



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

INTEGRAÇÃO DA ESTATÍSTICA ESPACIAL EM AMBIENTES GIS

Karla Donato Fook

Monografia de Qualificação em Computação Aplicada

INPE
São José dos Campos
2005

RESUMO

Este trabalho apresenta um panorama da integração da estatística espacial em ambientes GIS (*Geographical Information Systems*). Uma visão das necessidades existentes nas comunidades de GIS e de Estatística Espacial com relação a tal integração é destacada. Para tanto, o texto descreve, de forma introdutória, a análise estatística de dados espaciais. A estatística espacial é tratada dentro de um contexto mais voltado à utilização de ferramentas computacionais. Um levantamento do que já existe em termos de integração destas áreas é realizado e exemplos de integração são apresentados. Alguns desafios presentes na integração entre tecnologias são elencados no final do documento.

SUMÁRIO

RESUMO

SUMÁRIO

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE SIGLAS E ABREVIATURAS

1 INTRODUÇÃO	1
2 ESTATÍSTICA ESPACIAL.....	3
2.1 Análise de Pontos ou Eventos	5
2.2 Análise de Superfícies ou Distribuições	6
2.3 Análise de Áreas.....	8
2.4 Análise Espaço-Temporal	10
3 ESTATÍSTICA ESPACIAL E GIS	13
➤ Ligação GRASS e R	17
➤ Ligação Spring e SpaceStat	18
➤ Ligação TerraLib e Open Modeller	20
4 Integração entre R e TerraLib	21
4.1 aRT: API R-TerraLib	22
4.1.1 Porte da API aRT para a plataforma Windows.....	29
5 CONSIDERAÇÕES FINAIS	32
APÊNDICE A.....	34
REFERÊNCIAS BIBLIOGRÁFICAS	37

LISTA DE FIGURAS

FIGURA 1 – Estatística Espacial inserida no contexto da Análise Espacial [Fonte: (Reis, 2005)]	4
FIGURA 2 – Exemplo de dados pontuais [Fonte: (DPI/INPE, 2004a)]	5
FIGURA 3 – Padrões de dados pontuais [Fonte:(DPI/INPE, 2004a)]	6
FIGURA 4 – Exemplo de dados de superfície [Fonte: (Druck et al., 2004)]	8
FIGURA 5 – Exemplo de dados de área [Fonte: (Druck et al., 2004)]	9
FIGURA 6 – Exemplo de cadastro urbano [Fonte: (Casanova et al., 2005)]	11
FIGURA 7 – Estratégia acoplamento forte [Adaptado de: (Bivand and Neteler, 2000)]	14
FIGURA 8 – Estratégia de acoplamento fraco [Adaptado de: (Bivand and Neteler, 2000)]	16
FIGURA 9 – Possíveis Rotas de Transferência entre R/GRASS/PostgreSQL [Adaptado de: (Bivand, 2001; Bivand and Neteler, 2000)]	18
FIGURA 10 – Exemplo de integração: Spring e SpaceStat	19
FIGURA 11 – Ligação R/TerraLib/TerraView	22
FIGURA 12 – Esquema aRT, plataforma Linux	23
FIGURA 13 – Carregamento do pacote aRT no ambiente R	25
FIGURA 14 – Conexão com banco de dados	25
FIGURA 15 – Criação do objeto da classe aRTdb	25
FIGURA 16 – Criação de um layer no banco de dados	26
FIGURA 17 – Criação de um objeto da classe aRTtheme	26
FIGURA 18 – Comandos para Análise de Dados no R	27
FIGURA 19 – Resultado plotado no ambiente R [Fonte: (Neto et al., 2005)]	28
FIGURA 20 – Visualização dos dados no TerraView [Fonte: (Neto et al., 2005)]	28
FIGURA 21 – Esquema aRT, plataforma Windows	29
FIGURA 22 – Carregamento do pacote aRT no ambiente R, plataforma Windows	30
FIGURA 23 – Comando <i>require(aRT)</i> no ambiente R, plataforma Windows	30
FIGURA 24 – Conexão com banco de dados no ambiente R, plataforma Windows	31
FIGURA 25 – Criação do objeto da classe aRTdb, plataforma Windows	31
FIGURA 26 – Arquivo MakeVars	35

LISTA DE TABELAS

TABELA 1–Tipos de Dados na Estatística Espacial [Fonte:(Druck et al., 2004)]	12
TABELA 2 – Estratégias de Integração entre GIS e Estatística Espacial	17

LISTA DE SIGLAS E ABREVIATURAS

API	- <i>Application Programming Interface</i>
aRT	- API R-TerraLib
BUGS	- <i>Bayesian Inference Using Gibbs Sampler</i>
CRAN	- <i>Comprehensive R Archive Network</i>
CRIA	- Centro de Referência em Informação Ambiental
DBMS	- <i>Database Management Systems</i>
DPI	- Divisão de Processamento de Imagens
ESDA	- <i>Exploratory Spatial Data Analysis</i>
ESRI	- <i>Environmental Systems Research Institute</i>
FAPESP	- Fundação de Amparo à Pesquisa do Estado de São Paulo
GIS	- <i>Geographical Information Systems</i>
GRASS	- <i>Geographical Resources Analysis Support System</i>
GSLIB	- <i>Geostatistical Software Library</i>
GWR	- <i>Geographically Weighted Regression</i>
INPE	- Instituto Nacional de Pesquisas Espaciais
PUC/SP	- Pontifícia Universidade Católica de São Paulo
SIG	- Sistema de Informação Geográfica
STIS	- <i>Space-Time Intelligence System</i>
UFPR	- Universidade Federal do Paraná
USP	- Universidade de São Paulo

1 INTRODUÇÃO

A integração entre os Sistemas de Informação Geográfica (SIG ou GIS) e a Estatística Espacial é um desafio relevante para o melhor entendimento do espaço geográfico. Os GIS permitem o armazenamento e a apresentação de dados espaço-temporais e são aplicados na administração de recursos, telecomunicações, planejamento urbano e regional, e em todas as ciências que tratam da superfície da Terra (Câmara et al., 2003a; Goodchild, 2000). A estatística espacial possui técnicas de análise que modelam fenômenos em função de propriedades de processos estocásticos (Bailey and Gatrell, 1995). Desenvolvimentos esboçam GIS e os métodos de estatísticas espaciais como suportes teóricos essenciais para a análise de dados espacial (Goodchild and Haining, 2004). A junção das duas tecnologias permite a união da capacidade de gerenciamento de dados dos GIS com as técnicas de análise da Estatística Espacial.

Ferramentas de análise estatística de dados espaciais e GIS são aplicadas nos campos da saúde, meteorologia, monitorando ambiental e exploração mineral, dentre outros (Bailey, 2001; Druck et al., 2004). Muitas destas ferramentas ainda não operam em um único ambiente de software. Dessa forma, tem-se uma gama de pacotes de softwares estatísticos que implementam estatísticas espaciais, mas sem contemplar componentes essenciais providos por um GIS como um banco de dados espacial robusto, modelos espaciais e algoritmos de visualização (Krivoruchko, 2003). Por outro lado, usuários de GIS encontram dificuldades para fazer análises mais sofisticadas pelo fato do software não disponibilizar ferramentas para análise estatística de forma satisfatória.

Há diferentes mecanismos de integração entre GIS e softwares destinados à análise estatística. Algumas dessas ferramentas de análise espacial são ligadas a um GIS por mecanismos de acoplamento fraco, onde o GIS é usado para conversão de dados e exibição gráfica, e os modelos espaciais são executados fora das estruturas do GIS. Um exemplo é a ligação

entre SpaceStat e ArcView (Anselin and Bao, 1997). Esta estrutura permite o uso de programas existentes, mas requer trabalho significativo na conversão de dados e apresenta problemas de redundância e consistência, devido à criação de versões múltiplas dos mesmos dados.

Outra alternativa de integração é o chamado acoplamento forte, que permite que o GIS e o software de estatística espacial compartilhem um banco de dados comum. Há ainda a integração completa, onde há a incorporação de funções de estatística espacial dentro de GIS, por exemplo. Algumas formas de integração entre softwares estatísticos e GIS são apresentadas e exemplificadas neste trabalho, ressaltando-se aspectos positivos e não-positivos das mesmas.

Este documento apresenta a estatística espacial no contexto da análise de dados espaciais e aborda a integração entre a estatística espacial e GIS sob uma ótica computacional, mostrando alguns tipos de conexão existentes. Apresenta também o porte da integração entre a TerraLib e o software estatístico R para a plataforma Windows. Isto amplia a oferta de funcionalidades propiciada por esta integração que inicialmente era restrita a usuários da plataforma Linux. A necessidade de integração entre a tecnologia GIS e softwares de naturezas diversas surge no cenário atual como item motivador da busca de novas alternativas de integração ou ampliação da demanda na aplicação dos mecanismos já existentes.

O capítulo 2 aborda a estatística espacial e os tipos de dados analisados. No capítulo 3 um levantamento sobre a integração da estatística espacial em ambientes GIS é apresentado. Um exemplo de integração é mostrado no capítulo 4 e algumas considerações referentes ao estudo são ressaltadas no final do trabalho.

2 ESTATÍSTICA ESPACIAL

A estatística espacial está inserida em um contexto mais amplo, a análise espacial de dados. A análise espacial objetiva mensurar propriedades e relacionamentos considerando a localização espacial do fenômeno em estudo, ou seja, permite estudar, explorar e modelar fenômenos geográficos (Druck et al., 2004; Goodchild and Haining, 2004). Segundo Druck et al. (2004), “A Análise Espacial é composta por um conjunto de procedimentos encadeados cuja finalidade é a escolha de um modelo inferencial que considera explicitamente os relacionamentos espaciais presentes no fenômeno”.

Um fenômeno espacial é um processo que pode ser expresso por meio de uma distribuição no espaço, podendo ser contínuo ou discreto. É considerado discreto quando se trata de entidades do mundo real como municípios, quadras, escolas, hospitais, dentre outros. Um fenômeno é contínuo quando se refere à uma informação presente em todas as posições, como temperatura, pressão e salinidade do solo, por exemplo.

Para a análise espacial conceitos como *dependência espacial* e *autocorrelação espacial* são fundamentais. Entende-se por dependência espacial o fato de que a maior parte das ocorrências naturais ou sociais apresentam entre si uma relação que depende da distância (Druck et al., 2004). A expressão computacional do conceito de dependência espacial é a autocorrelação espacial. A idéia é verificar como a dependência espacial varia, a partir da comparação entre os valores de uma amostra e de seus vizinhos.

A Figura 1 exhibe os campos da Análise Espacial segundo Fischer et al. (1996) e as áreas da Estatística Espacial identificadas por Bailey e Gatrell (1995).

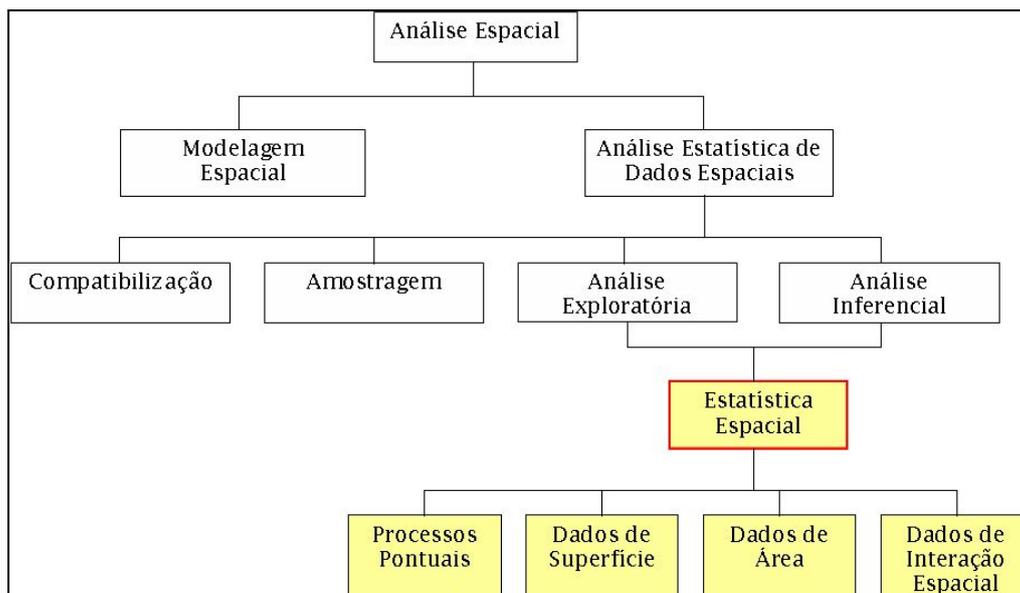


FIGURA 1 – Estatística Espacial inserida no contexto da Análise Espacial [Fonte: (Reis, 2005)]

A Figura 1 ilustra as duas principais áreas de estudo na análise espacial: modelagem espacial e análise estatística de dados espaciais (Fischer et al., 1996). A modelagem espacial se utiliza de modelos determinísticos e estocásticos para aplicações nas ciências ambientais e socioeconômicas. A análise estatística de dados espaciais engloba quatro sub-áreas que possuem estreita relação com a tecnologia GIS (Fischer et al., 1996):

1. Amostragem de objetos no banco de dados e escolha da escala de análise adequada;
2. Compatibilização de dados coletados, visto que há dados provenientes de fontes diversas, tais como dados censitários e imagens de satélite;
3. Análise exploratória de dados espaciais, definida como sendo uma “coleção de técnicas para descrever e visualizar distribuições espaciais, identificar situações atípicas (*outliers*), descobrir padrões de associação espacial, *clusters* e sugerir regimes espaciais ou formas de heterogeneidade espacial” (Anselin, 1998);
4. Análise inferencial de dados espaciais, que diz respeito à estimação de parâmetros de modelos e teste de hipóteses baseado em certas suposições;

As sub-áreas de análise exploratória e análise inferencial contemplam técnicas conhecidas como estatística espacial, onde a localização dos dados é relevante. A estatística espacial é subdividida em áreas de análise, de acordo com o tipo de dado tratado. Os tipos de dados compreendem dados pontuais ou eventos, superfícies ou distribuições, dados de área e dados de interação espacial¹ (Assunção, 2001; Bailey and Gatrell, 1995).

Há também a análise espaço-temporal que é uma área de crescente interesse por parte de pesquisadores da geoinformação e ainda envolve desafios nos âmbitos conceituais e de modelagem (Casanova et al., 2005; Hornsby and Egenhofer, 2000). A análise de cada tipo de dado tem particularidades próprias. A seguir serão abordadas as análises de pontos, superfícies e áreas, bem como a análise espaço-temporal.

2.1 Análise de Pontos ou Eventos

A análise de pontos (ou eventos) tem como objetivo estudar a distribuição espacial de fenômenos que são expressos através de ocorrências identificadas como pontos localizados no espaço, também chamados de processos pontuais (Assunção, 2001). A Figura 2 permite a visualização de exemplos de dados pontuais em uma determinada área de estudo.

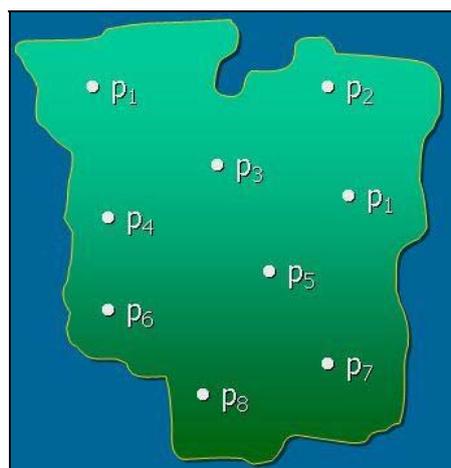


FIGURA 2 – Exemplo de dados pontuais [Fonte: (DPI/INPE, 2004a)]

¹ A análise de dados de interação consiste no estudo de fluxos a partir de um conjunto de dados de origem para um conjunto de destinos e, portanto, está relacionada a pares de pontos ou pares de áreas. Estes fluxos podem ser de pessoas ou bens de consumo.

Um padrão espacial de pontos é um conjunto de dados espaciais, compostos por coordenadas dos eventos e sua referência temporal. O objeto de interesse na análise deste tipo de dados é a própria localização espacial dos eventos em estudo. Pode-se testar hipóteses sobre o padrão observado, como por exemplo se o padrão é aleatório, se apresenta-se em aglomerados ou se os pontos estão regularmente distribuídos. Alguns dos padrões detectáveis neste tipo de análise são exibidos na Figura 3.



FIGURA 3 – Padrões de dados pontuais [Fonte:(DPI/INPE, 2004a)]

Em geral, os pontos não estão associados a valores, e sim à ocorrência dos eventos considerados. Entretanto, em alguns casos, os pontos podem estar associados a atributos de identificação. Outra característica é que a área dos eventos não é considerada como sendo uma medida válida nos dados de distribuições pontuais. A localização de crimes, localizações de núcleos celulares em uma parcela de tecido microscópico e a ocorrência de doenças são exemplos destes processos (Diniz, 2000).

2.2 Análise de Superfícies ou Distribuições

Nesta análise espacial o tipo de dado em estudo é a superfície. Esta é gerada a partir de um processo de interpolação² dos dados pontuais presentes na área de estudo. O objetivo desta análise é reconstruir a superfície

² Processo em que se determina o valor de uma função num ponto interno dum intervalo a partir dos valores da função nas fronteiras desse intervalo

na qual as amostras foram retiradas. A análise interessa-se agora em padrões nos valores dos atributos, e não mais nos padrões das localizações das observações, como na análise de dados pontuais (Bailey and Gatrell, 1995).

As superfícies contínuas podem ser estimadas a partir de um conjunto de amostras de campo, que podem estar regularmente ou irregularmente distribuídas. As amostras são valores representativos do fenômeno estudado e podem representar variáveis naturais (teor de argila no solo) e socio-econômicas (taxa de homicídios) (Druck et al., 2004). A análise espacial de superfícies permite que dados disponíveis sob a forma pontual sejam interpolados gerando uma superfície, na forma de grade regular, que representa o padrão de distribuição da variável estudada. Este tipo de dado espacialmente contínuo é geralmente referenciado como dado geoestatístico. Exemplos típicos de dados de superfície são medidas geológicas de depósitos minerais, salinidade e permeabilidade do solo, a concentração de algum poluente e medição de chuva, dentre outros (Bailey and Gatrell, 1995).

Os modelos trabalhados nesta análise são os modelos determinísticos de efeitos locais, que estimam cada ponto da superfície a partir da interpolação das amostras mais próximas; os modelos determinísticos de efeitos globais, nos quais se trabalha com uma variação em larga escala e desconsiderando-se a variabilidade local; e os modelos estatísticos de efeitos locais e globais (Geoestatística), onde cada ponto da superfície é estimado a partir da interpolação das amostras mais próximas, utilizando um estimador estatístico (Druck et al., 2004). Os métodos geoestatísticos provêm grande flexibilidade para interpolação e possuem como hipótese central o conceito de estacionariedade, que supõe um comportamento homogêneo da estrutura de correlação espacial na região em estudo (Burrough and McDonnel, 1998). Para Câmara et al. "a noção de processo estacionário indica uma área de trabalho na qual as relações entre as medidas são função da distância entre elas" (Câmara et al., 2003a). A Figura 4 exibe um dado de superfície.

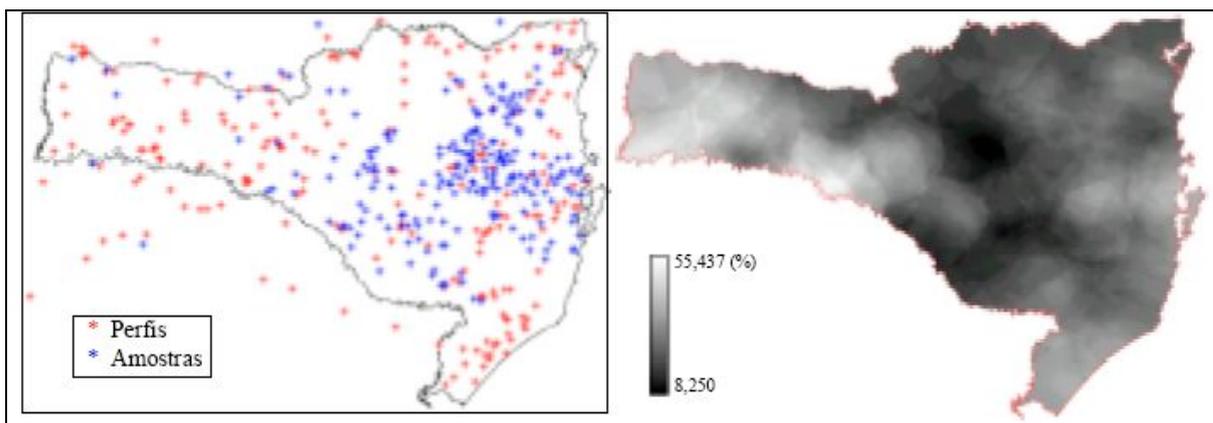


FIGURA 4 – Exemplo de dados de superfície [Fonte: (Druck et al., 2004)]

Na Figura 4 tem-se um exemplo com a distribuição de perfis e amostras de solo em Santa Catarina, Brasil (esquerda) e distribuição contínua estimada para a variável saturação por bases (direita). As cruces indicam a localização dos pontos de coleta de amostras no solo. A partir destas foi estimado um modelo de dependência espacial que permitiu a interpolação da superfície do mapa da figura. O modelo inferencial objetiva quantificar a dependência espacial entre os valores das amostras utilizando para isso técnicas de geoestatística.

2.3 Análise de Áreas

A análise de áreas é aplicada quando não se dispõe da localização exata dos eventos, mas de um valor por área, isto é, lida com dados associados a levantamentos populacionais como censos e estatísticas de saúde. Estes dados se referem a indivíduos localizados em pontos específicos do espaço e são, em geral, agregados em unidades de análise delimitadas por polígonos fechados. Esta análise usa atributos que não variam continuamente, mas que possuem valores específicos para subáreas que compõem uma dada região em estudo. O interesse está na detecção e possíveis explicações para o padrão espacial ou tendência de distribuição para valores de área (Diniz, 2000).

Como exemplo, veja-se na Figura 5 a distribuição espacial do índice de exclusão/inclusão social de São Paulo. O trabalho foi produzido pela equipe

liderada pela Prof^a Aldáiza Sposati (PUC/SP) (Sposati *apud* (Druck et al., 2004)). Os indicadores de exclusão/inclusão social foram obtidos a partir dos dados coletados nos 96 distritos de São Paulo, com base no censo de 1991. A Figura 5 exibe o tipo de dado de área (direita) e os agrupamentos de exclusão e inclusão social (direita). Visualiza-se agrupamentos de exclusão social nas Zonas Leste e Sul, e de inclusão social na área central de São Paulo.

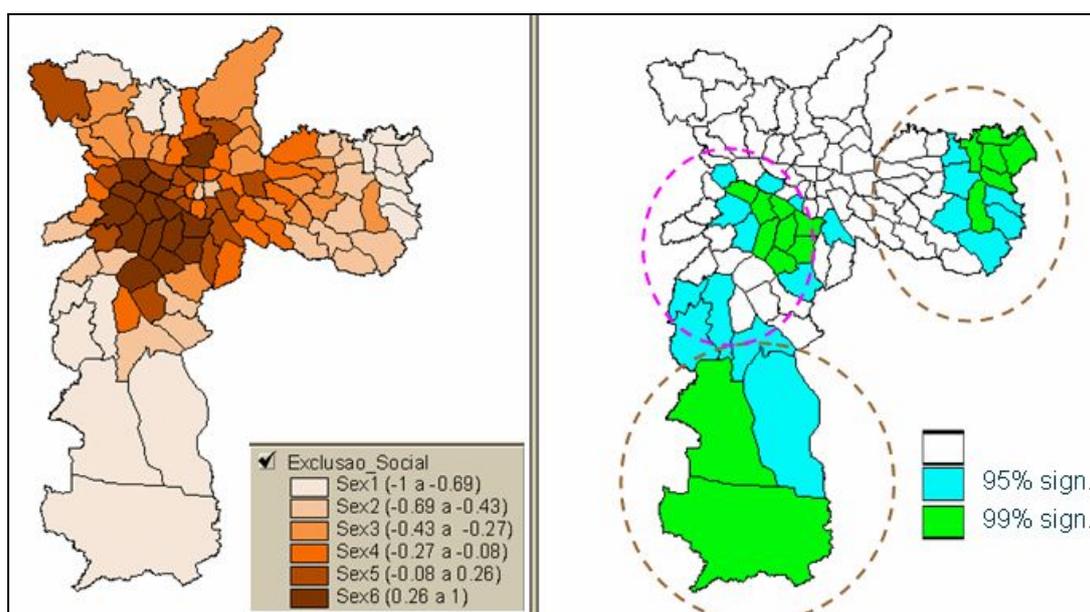


FIGURA 5 – Exemplo de dados de área [Fonte: (Druck et al., 2004)]

Os tipos de análise de área incluem:

- Análise exploratória, com produção de indicadores de autocorrelação espacial. Os métodos exploratórios envolvem a procura de boas descrições dos dados, a fim de ajudar ao analista a desenvolver algumas hipóteses sobre o assunto e modelos apropriados para tais dados (Bailey and Gatrell, 1995).
- Modelos de regressão espacial, que buscam estabelecer relações entre duas ou mais variáveis quantitativas, de modo que uma variável possa ser prevista a partir de outra, no caso de modelos univariados, ou a partir de outras, no caso de modelos multivariados (Neter e Wasserman *apud* (Almeida, 2003)). O GWR – *Geographically*

Weighted Regression é um exemplo de técnica de análise exploratória que busca ajustar um modelo de regressão a cada ponto observado, ponderando todas as demais observações como função da distância a este ponto (Druck et al., 2004; Fotheringham et al., 2002).

- Modelos inferenciais, que buscam estabelecer a distribuição estatística subjacente aos dados coletados. Este modelos incluem modelos bayesianos empíricos e completos (Câmara et al., 2003a). Na Modelagem Bayesiana Empírica estima-se os valores a partir dos dados, ou seja, um conhecimento prévio ou crença sobre os parâmetros de interesse e/ou dados observados são levados em conta durante estimação de seus valores. Na Modelagem Bayesiana Completa a distribuição *a posteriori* do parâmetro é obtida por um processo iterativo de simulação (Bailey and Gatrell, 1995; DPI/INPE, 2004c).

2.4 Análise Espaço-Temporal

A análise espaço-temporal lida com modelos preditivos que possuem representação espaço-temporal, onde a evolução do fenômeno é expressa através de representação funcional, como uso e ocupação da terra, cadastro urbano e poluição. O desenvolvimento de sistemas de bancos de dados espaço-temporais e GIS temporais envolve desafios nos âmbitos conceitual e de modelagem, uma vez que a representação do tempo se mostrou mais complexa do que a simples adição de uma terceira (ou quarta) dimensão nas representações espaciais (Burrough, 1998; Casanova et al., 2005; Goodchild and Haining, 2004; Hornsby and Egenhofer, 2000). A comunidade da geoinformação empreende esforços nesta área. Goodchild (2004) ressalta que "Peuquet (2002) provê uma análise do suporte filosófico e conceitual do problema, Langran (1992) apresenta uma perspectiva mais prática, enquanto Frank et al. (2001) endereça as questões de informar limites de zona que se deslocam através do tempo".

A Figura 6 ilustra exemplo do cadastro imobiliário de um centro urbano, onde determinada área se apresenta como na Figura 6.a em um instante t e, após um intervalo de tempo (instante $t+n$) a mesma área se apresenta como na Figura 6.b.

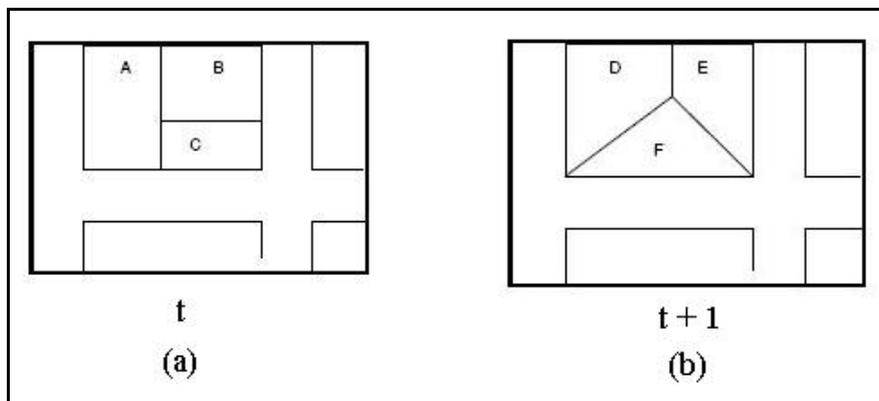


FIGURA 6 – Exemplo de cadastro urbano [Fonte: (Casanova et al., 2005)]

As técnicas de análise para fenômenos espaço-temporais incluem modelos multi-escala, que estabelecem conexões entre fenômenos de macroescala (tipicamente relacionados com fatores econômicos) e fenômenos de microescala (tipicamente associados a transições no uso da terra) para capturar as diferentes relações dinâmicas (Câmara et al., 2003b).

Este tipo de análise pode ser aplicada à avaliação de padrões que variam no tempo. Por exemplo, avaliar o padrão de procriação de pássaros, que varia de acordo com alterações ambientais pode ser primordial para a conservação de espécies; avaliar o padrão de doenças, que também se altera com o tempo, pode ser útil no julgamento da importância relativa de suas diferentes causas; e examinar o impacto do grau de dinâmica temporal no comportamento de um modelo de crescimento urbano é atividade fundamental para entidades de gestão pública (Goldstein et al., 2004; Liu and Andersson, 2004; Unwin, 1996). Assunção et al. propõem um método estatístico para testar a presença de agrupamentos espaço-temporais de dados pontuais. Sistemas de advertência para detectar o aparecimento de agrupamentos espaciais em fase inicial seriam úteis em áreas de vigilância. O método pode ser aplicado para análise de focos de fogo em áreas florestais

como na Amazônia e crimes que acontecem em centros urbanos (Assunção et al., 2003). O aplicativo STIS – *Space-Time Intelligence System (TerraSeer)* (AvRuskin et al., 2004; Jacquez et al., 2005) e a biblioteca de classes para GIS TerraLib (INPE) (Câmara et al., 2000) são exemplos de GIS que manipulam o componente tempo em seus *datasets*.

A tabela seguinte sumariza os tipos de análise de dados apresentados acima e suas principais características.

TABELA 1–Tipos de Dados na Estatística Espacial [Fonte: (Druck et al., 2004)]

Análises	Tipos de Dados	Objetivos	Problemas	Exemplos
Padrões Pontuais ou Eventos	Eventos localizados	Localização espacial dos eventos em estudo	Determinação de Padrões	Ocorrência de doenças
Superfícies ou Distribuições	Amostras de campo e Matrizes	Reconstrução da superfície na qual as amostras foram retiradas	Interpolação e Medidas de Incerteza	Depósitos Minerais
Áreas	Polígonos e Atributos	Casos onde não se dispõe da localização exata dos eventos, mas de um valor por área	Regressão e Distribuições Conjuntas	Dados Censitários
Espaço-Temporal	Representação funcional	Estabelecimento de conexões entre fenômenos de macroescala e microescala	Avaliação de padrões que variam com o tempo	Cadastro Urbano, Padrões de doenças

O capítulo 3 apresenta aspectos da integração entre softwares utilizados na estatística espacial e sistemas de informações geográficas.

3 ESTATÍSTICA ESPACIAL E GIS

A integração entre Estatística Espacial e GIS propicia o fortalecimento das áreas envolvidas e aumenta a eficácia individual e comum de ambas. A tecnologia GIS oferece um pronto acesso a rotinas complexas e sofisticadas da estatística espacial para uma grande comunidade de usuários (Goodchild and Haining, 2004; Marble, 2000). Segundo Goodchild, falta muito para o processo de interligação entre estas tecnologias estar completo (Goodchild, 2000). Entretanto, observa-se considerável progresso através de propostas de diferentes tipos de estratégias de integração entre as áreas. Pode-se destacar a integração completa (*full integration*) e os mecanismos acoplamento fraco (*loose coupling*) e acoplamento forte (*close coupling*, *seamless coupling* ou *strong coupling*) (Anselin, 1998; Bailey and Gatrell, 1995; Bao et al., 2000; Fischer et al., 1996). A ligação pode ainda ocorrer através da combinação dos tipos acoplamento fraco e acoplamento forte.

Na integração completa as ferramentas de análise estatística são incorporadas diretamente no software GIS, sendo esta sua principal vantagem. Esta estratégia tem sido utilizada em ferramentas como as de geoestatística (semivariograma e os diversos tipos de krigeagem) e índices de autocorrelação espacial. Algumas ferramentas foram incorporadas em sistemas não-comerciais/acadêmicos como por exemplo a geoestatística, a função-K e o mapa de densidade de kernel no GIS SPRING/versão 4.1.1, software livre desenvolvido pelo INPE – Instituto Nacional de Pesquisas Espaciais (Câmara et al., 1996). Neste caso, as funções do módulo geoestatístico são baseadas na biblioteca de software GSLIB – *Geostatistical Software Library* (Felgueiras et al., 1999). Outro exemplo é a ferramenta para a análise de conglomerados espaciais, estimador Bayesiano empírico global e local inseridos no software livre e código aberto TerraView/versão 3.0.3, também desenvolvido pelo INPE.

No entanto, para técnicas de análise de dados espaciais mais sofisticadas, este tipo de integração não parece ser a mais promissora. A

inclusão de grande número de funções de estatística espacial num GIS pode aumentar a complexidade do sistema para usuários comuns. Além disto, envolve um grande investimento de programação e manutenção de software, duplicando funcionalidades já disponíveis em outros sistemas.

As alternativas à integração completa são os acoplamentos fraco e forte. Nestas abordagens, os softwares de GIS e de análise estatística funcionam separadamente e compartilham os dados a serem analisados. Estas estratégias permitem que o software de análise de dados espaciais evolua independentemente do software GIS, incorporando técnicas que não seriam, a princípio, vantajosas de ser integradas completamente ao GIS do ponto de vista comercial.

Na integração através do acoplamento forte há uma ligação maior entre os softwares GIS e de análise estatística. Isto ocorre porque o mecanismo permite que rotinas do software estatístico sejam chamadas de dentro do GIS, ou vice-versa, ou ainda que o GIS e o software de análise compartilhem um banco de dados comum sem perda de estruturas mais altas como topologia, identidade do objeto, metadados e outros relacionamentos (Fischer et al., 1996). Esta ligação apresenta a vantagem de que, em geral, a análise pode ser feita no ambiente do GIS, a partir da chamada de uma rotina do software. O acoplamento forte é exibido na Figura 7.

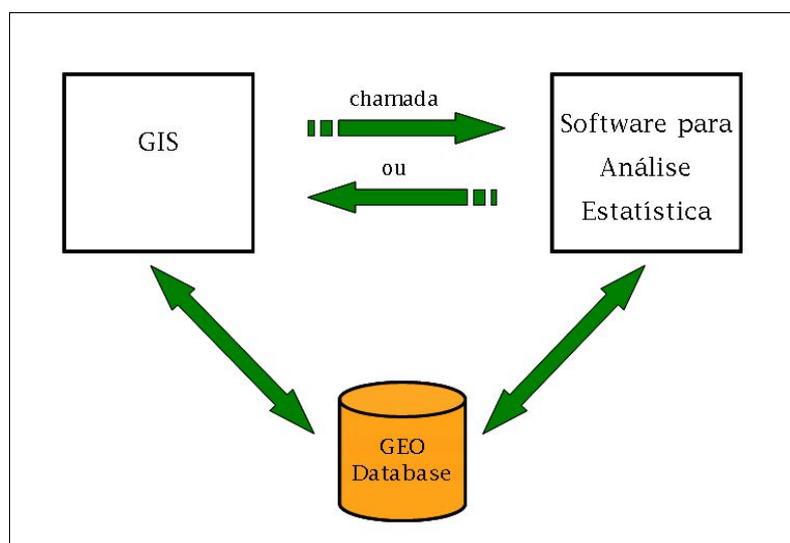


FIGURA 7 – Estratégia acoplamento forte [Adaptado de: (Bivand and Neteler, 2000)]

Alguns exemplos deste tipo de estratégia de acoplamento são:

- Integração entre os softwares gratuitos e código aberto GRASS (*Geographical Resources Analysis Support System*) e R. O ambiente de programação estatística do R é carregado dentro do GRASS através do pacote R denominado GRASS (Bivand, 2000; Bivand, 2001; Bivand and Neteler, 2000);
- A ligação entre o software para análise de dados S-PLUS (MathSoft) e do GIS desktop ArcView (ESRI), ambos proprietários, que ocorre por meio de uma ligação bi-direcional. O S-PLUS é implementado como uma extensão ao software ArcView. A interface do usuário é acrescida com 2 novos menus e a interação do usuário é levada para o meio externo através de um conjunto de diálogos. A interface gráfica “esconde” toda a complexidade dos mecanismos de acoplamento (Bao et al., 2000);
- Integração entre ArcView e o software estatístico proprietário SpaceStat. Neste caso, a extensão SpaceStat para ArcView acrescenta dois menus à interface do software ArcView, permitindo a exportação dos dados do ArcView para o SpaceStat e a visualização dos resultados das análises do SpaceStat no ArcView (Anselin and Bao 1997);
- A Integração entre o GIS proprietário IDRISI e o pacote geoestatístico Gstat (livre e código aberto) é semelhante à integração entre ArcView e SpaceStat. Aqui parte das funcionalidades do Gstat são disponibilizadas na interface do IDRISI através de uma versão modificada do Gstat (Pebesma, 2004);
- A ligação entre a ferramenta WinBUGS e o R. O WinBUGS é um software livre para análise Bayesiana de dados e versão para Windows do programa BUGS (*Bayesian Inference Using Gibbs Sampler*). A integração ocorre através do pacote R2WinBUGS,

permitindo que as simulações do WinBUGS sejam armazenadas diretamente num objeto R (Sturtz et al., 2005) (Lunn et. al *apud* (Reis, 2005)).

- A integração entre a TerraLib e o R, que acontece através do pacote aRT (Neto et al., 2005).

O mecanismo de acoplamento fraco é outra forma de integração entre GIS e softwares de estatística espacial. Neste mecanismo, os dados são exportados do GIS para dentro de um pacote de análise estatística de dados espaciais, e os resultados da análise são novamente retornados para o GIS, também por meio da exportação, para que sejam visualizados. Este tipo de integração pode ser esquematizado na Figura 8.

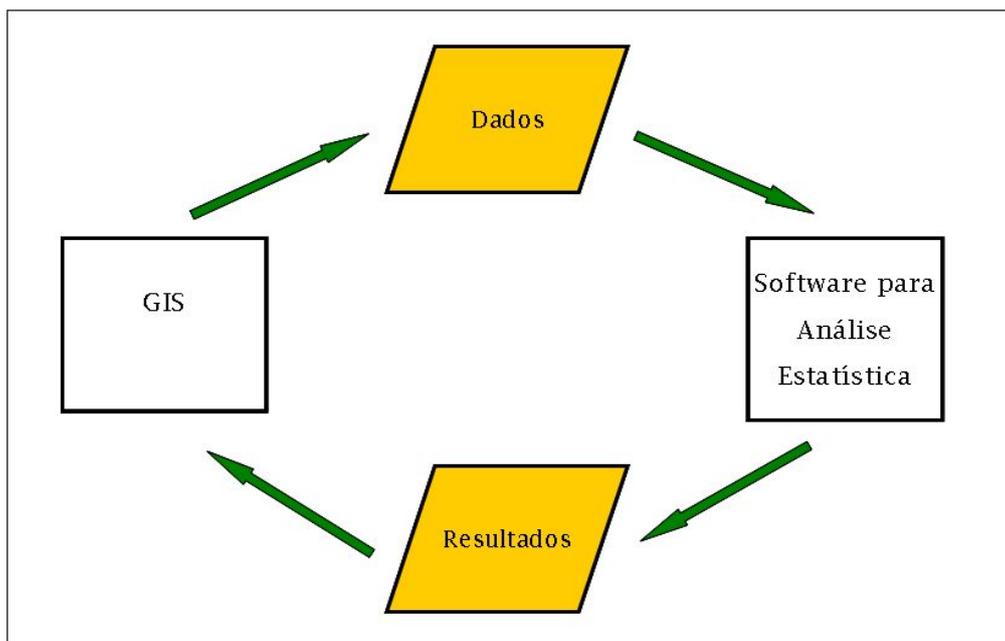


FIGURA 8 – Estratégia de acoplamento fraco [Adaptado de: (Bivand and Neteler, 2000)]

A estratégia de acoplamento fraco apresenta-se computacionalmente simples de ser implementada. Basta que os programas trabalhem com um formato de dados comum a ambos os softwares. Contudo, pode haver *overhead* nos processos de exportação e importação dos dados e/ou resultados quando muitos conjuntos de dados forem necessários para as análises estatísticas. Alguns exemplos da aplicação desta estratégia são:

- Integração do GIS SPRING com o programa SpaceStat. O SPRING exporta dados no formato vetorial para o formato SpaceStat (Anselin, 1992; Câmara et al., 1996; DPI/INPE, 2004b);
- Integração entre o GIS proprietário ArcInfo e o R. O pacote RarInfo do R permite a importação de arquivos no formato (padrão) ArcInfo *binary coverages* (Rubio, 2003);

A Tabela 2 sintetiza as estratégias de integração abordadas acima. As subseções seguintes abordam alguns dos exemplos acima elencados de forma mais detalhada.

TABELA 2 – Estratégias de Integração entre GIS e Estatística Espacial

Estratégia	Descrição	Vantagens	Problemas	Exemplos
Integração completa	GIS contempla funções de Estatística Espacial ou vice-versa	As ferramentas de análise estatística são incorporadas no GIS	<i>Overhead</i> no GIS devido à incorporação de ferramentas de análise estatística	Módulo de Geoestatística do SPRING
Acoplamento forte	GIS e Estatística Espacial acessam o mesmo banco de dados	Análise realizada no ambiente do GIS, através da chamada de uma rotina	Maior complexidade na implementação.	GRASS + R R+TerraLib
Acoplamento fraco	GIS e Estatística Espacial trocam dados através do processo de exportação	computacionalmente simples de ser implementada	<i>Overhead</i> na exportação/importação dos dados e/ou resultados	SPRING + SpaceStat

➤ Ligação GRASS e R

A integração entre o GRASS e o ambiente estatístico R envolve atualmente o acoplamento forte, executando o R de dentro do ambiente

GRASS. O R herda as configurações de todas as variáveis de ambiente necessárias pelos programas GRASS, que são executados de dentro do R através do comando `system()`. Isto inclui os programas que requerem interação de usuário. O carregamento do R no GRASS é realizado através do pacote GRASS (Bivand, 2000; Bivand and Lucas, 2000; Bivand and Neteler, 2000). A Figura 9 mostra as rotas que a transferência de dados podem ter. A versão corrente suporta a transferência binária direta e a transferência através de arquivos ASCII temporários.

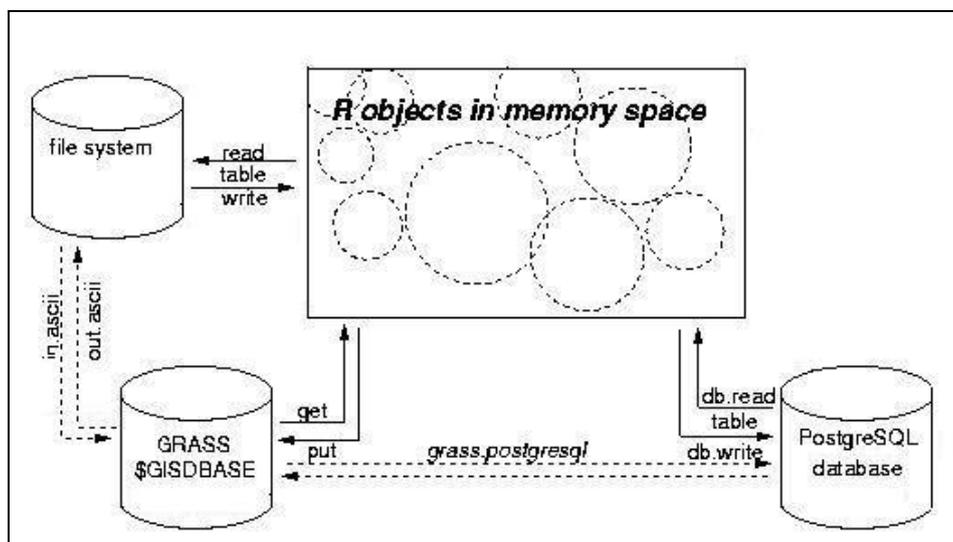


FIGURA 9 – Possíveis Rotas de Transferência entre R/GRASS/PostgreSQL [Adaptado de: (Bivand, 2001; Bivand and Neteler, 2000)]

A abordagem que utiliza a interface RPgSQL R-PostgreSQL com a interface de GRASSPostgreSQL será retomada quando o formato vector do GRASS se estabilizar (Bivand, 2001; Bivand and Neteler, 2000). Uma interface mais robusta entre GRASS (versão 6) e R está sendo elaborada por Bivand et al. Esta utiliza o pacote *spgrass6* e depende no momento de dois pacotes (*sp*, *rgdal*) que devem ser primeiramente instalados (Bivand, 2005).

➤ Ligação Spring e SpaceStat

A integração entre o Spring e o SpaceStat ocorre através do mecanismo de acoplamento fraco. O Spring é um GIS que contempla funções

de processamento de imagens, análise espacial, modelagem numérica de terreno e consulta a bancos de dados geográficos (Câmara et al., 1996). O SpaceStat é uma ferramenta utilizada para: Construção e manipulação de matrizes de pesos espaciais; Análise exploratória de dados espaciais (ESDA); e Análise econométrica espacial, que contempla testes de especificação e métodos de estimação para modelos de regressão que incorporam dependência e heterogeneidade espacial (Anselin, 1992).

O Spring exporta dados no formato vetorial para o formato SpaceStat, pois esse software não possui módulo gráfico e o Spring é utilizado para visualização. Neste processo de integração, exporta-se a tabela de atributos e a matriz de vizinhança, a partir do Spring, para o SpaceStat. As análises são realizadas no SpaceStat e as tabelas, atualizadas, são exportadas novamente para o Spring para que os novos atributos possam ser visualizados. A Figura 10 exibe o esquema do processo de exportação entre os softwares.

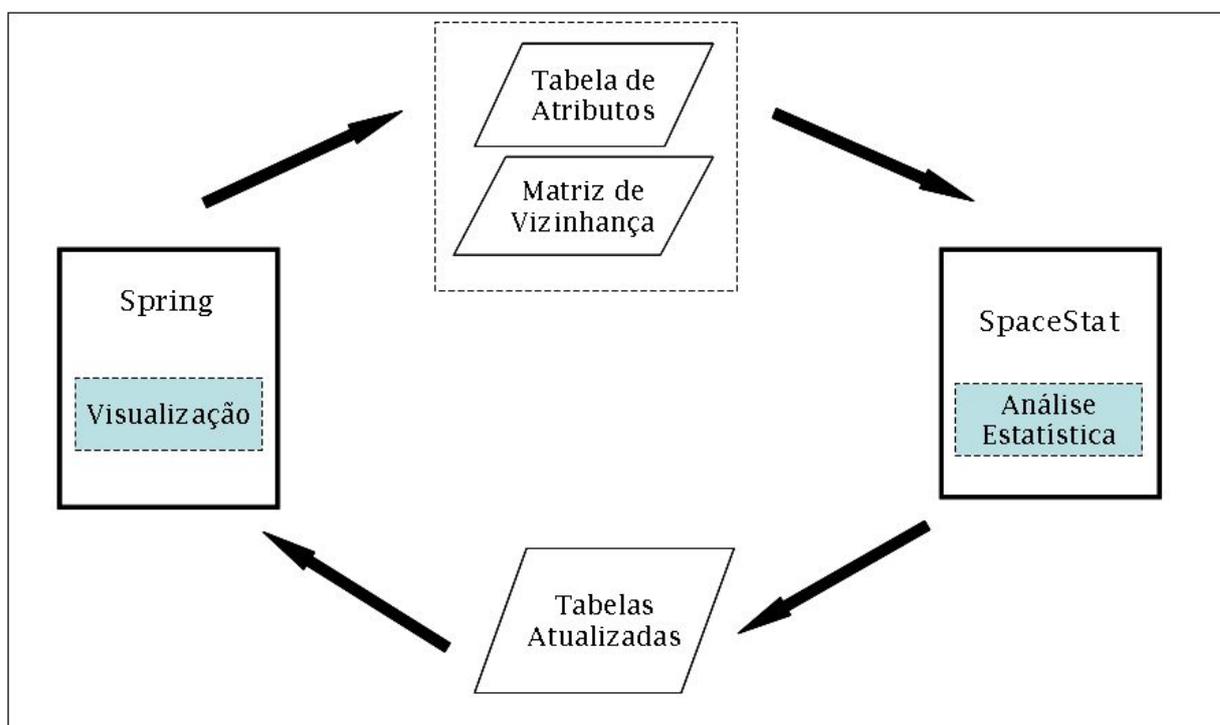


FIGURA 10 – Exemplo de integração: Spring e SpaceStat

➤ Ligação TerraLib e Open Modeller

As estratégias de integração também podem ser aplicadas a GIS e softwares de outras naturezas, como os de modelagem espacial. Um exemplo de conexão via acoplamento forte neste caso é a proposta do projeto TerraModeller, que objetiva combinar as funcionalidades da TerraLib com o ambiente OpenModeller. O OpenModeller é um ambiente computacional para o desenvolvimento de modelos e cenários. Um dos produtos do projeto speciesLink, é uma ferramenta para modelagem de nichos ecológicos, usada para prever distribuição de espécies (CRIA and FAPESP, 2005) (CRIA et al., 2005). Desenvolvida pelo CRIA (Centro de Referência em Informação Ambiental) possui código aberto e usa GDAL, uma biblioteca de tradução para formatos de dados geoespaciais raster, para ler vários formatos de arquivo de mapa. O TerraModeller se propõe a integrar ferramentas de Modelagem de Nicho Ecológico e GIS para assegurar o acesso a grandes volumes de dados ambientais (clima, vegetação, topografia, solos e uso da terra) necessários à arquitetura de modelagem, além de disponibilizar modelos de distribuição de espécies diferentes em um ambiente unificado (CRIA et al., 2005). A TerraLib é uma biblioteca de classes de distribuição gratuita e código aberto utilizada para o desenvolvimento de GIS (Câmara et al., 2000).

O projeto TerraModeller encontra-se em fase inicial, sendo financiado pela FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) e implementado pelas Instituições CRIA, INPE e Escola Politécnica da USP (Universidade de São Paulo). O período previsto para o projeto é de 4 anos (2005 a 2008).

4 Integração entre R e TerraLib

Neste capítulo, apresenta-se em maior detalhe a integração entre o pacote estatístico R e a biblioteca de classes TerraLib. O mecanismo de *acoplamento forte* realiza a integração e os resultados podem ser visualizados através do TerraView. Esta descrição objetiva identificar em maior detalhe os procedimentos de integração entre estatística espacial e GIS.

O R consiste de uma linguagem e de um ambiente que possui uma ampla variedade de técnicas gráficas e estatísticas, sendo um veículo para métodos interativos de análise de dados. É uma implementação gratuita da linguagem S, que forma a base do sistema comercial S-PLUS. Provê funções escritas por usuários para carregamento dinâmico do código compilado, sendo portanto extensível (Bivand, 2003; Grunsky, 2002; Venables et al., 2001). Muitas técnicas estatísticas clássicas e modernas foram implementadas dentro do ambiente R. Algumas destas técnicas foram construídas no ambiente básico do R e muitas foram providas por meio de pacotes como *gstat* (Pebesma, 2004; Pebesma and Wesseling, 1998), *geoR*, *geoRglm* (Ribeiro-Jr. et al., 2003) e *R2WinBUGS* (Sturtz et al., 2005). Há aproximadamente 25 pacotes fornecidos com o R (chamados “padrão” e “recomendados”) e muitos outros estão disponíveis através das famílias CRAN³ (*Comprehensive R Archive Network*) em sítios da Internet. O R administra toda a computação em memória, apresentando uma dificuldade potencial para análise de *datasets* muito grandes. Um modo para superar isto é utilizar um mecanismo de acesso a um sistema de banco de dados externo, como PostgreSQL ou MySQL (Bivand and Neteler, 2000).

O TerraView é um software livre desenvolvido pela equipe da DPI / INPE e construído sobre a TerraLib. Oferece à comunidade um fácil visualizador de dados geográficos com recursos de consulta a análise dos mesmos. Manipula dados vetoriais (pontos, linhas e polígonos) e matriciais

³ Rede de ftp e web servers que armazena versões de código e documentação R idênticas e atualizadas.

(grades e imagens), ambos armazenados em SGBD objeto-relacionais de mercado, incluindo ACCESS, PostgreSQL, MySQL e Oracle (Câmara et al., 2000; DPI/INPE, 2002). A integração do R com o GIS TerraView ocorre por meio da API (*Application Programming Interface*) aRT descrita a seguir.

4.1 aRT: API R-TerraLib

A API aRT provê a integração através do acoplamento forte entre os softwares R e TerraLib. Trata-se de um pacote do R que foi desenvolvido pela Universidade Federal do Paraná (UFPR). Ao efetivar o intercâmbio entre esses softwares, o aRT propicia ao analista de dados no R, os recursos de acesso aos Bancos de Dados Geográficos e às operações implementadas na TerraLib, além de incorporar os resultados gerados no R ao Banco de Bados de forma direta e transparente ao usuário. O analista pode executar um pacote espacial como *spdep*, *gstat* ou *geoR*, e inserir os resultados no banco de dados. Esses resultados podem então ser visualizados pelo TerraView (Neto et al., 2005). A ligação R/TerraLib/TerraView é exibida na Figura 11.

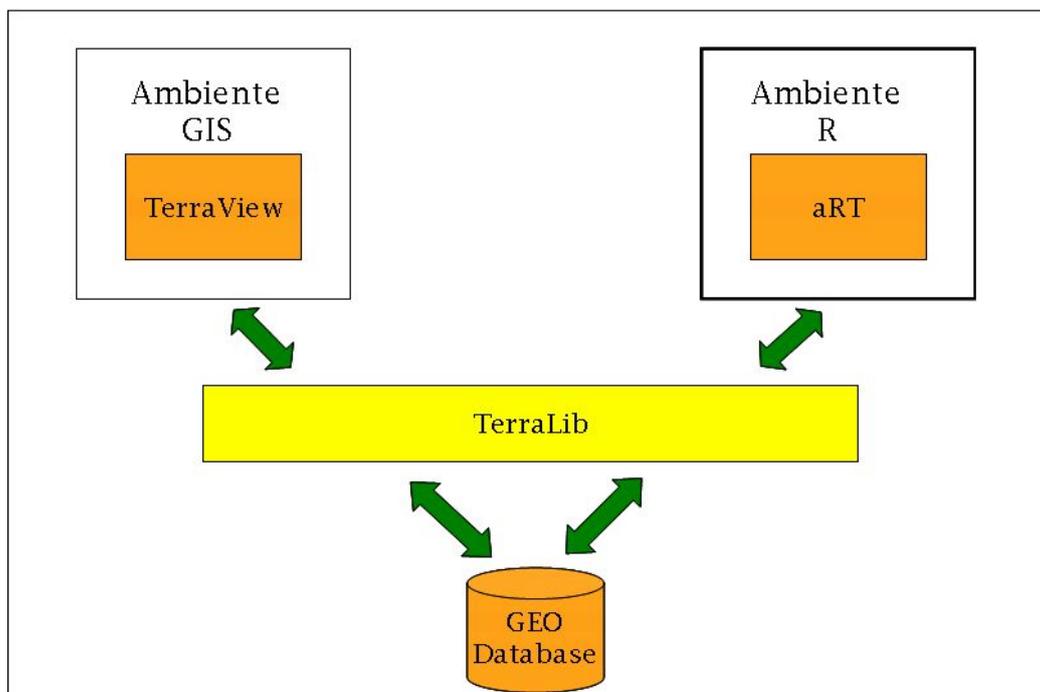


FIGURA 11 – Ligação R/TerraLib/TerraView

O aRT pode ser instalado no ambiente do R e carregado sempre que necessário. Uma vez carregado, tem-se acesso ao banco de dados e ferramentas TerraLib, objetos das classes da biblioteca podem ser criados e manipulados. Os resultados da análise estatística podem ser incorporados ao banco TerraLib. Inicialmente desenvolvido para o Linux, o esquema de funcionamento do aRT na referida plataforma pode ser visualizado na Figura 12.

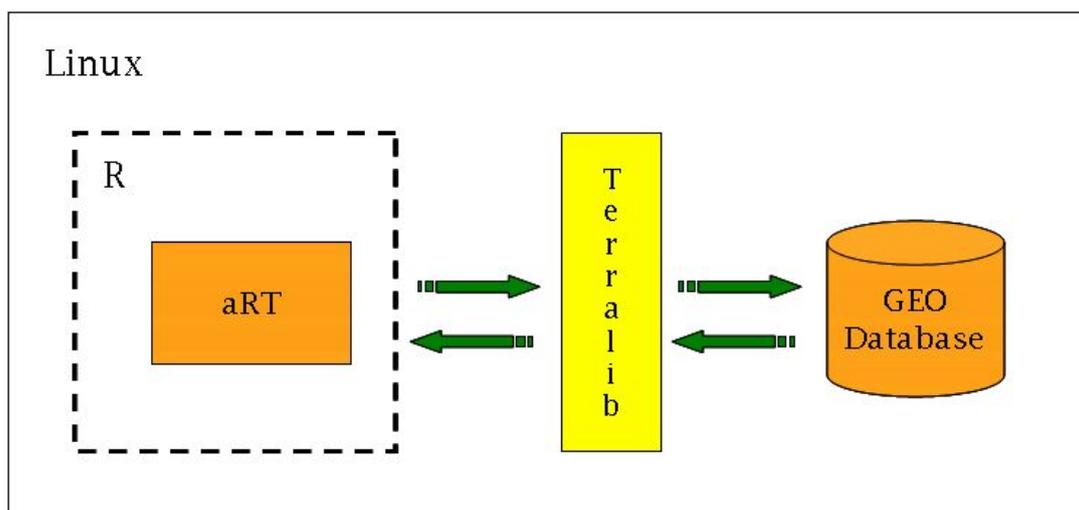


FIGURA 12 – Esquema aRT, plataforma Linux

O pacote aRT é constituído de 4 classes para manipular os dados e funções da TerraLib, sendo elas: aRT, aRTdb, aRTlayer e aRTtheme. Objetos da classe aRT permitem conexão com DBMS para a realização de funções de administração de banco de dados. Uma vez criado, o objeto armazena uma conexão virtual, ou seja, todas as vezes que o acesso ao DB é requerido, o objeto se conecta, executa alguma tarefa e então se desconecta (Neto et al., 2005).

Um objeto da classe aRTdb provê a criação e o acesso a um novo banco de dados e necessita que um objeto aRT seja criado (a conexão criada por esse objeto é passada como argumento para o construtor da classe). Os objetos de aRTdb mantêm todos os objetos de TerraLib necessários pelo aRT em memória. Se um objeto é removido do R, todos os seus filhos se tornam

objetos inválidos quando o coletor de lixo do R remove este objeto da memória (Neto et al., 2005).

A classe `aRTlayer` permite a manipulação de *layers*⁴ em aRT. Um *layer* pode armazenar qualquer geometria de um tipo (pontos, polígonos ou raster, linhas e células), e atributos. Os dados devem ser convertidos para o formato aRT antes de sua inserção no banco de dados. Existem algumas funções que realizam esta conversão de dados de outros pacotes espaciais, como *splancs* e *geoR* (Neto et al., 2005). A classe `aRTtheme` é voltada a usuários do TerraView e propicia a visualização de *themes*⁵ no referido GIS (Neto et al., 2005). Através das classes descritas, o pacote aRT contempla as seguintes operações básicas:

1. Criar novo banco de dados (a partir do R)
2. Criar novo *Layer*
3. Popular *Layer* com informações de um dado no R
4. Extrair informações geográficas
 - a. Geometrias
 - b. Atributos
 - c. Geometrias e Atributos
5. Alimentar o banco com resultado de operação R
6. Adicionar novo atributo

Alguns exemplos do aRT no R serão mostrados a seguir para um melhor entendimento do funcionamento do pacote. A princípio, o aRT deve ser carregado no ambiente R. O comando `require()` habilita o pacote para uso.

⁴ São abstrações da TerraLib que usam tabelas de dados e tabelas de controle em um banco de dados. Agregam informações espaciais que são localizadas sobre a mesma região geográfica e compartilham o mesmo conjunto de atributos, tais como mapas temáticos, mapas cadastrais de objetos geográficos e dados matriciais.

⁵ Representa um subconjunto dos objetos de um *layer*. Contém todos os objetos do layer ou somente aqueles selecionados segundo algum tipo de restrição.

```

> require (aRT)
Loading required package: aRT

-----
API R-TERRALIB
aRT version 0.2-6 (2005-04-14) is now loaded
-----

[1] TRUE

```

FIGURA 13 – Carregamento do pacote aRT no ambiente R

Uma vez habilitado o pacote, seus métodos podem ser chamados da linha de comando do R (*prompt*). O comando para conexão com o banco de dados é visualizado na Figura 14.

```

> con <- new("aRT", user = "root", host = "localhost", pass =
"")

> print (con)
Object of class aRT

DBMS = "MySQL"
User = "root"
Password = ""
Port = 3306
Host = "localhost"

> showDbs(con)
[1] "mysql" "parana" "test"

```

FIGURA 14 – Conexão com banco de dados

A criação do banco de dados é realizada através do comando exibido na Figura 15. O método `print()` permite que o usuário observe os objetos criados.

```

> db <- new("aRTdb", con, db = "parana", create = TRUE)
Creating a new database

> print (db)
Object of class aRTdb

Database = "parana"
This object has 0 children

```

FIGURA 15 – Criação do objeto da classe aRTdb

Os comandos do aRT que permitem a criação de um *layer* e de um *theme* no banco de dados TerraLib são exibidos nas Figuras 16 e 17, respectivamente.

```
> layer.points <- new("aRTlayer", db, layer = "points", create =
TRUE)
Creating a new layer

> print (layer.points)
Object of class aRTlayer

Layer = "points"
Number of polygons = 0
Number of lines = 0
Number of points = 0
Layer does not have raster data
Projection Name = "NoProjection"
Projection Datum = "Spherical"
Projection Longitude = 0
Projection Latitude = 0
```

FIGURA 16 – Criação de um layer no banco de dados

```
> theme.points <- new("aRTtheme", layer.points, th = "points", cr
= T, v = "view")
Creating a new theme
Creating a view
inserting...
reloading view set...
View created successfully
Saving...
Building collection...
Saved successfully
>
```

FIGURA 17 – Criação de um objeto da classe aRTtheme

Os comandos da Figura 18 executam um algoritmo para análise espacial, a krigagem, sobre as amostras do banco de dados “parana”. Para isto, os pacotes *geoR* e *splancs* são carregados.

```

> require(geoR)

> ml <- likfit(parana, trend="1st", ini=c(1000, 100))
-----
likfit: likelihood maximisation using the function optim.
likfit: Use control() to pass additional
        arguments for the maximisation function.
        For further details see documentation for optim.
likfit: It is highly advisable to run this function several
        times with different initial values for the parameters.
likfit: WARNING: This step can be time demanding!
-----

likfit: end of numerical maximisation.

> gx <- seq(50,800, by=10)
> gy <- seq(-100,650, by=10)
> loc0 <- expand.grid(gx,gy)
> points(parana, bor=borders)
> points(loc0, pch=".", col=2)
> loc1 <- polygrid(loc0, bor=parana$bor)

Loading required package: splancs
Spatial Point Pattern Analysis Code in S-Plus
Version 2 - Spatial and Space-Time analysis
Attaching package: 'splancs'
The following object(s) are masked from package:sp :bbox

> points(loc1, pch="+", col=4)
> KC <- krige.control(trend.d="1st", trend.l="1st", obj=ml)
> kc <- krige.conv(parana, loc=loc0, krige= KC, bor=parana$borders)
krige.conv: results will be returned only for prediction locations
inside the borders
krige.conv: model with mean given by a 1st order polynomial on the
coordinates
krige.conv: Kriging performed using global neighbourhood
>
> save.image()
>
> image(kc, col=terrain.colors(15))
>

```

FIGURA 18 – Comandos para Análise de Dados no R

Os dados processados no R podem ser plotados no ambiente R e visualizados no TerraView como mostram as Figuras 19 e 20.

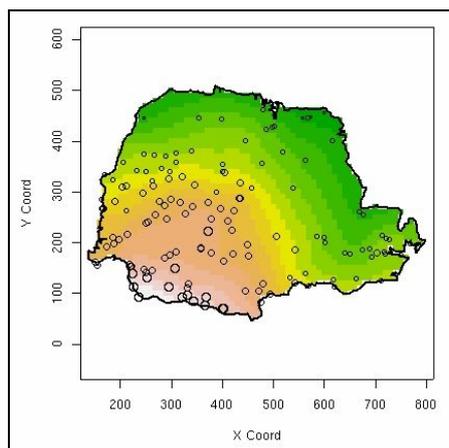


FIGURA 19 – Resultado plotado no ambiente R [Fonte: (Neto et al., 2005)]

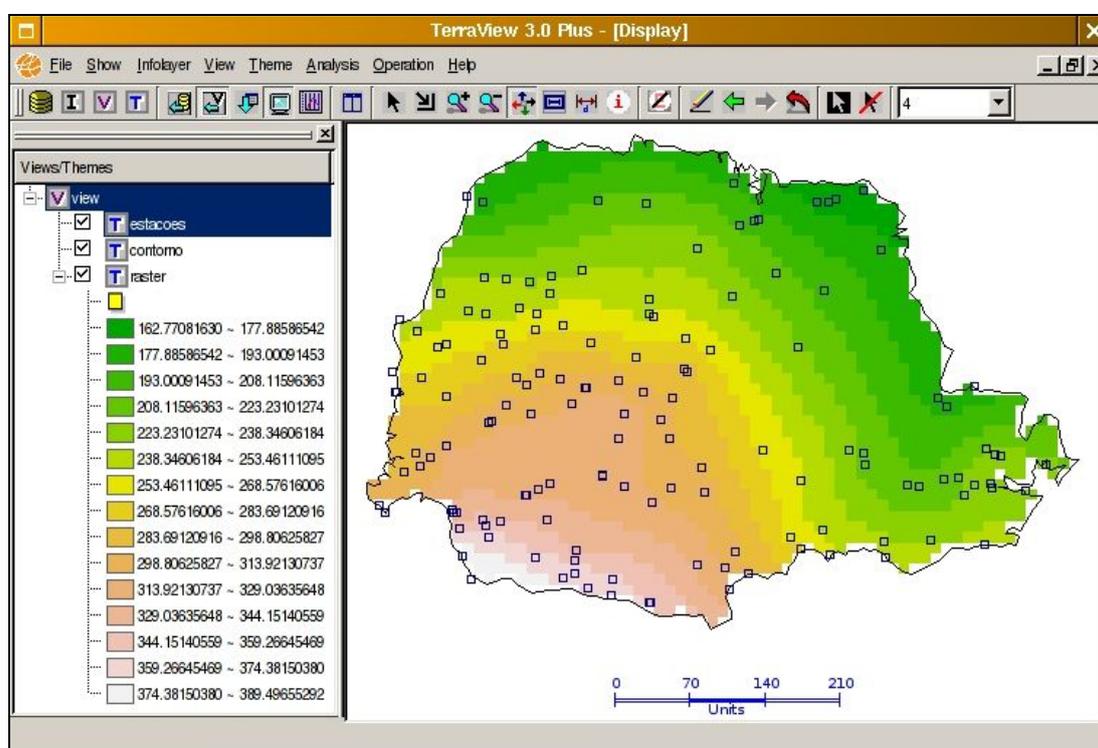


FIGURA 20 – Visualização dos dados no TerraView [Fonte: (Neto et al., 2005)]

O R e a TerraLib são softwares que operam tanto na plataforma Linux quanto na plataforma Windows, atingindo usuários de ambos os Sistemas Operacionais. Por outro lado, a API aRT até então só estava disponível sob o ambiente Linux. Cientes da relevância da integração entre os softwares R e TerraLib, buscou-se disponibilizar tal integração também para o Sistema Operacional Windows. A seguir o porte do aRT para a plataforma Windows é descrito.

4.1.1 Porte da API aRT para a plataforma Windows

O porte do aRT para o Windows é algo desejado por usuários do pacote que trabalham na referida plataforma. Foi realizado de forma conjunta por equipes do INPE e da UFPR com o intuito de migrar a integração dos softwares R e TerraLib, conseqüentemente R e TerraView, e disponibilizar as funcionalidades do aRT para o nicho de usuários do Windows, disseminando ainda mais sua utilização. O esquema de funcionamento do aRT na plataforma Windows é exibido na Figura 21.

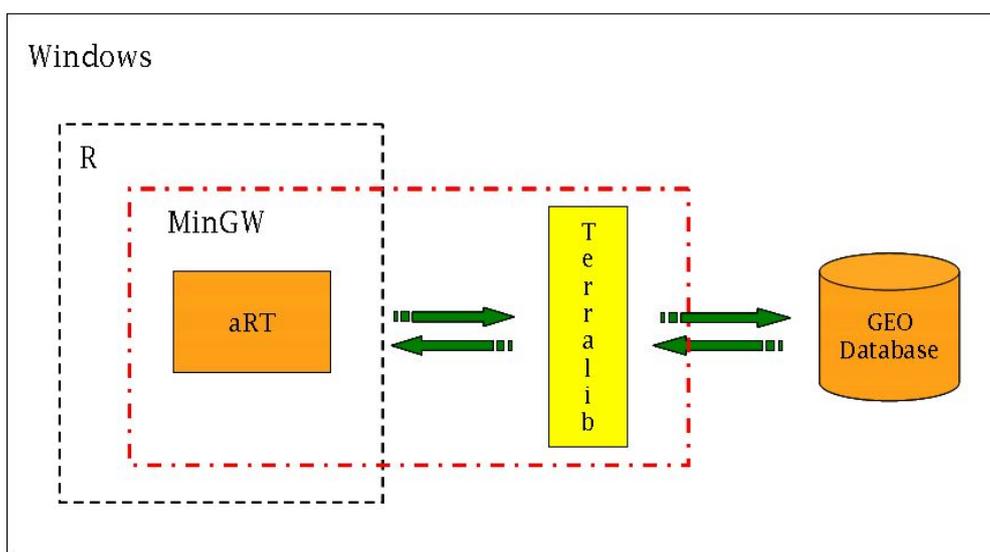


FIGURA 21 – Esquema aRT, plataforma Windows

Neste caso, o pacote aRT foi compilado sob um ambiente que simula, de forma mínima, o ambiente operacional do Linux. Para tanto, uma gama de softwares e utilitários foram instalados no Windows. Tanto a documentação da instalação do R (Murdoch, 2005) quanto a documentação da criação e instalação de pacotes do R para a plataforma Windows (Rossi, 2005) foram consultadas e seguidas para a montagem do ambiente requerido sob o referido Sistema Operacional. Alguns exemplos da execução de funções do aRT no R, sob ambiente Windows serão mostrados. A Figura 22 exibe a lista de pacotes disponíveis no R após o porte.



FIGURA 22 – Carregamento do pacote aRT no ambiente R, plataforma Windows

Inicialmente o pacote aRT é carregado no ambiente R, selecionando-o como na Figura 22 e clicando-se no botão OK. O *prompt* do ambiente R é disponibilizado novamente para que o usuário digite os comandos desejados. O comando *require()* é mostrado na Figura 23.



FIGURA 23 – Comando *require(aRT)* no ambiente R, plataforma Windows

Uma vez carregado o aRT, suas funções podem ser executadas de forma similar à da plataforma Unix como mostrado nas Figuras 24 e 25 que exibem, respectivamente, a conexão do R com um banco de dados TerraLib e a criação de um banco de dados.

```
> con <- new("aRT", user = "mysql", host = "localhost", pass = "")
Trying to connect ... yes
> print (con)
Object of class aRT

DBMS: "MySQL"
User: "mysql"
Password: ""
Port: 3306
Host: "localhost"
> showDbs (con)
[1] "mysql" "parana" "test"
>
```

FIGURA 24 – Conexão com banco de dados no ambiente R, plataforma Windows

```
> db <- new("aRTdb", con, db = "parana", create = TRUE)
Connecting with database 'parana' ... no
Creating database 'parana' ... yes
Creating conceptual model of database 'parana' ... yes
Loading layer set of database 'parana' ... yes
Loading view set of database 'parana' ... yes
> print (db)
Object of class aRTdb

Database: "parana"
External tables: (none)

This object has 0 children
>
```

FIGURA 25 – Criação do objeto da classe aRTdb, plataforma Windows

Um roteiro contemplando um maior detalhamento do porte da API aRT para a plataforma Windows encontra-se no apêndice deste documento.

5 CONSIDERAÇÕES FINAIS

O estudo realizado destacou a necessidade de integração entre softwares de estatísticas espaciais e GIS. De um lado, tem-se uma gama de pacotes de softwares que implementam estatísticas espaciais, mas sem contemplar componentes essenciais providos por um GIS, como um banco de dados espacial robusto, modelos espaciais e algoritmos de visualização (Goodchild, 2000; Goodchild and Haining, 2004; Krivoruchko, 2003). Do outro lado, usuários de GIS encontram dificuldades para fazer análises mais sofisticadas, pois parte dos softwares não disponibilizam ferramentas para análise estatística mais complexas. A junção destas duas tecnologias propicia o fortalecimento das mesmas e aumenta a eficácia individual e comum de ambas.

A interligação da estatística espacial e GIS pode ser completa ou por meio de mecanismos de ligação (acoplamento forte e acoplamento fraco). A integração via acoplamento forte entre a biblioteca de classes TerraLib e o software R através do pacote aRT foi apresentada em maior detalhe, destacando-se o trabalho realizado para efetuar o porte deste pacote para a plataforma Windows. Este trabalho permite que um maior número de usuários tenha acesso às funcionalidades oferecidas pelo aRT.

Acredita-se que a integração possa tornar tanto GIS quanto as estatísticas espaciais mais robustos, elevando o grau de complexidade das análises e a qualidade dos resultados. Alguns pontos devem ser destacados como sendo desafios na área de conexão entre softwares:

- Novas perspectivas de conexão entre os softwares de estatística espacial e GIS emergem com a abordagem espaço-tempo e com a alta demanda da utilização de estatística espacial e GIS em áreas como saúde, gerenciamento ambiental, mineração e planejamento urbano (Burrough, 1998; Casanova et al., 2005; Goodchild and Haining, 2004; Hornsby and Egenhofer, 2000).

- A necessidade de integração entre a tecnologia GIS e softwares de outras naturezas, como os de modelagem por exemplo, também se faz presente no cenário atual e age como item motivador da busca de novas alternativas de integração ou como áreas para a aplicação dos mecanismos já existentes.
- A sistematização do ambiente de acoplamento atual demanda maior atenção por parte dos desenvolvedores.
- A integração de softwares de arquiteturas distribuídas surge como desafio tecnológico, pois requer todo um arcabouço para a utilização conjunta dos sistemas de forma transparente para o usuário.

Na continuidade dos estudos nesta área, pretende-se averiguar as possíveis formas de integração entre ambientes computacionais com arquitetura distribuída.

APÊNDICE A

Roteiro do porte do aRT (API R-TerraLib) para a plataforma Windows

O pacote aRT (API R-Terralib) foi desenvolvido inicialmente para a plataforma UNIX. Entretanto, o porte para a plataforma Windows foi proposto para disponibilizar as funcionalidades do aRT também para os usuários do R no Windows. As etapas do porte serão descritas a seguir.

1. Instalar o R;
2. Checar as dependências do aRT:
 - 2.1. MySQL Server 4.1;
 - 2.2. QT Library;
 - 2.3. TerraLib (para UNIX);
 - 2.4. *Package sp*;
3. Preparar o ambiente (Sistema Operacional):
 - 3.1. Baixar e instalar softwares:
 - Rtools: conjunto de utilitários Unix;
 - MinGW: provê um ambiente minimalista Unix sob o Windows;
 - Perl (*Practical Extraction and Report Language*): funções para manipulação de texto;
 - Microsoft HTML Compiler: necessário para help de pacotes do R;
 - 3.2. Baixar o arquivo Libmysql.a para MinGW para a pasta de instalação do MySQL;
 - link: <http://www.dvrsol.com/programming.html#mymingw>
 - 3.3. Copiar a libmysql.dll para c:\...\R\rw2010\bin;
 - 3.4. Ajustar variável de ambiente PATH inserindo os caminhos dos seguintes softwares instalados:
 - C:\Arquiv~1\R\rw2011\bin;
 - C:\Arquiv~1\HTMLHE~1;
 - C:\Perl\bin;c:\Rtools\bin;

- C:\MinGW\bin;
- C:\Arquivos de programas\MySQL\MySQL Server 4.1\bin

3.5. Criar as variáveis de ambiente:

- TERRALIB_DIR = <pasta de instalação do R>;
- R_HOME = <pasta de instalação do R>;
- QTDIR = <pasta de instalação do QT>;
- QMAKESPEC = linux-g++;

4. Preparar pacote aRT para instalação:

4.1. Baixar o pacote aRT e extrair para C:/aRT;

4.2. Copiar fontes TerraLib necessários para C:/aRT/src;

4.3. Alterar Makevars (C:/aRT/src) segundo Figura 26:

```
TERRALIB_DIR = c:/terralib

PKG_CPPFLAGS = -DHAVE_MYSQL -
Ic:/terralib/src/terralib/drivers/MySQL -
Ic:/terralib/src/terralib/drivers/mysql/include \
-I$(TERRALIB_DIR)/src/terralib/kernel \
-I$(TERRALIB_DIR)/src/terralib/functions \
-I$(TERRALIB_DIR)/src/terralib/application \
-I$(TERRALIB_DIR)/src/terralib/stat \
-I$(TERRALIB_DIR)/src/shapelib \
-I$(R_HOME)/include -I$(R_HOME)/include/R_ext \
-pipe -Wall -g -DQT_SHARED -DQT_NO_DEBUG

PKG_LIBS =
$(TERRALIB_DIR)/terralibx/terralib/terralib.a \
$(TERRALIB_DIR)/terralibx/tiff/tifflib.a \
$(TERRALIB_DIR)/terralibx/shapelib/shapelib.a \
c:/Arquiv~1/mysql/libmysql.a
```

FIGURA 26 – Arquivo MakeVars

5. Preparar/Compilar o Ambiente da TerraLib:

5.1. Copiar o MakeTerralib (do linux) para a pasta C:\...\terralibx\terralib;

5.2. Executar o mingw32-make -f makeTerralib (via linha de comando);

- 5.3. Gerar o arquivo terralib.a (via linha de comando):
 - 5.3.1. ar cr terralib.a *.o;
 - 5.3.2. ranlib terralib.a;
- 5.4. Mover terralib.a para a pasta ...\terralibx\terralib (indicada no arquivo Makevars – vide Figura 26);
- 5.5. Gerar tiff.lib.a:
 - 5.5.1. Copiar o MakeTiff (do linux) para a pasta C:\terralib\terralibx\tiff;
 - 5.5.2. Executar o mingw32-make –f makeTiff (via linha de comando);
 - 5.5.3. Gerar o arquivo tiff.lib.a (via linha de comando, no diretório .obj):
 - 5.5.3.1. ar cr tiff.lib.a *.o;
 - 5.5.3.2. ranlib tiff.lib.a;
 - 5.5.4. Mover tiff.lib.a para a pasta ...\terralibx\tiff\ (indicada no arquivo Makevars – vide Quadro 1);
- 5.6. Repetir os passos de 5.5 para shapelib;
6. Executar o R CMD INSTALL C:\aRT via linha de comando;
7. Carregar o pacote aRT no ambiente R;

REFERÊNCIAS BIBLIOGRÁFICAS

- ALMEIDA, C. **Modelagem Espacial Dinâmica como Ferramenta de Planejamento: Simulação de Mudanças de Uso do Solo em Bauru e Piracicaba (SP)**. São José dos Campos: INPE, 2003. Doutorado em Sensoriamento Remoto, 2003.
- ANSELIN, L., 1992, SpaceStat tutorial: a workbook for using SpaceStat in the analysis of spatial data, Santa Barbara, NCGIA (National Center for Geographic Information and Analysis).
- ANSELIN, L., 1998. Exploratory Spatial Data Analysis in a Geocomputational Environment. In: LONGLEY, P.; BROOKS, S.; MCDONNELL, R.; MACMILLAN, B., eds., **Geocomputation: A Primer**: New York, John Wiley & Sons.
- ANSELIN, L.; BAO, S., 1997. Exploratory spatial data analysis linking SpaceStat and ArcView. In: FISCHER, M.; GETIS, A., eds., **Recent developments in spatial analysis**: Berlin, Springer-Verlag, p. 35-59.
- ASSUNÇÃO, R. **Estatística Espacial com Aplicações em Epidemiologia, Economia e Sociologia**. São Carlos, SP: UFScar, 2001.
- ASSUNÇÃO, R. M.; TAVARES, A. I.; KULLDORFF, M. An early warning system for space-time cluster detection. In: V Workshop Brasileiro em Geoinformática - GeoInfo2003. Campos do Jordão - São Paulo, 2003. p.
- AVRUSKIN, G. A.; JACQUEZ, G. M.; MELIKER, J. R.; SLOTNICK, M. J.; KAUFMANN, A. M.; NRIAGU, J. O. Visualization and exploratory analysis of epidemiologic data using a novel space time information system. **International Journal of Health Geographics**, 2004.
- BAILEY, T. Spatial Statistics Methods in Health. **Cadernos de Saúde Pública**, v. 17, n.5, 2001.
- BAILEY, T.; GATRELL, A. **Interactive Spatial Data Analysis**. London: Longman Scientific and Technical, 1995.
- BAO, S.; ANSELIN, L.; MARTIN, D.; STRALBERG, D. Seamless integration of spatial statistics and GIS: The S-PLUS for ArcView and the S+Grassland Links. **Journal of Geographical Systems**, v. 2, n.3, 2000.
- BIVAND, R. Using the R statistical data analysis language on GRASS 5.0 GIS data base files. **Computers and Geosciences**, v. 26, p. 1043-1052, 2000.
- BIVAND, R. R and geographical information systems, especially GRASS. In: 2nd

- International Workshop on Distributed Statistical Computing - DSC. Vienna, Austria, 2001. p.
- BIVAND, R. Approaches to Classes for Spatial Data in R. In: Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003). Vienna, Austria, 2003. p.
- BIVAND, R., 2005, Interfacing GRASS 6 and R: Status and development directions, GRASS-News, Germany, p. 11-16.
- BIVAND, R.; LUCAS, A. E., 2000. Integrating models and geographical information systems. In: OPENSHAW, S.; ABRAHART, R., eds., **Geocomputation**: London, Taylor & Francis.
- BIVAND, R.; NETELER, M. Open Source geocomputation: using the R data analysis language integrated with GRASS GIS and PostgreSQL data base systems. In: The Fifth International Conference on GeoComputation. Kent, UK, 2000. p.
- BURROUGH, P., 1998. Dynamic Modelling and Geocomputation. In: LONGLEY, P.; BROOKS, S.; MCDONNELL, R.; MACMILLAN, B., eds., **Geocomputation: A Primer**: New York, John Wiley.
- BURROUGH, P.; MCDONNELL, R. **Principles of Geographic Information Systems**. New York: Oxford University Press, 1998.
- CÂMARA, G.; MONTEIRO, A. M.; DAVIS, C. **Introdução à Ciência da Geoinformação**. São José dos Campos: INPE, 2003a.
- CÂMARA, G.; MONTEIRO, A. M.; MENEZES, J. S. Representações Computacionais do Espaço: Fundamentos Epistemológicos da Ciência da Geoinformação. **Revista Geografia (UNESP)**, v. 28, n.1, p. 83-96, 2003b.
- CÂMARA, G.; SOUZA, R.; FREITAS, U.; GARRIDO, J. SPRING: Integrating Remote Sensing and GIS with Object-Oriented Data Modelling. **Computers and Graphics**, v. 15, n.6, p. 13-22, 1996.
- CÂMARA, G.; SOUZA, R.; PEDROSA, B.; VINHAS, L.; MONTEIRO, A. M.; PAIVA, J.; CARVALHO, M. T.; GATTASS, M. TerraLib: Technology in Support of GIS Innovation. In: II Brazilian Symposium on Geoinformatics, GeoInfo2000. São Paulo, 2000. p.
- CASANOVA, M.; CAMARA, G.; DAVIS, C.; VINHAS, L.; QUEIROZ, G., eds., 2005, Bancos de Dados Geograficos (Spatial Databases): Curitiba, Editora MundoGEO.
- CRIA; FAPESP. **openModeller: Static Spatial Distribution Modelling Tool**. CRIA, 2005. <http://openmodeller.cria.org.br/>.

- CRIA; INPE; USP, 2005, Project Proposal: A framework for species distribution modeling, INPE, p. 20.
- DINIZ, A., 2000, Estatística Espacial, UFMG - Universidade Federal de Minas Gerais, p. 15.
- DPI/INPE. TerraView. INPE, 2002. <http://www.dpi.inpe.br/terraview/index.php>.
- DPI/INPE, 2004a, Estatística Espacial - Análise de Padrões de Distribuição de Pontos, Notas de Aula / Análise Espacial - INPE.
- DPI/INPE. **Intercâmbio de Dados**. INPE, 2004b. <http://www.dpi.inpe.br/spring/portugues/funcoes.html>.
- DPI/INPE, 2004c, Uma Breve Introdução à Inferência Bayesiana, Notas de Aula / Análise Espacial - INPE.
- DRUCK, S.; CARVALHO, M. S.; CÂMARA, G.; MONTEIRO, A. M. V. **Análise Espacial de Dados Geográficos**. Brasília: EMBRAPA (ISBN 85-7383-260-6), 2004.
- FELGUEIRAS, C.; CAMARGO, E.; MONTEIRO, A.; CÂMARA, G.; FUKS, S. Integrating Geostatistical Tools in Geographical Information Systems. In: I Brazilian Workshop on Geoinformatics. Brazilian Computing Society, Campinas, 1999. p.
- FISCHER, M. M.; SCHOLTEN, H. J.; UNWIN, D., 1996. Geographic Information Systems, spatial data analysis and spatial modelling: an introduction. In: FISCHER, M. M.; SCHOLTEN, H. J.; UNWIN, D., eds., **Spatial analysis perspectives in GIS**, Taylor and Francis.
- FOTHERINGHAM, A. S.; BRUNSDON, C.; CHARLTON, M. **Geographically Weighted Regression: The Analysis of Spatially Varying Relationships**. Chichester: Wiley, 2002.
- GOLDSTEIN, N. C.; CANDAU, J. T.; CLARKE, K. C. Approaches to simulating the "March of Bricks and Mortar". **Computers, Environment and Urban Systems**, v. 28, p. 125-147, 2004.
- GOODCHILD, M. F. The current status of GIS and spatial analysis. **Journal of Geographical Systems**, v. 2, p. 5-10, 2000.
- GOODCHILD, M. F.; HAINING, R. P. GIS and spatial data analysis: covering perspectives. **Papers in Regional Science**, v. 83, p. 363-385, 2004.
- GRUNSKY, E. C. R: a data analysis and statistical programming environment— an emerging tool for the geosciences. **Computers & Geosciences**, v. 28, p. 1219–1222, 2002.

- HORNSBY, K.; EGENHOFER, M. Identity-Based Change: A Foundation for Spatio-Temporal Knowledge Representation. **International Journal of Geographical Information Science**, v. 14, n.3, p. 207-224, 2000.
- JACQUEZ, G. M.; GREILING, D. A.; KAUFMANN, A. M. Design and implementation of a Space-Time Intelligence System for disease surveillance. **Journal of Geographical Systems**, v. 7, p. 7-23, 2005.
- KRIVORUCHKO, K. Using Spatial Statistics In GIS. In: International Congress on Modelling and Simulation. Townsville, Australia, 2003. p.
- LIU, X. H.; ANDERSSON, C. Assessing the impact of temporal dynamics on land-use change modeling. **Computers, Environment and Urban Systems**, v. 28, p. 107-124, 2004.
- MARBLE, D. Some thoughts on the integration of spatial analysis and Geographic Information Systems. **Geographical Systems**, v. 2, p. 31-35, 2000.
- MURDOCH, D. **Building R for Windows**. University of Western Ontario, 2005. <http://www.murdoch-sutherland.com/Rtools/>.
- NETO, P. R. D. A.; CARRERO, M. A.; JR., P. J. R., 2005, aRT: API R-TerraLib, Paraná, Brasil, Universidade Federal do Paraná, p. 19.
- PEBESMA, E. J. Multivariable geostatistics in S: the gstat package. **Computers & Geosciences**, v. 30, p. 683-691, 2004.
- PEBESMA, E. J.; WESSELING, C. G. Gstat: a program for geostatistical modelling, prediction and simulation. **Computers & Geosciences**, v. 24, n.1, p. 17-31, 1998.
- REIS, I. A., 2005, O estado da arte da integração entre Sistemas de Informação Geográfica e Modelos Inferenciais Bayesianos, São José dos Campos, INPE, p. 27.
- RIBEIRO-JR., P. J.; CHRISTENSEN, O. F.; DIGGLE, P. J. geoR and geoRglm: Software for Model-Based Geostatistics. In: 3rd International Workshop on Distributed Statistical Computing (DSC 2003). Vienna, Austria, 2003. p.
- ROSSI, P., 2005, Making R Packages Under Windows: A Tutorial, Chicago, U.S.A., University of Chicago - Graduate School of Business, p. 15.
- RUBIO, V. G. RArInfo: Using GIS Data with R. In: Proceedings of the 3rd Internacional Workshop on Distributed Statistical Computing (DSC 2003). Vienna, Austria, 2003. p.
- STURTZ, S.; LIGGES, U.; GELMAN, A. R2WinBUGS: A Package for Running

WinBUGS from R. **Journal of Statistical Software**, v. 12, n.3, 2005.

UNWIN, A., 1996. Exploring spatio-temporal data. In: FISCHER, M. M.; SCHOLTEN, H. J.; UNWIN, D., eds., **Spatial analysis perspectives in GIS**, Taylor and Francis.

VENABLES, W. N.; SMITH, D. M.; TEAM, R. D. C., 2001, **An Introduction to R: Notes on R: A Programming Environment for Data Analysis and Graphics**, Bristol, UK, R Development Core Team.