

# Integração entre SIG e Modelos Inferenciais Bayesianos

Ilka Afonso Reis<sup>1,2</sup>

<sup>1</sup>Programa de Pós-graduação - Instituto Nacional de Pesquisas Espaciais (INPE)  
Av. dos Astronautas, 1.758 - CEP 12227-010  
São José dos Campos - SP - Brasil

<sup>2</sup>Departamento de Estatística  
Universidade Federal de Minas Gerais – Belo Horizonte, MG – Brasil

ilka@dpi.inpe.br

***Abstract.** GIS are an important tool to deal with spatial data as well the Bayesian method is a powerful technique to statistical data analysis. This paper describes the state of the art on GIS and Bayesian Inferential Models softwares integration and shows alternatives to improve it.*

***Resumo.** O objetivo deste artigo é descrever o estado da arte da integração entre SIG e Modelos Inferenciais Bayesianos, considerando os softwares utilizados nas duas áreas. Para isto, aborda a natureza da análise estatística de dados espaciais, em particular, a estatística espacial, inserindo neste contexto os Modelos Inferenciais Bayesianos. Por fim, descreve o que já é possível em termos de integração entre os softwares de SIG e os de estatística espacial e entre os softwares de SIG e os de Modelos Inferenciais Bayesianos, apontando lacunas a serem preenchidas.*

## 1. Introdução

A integração entre os Sistemas de Informação Geográfica (SIG) e a Análise Espacial é um assunto bastante discutido em ambas as comunidades. Uma procura com estes termos nos mecanismos de busca por livros e artigos na Internet leva a uma quantidade gigantesca de documentos. A importância da Análise Espacial, em particular, da análise estatística de dados espaciais, tem sido amplamente reconhecida na comunidade de SIG (Goodchild e Haining, 2004). Paralelamente, os Modelos Inferenciais Bayesianos ocupam um espaço cada vez maior na análise estatística de dados espaciais.

Os Sistemas de Informação Geográfica são uma importante ferramenta para lidar com dados espaciais, assim como o método bayesiano é uma poderosa técnica para análise de dados. Deste modo, a integração entre os dois deveria ser assunto de interesse nas comunidades de SIG e de Análise Espacial. No entanto, até onde é de conhecimento da autora, não existem trabalhos sobre este tópico especificamente, o que pode ser devido ao fato de que o foco das discussões esteja na integração entre SIG e Análise Espacial como um todo.

Este trabalho visa a preencher esta lacuna, descrevendo o estado da arte da integração entre SIG e Modelos Inferenciais Bayesianos. A seção 2 apresenta o conceito de estatística espacial e uma breve introdução às idéias da inferência bayesiana, situando a aplicação de seus modelos inferenciais nas áreas da estatística espacial. A seção 3

discute os tipos de integração entre os *softwares* de SIG e os de estatística espacial atualmente existentes, destacando o que há em termos de integração entre SIG e Modelos Inferenciais Bayesianos. Finalmente, na seção 4, são feitas as considerações finais sobre o tema. Uma versão estendida deste artigo pode ser encontrada em Reis (2005).

## 2. Estatística Espacial e Modelos Inferenciais Bayesianos

O conjunto de métodos estatísticos para análise exploratória e inferencial de dados espaciais é frequentemente chamado de estatística espacial. Segundo Bailey e Gatrell (1995), a estatística espacial divide-se em quatro grandes áreas, de acordo com o tipo de dado analisado: análise de *processos pontuais*, análise de *dados espacialmente contínuos* (geoestatística), análise de *dados de área* e análise de *dados de interação espacial*. A estatística espacial está inserida num contexto bem mais amplo, o da Análise Espacial, como bem destacam Fischer et al. (1994) e Bailey e Gatrell (1995).

Dentre todos os programas de análise estatística de dados espaciais atualmente existentes, talvez os mais promissores quanto à integração com SIG sejam aqueles desenvolvidos como implementações da linguagem S: o S-Plus e o R. Em especial, o **R** é um programa de código aberto e possui uma série de pacotes com técnicas de estatística espacial, entre eles o **geoR** e **geoRglm** (geoestatística), **spdep** (dados de área) e **splancs** e **spatstat** (análise de padrões pontuais) (<http://gauss.est.ufpr.br/CRAN/>). O projeto **R-spatial** desenvolve métodos para lidar com dados espaciais. As contribuições ao desenvolvimento do R são intensas, com novos e variados pacotes tornando-se disponíveis com muita frequência. Além do S-Plus e do R, existem outros *softwares* para análise estatística de dados espaciais, sejam eles comerciais (ex: SpaceStat) ou não comerciais (ex: Gstat e GeoDa).

A estatística bayesiana pode ser utilizada como método de inferência na análise estatística de dados espaciais. Em geral, o objetivo da inferência estatística é estimar parâmetros de uma certa população e/ou testar hipóteses sobre eles, através de uma amostra extraída desta população. Para a inferência clássica, toda informação sobre o parâmetro a ser estimado está contida na amostra e expressa através da chamada função de verossimilhança.

Na inferência bayesiana, nem toda informação sobre o parâmetro vem da amostra. Parte dela está contida na amostra (função de verossimilhança). Outra parte reside num conhecimento *a priori* sobre o parâmetro (antes de se conhecer dados) e pode ser expressa através de uma distribuição de probabilidade (distribuição *a priori* do parâmetro). O peso de cada informação, amostral ou da *priori*, é regulado através de quão *informativa* é a distribuição *a priori* atribuída ao parâmetro. As duas partes da informação são combinadas com o auxílio do teorema de Bayes, resultando na distribuição *a posteriori* do parâmetro, utilizada nas inferências sobre ele. De posse desta distribuição, é possível calcular probabilidades, percentis, média, entre outras propriedades (Carlin e Louis, 1996). Em problemas mais complexos, o cálculo analítico das distribuições *a posteriori* é inviável. As inferências são feitas através de amostras destas distribuições, obtidas por simulação via *Monte Carlo Markov Chain* (MCMC) (Gamerman, 1997).

Com o avanço na análise bayesiana de dados proporcionada pelos métodos MCMC, pesquisadores em diversas áreas do conhecimento foram encorajados a incorporar as técnicas inferenciais bayesianas na análise de seus problemas, entre eles, os que utilizam dados georreferenciados. Os estudos com dados de área foram os que primeiro se beneficiaram do uso de técnicas bayesianas (Tsutakawa et al., 1985), seguidos de perto pelos que usam dados espacialmente contínuos (Kitanidis, 1986) e processos pontuais (Akman e Raftery, 1986). Um pouco mais recente é a discussão de métodos bayesianos aplicados à análise de dados de interação espacial (Banerjee et al., 2000).

Os dois *softwares* de análise bayesiana de dados mais conhecidos (WinBUGS e BayesX) possuem algumas funções para lidar com dados espaciais. O WinBUGS (*Bayesian Inference Using Gibbs Sampler*, versão Windows) possui funções para modelagem de dados de área e espacialmente contínuos. As versões mais recentes incluem o módulo **GeoBUGS**, que permite a importação de mapas de polígonos em vários formatos, construção e manipulação da matriz de vizinhança e visualização dos resultados nestes mapas (<http://www.mrc-bsu.cam.ac.uk/bugs>). O BayesX foi criado para lidar com modelos de regressão complexos e permite a análise de dados de área, incluindo a importação de mapas de polígonos. (<http://www.stat.uni-muenchen.de/~bayesx/bayesx.html>)

### 3. Integração entre SIG e Modelos Inferenciais Bayesianos

A integração entre SIG e Modelos Inferenciais Bayesianos passa pela integração entre SIG e estatística espacial. Muito progresso tem sido feito na direção desta integração, com diferentes tipos de estratégias (Ungerer e Goodchild, 2002 ; Bailey e Gatrell, 1995 ; Goodchild e Haining, 1992). Na **integração do tipo completa** (*full integration*), as ferramentas de análise estatística são incorporadas diretamente no *software* SIG. Esta estratégia tem sido utilizada no caso de ferramentas mais consagradas, como as de geoestatística e índices de autocorrelação espacial. Alguns exemplos de SIG que utilizam esta estratégia com uma gama variada de ferramentas são o SPRING e TerraView.

Na estratégia de **integração do tipo acoplamento**, os *softwares* de SIG e de estatística espacial existem separadamente e, de algum modo, compartilham os dados a serem analisados. Esta estratégia permite que o *software* de análise de dados espaciais evolua independentemente do SIG. No **acoplamento fraco**, os dados são exportados do SIG para serem analisados por um *software* de estatística espacial e os resultados são novamente exportados para o SIG com a finalidade de visualização. Exemplos desta estratégia são as integrações entre SPRING e SpaceStat e entre ArcInfo e R. O **acoplamento de tipo forte** permite que o SIG e o *software* de estatística espacial compartilhem um banco de dados comum, com a vantagem de que a análise pode ser feita no ambiente do SIG, a partir da chamada de uma rotina do *software* de estatística espacial. Este é o caso das integrações entre GRASS e R, ArcView e S-plus , ArcView e SpaceStat, IDRISI e Gstat, Grassland e S-Plus.

A estratégia de acoplamento entre o TerraView e o R também é do tipo forte, com a diferença de que as análises são feitas no ambiente R, via o pacote aRT (API

R\_Terralib) (<http://www.est.ufpr.br/aRT>). Uma vez aberta a sessão do programa R e carregado o pacote aRT, tem-se acesso ao banco de dados e ferramentas da biblioteca TerraLib, assim como podem ser criados e manipulados objetos de suas classes. As funções do R são utilizadas na análise estatística e os resultados são incorporados ao banco TerraLib e podem ser visualizados no TerraView.

A integração entre os *softwares* de SIG e os de estatística espacial é extensamente discutida em Fook (2005).

### 3.1. O presente

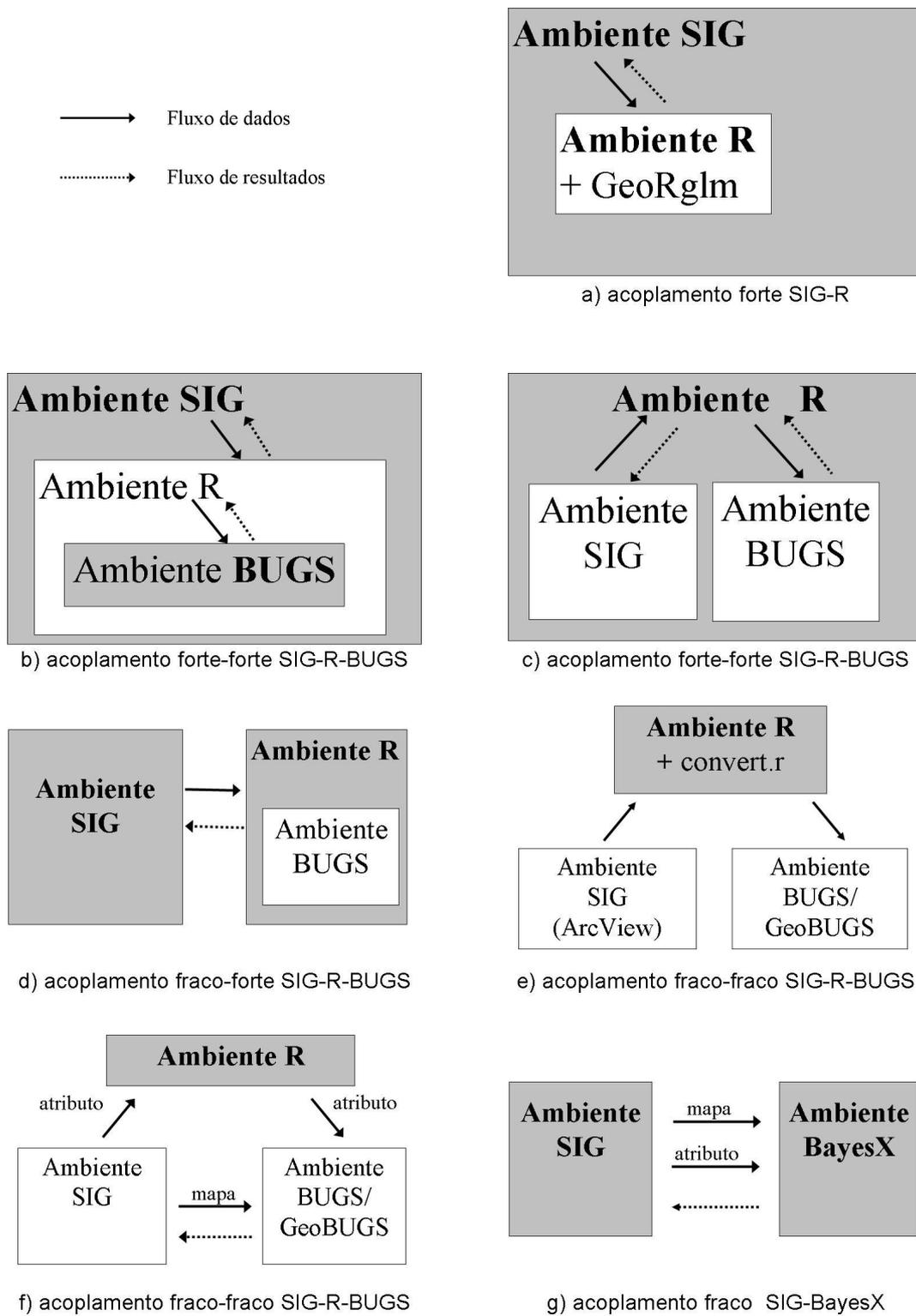
Considerando o atual estágio de integração entre SIG e os *softwares* de estatística espacial, a integração entre SIG e modelagem bayesiana de dados pode ser feita de cinco maneiras (Figura 1).

Atualmente, o **acoplamento forte** (Figura 1a) está limitada aos modelos de krigagem bayesiana, através do pacote GeoRglm do R. A alternativa para a análise de outros tipos de dados seria a integração em **duas etapas de acoplamento do tipo forte**: entre SIG e R e entre R e WinBUGS, via o pacote R2WinBUGS. Estas duas alternativas estão disponíveis para os SIG que permitem acoplamento forte com o R: GRASS (figuras 1a e 1b) e TerraView (figuras 1a e 1c).

Uma terceira alternativa de integração acontece em **duas etapas: acoplamento fraco** (o SIG exporta a tabela de atributos num formato lido no ambiente R); **acoplamento forte** (o R chama o WinBUGS para analisar os dados). Os resultados são exportados num formato acessível ao SIG, onde o banco de dados seria atualizado (Figura 1d). Esta estratégia é limitada à exportação da tabela de atributos com a informação geográfica truncada no caso de dados de área (apenas os centróides dos polígonos são exportados, perdendo-se a topologia). Qualquer informação referente à topologia teria que ser fornecida ao R separadamente. Esta alternativa de integração é bem mais trabalhosa e pouco elegante, mas possibilita que os Modelos Inferenciais Bayesianos sejam integrados a uma gama maior de *softwares* SIG.

O módulo GeoBUGS do WinBUGS permite importação dos mapas de polígonos nos formatos ArcInfo, S-Plus e EpiMap (módulo do EpiInfo, *software* para análise de dados epidemiológicos). Existe uma rotina em R (**convert.r**) para converter os arquivos em formato *shape* do ArcView para o formato S-plus, possibilitando sua leitura no GeoBUGS. Deste modo, o ArcView pode ser parcialmente integrado ao WinBUGS/GeoBUGS via R, numa combinação de **dois acoplamentos fracos**: SIG-R e R-WinBUGS/GeoBUGS (Figura 1e). Considerando que o SPRING e o TerraView exportam dados no formato *shape*, estes dois SIG poderiam ser integrados ao BUGS. O ArcInfo e o EpiMap já exportam seus mapas no formato lido pelo BUGS, mas a tabela de atributos tem que ser exportada para o R a fim de ser convertida (Figura 1f).

O programa BayesX permite a importação de mapas de polígonos num formato bastante semelhante ao formato exportado pelo EpiMap. Uma vez importado o mapa, a estrutura de vizinhança dos polígonos é construída e os pesos necessários às análises de regressão espacial são calculados. A tabela de atributos é importada como arquivo texto (Figura 1g). Os resultados das análises podem ser visualizados no próprio BayesX.



**Figura 1 . Ilustração dos tipos de integração entre SIG e Modelos Inferenciais Bayesianos.**

Desta forma, se um SIG exporta seus polígonos num formato semelhante ao requerido pelo BayesX e sua tabela no formato texto, ele pode ser integrado ao BayesX.

### 3.1 – O futuro

Atualmente, todas as alternativas de integração ente SIG e Modelos Inferenciais Bayesianos passam pelo programa R, com exceção da que utiliza o programa BayesX. O programa R é uma ferramenta promissora para a análise bayesiana de dados espaciais. Assim como já existe um pacote R para krigagem bayesiana, pacotes para análise bayesiana de outros tipos de dados espaciais podem surgir a qualquer momento. Deste modo, as alternativas de integração SIG e R deveriam ser cuidadosamente consideradas na discussão da integração entre SIG e Modelos Inferenciais Bayesianos.

Considerando o BUGS e o BayesX os principais *softwares* para análise bayesiana de dados, como fazer a integração entre SIG e Modelos Inferenciais Bayesianos via BUGS/BayesX? Duas alternativas são apresentadas a seguir, em ordem crescente de dificuldade de implementação:

#### 1) **Acoplamento fraco:**

1.1) SIG-BayesX: o SIG exportaria o mapa de polígonos no formato requerido pelo BayesX e a tabela de atributos no formato texto. Os dois arquivos seriam importados no BayesX, que já calcula a matriz de vizinhança e pesos. Após as análises, os resultados seriam exportados num formato que permita ao SIG atualizar sua tabela de atributos.

1.2) SIG-BUGS: o SIG exportaria sua tabela de atributos no formato requerido pelo BUGS, assim como a matriz de vizinhança e os pesos necessários à análise de regressão espacial, se for o caso (O BUGS calcula a matriz de adjacência, mas não os pesos). Os arquivos gerados seriam importados no BUGS. Os resultados das análises seriam exportados no formato permitido pelo BUGS e seriam importados pelo SIG, que atualizaria sua tabela de atributos e geraria a visualização. As adaptações necessárias para este tipo de integração devem ser feitas no SIG, tornando possível a leitura e escrita das tabelas de atributos no formato requerido pelo BUGS.

2) **Acoplamento forte:** o SIG chamaria o BUGS/BayesX em seu ambiente. A exemplo do que acontece na integração GRASS-R e R-BUGS, o SIG usaria uma macro linguagem e carregaria o BUGS/BayesX dentro de seu ambiente através de uma API BUGS-to-SIG ou BayesX-to-SIG. O modo “linha de comando” implícito pelo uso de uma macro linguagem seria o mais natural, já que o WinBUGS e o BayesX fazem uso de comandos para explicitar o modelo e entrada de dados, embora uma interface gráfica pode ser pensada, semelhantemente as que são utilizadas na integração ArcView-SPlus e ArcView-SpaceStat.

Existe ainda a alternativa de **integração completa**, incorporando os métodos inferenciais bayesianos nas opções de análise do SIG. O método bayesiano empírico para estimação de taxas (dados de área) já está implementado no TerraView 3.0.3, por exemplo. Apesar de ser extremamente confortável para o usuário, esta alternativa pode ser custosa do ponto de vista operacional no caso da implementação da abordagem

bayesiana completa. Isto porque o desenvolvimento e atualização dos algoritmos para geração e avaliação dos resultados exigiriam um esforço do desenvolvedor de SIG, que desviaria sua atenção de outras funções mais específicas do sistema. Além disto, esta não parece ser a estratégia mais promissora, quando comparada às estratégias de acoplamento (Ungerer e Goodchild, 2002).

É necessário ressaltar que a avaliação dos resultados fornecidos pelo WinBUGS (diagnósticos de convergência das cadeias geradas) deve ser feita à parte pelos programas específicos para tal (ex: CODA, BOA). Esta observação também vale para as alternativas de integração já existentes e relacionadas na seção 3.1.

#### **4. Considerações Finais**

A análise estatística de dados espaciais é considerada a mais sofisticada aplicação de SIG, embora não seja a mais significativa comercialmente (Goodchild e Haining, 2004). Também sob o aspecto comercial, a análise bayesiana de dados ainda é considerada uma técnica pouco interessante, visto que não existe nenhum *software* estatístico comercial com métodos bayesianos implementados. Neste contexto, convém notar que a maioria dos programas envolvidos nas atuais alternativas de integração são programas *freeware* e/ou *open source*.

A não existência de uma integração direta entre os *softwares* de SIG e os de análise bayesiana de dados é uma lacuna a ser preenchida. Para esta tarefa, o desenvolvimento da integração com os SIG *open source/freeware* parece ser o caminho mais promissor.

A integração entre R e BUGS representa uma ferramenta poderosa na análise bayesiana de dados espaciais, visto que o R está integrado a alguns *softwares* de SIG. Na verdade, a integração SIG-R tem um grande potencial na integração entre SIG e Modelos Inferenciais Bayesianos. O projeto R tem ganhado muitos adeptos na comunidade de análise estatística de dados. É natural que os métodos bayesianos para análise de dados espaciais sejam incorporados às funcionalidades do ambiente R, como já é o caso dos pacotes GeoR e GeoRglm.

A integração com Modelos Inferenciais Bayesianos oferece aos Sistemas de Informação Geográfica mais uma oportunidade de consolidar seu potencial como importante ferramenta para Análise de Dados Espaciais.

#### **7. References**

- Akman, V. E. and A. E. Raftery (1986). Bayes Factors for Nonhomogeneous Poisson Processes with Vague Prior Information. *Journal of the Royal Statistical Society Series B-Methodological*, vol. 48, n. 3, pp. 322-329.
- Bailey, T. e A. Gatrell (1995) *Interactive Spatial Data Analysis*. London, Longman Scientific and Technical, 413 p.
- Banerjee, S. , Gelfand, A. E. , Polasek, W. (2000). Geostatistical modelling for spatial interaction data with application to postal service performance. *Journal of Statistical Planning and Inference*, vol. 90, número 1, pp. 87-105.

- Carlin, B. e Louis, T. A. (1996) Bayes and Empirical Bayes Methods for Data Analysis. Ed. Chapman & Hall, London, 399 p.
- Fischer, M. M. , Scholten, H. J. e Unwin, D. (1996). Geographic Information Systems, spatial data analysis and spatial modelling: an introduction. In: Spatial analysis perspectives in GIS, ed. Fischer, M. M. , Scholten, H. J. e Unwin, D., Taylor and Francis.
- Fook, K. D. (2005) Integração da Estatística Espacial em ambientes GIS. Monografia de Qualificação, Pós-Graduação em Computação Aplicada, INPE, São José dos Campos, SP. Disponível em: <http://www.dpi.inpe.br/~karla>
- Gamerman, D. (1997) Markov Chain Monte Carlo: stochastic simulation for Bayesian inference. Chapman and Hall, London, 245 p.
- Goodchild, M., Haining, R. et al. (1992). Integrating GIS and Spatial Data-Analysis - Problems and Possibilities. International Journal of Geographical Information Systems, vol. 6, n. 5, pp 407-423.
- Goodchild, M. F. e Haining, R. P. (2004) GIS and spatial data analysis: covering perspectives. Papers in Regional Science, vol. 83, pp. 363-385.
- Kitanidis, P. K. (1986). Parameter Uncertainty in Estimation of Spatial Functions - Bayesian-Analysis. Water Resources Research, vol 22. n. 4, pp 499-507.
- Reis, I. A. (2005). O estado da arte da integração entre Sistemas de Informação Geográfica e Modelos Inferenciais Bayesianos. Monografia da disciplina Estudo Orientado I, Pós-Graduação em Sensoriamento Remoto, INPE, São José dos Campos, SP. Disponível em: <http://www.dpi.inpe.br/~ilka>
- Tsutakawa, R. K. , Shoop, G. L. e Marienfeld, C. J (1985). Empirical Bayes Estimation of Cancer Mortality-Rates. Statistics in Medicine, vol. 4, n. 2, pp 201-212.
- Ungerer, M. J. e Goodchild, M. F. (2002). Integrating spatial data analysis and GIS: a new implementation using the Component Object Model (COM). International Journal of Geographical Information Science, vol. 16, n.1, pp. 41-53.