoria das variáveis regionalizadas. A modelagem dos atributos se baseia nos estimadores de krigeagem linear, principalmente a krigeagem ordinária. Neste modelo, o valor estimado e a variância do erro de estimação são considerados como parâmetros de média e de variância, respectivamente, de um modelo de distribuição gaussiana associado ao atributo. Assim, a variância do erro de estimação, ou o desvio padrão correspondente, é usada como unidade de medida de incerteza da modelagem do atributo. Também, uma combinação da média e da variância do erro de estimação é utilizada para definição de intervalos de confiança. Além disso, nesta etapa da modelagem são introduzidas, formalizadas e implementadas teorias de propagação de incertezas que possibilitam qualificar os produtos da modelagem. O trabalho atualmente mais expressivo e completo desta etapa é sistematizado e apresentado por Heuvelink, 1998. Porém, existem problemas com este tipo de modelagem. Primeiramente, a variância de krigeagem linear está condicionada ao tipo de estimador utilizado e não é dependente dos valores dos atributos (Isaaks e Srivastava, 1989). Além disso, a hipótese de distribuição paramétrica multigaussiana para o atributo deve ser verificada, o que é praticamente impossível (Deutsch e Journel, 1998). No capítulo 3 será apresentado uma análise crítica do uso dessa variância como medida de incerteza. Apesar destas restrições, essa medida de incerteza vem sendo usada, com frequência, por vários usuários de SIG, por falta de conhecimento ou de outra alternativa. O segundo problema é o fato de a krigeagem linear permitir a modelagem espacial apenas para atributos numéricos. Assim, o problema de atribuição de incerteza para dados temáticos não é resolvido e, portanto, também não se resolve o problema de propagação de incertezas para modelos computacionais lógicos que envolvam variáveis temáticas.

Do conhecimento do autor, a *modelagem estocástica por indicação* é mais recente e não havia sido, até então, sistematizada para uso em modelagem computacional desenvolvida em SIG. Esta tese é o primeiro esforço neste sentido. Este tipo de modelagem computacional se caracteriza por utilizar os procedimentos não lineares da geoestatística, a krigeagem por indicação e a simulação estocástica por indicação, para modelar a variabilidade dos atributos espaciais. Estes procedimentos possibilitam a inferência de uma aproximação discretizada do modelo de distribuição de probabilidade do atributo que é, então, utilizada para modelagem da incerteza sobre seus valores. Assim, tem-se uma modelagem espacial não paramétrica que pode, portanto, ser usada sem restrições ao tipo de distribuição do atributo. Os modelos de incerteza são obtidos diretamente da distribuição, independem de um estimador escolhido e estão relacionados ao comportamento de variabilidade do atributo. Metodologias de estimativa de incertezas para procedimentos geoestatísticos por indicação são apresentados e formalizados nos capítulos 3, 4 e 5 deste trabalho. Outra vantagem importante destes procedimentos é a possibilidade de se modelar dados temáticos, além dos dados de natureza numérica. Assim, pode-se trabalhar com propagação de incertezas para modelos computacionais que envolvam atributos numéricos e temáticos.

2.5 Conclusões

O presente capítulo apresentou uma introdução aos conceitos de modelagem computacional, no ambiente de um SIG, para modelagem de processos ambientais. Ressaltou-se que, a modelagem dos dados espaciais, envolvidos em um modelo computacional, pode ser feita segundo paradigmas determinístico ou estocástico.

O paradigma estocástico, que utiliza procedimentos da geoestatística para espacialização de atributos de natureza ambiental, possibilita a estimação das incertezas associadas a esses atributos. Este capítulo descreve e propõe o uso de uma formulação estocástica mais rigorosa para modelar incerteza de atributos numéricos e temáticos, a modelagem estocástica por indicação. As incertezas inferidas por esta modelagem são utilizadas por metodologias de propagação de incerteza que devem, necessariamente, ser incluídas no processo da modelagem ambiental.

A incerteza propagada para o resultado de uma modelagem computacional fornece informação quantitativa sobre o produto obtido. A qualidade do produto gerado é utilizada para avaliação da qualidade do modelo, para análises de sensibilidade do modelo aos dados de entrada e para a quantificação dos riscos assumidos ao se tomar decisões apoiadas nesse produto.

Entretanto, apesar da sua importância, as informações de incerteza ainda são pouco, ou quase nunca, consideradas nas aplicações desenvolvidas no ambiente de SIG. Isto ocorre, principalmente, devido à falta de ferramentas disponibilizadas nos SIG atualmente em uso. Este trabalho é uma contribuição para a mudança desse cenário, uma vez que, apresenta, analisa e propõe o uso de novas ferramentas e de novos procedimentos em modelagens computacionais implementadas em SIG.