

ANÁLISE DA PRECISÃO DOS MODELOS TEMÁTICOS GERADOS PELOS SISTEMAS DE INFORMAÇÃO GEOGRÁFICA.

Margareth SIMÕES e Suzana DRUCK

Centro Nacional de Pesquisa de Solos / EMBRAPA
Rua Jardim Botânico, 1024 - Cep 22.460-000 - Rio de Janeiro

Resumo

A utilização dos dados geo-referenciados existentes nos bancos de dados ambientais disponíveis vêm se processando com a utilização de Sistemas de Informação Geográfica (SIGs), que reprocessam as observações originais gerando um novo conjunto de dados, através do processo de análise e manipulação da informação. A validade dos resultados obtidos depende de se ter alcançado um produto com um certo nível de confiabilidade. As diversas etapas de utilização de um SIG vão acumulando erros e os propagando a cada operação de manipulação, comprometendo cada vez mais a qualidade do produto final a ser gerado pelas ferramentas disponíveis nos SIGs. Este trabalho visa levantar algumas questões sobre a propagação de erros nos SIGs.

Abstract

Spatial data available in natural resource databases are used through Geographic Information Systems (GIS), which may manipulate the original information creating a new data set. Validation of results depends on the reliability of the obtained product. The various steps involved during GIS manipulation lead to an accumulation of error and further propagation. Hence, the quality of the final product is probably affected. This research work highlights some aspects of error propagation on GIS procedures.

1 Introdução

Normalmente os sistemas de informação espacial assumem implicitamente que os fenômenos do mundo real podem ser modelados por entidades discretas, tais como: ponto, linhas, polígonos ou *pixels*, que possuem atributos exatos. As informações a cerca dos atributos de uma entidade podem ser recuperados através da utilização de um método apropriado de tradução das questões do usuário para uma linguagem inteligível pelo computador.

Quando um dado ambiental é manipulado por um modelo lógico, assume-se que os resultados são corretos. No entanto, na realidade os resultados conterão erros porque os valores dos atributos de entrada não podem ser determinados de forma exata. A validade dos resultados obtidos depende

da precisão com que a base de dados é montada.

Os SIGs, entretanto, no atual estágio de desenvolvimento, não possuem mecanismos para estabelecer a validade dos produtos que geram, isto é, os SIGs fornecem os mecanismos para transformar a informação obtida sem simultaneamente fornecer e quantificar o grau de confiabilidade desta informação. A literatura relativa ao desenvolvimento de aplicações utilizando-se SIGs reflete a omissão na estimativa da propagação de erros. Tal omissão não se efetua simplesmente porque em geral os erros nestes processos são pequenos, mas sim porque a estrutura de modelagem nos SIGs não possibilita este procedimento. As diversas etapas de utilização de um Sistema de Informação Geográfica vão acumulando erros e os propagando a cada operação de manipulação, comprometendo cada vez mais

a qualidade do produto final a ser gerado pelas ferramentas disponíveis nos SIGs.

2 Fontes de Erros nos SIGs

Pode-se considerar que existem dois grupos de erros na utilização de mapeamentos automatizados. O primeiro são os erros inerentes aos dados, ou seja, os erros presentes nos documentos fontes utilizados. O segundo grupo, diz respeito à manipulação da informação em si e são chamados de erros operacionais. Estes erros são produzidos através de procedimentos nos SIGs, seja na captura, no gerenciamento ou na análise da informação.

Pode-se, de uma forma geral, enumerar os seguintes erros operacionais na utilização de um SIG: erros na aquisição de dados (por exemplo, na digitalização), erros numéricos no armazenamento dos dados, erros de conversão (vetor x *raster*, *raster* x vetor) e erros advindos dos modelos lógicos.

Este último tipo de erro será tratado com mais atenção, uma vez que, por relacionar diversos níveis de informações e empregar uma seqüência de transformações, cruzamentos, etc, acabam acumulando os erros das demais etapas, que são propagados através destas manipulações.

3 Erros advindos do Modelo Lógico

Se um usuário desejar realizar, por exemplo, um estudo de avaliação da terra e fornecer os seguintes parâmetros, na classificação de risco de erosão:

Serão consideradas áreas de alto risco se:

Se declividade > 10% AND Textura do Solo = areia AND Cobertura Vegetal < 25%

Neste caso o sistema fará uma interseção *booleana* (AND), através da sobreposição dos diversos Planos de Informação.

Cada polígono ou *pixel* representando uma área será testado em seus atributos fornecendo como resposta TRUE ou FALSE.

Todas as entidades que não satisfizerem as três condições serão rejeitadas.

Este tipo de modelagem é equivocada em diversos aspectos. Primeiramente porque assume que o fenômeno estudado pode ser descrito por uma simples expressão *booleana*, relacionando declividade, vegetação e textura. Deve-se considerar, por exemplo, que o risco de erosão continuará existindo se a declividade for só um pouco menor que os 10% dados na modelagem (por exemplo 9%). E no entanto, de acordo com o modelo, utilizando-se AND, a área seria considerada segura, pois um dos parâmetros não satisfiz a condição de risco.

Um outro equívoco inerente a este modelo é que este assume que todas as medidas são extremamente precisas, o que sabe-se que na prática não é verdade. Existem erros de medida e de variação espacial.

Em geral, na maioria das vezes é impossível se afirmar que os atributos armazenados numa base de dados estão livres de erros. Consequentemente, quando se utiliza dados com um certo grau de incerteza nos modelos lógicos ou quantitativos, os resultados advindos também irão conter erros.

Uma melhoria na precisão dos modelos através da eliminação das imprecisões advindas da lógica *booleana*, ou seja, eliminação do problema da fronteira rígida, pode ser obtida através da utilização da lógica *fuzzy*. Para uma melhor compreensão, uma breve descrição dos processos de classificação *booleana* e *fuzzy* será dada seguir.

4 Classificação Booleana

Os métodos de se verificar o ajuste de uma área ao conjunto de requisitos que irá classificá-la é muito simples quando se utiliza a lógica *Booleana*. Características complexas da terra e suas classes podem ser definidas utilizando-se os operadores AND, OR, NOT, XOR, para especificar que combinações de valores de atributos são necessárias para um dado propósito de classificação. As características que irão definir uma classe

são normalmente definidas através da especificação das variações de um certo número de valores de propriedades chaves que um indivíduo deve possuir. Para se classificar como membro de um dado domínio, as características de uma dada área devem se encaixar em todas as especificações, o que se faz, através do uso múltiplo do *AND Booleano*, ou seja, através de uma interseção:

$$R = \text{true (pertence a uma dada classe)} \\ \text{se } A \text{ AND } B \text{ AND } C \text{ AND } D$$

onde A,B,C... representam os intervalos especificados das propriedades que norteiam a classificação. Os valores que se obtém aqui são, portanto, binários: *verdadeiro* quando o resultado da interseção é 1 e *falso* quando o resultado é 0. Neste caso, quando se está classificando um determinado fenômeno que depende de diversas variáveis, o resultado desta classificação se baseará no *menor* valor determinado na região a ser classificada, ou seja:

$$R = \text{MIN} (Q_1, Q_2, Q_3 \dots)$$

onde os Q_i são os valores da classificação de cada parâmetro que influencia o fenômeno estudado. Esta classificação é chamada de *estrita* ou *exata* (BURROUGH et al., 1992).

5 Classificação Contínua

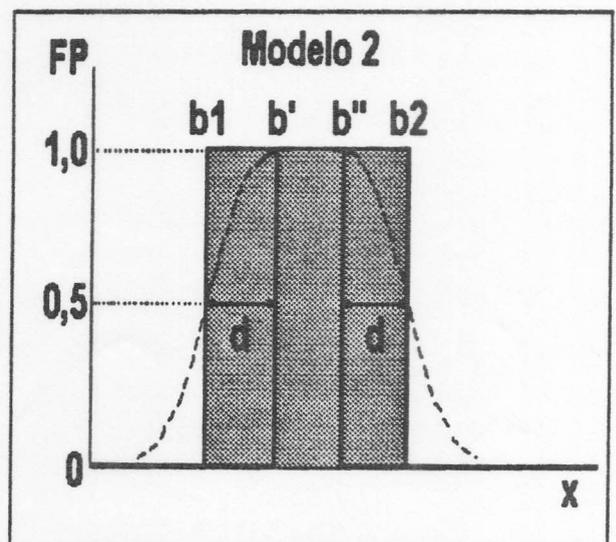
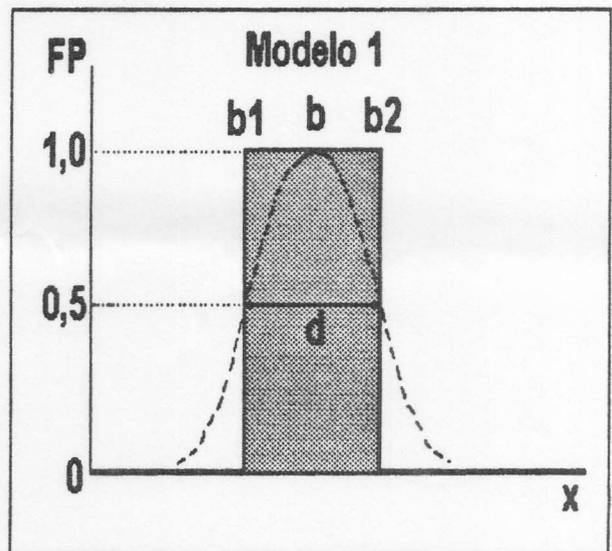
A teoria do conjunto *fuzzy*, apresentada por ZADEH (1965), é uma generalização da teoria convencional de conjunto que utiliza a lógica *booleana* em que as classes de objetos são precisamente definidas. Os conjuntos *fuzzy* lidam com a imprecisão e a sua utilização é apropriada em sistemas complexos em que os conceitos se caracterizam pela : **generalidade, ambigüidade e imprecisão.**

Aos indivíduos que se encaixam exatamente a classes estritamente definidas, são atribuídos um valor de pertinência igual a 1. Os indivíduos que estão fora da variação que caracteriza uma determinada classe, terá

seu valor de pertinência dependente do seu grau de proximidade com a classe definida.

O grau de pertinência é fornecido através de uma função. A título ilustrativo na Figura 1 são apresentados 5 modelos de funções de pertinência. De acordo com o fenômeno estudado, determina-se qual o tipo mais apropriado. Estes modelos de classificação *fuzzy*, geralmente utilizados para dados de recursos naturais, são extensões das funções geradas por KANDEL (1986).

Na Figura 1, o eixo horizontal indica o valor do atributo e o vertical o valor de pertinência (FP). A área hachurada representa o conjunto *booleano*, enquanto as tracejadas o conjunto *fuzzy*.



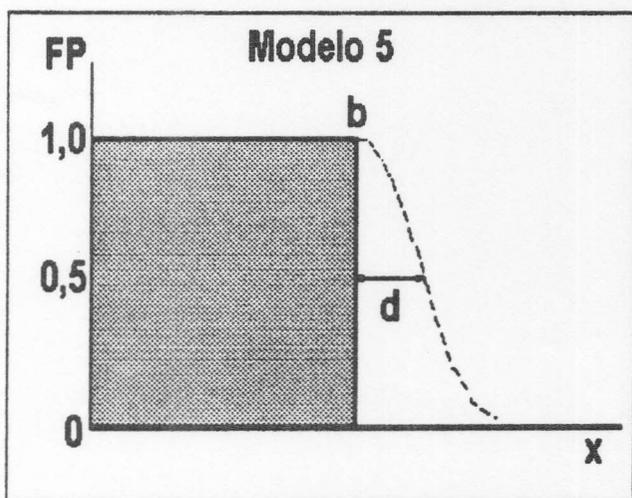
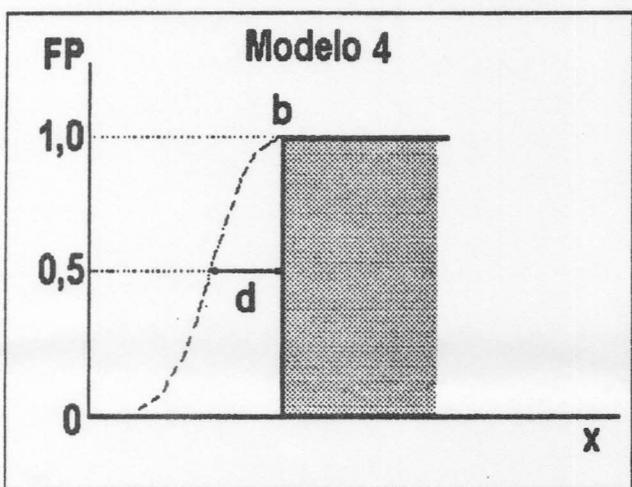
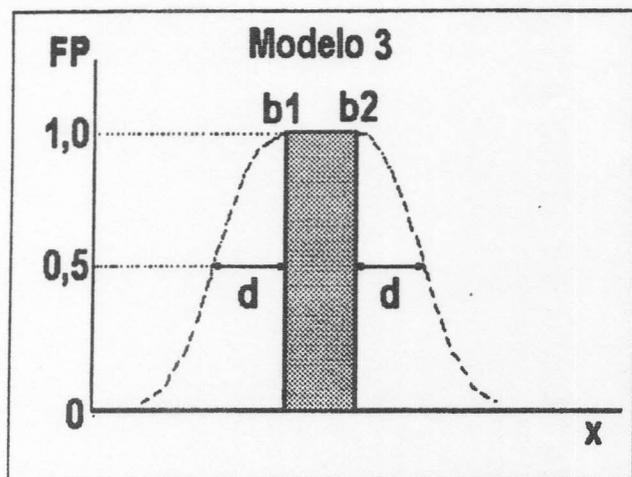


Figura 1 - Modelos de Classificação *fuzzy* e *booleana*.

Devido as suas características, conclui-se que a utilização da classificação contínua pode reduzir a propagação de erros nos modelos lógicos, proporcionando resultados

mais confiáveis, comparados com a utilização da classificação *booleana*.

Para um maior aprofundamento na teoria dos conjuntos fuzzy sugere-se a leitura de: ZADEH, 1965, 1975, CHANG e BURROUGH, 1987; BURROUGH, 1989, KLIR e FOLGER, 1988, dentre outros.

6 Sistema Integrado de Geoestatística, Classificação Contínua e Propagação de Erro

A Figura 2 em anexo resume alguns procedimentos computacionais que serão desenvolvidos no Centro Nacional de Pesquisa de Solos da EMBRAPA, juntamente com o Instituto de Pesquisas Espaciais (INPE). Estes procedimentos envolvem métodos geoestatísticos, lógica contínua e propagação de erros.

A partir dos dados de entrada do SIG, procede-se uma análise exploratória dos dados onde os valores discrepantes do conjunto são averiguados. Estes valores sofrerão transformações para que os dados se tomem simétricos e bem comportados. A seguir, realiza-se um ajuste entre o variograma experimental e o teórico. Executa-se, então, o procedimento de krigagem. A krigagem é um processo estatístico que faz inferências de valores não observados, podendo, por exemplo, ser utilizado para interpolar dados de amostras de solos.

A partir da krigagem obtém-se uma grade de valores e uma grade de variância. A grade de valores será utilizada no processo de classificação e a de variância no processo de propagação de erros. Cada ponto da grade possui o valor inferido e a variância. Pode-se a partir daí produzir dois tipos de mapas: o mapa de isovalores e o de isovariância.

Se forem utilizadas funções diferenciáveis, tais como: regressão, soma ponderada etc, serão usadas expansões de Taylor. Caso sejam utilizadas funções não diferenciáveis, tais como: operações de mínimo e máximo, serão usadas simulações.

No final destes procedimentos, tem-se disponível uma grade de valores interpolados (proveniente diretamente da krigagem), uma grade de valores *fuzzy* e uma grade de variância, com a qual pode-se obter uma quantificação do erro manipulado.

7 Bibliografia

BURROUGH, P.A. Five Reasons Why Geographical Information Systems are not Being Used Effectively for Land Resources Assessment. In Proc. Autocarto London (ed. M. Blakemore), pp. 145-154. Atocarto, London, 1986.

BURROUGH, P.A.; MacMILLIAN, R.A.; DEURSEN, W. Fuzzy Classification Methods for Determining Land Suitability from Soil Profile Observations and Topography. Journal of Soil Science, (43):193-210, 1992.

CHANG, L.; BURROUGH, P.A. Fuzzy Reasoning: A New Quantitative Aid for Land Evaluation. Soil Survey and Land Evaluation 7:69-80, 1987.

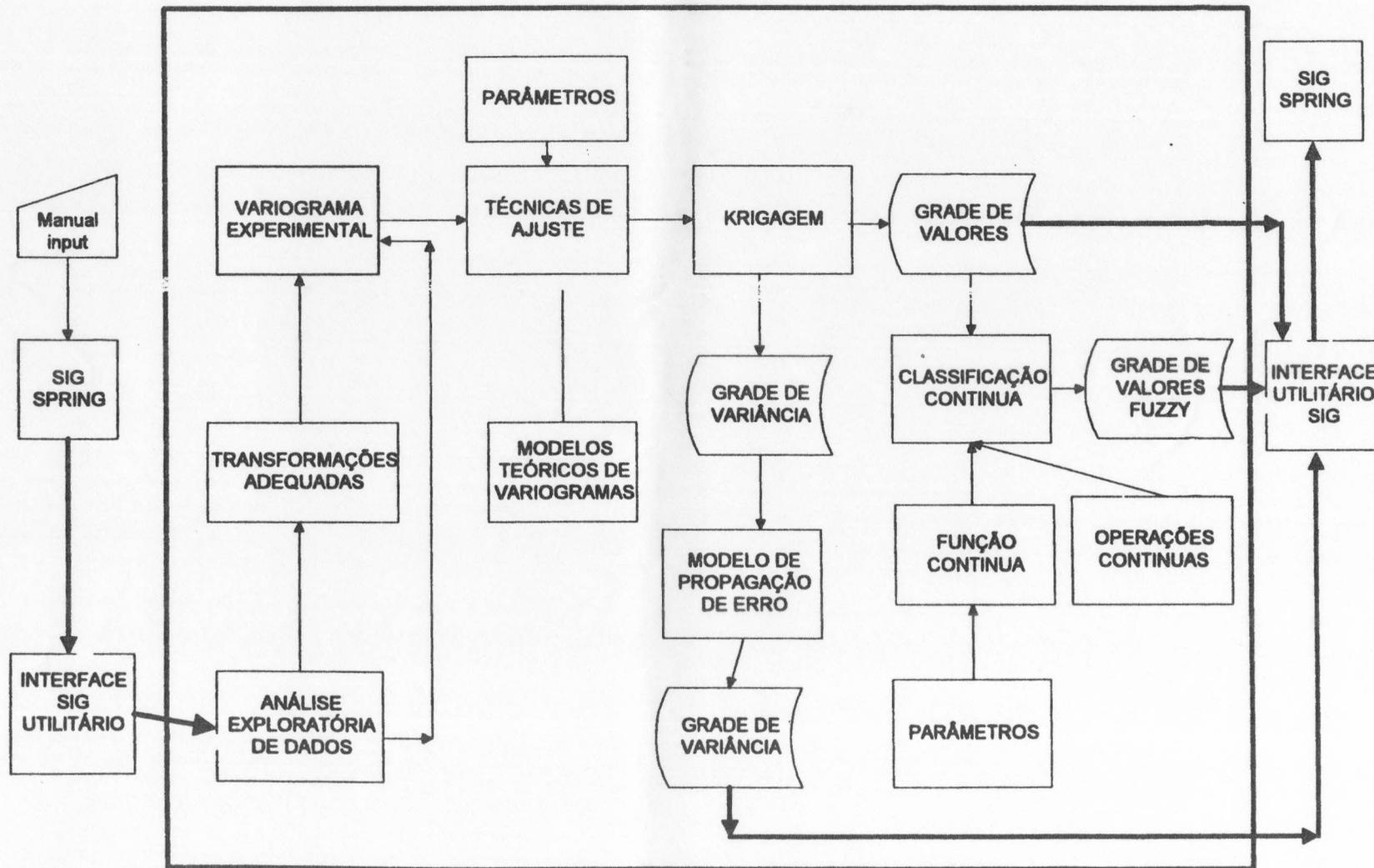
KANDEL, A. Fuzzy Mathematical Techniques With Applications. Addison-Wesley, Massachusetts, 1986.

KLIR, G.J.; FOLGER, T.A. Fuzzy Sets, Uncertainty and Information, Engelwood Cliffs, Prentice-Hall, 1988.

ZADEH, L.A. Fuzzy Sets. Information and Control 8:338-353, 1965..

Figura 2

UTILITÁRIO DE GEOESTATÍSTICA, CLASSIFICAÇÃO CONTÍNUA E PROPAGAÇÃO DE ERRO



001013