



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

**O ESTADO DA ARTE DA INTEGRAÇÃO ENTRE
SISTEMAS DE INFORMAÇÃO GEOGRÁFICA E
MODELOS INFERENCIAIS BAYESIANOS**

Ilka Afonso Reis

Trabalho realizado como parte das exigências da disciplina Estudo Orientado
do curso de Doutorado em Sensoriamento Remoto

INPE
São José dos Campos
2005

RESUMO

O objetivo deste trabalho é descrever o estado da arte da integração entre Sistemas de Informação Geográfica (GIS) e Modelos Inferenciais Bayesianos. Para isto, aborda a natureza dos GIS e da análise espacial de dados, em especial, a análise estatística de dados espaciais, inserindo neste contexto os Modelos Inferenciais Bayesianos. Por fim, analisa o que já é possível em termos de integração entre GIS e Modelos Bayesianos, apontando lacunas a serem preenchidas.

Palavras-chave: Análise Bayesiana, GIS, Estatística Espacial, Geoestatística, BUGS, BayesX, R.

SUMÁRIO

1 – Introdução	3
2 – Sistemas de Informação Geográfica (GIS)	4
2.1 – A Evolução do GIS	5
3 – Análise Espacial	6
4 – Modelos Inferenciais Bayesianos	9
4.1 – A Evolução dos Modelos Inferenciais Bayesianos	10
4.2 - Modelos Inferenciais Bayesianos e Análise Estatística de Dados Espaciais	11
5 – A Integração entre GIS e Análise Estatística de Dados Espaciais	15
5.1 - A Integração entre GIS e Modelos Inferenciais Bayesianos: o presente	17
5.2 - A Integração entre GIS e Modelos Inferenciais Bayesianos: o futuro	19
6 – Considerações Finais	21
7 – Referências Bibliográficas	22

1 – Introdução

A integração entre os Sistemas de Informação Geográfica (GIS) 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 especial, da Análise Estatística de Dados Espaciais, tem sido amplamente reconhecida na comunidade de GIS¹.

Paralelamente a estas discussões, os modelos Bayesianos vêm ocupando um espaço cada vez maior na análise estatística de dados. Em 1991, foi inaugurada a *Bayesian Statistical Science Section* dentro da *American Statistical Association* (ASA) e, em janeiro de 2005, aconteceu a *International Workshop/Conference on Bayesian Statistics and its Applications* (IWCBSA), que tinha como um dos tópicos a Estatística Espacial e Ambiental. Este crescimento também ser constatado pelo considerável número de publicações de trabalhos com análise Bayesiana de dados nos periódicos das mais diversas áreas do conhecimento.

Entre os que trabalham com dados espaciais, o apelo da alternativa Bayesiana não poderia ser menor, principalmente para resolver problemas onde a Estatística Clássica não consegue soluções satisfatórias, como, por exemplo, no caso de mapeamento de taxas em pequenas áreas. Somente para exemplificar, uma busca por artigos foi feita no PubMed, biblioteca que congrega uma base de dados com os principais jornais da área biomédica, e em algumas revistas selecionadas². As palavras utilizadas foram “spatial”, “space-time”, “epidemiology” e “bayesian” combinadas com as expressões “bayesian model” e “hierarchical model” e suas variações e adaptações. Esta busca gerou 121 registros entre 1995 e 2005, número que certamente subestima a verdadeira quantidade de trabalhos que fazem uso de dados espaciais e técnicas Bayesianas publicados neste período. Uma pequena amostra destes trabalhos pode ser encontrada na seção de referências deste trabalho (marcadas com *)

¹ Exemplos de alguns encontros recentes para a discussão sobre a integração entre GIS e Análise Espacial:
- Conferência **GIS and Spatial Analysis** da IAMG (International Association of Mathematical Geology), em junho de 2005.
- **StatGIS 2003** : Interfacing Geostatistics, GIS and Spatial Databases – Setembro de 2003.
- Reunião do CSISS: **Specialist Meeting on Spatial Data Analysis Software Tools**, em maio de 2002.
- Reunião da American Statistical Association **Spatial Statistics: Integrating Statistics, GIS, and Statistical Graphics**, em outubro de 2001.
- GISDATA **Specialist Meeting on GIS and Spatial Analysis**, em dezembro de 1993.

² Algumas revistas com trabalhos de Análise Estatística de Dados Espaciais: *International Journal of Geographical Health*, *Geographical Analysis*, *Environmental and Ecological Statistics*, *Mathematical Geology*, *Envirometrics*, *Journal of Statistical Inference and Planning*, *Journal of Royal Statistical Society – Serie C (Applied Statistics)* e periódicos publicados pela ASA. Algumas buscas foram limitadas ao período de 2000 a 2005.

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 Sistemas de Informação Geográfica e de Análise Espacial. No entanto, não são muitas as evidências de que haja preocupação específica com este tópico atualmente, o que pode ser devido ao fato de que o foco das discussões esteja na Análise Espacial como um todo.

Este trabalho visa a preencher esta lacuna, descrevendo o estado da arte da integração entre Sistemas de Informação Geográfica e Modelos Inferenciais Bayesianos. Para isto, o texto foi estruturado em seções, as quais são descritas a seguir. Na Seção 2, são apresentadas algumas definições de Sistemas de Informação Geográfica e seus modelos de dados, assim como um breve histórico de sua evolução e alguns dos *softwares* atuais que consideramos importantes na integração com modelos Bayesianos. A Seção 3 apresenta o conceito de Análise Espacial, com enfoque especial na Análise Estatística de Dados Espaciais. Na Seção 4, é feita uma breve introdução às idéias da Estatística Bayesiana, situando a aplicação de seus modelos inferenciais em cada uma das áreas da Estatística Espacial explicitadas na seção anterior. A Seção 5 discute os tipos de integração entre os *softwares* de Sistemas de Informação Geográfica e os de Análise Estatística de Dados Espaciais atualmente existentes, destacando o que há em termos de integração entre Sistemas de Informação Geográfica e Modelos Inferenciais Bayesianos. Finalmente, na Seção 6, são feitas as considerações finais sobre o tema.

2 – Sistemas de Informação Geográfica (GIS)

Um Sistema de Informação Geográfica (GIS) pode ser definido como um conjunto poderoso de ferramentas para coletar, armazenar, recuperar, transformar e apresentar dados espaciais sobre o mundo real (Burrough, 1986). A esta definição, Goodchild (1987) acrescenta a função de analisar a informação geográfica, observando que o elemento que distingue um GIS dos outros tipos de sistemas que lidam com dados espaciais é sua ênfase na análise. Estas são visões de GIS como um produto, pois o termo pode significar também uma indústria, uma tecnologia e uma ciência (Reader, 1995)

Há dois modelos de dados fundamentais para representar o componente espacial da informação geográfica: o **modelo vetorial** e o **modelo matricial** (Fischer et al., 1996). O modelo vetorial enxerga o espaço geográfico como uma coleção de objetos que podem ser representados como pontos, linhas ou polígonos. Numa determinada escala, por exemplo, os estados de um país podem ser representados como polígonos e suas capitais, como pontos.

O segundo modelo subdivide o espaço em campos, que podem ser modelados de diferentes maneiras, como, por exemplo, uma grade de células (pixels) numa imagem de satélite, uma grade de pontos regularmente espaçados no modelo de elevação digital, um conjunto de polígonos não-sobrepostos em um mapa de solos ou isolinhas digitalizadas em um mapa de contorno.

Enquanto o modelo vetorial é mais utilizado nas áreas de ciências sociais e epidemiologia, o modelo de campos é mais comum nas ciências ambientais (Fischer et al., 1996).

Câmara e Monteiro (2004) fazem uma extensa exposição do problema de representar o mundo real no ambiente computacional de um GIS, trazendo à discussão o “paradigma dos quatro universos”.

2.1 - Evolução do GIS

O surgimento dos primeiros Sistemas de Informação Geográfica foi motivado pela necessidade de processar computacionalmente informações cartográficas já disponíveis em papel (Goodchild, 2000; Burrough, 1998). Na década de 60, foi construído o *Canada Geographic Information System*, considerado o primeiro GIS. Paralelamente, o *U. S. Bureau of the Census*, órgão americano responsável pelo censo, construiu um sistema de informação geográfica rudimentar para processar e tabular os dados de agregação espacial do censo de 1970 (Goodchild, 2000).

Em 1982, o *U.S. Army Corps of Engineers Construction Engineering Research Laboratory* (CERL) iniciou desenvolvimento de um sistema de informação geográfica para auxiliar no gerenciamento de instalações militares, o GRASS (*Geographic Resources Analysis Support System*) (Mitasova e Neteler, 2004) (<http://grass.baylor.edu/>). Este é um sistema de código aberto e atualmente suporta os formatos vetorial e matricial. Seu desenvolvimento fomentou a criação da *Open GRASS Foundation* (OGF) que, mais tarde, se tornou o *Open GIS Consortium* (OGC), cuja missão é “disponibilizar abertamente especificações de interfaces espaciais para uso global” (<http://www.opengeospatial.org/>).

Atualmente, existe uma variedade de GIS de distribuição gratuita desenvolvidos para aplicações gerais e também para as mais específicas (<http://freegis.org/>).

Os primeiros GIS comerciais surgiram no final da década de 70. Com a queda do preço dos computadores no início dos anos 80, estes sistemas puderam ser adquiridos por universidades, corporações e agências governamentais regionais. Hoje, a indústria do GIS é multibilionária. As aplicações de GIS estão presentes em quase todas as disciplinas que lidam com dados geograficamente referenciados, desde as ciências atmosféricas, passando pela epidemiologia até a criminologia e história (Goodchild e Haining, (2004). Entre os GIS comerciais, destacam-se os da família **ArcGIS** (ArcInfo e ArcView) e o **IDRISI**.

No Brasil, o desenvolvimento dos primeiros GIS teve início na década de 80, com iniciativas acadêmicas e comerciais. No setor acadêmico, destaca-se o **SAGA** (Sistema de Análise Geo-Ambiental), desenvolvido pelo Laboratório de Geoprocessamento do Departamento de Geografia da UFRJ (<http://www.lageop.ufrj.br/saga/>). Em 1992, o Instituto Nacional de Pesquisas Espaciais (INPE) começou o desenvolvimento do Sistema de Processamento de Informações Georreferenciadas (**SPRING**) (Câmara et al., 2002). O SPRING é um GIS gratuito e está disponível na Internet desde 1996 (www.dpi.inpe.br/spring).

Com base na experiência do SPRING, seus autores estão construindo a **TerraLib**, uma biblioteca de classes de código aberto para desenvolvimento de aplicações de GIS (www.terralib.org). Até o momento, uma das mais importantes aplicações da TerraLib é o *software* **TerraView**, um GIS gratuito e de código aberto (<http://www.dpi.inpe.br/terraview>). O TerraView trabalha com dados vetoriais e matriciais, permitindo a visualização de dados geográficos, consultas espaciais e análises estatísticas (Câmara et al., 2002).

Em adição a sua importante função de gerenciar dados espaciais, os Sistemas de Informação Geográfica representam uma base sólida para a Análise Espacial (Goodchild e Haining, 2004).

3 - Análise Espacial

O conceito de Análise Espacial é definido com enfoques diferentes por diversos grupos de autores, mas com um ponto em comum: a utilização ou não de técnicas estatísticas define dois campos de estudo dentro da Análise Espacial.

Bailey e Gatrell (1995) chamam a atenção para a distinção entre Análise Espacial e Análise de Dados Espaciais. Ambas as análises utilizam dados geograficamente referenciados. Contudo, a Análise de Dados Espaciais envolve dados observacionais, com o objetivo de entender, explicar e prever estes fenômenos, enquanto a Análise Espacial também lida com outros problemas como, por exemplo, minimização de custos de transporte, localização ótima de infra-estrutura e problemas de rota, cuja solução envolve técnicas matemáticas de otimização. Também para Fischer et al. (1996), Análise Espacial é mais do que Análise Estatística de Dados Espaciais. Em seu trabalho, os autores identificaram duas principais áreas de estudo: Modelagem Espacial e Análise Estatística de Dados Espaciais.

A Modelagem Espacial inclui uma ampla diversidade de modelos determinísticos e estocásticos utilizados nas ciências ambientais, sociais e de economia regional, entre outras. Os modelos determinísticos utilizam leis físicas conhecidas, enquanto a abordagem estocástica utiliza a teoria das probabilidades para descrever processos como erosão, movimento de reservatórios aquíferos subterrâneos e absorção de poluentes, por exemplo.

Na área de Análise Estatística de Dados Espaciais (AEDE), Fischer et al. (1996) identificaram quatro grandes sub-áreas nas quais a prática atual de GIS pode se fortalecer:

- 1) Amostragem de objetos no banco de dados e escolha da escala de análise adequada;
- 2) Compatibilização de dados coletados na mesma área de estudo, porém em conjuntos de zonas diferentes, visando à comparação de variáveis;
- 3) Análise exploratória de dados espaciais, incluindo a procura por características dos dados tais como tendências, valores discrepantes (*outliers*), padrões e associações espaciais;
- 4) Análise inferencial de dados espaciais, que diz respeito à estimação de parâmetros de modelos e teste de hipóteses baseado em certas suposições.

As duas últimas sub-áreas englobam técnicas que freqüentemente são chamadas de Estatística Espacial. Bailey e Gatrell (1995) dividem a Estatística Espacial em quatro grandes áreas, de acordo com o tipo de dado analisado:

- 1) Análise de **processos pontuais**, na qual a localização do ponto é o componente aleatório, com a existência ou não de observações associadas a este ponto. Exemplo: estudo da segregação espacial entre os indivíduos masculinos e femininos de uma espécie de planta a partir do mapeamento da posição destes indivíduos (Marques et al., 2002);
- 2) Análise de **dados espacialmente contínuos** (Geoestatística), na qual a localização do ponto é fixa (conhecida) e o componente aleatório são as observações feitas neste ponto. Exemplo: estudo da emissão de enxofre numa região a partir de medição em determinados locais (Host, 1999);
- 3) Análise de **dados de área**, na qual as observações (componente aleatório) estão associadas a uma área (polígono). Alguns autores denominam este tipo de dado como dado de bloco (Gelfand et al, 2001). Exemplo: estimação de taxas de uma doença nos distritos sanitários de uma cidade (Assunção et al., 2001);
- 4) Análise de dados **de interação espacial**, que consiste no estudo dos fluxos a partir de um conjunto de origens para um conjunto de destinos e, portanto, está relacionada a pares de pontos ou pares de áreas. Estes fluxos podem ser pessoas ou bens de consumo, por exemplo (Bailey e Gatrell, 1995).

A Figura 1 mostra o esquema de classificação das áreas da Análise Espacial, combinando-se as definições de Fischer et al. (1996) e Bailey e Gatrell (1995).

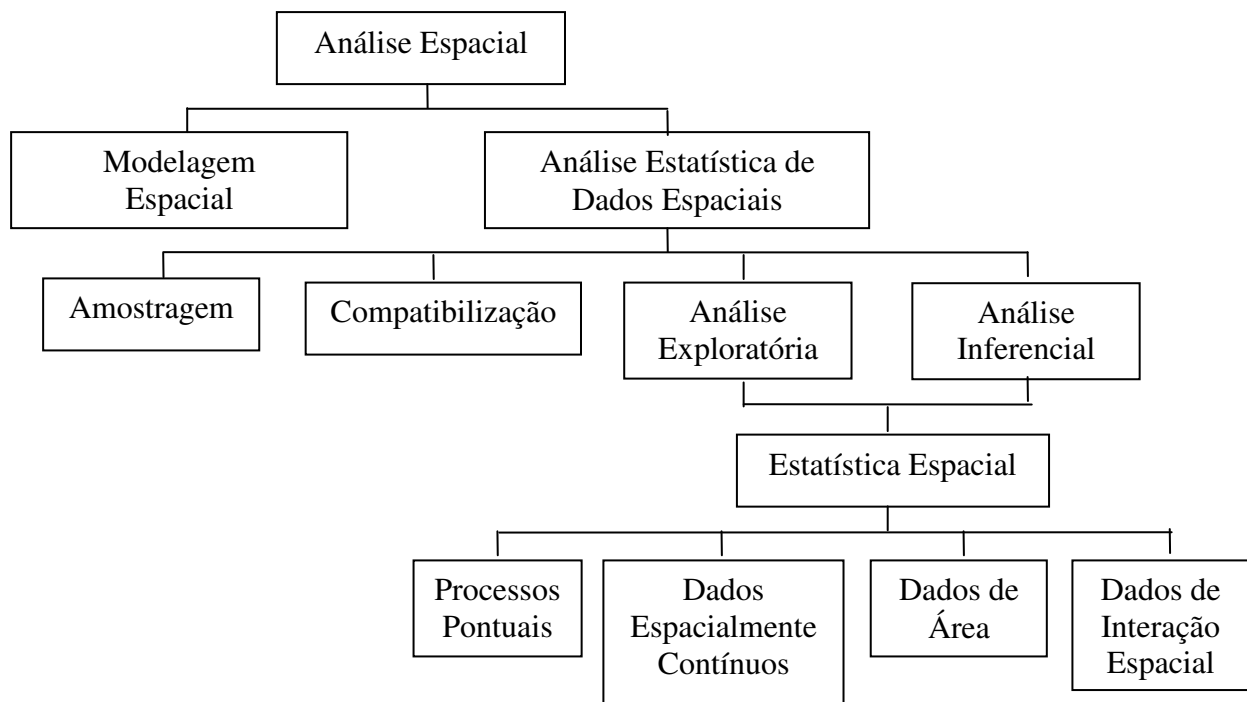


Figura 1 – Campos da Análise Espacial e da Análise Estatística de Dados Espaciais identificados por Fischer et al. (1996) e as áreas da Estatística Espacial identificadas por Bailey e Gatrell (1995).

Cabe ressaltar aqui que existe uma certa confusão entre os termos Geoestatística e Estatística Espacial na literatura de Análise Espacial. Alguns textos falam dos dois conjuntos de técnicas como se fossem excludentes, ou ainda incluem técnicas de análise de dados que não são os espacialmente contínuos na definição do termo Geoestatística. Isto pode ser devido à grande popularidade das técnicas de Geoestatística, principalmente no meio em que foram criadas, a Geologia. Este é o termo consagrado para denominar o conjunto de ferramentas estatísticas aplicadas às Geociências, assim como o termo Bioestatística é utilizado na área Biomédica. No entanto, convém lembrar que o tipo de dado espacial com o qual lida a Geoestatística é apenas um dos tipos de dados espaciais, ou seja, a Geoestatística é uma das áreas da Estatística Espacial, como bem estabelece Bailey e Gatrell (1995) e Cressie (1993).

Os Modelos Bayesianos podem ser utilizados como método de inferência nas técnicas de Estatística Espacial. Sendo assim, este texto dará maior ênfase a esta parte da Análise Espacial.

4 - Modelos Inferenciais Bayesianos

O objetivo da Inferência Estatística é estimar parâmetros de uma certa população e/ou testar hipóteses sobre eles. Uma população pode ser o conjunto de valores de temperatura de uma região, por exemplo, e um parâmetro pode ser uma média, uma proporção ou um coeficiente de regressão. Geralmente, a população não pode ser inteiramente estudada, devido ao seu tamanho, dificuldade de acesso ou de medição. Nestes casos, recorre-se a uma amostra desta população.

Para a Inferência Clássica, toda informação sobre o parâmetro a ser estimado está contida na amostra. O modelo probabilístico é estabelecido sobre os dados amostrais, imaginando-se conhecidos os parâmetros. Este modelo é o que se chama Função de Verossimilhança, uma função escrita em termos dos dados e dos parâmetros. O método de estimação conhecido como Método da Máxima Verossimilhança consiste em encontrar os valores dos parâmetros que maximizam a função de verossimilhança estabelecida com aqueles dados amostrais. Isto é, observada a amostra, quais os valores dos parâmetros tornam mais prováveis os dados coletados? Estes valores são chamados estimativas de máxima verossimilhança. A estas estimativas, podem ser associados *intervalos de confiança*, expressando a incerteza embutida no processo de estimação.

Na Inferência Bayesiana, nem toda informação sobre o parâmetro vem da amostra. Parte dela está contida na amostra e é expressa através da 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, a chamada distribuição *a priori* do parâmetro, ou simplesmente, *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 (Bayes, 1763), resultando na distribuição *a posteriori* do parâmetro. Esta distribuição é utilizada nas inferências sobre o parâmetro. De posse da distribuição do parâmetro conhecidos os dados (*a posteriori*), é possível calcular probabilidades, percentis, média, entre outras propriedades.

A distribuição *a priori* dos parâmetros de interesse também depende de outros parâmetros, os *hiperparâmetros*, denominados assim para serem diferenciados dos parâmetros principais. A estes hiperparâmetros, podem ser atribuídos valores ou outras distribuições *a priori*, as chamadas *hiperprioris*. Este processo continua até um nível em que valores são atribuídos aos hiperparâmetros. A referência a este processo como Modelagem Hierárquica é bastante comum.

Se existe nenhum ou pouco conhecimento *a priori* sobre o parâmetro, costuma-se expressar esta ignorância através da atribuição de uma distribuição *a priori* conhecida como *priori vaga*, na qual a variância é grande, deixando um amplo domínio para a variação dos valores do parâmetro.

Uma das abordagens para lidar com os parâmetros ou com os hiperparâmetros é estimá-los usando os dados da amostra. Esta abordagem é conhecida como Bayesiana Empírica, na qual são obtidas estimativas pontuais do parâmetro e de sua variância. Isto diferencia a abordagem Bayesiana Empírica da chamada Bayesiana Completa, na qual a distribuição *a posteriori* do parâmetro é obtida de forma completa.

4.1 - Evolução dos Modelos Inferenciais Bayesianos

Os métodos Bayesianos datam de 1763, quando o reverendo Thomas Bayes publicou seu artigo original, explicitando a relação hoje conhecida como Teorema de Bayes. As técnicas Bayesianas geraram algum interesse em Laplace, Gauss e outros cientistas no século 19, mas foram ignoradas pelos estatísticos do começo do século 20. Os chamados clássicos ou frequentistas argumentavam que os métodos Bayesianos só poderiam resolver problemas simples, onde a distribuição *a posteriori* pudesse ser escrita analiticamente. Isto exigia a escolha de distribuições *a priori* muitas vezes convenientes e, por isto, nem sempre representantes do verdadeiro pensamento do analista. Além disto, problemas mais complicados requereriam modelos mais complexos e, assim, o cálculo de integrais também mais complicadas (Carlin e Louis, 1996).

Na década de 80, com o advento de computadores mais rápidos e mais baratos, a análise Bayesiana de dados pôde avançar e eliminar as restrições sobre *prioris* e modelos. Os métodos de integração de Monte Carlo tornaram-se populares entre os Bayesianos, assim como os métodos iterativos de Monte Carlo, entre eles, o *Markov Chain Monte Carlo* (MCMC). As técnicas de MCMC permitem a geração de cadeias de Markov, que podem ser definidas como processos descrevendo trajetórias onde quantidades sucessivas são descritas probabilisticamente de acordo com o valor de seu predecessor imediato (Geman, 1997). Em muitos casos, estes processos tendem a um equilíbrio e as quantidades limites seguem uma distribuição probabilística que não varia mais (distribuição de equilíbrio). As técnicas MCMC possibilitam a simulação de uma distribuição através da sua incorporação como distribuição de equilíbrio de uma cadeia de Markov, simulando desta cadeia até que ela atinja seu equilíbrio. Neste estado, os valores da cadeia de Markov podem ser vistos como uma amostra de valores (correlacionados) da distribuição *a posteriori* conjunta dos parâmetros. De posse desta amostra, pode-se calcular médias, medianas, percentis, probabilidades, enfim, qualquer propriedade desejada da *posteriori*. O algoritmo de Metropolis-Hastings (Hastings, 1970) e o Gibbs Sampler (Geman e Geman, 1984; Gelfand e Smith, 1990; Casella e George, 1992) são os métodos mais utilizados para geração das cadeias de Markov.

O programa **BUGS** (*Bayesian Inference Using Gibbs Sampler*) e sua versão para Microsoft Windows (**WinBUGS**) (Lunn et. al, 2000) são os mais conhecidos *softwares* para análise Bayesiana

de dados. Estes programas são gratuitos e muito bem documentados, com um manual repleto de exemplos e uma lista de discussão bastante ativa na Internet (<http://www.mrc-bsu.cam.ac.uk/bugs>). No BUGS (WinBUGS), é preciso fornecer o modelo probabilístico para a função de verossimilhança e também para a distribuição *a priori* de cada parâmetro e hiperparâmetro (Figura 2a). A entrada de dados utiliza um formato semelhante ao utilizado pelo *software* estatístico R. O BUGS então fornece as cadeias de valores simulados, no tamanho especificado pelo usuário. Estes valores podem ser usados nos diagnósticos de convergência das cadeias, que podem ser feitos nos *softwares* CODA (*Convergence Diagnostic and Output Analysis*) e BOA (*Bayesian Output Analysis*), disponíveis na página do BUGS. Medidas-resumo como média, mediana e percentis, para cada parâmetro, também são fornecidas.

Outro *software* disponível para análise Bayesiana de dados é o **BayesX**, que tem distribuição gratuita na Internet e roda sobre a plataforma MicroSoft Windows (<http://www.stat.uni-muenchen.de/~bayesx/bayesx.html>). O BayesX foi criado para lidar com modelos de regressão complexos e funciona no modo “linha de comando” (Figura 2b). A entrada de dados pode ser feita através da leitura de tabelas em formato texto.

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.

4.2 Modelos Inferenciais Bayesianos e Análise Estatística de Dados Espaciais

Tendo em mente a divisão da Estatística Espacial estabelecida por Bailey e Gatrell (1995), os modelos que utilizam dados de área foram os que primeiro se beneficiaram do uso de técnicas Bayesianas. Um dos primeiros trabalhos utilizando estes métodos foi o de Tsutakawa et al. (1985), no qual os autores estimaram taxas de mortalidade por câncer através do método Bayesiano empírico. Em 1989, Clayton abordou o uso modelos Bayesianos hierárquicos (abordagem completa) em epidemiologia descritiva. Com base nestes e em outros trabalhos, Bernardinelli e Montomoli (1992) fizeram a comparação entre métodos Bayesianos empíricos e completos na análise da variação geográfica de risco de doenças, utilizando o método MCMC nas análises Bayesianas completas. Desde então, os modelos Bayesianos para análise de dados de área tornaram-se muito populares, principalmente entre os pesquisadores que lidam com mapeamento de taxas. Para citar apenas alguns trabalhos, diversificando os anos e as fontes de publicação, temos Clayton e Kaldor (1987), Heisterkamp et. al. (1993), Smith et al. (1995), Reis e Assunção (1996), Rytkonen et al. (2001), Johnson (2004) e Yang et al. (2005). Uma das razões para esta popularidade está no

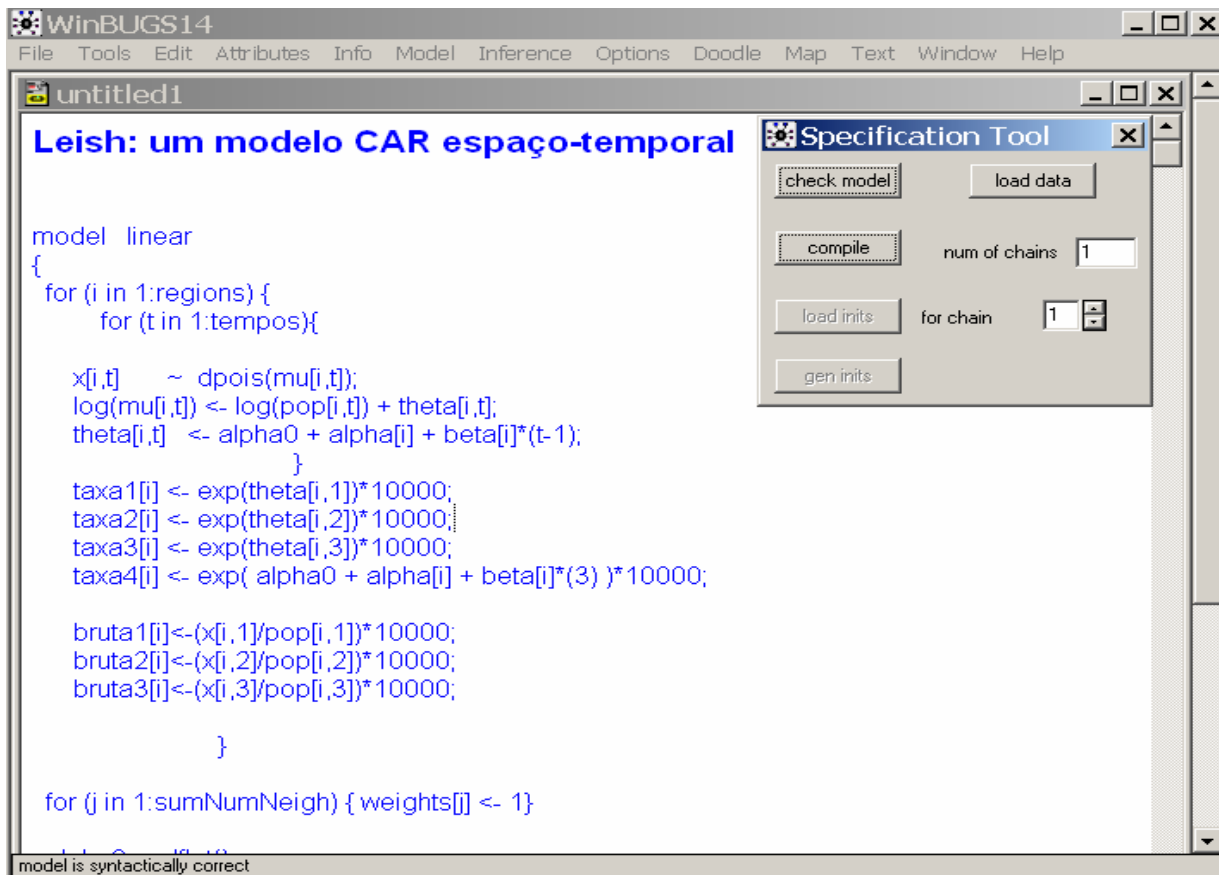


Figura 2a – Tela do WinBUGS 1.4 : janelas onde o modelo é especificado e compilado e os dados são carregados

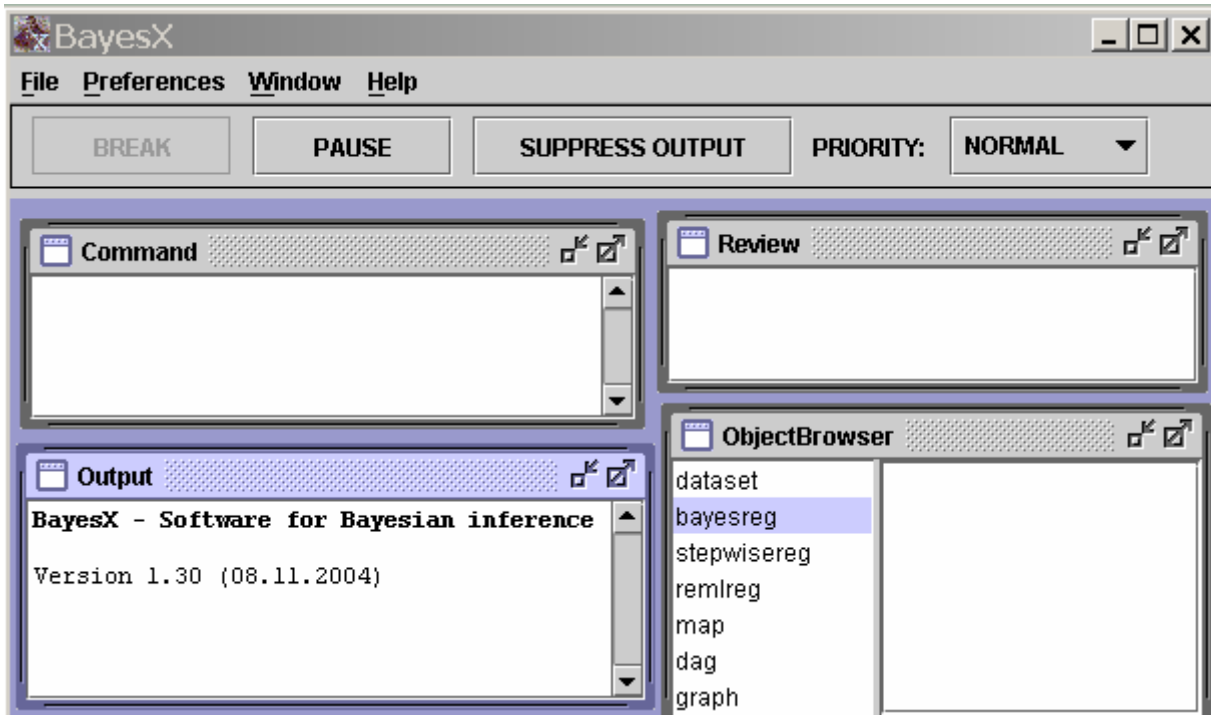


Figura 2b – Tela do BayesX 1.30 : janelas de comando, saídas e visualização de listas de objetos

fato de que os métodos Bayesianos permitem a estimação de taxas para pequenas áreas sem a grande instabilidade presente nas estimativas clássicas. Em 2000, Cressie et. al. fizeram uma revisão dos modelos utilizados na análise Bayesiana hierárquica de dados de área, abordando modelos espaço-temporais e estimação de taxas extremas.

Considerando dados espacialmente contínuos, em 1986, Kitanidis publicou um trabalho sobre análise Bayesiana da incerteza na estimação de parâmetros em funções espaciais. Neste trabalho, o autor mostrou que a krigagem ordinária e a universal, principais técnicas para análise de dados espacialmente contínuos, podem ser vistas como técnicas de predição Bayesiana onde existe uma ignorância *a priori* sobre os parâmetros da superfície de tendência (atribuição de *prioris* vagas para a média da superfície). Omre (1987) explicita o termo krigagem Bayesiana, discussão que continua em Omre (1989), trabalho no qual o autor utiliza o método Bayesiano empírico. Diggle et al (1998) fazem uma revisão dos métodos de krigagem para dados gaussianos e propõem um modelo para dados não gaussianos, utilizando métodos bayesianos na estimação dos parâmetros envolvidos. Muitos são os trabalhos que já fazem uso da krigagem Bayesiana. Para citar apenas alguns, Host (1999) estudou a emissão de enxofre numa região a partir de medição em determinados locais, utilizando krigagem Bayesiana; Diggle e Ribeiro Jr.(2002) discutiram a geoestatística baseada em modelos gaussianos e, aplicando métodos Bayesianos como em Kitanidis (1986) e Handcock e Stein (1993), analisaram dados de precipitação pluviométrica no Paraná, Brasil; Cowles e Zimmerman (2003) fizeram uma análise espaço-temporal dos dados de deposição de ácido, utilizando a abordagem Bayesiana. Recentemente, a precipitação pluviométrica no Pará (Brasil) também foi modelada através da krigagem Bayesiana, na tentativa de entender seu papel na incidência de malária no estado (Nobre et al., 2005); .

Schmidt et al. (2003) fazem uma revisão dos modelos para análise de dados de área e espacialmente contínuos, com foco especial na análise Bayesiana destes tipos de dados espaciais.

Quanto aos processos pontuais, uma das referências mais antigas é o trabalho de Akman e Raftery (1986), que abordaram o processo de Poisson não-homogêneo e fatores de Bayes. Em outros trabalhos, a incorporação dos métodos Bayesianos à análise de processos pontuais se dá na modelagem da função intensidade do processo, utilizando-se, por exemplo, o processo de Cox log-gaussiano (Moller et al., 1998). Entre os trabalhos publicados, podemos citar: Syversveen e Omre (1997); Ogata (1996) ; Cressie e Lawson (2000) e Brix e Diggle (2001).

A análise dos dados de interação espacial não é um assunto tão extensamente estudado quanto a análise dos outros tipos de dados espaciais, sob o ponto de vista Bayesiano. Alguns autores discutiram o problema, entre eles, Banerjee et al. (2000) e Gill e Swartz (2004).

Os *softwares* atualmente disponíveis para análise Bayesiana de dados (WinBUGS e BayesX) possuem algumas funções para lidar com dados espaciais. O WinBUGS possui funções para modelagem de dados de área e especialmente contínuos. As versões mais recentes (1.4.1 em diante) incluem um módulo chamado **GeoBUGS**, permite a importação de mapas de polígonos em vários formatos, construção e manipulação da matriz de vizinhança e visualização dos resultados nestes mapas.

O BayesX permite a análise de dados de área, incluindo a importação de mapas de polígonos.

Com respeito aos *softwares* disponíveis para análise estatística de dados espaciais, existem opções comerciais e não comerciais. Neste texto, serão apresentadas somente aquelas que julgamos importantes na integração com os GIS.

Dentre todos os programas de análise estatística de dados espaciais atualmente existentes, talvez os mais promissores quanto a integração com GIS sejam os que foram desenvolvidos como implementações da linguagem S: o S-Plus e o R. A linguagem S foi criada especialmente para visualização e exploração de dados, modelagem estatística e programação, permitindo uma grande flexibilidade na análise de dados (<http://cm.bell-labs.com/cm/ms/departments/sia/S/>).

O **S-Plus** é um sistema comercial e possui um módulo para lidar com os diversos tipos de dados espaciais (dados pontuais, especialmente contínuos e de área), o **S+SpatialStats**. Por sua vez, 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 **spancs** e **spatstat** (análise de padrões pontuais) (<http://gauss.est.ufpr.br/CRAN/>). Em particular, o projeto **R-spatial** desenvolve métodos para lidar com dados espaciais (<http://sourceforge.net/projects/r-spatial/>). As contribuições ao desenvolvimento do R são bastante intensas, com novos e variados pacotes tornando-se disponíveis com muita frequência.

Também um programa de código aberto, o **Gstat** foi desenvolvido para modelagem geoestatística multivariada, predição e simulação (<http://www.gstat.org/>). Recentemente, o Gstat foi disponibilizado como uma extensão na linguagem S, tanto para o R quanto para o S-Plus (Pebesma, 2004).

O **GeoDA** é um *software* gratuito construído para análise de dados espaciais do tipo espacialmente contínuo e de área. Possui funções de análise de regressão espacial e análise exploratória de dados, incluindo índices de autocorrelação espacial (<https://geoda.uiuc.edu/default.php>).

Como o GeoDA, o programa **SpaceStat** também possui funções de análise exploratória de dados espaciais e de regressão espacial, mas é um programa comercial (<http://www.terraseer.com/products>). Na verdade, os dois *softwares* tiveram a mesma origem, o *Spatial Analysis Laboratory* da *Illinois University*, USA.

5 – A Integração entre GIS e Análise Estatística de Dados Espaciais

O processo de convergência entre GIS e Análise Estatística de Dados Espaciais está longe de estar completo (Goodchild e Haining, 2004). No entanto, muito progresso tem sido feito nesta direção, com diferentes tipos de estratégias de integração (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* GIS. Esta estratégia tem sido utilizada no caso de ferramentas mais consagradas, como as de geoestatística (semivariograma e os diversos tipos de krigagem) e índices de autocorrelação espacial. Alguns sistemas não-comerciais possuem ainda outras ferramentas: simulação seqüencial (geoestatística), função K e mapa de densidade de kernel (análise univariada de pontos) (SPRING 4.4.1); análise de conglomerados espaciais, estimador Bayesiano empírico global e local (TerraView 3.0.3). A integração do tipo completa tem vantagem de permitir a análise dos dados diretamente no GIS. No entanto, para técnicas de análise de dados espaciais mais sofisticadas, e, talvez por isto, menos comercialmente interessantes, este tipo de integração não parece ser a mais promissora. Neste caso, duas alternativas de integração têm sido utilizadas, principalmente nos sistemas comerciais: *loose coupling* e *close coupling*.

Na integração do tipo *coupling*, os *softwares* de GIS e de análise estatística de dados espaciais 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 GIS, incorporando técnicas que, a princípio, não seriam tão vantajosas de serem integradas completamente do ponto de vista comercial.

Na estratégia do tipo *loose coupling*, os dados são exportados do GIS para serem analisados dentro de um pacote de análise estatística de dados espaciais e os resultados são novamente exportados para o GIS com a finalidade de visualização. Exemplos desta estratégia são a integração dos GIS ArcView e TerraView com o programa de análise de dados espaciais GeoDA e do GIS SPRING com o programa SpaceStat. O SPRING exporta dados no formato vetorial para o formato SpaceStat e o GeoDA aceita o formato *shape* exportado pelo ArcView e TerraView. Outro exemplo da estratégia *loose coupling* é a integração entre ArcInfo e R. O pacote R denominado **RArcInfo** permite a importação de arquivos no formato ArcInfo *binary coverages* (Rubio, 2003). A atual versão do RArcInfo ainda não permite a exportação para o formato de leitura do ArcInfo.

A estratégia *loose coupling* pode parecer simples do ponto de vista computacional, já que os programas precisam apenas ser capazes de trabalhar com um formato de dados comum a ambos. No entanto, pode se revelar ineficiente no caso de serem utilizados muitos conjuntos de dados e/ou

serem necessárias muitas análises, pois o tempo e esforço gastos na importação e exportação dos dados pode ser grande.

A estratégia *close coupling* permite que o GIS e o *software* de análise compartilhem um banco de dados comum, com a vantagem de que a análise pode ser feita no ambiente do GIS, a partir da chamada de uma rotina do *software*. Algumas integrações deste tipo exploram as facilidades das linguagens de programação (*macro languages*) que estão disponíveis em vários GIS, como é caso da integração entre o GRASS 5.0 e o programa estatístico R (Bivand, 2000). O ambiente de programação estatística do R é carregado dentro do GRASS através do pacote R denominado GRASS. Para a versão GRASS 6.0, o pacote R chama-se **spgrass6** (<http://sourceforge.net/projects/r-spatial>).

Já no caso ArcView e S-Plus, a integração entre os dois programas acontece por meio de uma ligação bi-direcional. O *S-Plus for ArcView* funciona com uma extensão do ArcView padrão. Uma vez que esta extensão é carregada no ArcView, dois menus (SPLUS e Spatial Statistics) são adicionados à interface usual e a conexão entre S-Plus e ArcView é estabelecida.

A integração entre ArcView e SpaceStat e entre o GIS IDRISI e o *software* de geoestatística Gstat são feitas de maneira semelhante. A extensão *SpaceStat for ArcView* adiciona dois menus a interface do Arcview, permitindo a exportação dos dados do ArcView para o SpaceStat e a visualização dos resultados das análises do SpaceStat no ArcView. No caso do IDRISI, muitas funcionalidades do Gstat podem ser acessadas diretamente na interface do IDRISI, através de uma versão modificada do Gstat C.

A ligação entre o GIS **Grassland**, uma versão comercial do GRASS, e o S-Plus também é do tipo *close coupling* com uma conversação bi-direcional, promovida pela API³ S+ (Bao e Anselin, 2000). Uma vez estabelecida a conexão, os objetos do S-Plus e os resultados da análise estatística podem ser acessados diretamente do Grassland.

A integração entre o *software* TerraView e o R é do tipo *close coupling* e acontece via o pacote **aRT** (API R_Terralib) (<http://www.est.ufpr.br/aRT>). O aRT possui quatro classes para manipular os dados e funções da TerraLib, a biblioteca de funções sobre as quais foi construído o TerraView. Uma vez aberta a sessão do programa R e carregado o pacote aRT, tem-se acesso ao banco de dados e ferramentas TerraLib, assim como podem ser criados e manipulados objetos das classes do TerraLib. As funções do R são utilizadas na análise estatística e os resultados são incorporados ao banco TerraLib. Atualmente, esta forma de integração está disponível apenas no

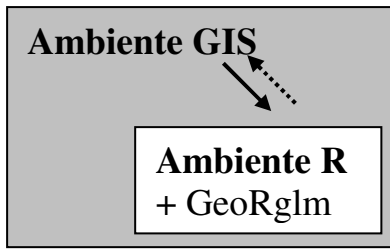
³ API é a abreviatura em inglês do termo *Application Programming Interface*. Na prática, uma API é qualquer interface que permita a um programa o uso das funcionalidades de outro, seja chamando este programa ou sendo chamado por ele.

ambiente LINUX, mas há projetos em andamento para que ela seja possível no ambiente MicroSoft Windows.

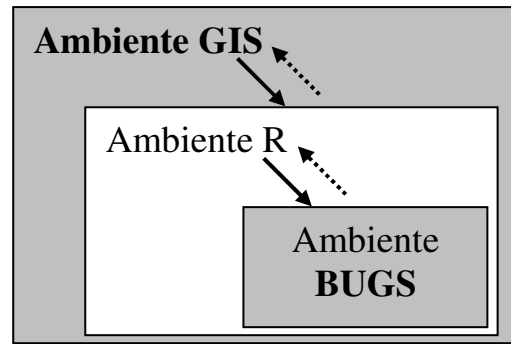
5.1 Integração entre GIS e Modelos Inferenciais Bayesianos: o presente

Considerando o atual estágio de integração entre GIS e os programas de análise estatística de dados espaciais, a integração entre GIS e modelagem Bayesiana de dados pode ser feita de cinco maneiras:

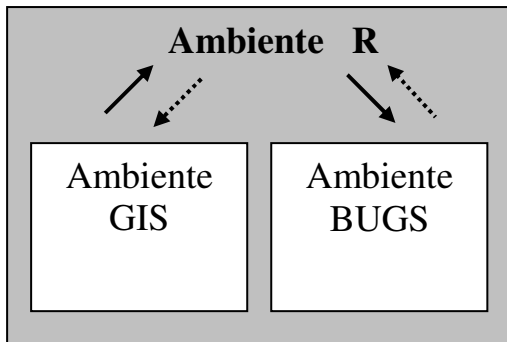
- 1) Integração do tipo *close coupling* do GIS com o programa R e seu pacote GeoRglm, limitada aos modelos de krigagem Bayesiana (Figura 3a). Atualmente, os GIS que possuem integração deste tipo com o R são o GRASS e TerraView;
- 2) Integração em duas etapas do tipo *close coupling*: para outros tipos de modelos, a alternativa seria usar o programa BUGS ou WinBUGS. O WinBUGS possui integração do tipo *close coupling* com o programa R, através do pacote R2WinBUGS, permitindo que as simulações do WinBUGS sejam armazenadas diretamente num objeto R. Assim, a integração entre GIS e BUGS aconteceria via R: funções do WinBUGS podem ser chamadas a partir do R, que, sua vez, seria chamado dentro do GIS (Figura 3b). Seria o caso da integração entre o GRASS e o WinBUGS, no qual o processo ocorre no ambiente GRASS. No caso da integração TerraView e WinBUGS, o processo acontece no ambiente R, que lê os dados do TerraView, passa-os para o WinBUGS, onde são gerados os resultados. Estes são automaticamente armazenados num objeto R e escritos no formato TerraLib via aRT (Figura 3c).
- 3) Integração em duas etapas do tipo *loose e close coupling*: o GIS exportaria a tabela de atributos num formato lido no ambiente R, por exemplo, texto separado por vírgulas. O R chamaria o WinBUGS para analisar os dados e exportaria os resultados num formato acessível ao GIS, onde o banco de dados seria atualizado (Figura 3d). Esta estratégia é limitada à exportação da tabela de atributos com a informação geográfica truncada no caso de dados de área, pois seriam exportados apenas os centróides dos polígonos, 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. No entanto, possibilita que os modelos inferenciais Bayesianos sejam integrados a uma gama maior de softwares GIS.



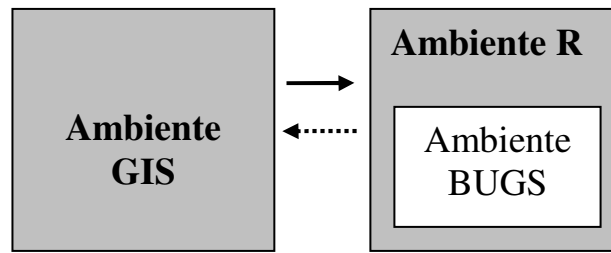
a) *close coupling* GIS-R



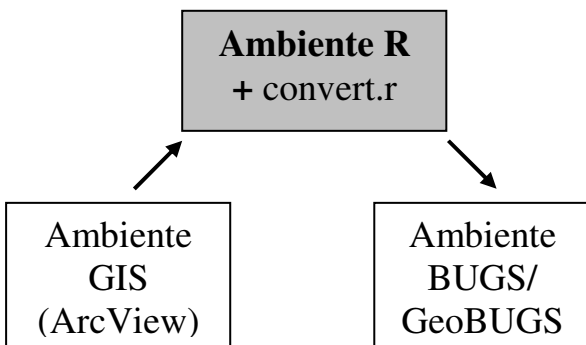
b) *close-close coupling* GIS-R-BUGS



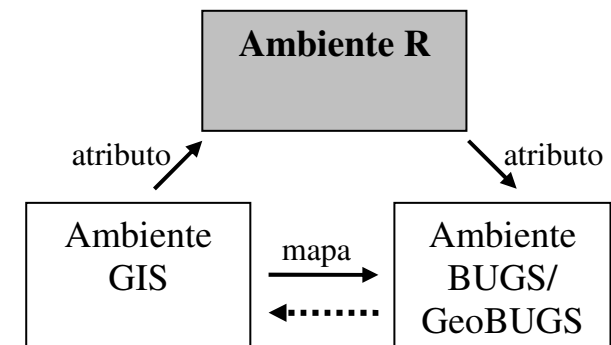
c) *close-close coupling* GIS-R-BUGS



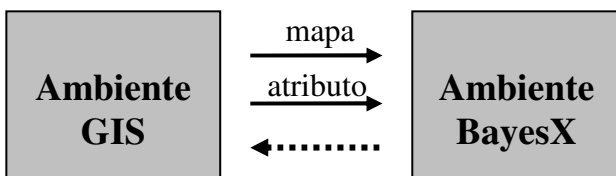
d) *loose-close coupling* GIS-R-BUGS



e) *loose-loose coupling* GIS-R-BUGS



f) *loose-loose coupling* GIS-R-BUGS



g) *loose coupling* GIS-BayesX

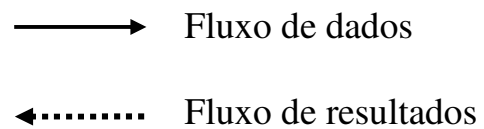


Figura 3 – Ilustração dos tipos de integração entre GIS e modelos inferenciais Bayesianos.

- 4) O módulo GeoBUGS do WinBUGS permite importação dos mapas de polígonos nos formatos ArcInfo, S-Plus e EpiMap⁴. 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 (<http://www.biostat.umn.edu/~yuecui/>). Esta rotina necessita que o arquivo *shape* seja transformado num arquivo CGM *Clear Text*, opção disponível no ArcView. Os atributos têm que ser exportados separadamente. Uma opção é exportá-los no formato CSV e lê-los no R, que é capaz de gerar o formato de entrada de dados requerido pelo WinBUGS. Deste modo, o ArcView pode ser parcialmente integrado ao WinBUGS/GeoBUGS via R, numa combinação de duas *loose coupling* : GIS-R e R-WinBUGS/GeoBUGS (Figura 3e). Considerando que o SPRING e o TerraView exportam dados no formato *shape*, estes dois GIS poderiam ser integrados ao BUGS, desde que haja a conversão do formato *shape* para CGM Clear Text. 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 3f).
- 5) 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 3g). Os resultados das análises podem ser visualizados no próprio BayesX. Desta forma, se um GIS exporta seus polígonos num formato semelhante ao requerido pelo BayesX e sua tabela o formato texto, ele pode ser integrado ao BayesX. Convém lembrar que o arquivo de polígonos terá que ser editado para se adequar ao formato BayesX.

Como na alternativa 3, os resultados gerados nas alternativas 4 e 5 podem ser exportados num formato acessível ao GIS.

5.2 - Integração entre GIS e Modelos Inferenciais Bayesianos: o futuro

Atualmente, todas as alternativas de integração ente GIS 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 GIS e R deveriam ser

⁴ O EpiMap faz parte do EpiInfo, *software* gratuito desenvolvido pela CDC (Center of Disease Control) para entrada e análise de dados epidemiológicos. O EpiMap permite que os dados do EpiInfo sejam visualizados em mapas (<http://www.cdc.gov/epiinfo/>).

cuidadosamente consideradas na discussão da integração entre GIS e Modelos Inferenciais Bayesianos.

Considerando o BUGS e o BayesX os principais *softwares* para análise Bayesiana de dados, como fazer a integração entre GIS e Modelos Inferenciais Bayesianos via BUGS/BayesX? Podemos conceber duas maneiras de se fazer esta integração. Estas alternativas são apresentadas a seguir em ordem crescente de dificuldade de implementação:

1) Integração *loose coupling*:

1.1) GIS-BayesX: o GIS 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 GIS atualizar sua tabela de atributos.

1.2) GIS-BUGS: o GIS 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 GIS, que atualizaria sua tabela de atributos e geraria a visualização. As adaptações necessárias para este tipo de integração devem ser feitas no GIS, tornando possível a leitura e escrita das tabelas de atributos no formato requerido pelo BUGS.

2) Integração *close coupling*: o GIS chamaria o BUGS/BayesX em seu ambiente. A exemplo do que acontece na integração GRASS-R e R-BUGS, o GIS usaria uma macro linguagem e carregaria o BUGS/BayesX dentro de seu ambiente através de uma API *BUGS-to-GIS* ou *BayesX-to-GIS*. O modo “linha de comando” implícito pelo uso de uma macro linguagem seria o mais natural, já que o BUGS, o WinBUGS e o BayesX fazem uso de comandos para explicitar o modelo e entrada de dados. Ainda assim, pode-se pensar numa interface gráfica como a que é utilizada na integração ArcView-SPlus e ArcView-SpaceStat. Isto, porém, não elimina a necessidade do modo “linha de comando”, que é como o BUGS, o WinBUGS e o BayesX trabalham.

Existe ainda a alternativa de integração completa (*full integration*), incorporando os métodos inferenciais Bayesianos nas opções de análise do GIS. 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 GIS 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 do tipo *coupling* (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 5.1.

6 – Considerações Finais

A análise estatística de dados espaciais é considerada a mais sofisticada aplicação de GIS, 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 GIS e os de Análise Bayesiana de Dados é uma lacuna a ser preenchida. Para esta tarefa, o desenvolvimento da integração com os GIS *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 GIS. Na verdade, a integração GIS-R tem um grande potencial na integração entre GIS 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 do 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 – Referências Bibliográficas

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** 48(3): 322-329.

* Assunção, R.M.; Reis, I. A.; C. L. Oliveira (2001) Diffusion and Prediction of Leishmaniasis in a Large Metropolitan Area in Brasil with a Space-Time Model. **Statistics in Medicine**, vol. 20, pp. 2319- 2335

Bailey, T.; Gatrell, A. **Interactive Spatial Data Analysis**. London, Longman Scientific and Technical, 1995, 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.

Bao, S. ; Anselin, L. ; Martin, D. ; Stralberg, D (2000). Seamless integration of spatial statistics and GIS: The S-PLUS for ArcView and the S+Grassland Links. **Journal of Geographical Systems**, vol. 2, pp 287 - 306

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. Philosophical Transactions of Royal Society of London, 53, 370-418. Reprinted, with an introduction by George Barnard, in 1958 in **Biometrika**, vol. 45, 293-315.

* Bernardinelli, L. and Montomoli, C. (1992) Empirical Bayes versus fully Bayesian analysis of geographical variation in disease risk. **Statistics in Medicine** 11, 983-1007.

Bivand, R. S. (2000). Using the R statistical data analysis language on GRASS 5.0 GIS database files. **Computers & Geosciences**, vol. 26, número 9-10, pp. 1043-1052.

Brix, A. e Diggle, P. J. (2001). Spatiotemporal prediction for log-Gaussian Cox processes. **Journal of the Royal Statistical Society Series B-Statistical Methodology**, vol. 63, pp. 823-841.

Burrough, P. A. (1986) **Principles of Geographical Information Systems for Land Assessment**, Oxford University Press, 194 p.

Burrough, P. A.; McDonnell, R. A. (1998) **Principles of Geographical Information Systems**. Oxford University Press, 333 p.

Câmara, G.; Neves, M ; Monteiro, A. M. V.; Souza, R. C. M ; Paiva, J. A.; Vinhas, L. (2002) SPRING and TerraLib: Integrating Spatial Analysis and GIS. In : **Proceedings of the SCISS Specialist Meeting “New Tools for Spatial Data Analysis”**, Santa Bárbara, CA, USA, May, 10-11.

Câmara, G.; Monteiro, A. M. V. (2004) Conceitos Básicos em Ciência da Geoinformação. Em: **Introdução à Ciência da Geoinformação**, Ed: Câmara, G.; Davis, C.; Monteiro, A. M. V. – INPE - São José dos Campos. Disponível na URL: <http://www.dpi.inpe.br/gilberto/livro/introd/>.

Carlin, B.; Louis, T. A. (1996) **Bayes and Empirical Bayes Methods for Data Analysis**. Ed. Chapman & Hall, London, 399 p.

Casella, G. ; George, E. I. (1992). Explaining the Gibbs Sampler. **American Statistician** vol. 46, número 3, pp 167-174.

* Clayton, D. (1989). Hierarchical Bayesian models in descriptive epidemiology. **Proceedings of the XIVth International Biometrics Conference**, pp. 201-213

* Clayton, D. ; J. Kaldor (1987). Empirical Bayes Estimates of Age-Standardized Relative Risks for Use in Disease Mapping. **Biometrics**. 43: 671-681.

* Cowles, M. K. e Zimmerman, D. L. (2003). A Bayesian space-time analysis of acid deposition data combined from two monitoring networks. **Journal of Geophysical Research-Atmospheres**, 108(D24), art. no. 9006.

Cressie, N. (1993) **Statistics for Spatial Data**. Revised Edition. John Wiley and Sons, Inc. 900 p.

* Cressie, N. ; Stern, H. S.; Wright, D. R (2000) Mapping rates associated with polygons. **Journal of Geographical Systems**, 2, pp. 61-69.

* Cressie, N. ;Lawson, A. B. (2000). Hierarchical probability models and Bayesian analysis of mine locations. **Advances in Applied Probability** 32(2): 315-330.

* Diggle, P. J., J. A. Tawn, et al. (1998). Model-based geostatistics. **Journal of the Royal Statistical Society Series C-Applied Statistics**, vol. 47, pp. 299-326.

* Diggle, P. J.; Ribeiro Jr, P. J. (2002) Bayesian Inference in Gaussian Model-based Geostatistics. **Geographical and Environmental Modelling**, volume 6, number 2 pp. 129 - 146

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.

Gamerman, D. (1997) **Markov Chain Monte Carlo: stochastic simulation for Bayesian inference**. Chapman and Hall, London, 245 p.

Gelfand, A. E. ; Smith, A. F. M. (1990) Sampling-based approaches to calculating marginal densities. **Journal of American Statistician Association**, 85, 398-409

Gelfand, A. E.; Zhu, L.; Carlin, B.(2001) On change of support problem for spatial-temporal data. **Biostatistics**, vol. 2, número 1, pp. 31-45.

- Geman, S ; Geman, D.(1984) Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, 6, 721-741.
- * Gill, P. S.; Swartz, T. B. (2004). Bayesian analysis of directed graphs data with applications to social networks. **Journal of the Royal Statistical Society Series C-Applied Statistics** vol. 53, pp. 249-260.
- Goodchild, M. F. (1987) A spatial analytical perspective on geographical information systems **International Journal of Geographical Systems**, vol. 1, número 4, pp. 5-10.
- Goodchild, M. F. (2000) The current status of GIS and spatial data analysis. **Journal of Geographical Systems**, 2, pp. 5-10.
- Goodchild, M., R. Haining, et al. (1992). Integrating GIS and Spatial Data-Analysis - Problems and Possibilities. **International Journal of Geographical Information Systems** 6(5): 407-423.
- Goodchild, M. F.; Haining, R. P. (2004) GIS and spatial data analysis: covering perspectives. **Papers in Regional Science**, **83**, 363-385.
- Hastings, W. K. (1970) Monte Carlo sampling methods using Markov chains and their applications. **Biometrika**, 57, 97-109.
- * Handcock, M. S.; Stein, M. L. (1993). A Bayesian-Analysis of Kriging. **Technometrics** 35(4): 403-410.
- * Heisterkamp S. H. ; Doornbos G. ; Gankema, M. (1993). "Disease Mapping Using Empirical Bayes and Bayes Methods on Mortality Statistics in the Netherlands." **Statistics in Medicine** 12(19-20): 1895-1913.
- * Host, G. (1999) Bayesian estimation of European sulphur emissions using monitoring data and an acid deposition model. **Environmental and Ecological Statistics**, vol. 6, pp. 381-399
- * Johnson, G. (2004). Small area mapping of prostate cancer incidence in New York State (USA) using fully Bayesian hierarchical modelling. **International Journal of Health Geographics** 3(1): 29.
- * Kitanidis, P. K. (1986). Parameter Uncertainty in Estimation of Spatial Functions - Bayesian-Analysis. **Water Resources Research** 22(4): 499-507.
- Lunn, D. J. ; Thomas, A.; Best, N. G. ; Spiegelhalter, D. J. (2000) WinBUGS - a Bayesian modelling framework: concepts, structure and extensibility. **Statistics and Computing**, volume 10, pp. 321-333
- * Marques, A.R.; Fernandes, G. W.; Reis, I. A. , Assunção, R.M. (2002) Distribution of Adult Male and Female *Baccharis concinna* (Asteraceae) in the Rupestrian Fields of Serra do Cipó, Brazil. **Plant Biology**, vol. 4, pp. 94-103

Mitasova, H. ; Neteler, M. (2004) GRASS as Open Source Free Software GIS: Accomplishments and Perspectives, **Transactions in GIS**, vol. 8, numero 2, pp 145-154

* Møller, J., Syversveen, A.-R., and Waagepetersen, R. (1997). Log Gaussian Cox processes: A statistical model for analyzing stand structural heterogeneity. In Proceedings of First European Conference for Information Technology in Agriculture (editors Kure, H., Thysen, I., and Kristensen, A. R.), Department of Mathematics and Physics, The Royal Veterinary and Agricultural University, Denmark, 1997, 339-342.

* Nobre, A. A.; Schmidt, A. M.; Lopes, H. F..(2005) Spatio-temporal models for mapping the incidence of malaria in Pará, **Environmetrics**, volume 16, número 3, pp. 291 – 304

Ogata, Y. (1996). Evaluation of spatial Bayesian models - Two computational methods. **Journal of Statistical Planning and Inference**, vol. 51, número 1, pp.1-18.

* Omre, H. (1987). Bayesian Kriging-Merging Observations and Qualified Guesses in Kriging. **Mathematical Geology** 19(1): 25-39.

* Omre, H. and K. B. Halvorsen (1989). The Bayesian Bridge between Simple and Universal Kriging. **Mathematical Geology** 21(7): 767-786.

Pebesma, E.J. (2004). Multivariable geostatistics in S: the gstat package. **Computers & Geosciences**, vol. 30, pp. 683-691.

Reader, S. (1995) The present state of GIS and future trends. In: **GIS for Health and the Environment**, ed. Savigny, D.; Wijeyaratne, P. IDRC, pp. 172 (http://web.idrc.ca/es/ev-9357-201-1-DO_TOPIC.html)

* Reis, E. A. ; Assunção, R. M.. Mapeamento da Malária em Rondônia usando o Estimador Empírico de Bayes. In: X Encontro Nacional de Estudos Populacionais, 1996, Belo Horizonte. **Anais do X Encontro Nacional de Estudos Populacionais**. Belo Horizonte: ABEP, 1996. v. 2. p. 1077-1097.

* Ryttonen, M., J. Ranta, et al. (2001). Bayesian analysis of geographical variation in the incidence of Type I diabetes in Finland. **Diabetologia** 44: B37-B44.

Rubio, V. G. (2003) RArInfo: Using GIS Data with R. **Proceedings of the 3rd Internacional Workshop on Distributed Statistical Computing (DSC 2003)**, march 20-22, Vienna, Austria. (<http://www.ci.tuwien.ac.at/Conferences/DSC-2003/>)

Schmidt, A. M.; Nobre, A. A.; Ferreira, G. S. (2002) Alguns Aspectos da Modelagem de Dados Espacialmente referenciados, **Revista Brasileira de Estatística**, vol. 63, número 220, pp. 59-88.

* Smith, T.; Charlwood, J.D.; Takken, W.; Tanner, M.; Spiegelhalter, D.J.(1995) Mapping the densities of malaria vectors within a single village. **Acta Tropica**. Vol. 59, número 1, pp. 1-18.

* Syversveen, A. R. and H. Omre (1997). Conditioning of marked point processes within a Bayesian framework. **Scandinavian Journal of Statistics** 24(3): 341-352.

* Tsutakawa, R. K.; Shoop, G. L.; Marienfeld, C. J (1985). Empirical Bayes Estimation of Cancer Mortality-Rates. **Statistics in Medicine** 4(2): 201-212.

* Ungerer, M. J. 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** 16(1): 41-53.

* Yang, G.J.; Vounatsou P.; Zhou, X.N. ; Tanner M. ; Utzinger, H. (2005). A Bayesian-based approach for spatio-temporal modeling of county level prevalence of *Schistosoma japonicum* infection in Jiangsu province, China. **International Journal for Parasitology** 35(2): 155-162.