

ISSN 1677-9274

**Uso do Módulo “Análise” do
SPRING (Versão 3.6.02) para
Estudo de Dados Climáticos**



República Federativa do Brasil

Fernando Henrique Cardoso
Presidente

Ministério da Agricultura, Pecuária e Abastecimento

Marcus Vinicius Pratini de Moraes
Ministro

Empresa Brasileira de Pesquisa Agropecuária - Embrapa

Conselho de Administração

Márcio Fortes de Almeida
Presidente

Alberto Duque Portugal
Vice-Presidente

Dietrich Gerhard Quast
José Honório Accarini
Sérgio Fausto
Urbano Campos Ribeiral
Membros

Diretoria Executiva da Embrapa

Alberto Duque Portugal
Diretor-Presidente

Bonifácio Hideyuki Nakasu
Dante Daniel Giacomelli Scolari
José Roberto Rodrigues Peres
Diretores-Executivos

Embrapa Informática Agropecuária

José Gilberto Jardine
Chefe-Geral

Tércia Zavaglia Torres
Chefe-Adjunto de Administração

Kleber Xavier Sampaio de Souza
Chefe-Adjunto de Pesquisa e Desenvolvimento

Álvaro Seixas Neto
Supervisor da Área de Comunicação e Negócios



ISSN 1677-9274
Dezembro, 2002

*Empresa Brasileira de Pesquisa Agropecuária
Embrapa Informática Agropecuária
Ministério da Agricultura, Pecuária e Abastecimento*

Documentos 24

Uso do Módulo “Análise” do SPRING (Versão 3.6.02) para Estudo de Dados Climáticos

Amarindo Fausto Soares

Campinas, SP
2002

Embrapa Informática Agropecuária
Área de Comunicação e Negócios (ACN)

Av. André Tosello, 209

Cidade Universitária "Zeferino Vaz" – Barão Geraldo

Caixa Postal 6041

13083-970 – Campinas, SP

Telefone (19) 3789-5743 - Fax (19) 3289-9594

URL: <http://www.cnptia.embrapa.br>

e-mail: sac@cnptia.embrapa.br

Comitê de Publicações

Amarindo Fausto Soares

Ivanilde Dispato

José Ruy Porto de Carvalho (Presidente)

Luciana Alvim Santos Romani

Marcia Izabel Fugisawa Souza

Suzilei Almeida Carneiro

Suplentes

Adriana Delfino dos Santos

Fábio Cesar da Silva

João Francisco Gonçalves Antunes

Maria Angélica de Andrade Leite

Moacir Pedroso Júnior

Supervisor editorial: *Ivanilde Dispato*

Normalização bibliográfica: *Marcia Izabel Fugisawa Souza*

Capa: *Intermídia Produções Gráficas*

Editoração eletrônica: *Intermídia Produções Gráficas*

1ª. edição

on-line - 2002

Todos os direitos reservados

Soares, Amarindo Fausto.

Uso do módulo "Análise" do Spring (versão 3.6.02) para estudo de dados climáticos / Amarindo Fausto Soares. – Campinas : Embrapa Informática Agropecuária, 2002.

32 p. : il. – (Documentos / Embrapa Informática Agropecuária ; 24).

ISSN 1677-9274

1. Geostatística. 2. Sistema de informação geográfica. 3. Análise espacial. 4. Software Spring. I. Título. II. Série.

CDD – 551.015195 (21st ed.)

551.021 (21st ed.)

551.0727 (21st ed.)

Autor

Amarindo Fausto Soares

Eng. Agr., M.Sc. em Agronomia, Pesquisador Embrapa informática Agropecuária, Caixa Postal 6041, Barão Geraldo, 13083-970 – Campinas, SP.

Telefone (19) 3789-5780 - e-mail: fausto@cnptia.embrapa.br

Apresentação

O Sistema de Processamento de Informações Georreferenciadas (SPRING) foi desenvolvido pelo Instituto Nacional de Pesquisas Espaciais (INPE) com a participação da Embrapa Informática Agropecuária, desempenhando funções de processamento de imagens, análise espacial com ênfase em geoestatística, modelagem numérica de terreno, consulta de bancos de dados espacial.

A Embrapa Informática Agropecuária está desenvolvendo o projeto “Desenvolvimento e evolução de um sistema de monitoramento agroclimatológico para o Estado de São Paulo” através da criação de um banco de dados agroclimatológico envolvendo, entre outras funções, a modelagem numérica, modelagem estocástica com a espacialização das variáveis de clima, através do SPRING.

O processo de modelagem atualmente usado em análise ambiental é o resultado de operações em diversos níveis de representação sem se preocupar com a distorção da realidade geográfica. Dessa maneira muito se tem feito para que esse processo forneça também informação quantitativa a respeito da qualidade dos produtos gerados. O presente documento aborda a utilização do módulo Análise do SPRING, utilizando dados do próprio projeto, com a aplicação de interpoladores estocásticos para construção de modelos probabilísticos de incertezas relacionados aos valores não amostrados.

José Gilberto Jardine
Chefe-Geral

Sumário

Introdução	9
Análise Espacial	11
Procedimentos	13
Material	18
Resultados	18
Considerações Finais.....	31
Referências Bibliográficas.....	32

Uso do Módulo “Análise” do SPRING (Versão 3.6.02) para Estudo de Dados Climáticos

Amarildo Fausto Soares

Introdução

A informatização de processos em diversos setores da atividade humana vem se tornando bastante freqüente, popularizando e diversificando ainda mais sua utilização. Técnicas de geoprocessamento foram implementadas devido ao avanço da informática e da tecnologia orbital dando origem aos Sistemas de Informação Geográfica (SIGs), que são tecnologias voltadas ao tratamento de informações georreferenciadas.

O aumento na adoção dessas tecnologias vem norteando várias pesquisas na área da caracterização ambiental e análise de suas variáveis. Essas tecnologias referem-se ao uso dos SIGs, como base efetiva de informações para sistemas de apoio à decisão gerencial em seus diversos níveis, tendo motivado usuários e pesquisadores a buscar uma caracterização quantitativa para avaliação qualitativa dos produtos gerados (Felgueiras, 1999).

Segundo INPE (2002), as operações de consulta e manipulação de dados geográficos constituem a essência de um SIG, que o diferencia de tecnologias como Cartografia Automatizada e Projeto Auxiliado por Computador. O que distingue um SIG de outros tipos de sistemas de informação são aquelas funções que possibilitam a realização de análises

espaciais (geográficas). Tais funções utilizam os atributos espaciais e os não-espaciais das entidades gráficas armazenadas na base de dados, buscando fazer simulações (modelos) sobre os fenômenos do mundo real, seus aspectos ou parâmetros.

Os SIGs têm como produto uma manipulação de dados espaciais que sofrerão processo de modelagem, com a finalidade de representar uma realidade geográfica. Um grande número de situações podem ser simuladas e representadas cada uma envolvendo diferentes variáveis, fornecendo diversos níveis de representação que podem não corresponder à realidade por não ser levada em consideração a qualidade e a variabilidade espacial de cada dado isoladamente.

O presente documento objetiva descrever as experiências na utilização do módulo "Análise", contido no Sistema de Processamento de Informações Georreferenciadas (SPRING), desenvolvido pelo Instituto Nacional de Pesquisas Espaciais (INPE), para o estudo de variáveis climáticas de aproximadamente 110 estações climatológicas do Estado de São Paulo (Fig. 1), visando avaliar seu potencial para análises dessa natureza.



Fig. 1. Divisão Municipal do Estado de São Paulo. Os pontos pretos mostram a localização geográfica das estações meteorológicas usadas no trabalho.

Análise Espacial

A Análise Espacial estuda um fenômeno sob uma perspectiva abrangente, baseada em consultas e manipulações de seus dados espaciais, utilizando produtos dos SIGs, quanto a condição de localização, tendência, roteamento, padrões e muitos outros, proporcionados pela dupla natureza dos dados geográficos: localização espaço-temporal e atributos descritivos, através de consultas e manipulações de seus dados espaciais.

A localização espaço-temporal é a posição geográfica que ele possui, através de suas coordenadas (x, y) e a variação que ele sofre ao longo do tempo. Os atributos dizem respeito à natureza que os mesmos podem assumir: a) categórica ou temática, como vegetação, solos, textura, etc.; e b) contínua ou numérica, expressando valores de temperatura, altitude, teor de ferro, etc.

Segundo Felgueiras (1999), o produto de um SIG é sempre o resultado da manipulação de dados espaciais armazenados em meio digital. A informação é o resultado da modelagem de dados espaciais, buscando representar uma realidade geográfica, e das operações possíveis sobre estas representações. Este procedimento de modelagem computacional usando SIG é aplicado a modelos computacionais que representam fenômenos ambientais. Os diversos aspectos da realidade geográfica conduzem à idéia de camadas para representar a informação espacial em meio digital. Cada camada é uma representação de um conjunto distinto de elementos dessa realidade geográfica, isto é, um nível de informação espacial sob modelagem.

Em vista disso, prossegue o autor, um SIG é quase sempre o resultado de operações sobre diferentes camadas ou níveis de representação. Propriedades ambientais, tais como, temperatura, teor de ferro em uma rocha, textura do solo, relevo, e outras, são "camadas" de dados importantes em análises utilizando SIG. Deste modo, a capacidade de integração de dados espaciais dos SIGs significa operar sobre níveis de informação espacial diferenciados. Como a cada nível cabe a representação de um conjunto de elementos da realidade modelada, o resultado de uma operação multinível precisa se preocupar com a distorção da realidade geográfica em modelagem, que pode ocorrer como resultado da "integração" de representações diferenciadas e com características particulares.

Segundo Felgueiras (1999), muitos esforços têm sido empregados na formulação de uma estratégia para que o SIG, com sua capacidade de análise multinível, forneça, também, informação quantitativa com respeito a qualidade dos produtos gerados pelas operações envolvidas na integração supracitada. A qualidade desses produtos depende da qualidade dos dados espaciais, em primeira instância, da qualidade de cada nível de informação utilizado na análise e, também, do tipo e dos parâmetros da equação utilizada para integrar as diferentes "camadas". Informação especializada é a base para a tomada de decisão e informação com qualidade é a base para uma decisão mais acertada.

Conseqüentemente, os produtos da modelagem computacional não contêm informação de qualidade. A qualidade dos resultados é negligenciada ou é inferida subjetivamente pelos usuários do sistema. Este tipo de modelagem ainda é praticado nos tempos atuais, mas obviamente é pobre e tende a ser substituído por modelos mais representativos. Em função de um melhor entendimento, a modelagem computacional em ambiente de SIG será classificada segundo 3 etapas de evolução distintas: a modelagem determinística; a modelagem estocástica¹ linear; e, a modelagem estocástica não linear por indicação. A modelagem ambiental determinística se caracteriza pela modelagem dos dados espaciais segundo o paradigma determinístico. Essa modelagem é praticada desde os primeiros modelos computacionais implementados em um SIG. Neste tipo de modelagem, os modelos de dados espaciais não contêm componente aleatório e são, portanto, considerados exatos.

O SPRING vem aperfeiçoando, a cada versão, o seu módulo Análise objetivando incrementar a performance dos métodos de modelagem existente em seu pacote. Segundo Felgueiras (1999) a Modelagem Numérica de Terreno (MNT), empregando modelagem determinística, executa uma análise multinível diferenciada sem se preocupar com a distorção da realidade geográfica em modelagem. Dessa maneira, prossegue o autor, muitos esforços têm sido empregados na formulação de uma estratégia para que o SIG, com sua capacidade de análise multinível, forneça, também, informação quantitativa com respeito a qualidade dos produtos gerados pelas operações envolvidas na integração.

¹ O termo modelagem estocástica é utilizado, neste texto, como referência exclusiva à modelagem que se utiliza de procedimentos geoestatísticos para representação de atributos espaciais

Desse modo, o módulo Análise emprega processos analíticos de modelagem estocástica, oriundos de processos contidos na geoestatística.

Procedimentos

O módulo Análise, contido no Spring, executa um elenco de tarefas disponíveis, listadas a seguir, abrangendo operações no ambiente, nas variáveis que nele estão incidindo, medindo as distâncias entre eles e cruzando essas informações. De acordo com INPE (2002) são usadas:

1. Medidas geográficas;
 - 1.1. por apontamento como medida de área/perímetro de polígono; comprimento de linhas; distância entre entidades;
 - 1.2. por edição de polígonos para medidas de área/perímetro; linhas para medir comprimento; medida de distância entre dois pontos;
2. Medidas de Classe;
3. Mapa de Distâncias;
4. Tabulação Cruzada: permite calcular a área das interseções entre as classes dos Planos de Informação (Pis) temáticos, como classes de vegetação que ocorrem em determinadas classes de solos;
5. Álgebra de Mapas: executa uma seqüência de operações definidas pelo usuário, através de um conjunto de sentenças descritivas, envolvendo operadores, funções e dados espaciais, constituindo a Linguagem Espacial para Geoprocessamento Algébrico (LEGAL);
6. Análise de Padrões Pontuais: permite analisar como estão distribuídas as variáveis, através de métodos e modelagem, e explicar o comportamento do processo espacial com relação a algum outro fenômeno. Nessa distribuição, as amostras podem estar: a) agrupadas; b) espalhadas; e c) aleatoriamente distribuídas;
7. Estimativa de Densidade, permite estimar a densidade do padrão de pontos, através de um estimador Kernel quártico INPE (2002);
8. Geoestatística: importante ferramenta de Análise Espacial, que vem sendo muito utilizada em processos de modelagem e que será enfatizada no decorrer do trabalho, com apresentação de alguns testes utilizando dados citados no objetivo do trabalho.

Um fato muito importante que deve ser levado em consideração, é a variação espacial que o dado sofre. Durante muito tempo estudou-se essa variabilidade através de cálculos utilizando parâmetros como média, desvio padrão, baseados na estatística clássica, considerando que esse tipo de variação era aleatória.

Krige (1951), citado por INPE (2002), trabalhando com dados de concentração de ouro, concluiu que somente a informação dada pela variância seria insuficiente para explicar o fenômeno em estudo. Para tal, seria necessário levar em consideração a distância entre as observações.

A partir daí surge o conceito da Geoestatística, que leva em consideração a localização geográfica e a dependência espacial. Matheron (1963, 1971), citados por INPE (2002), baseado nas observações de Krige, desenvolveu a teoria das variáveis regionalizadas, a partir dos fundamentos da Geoestatística. Segundo Olea (1975) citado por INPE (2002), uma variável regionalizada é uma função numérica com distribuição espacial, que varia de um ponto a outro com continuidade aparente, mas cujas variações não podem ser representadas por uma função matemática simples.

De acordo com Felgueiras (1999) a geoestatística trata os valores dos atributos ambientais segundo a teoria das variáveis regionalizadas. Neste contexto os valores dos atributos são representados como variáveis ou campos aleatórios, cujas funções de distribuição de probabilidade modelam a incerteza² dos valores assumidos. Estas incertezas devem ser consideradas como parte da representação do atributo e se propagam em análises realizadas no ambiente de modelagem computacional de um SIG.

A geoestatística modela os valores de um atributo dentro de uma região da superfície terrestre $A \subset \mathfrak{R}^2$, como uma função aleatória. Para cada posição $\mathbf{u} \in A$ o valor do atributo de um dado espacial é modelado como uma variável aleatória $Z(\mathbf{u})$ ³. Isto significa que, na posição \mathbf{u} , $Z(\mathbf{u})$ pode

² O termo incerteza é usado como uma medida indicativa da magnitude do erro, uma vez que não se pode determinar com exatidão o erro de estimação do atributo. Para isto seria necessário o conhecimento do valor real do atributo em cada posição estimada.

³ As definições e convenções adotadas seguem os padrões da geoestatística, funções e variáveis aleatórias VA são denotadas com caracteres maiúsculos (exemplo: $Z(\mathbf{u})$ e Z respectivamente), valores observados são representados por caracteres minúsculos (exemplo: valor da variável aleatória - VA Z medido na posição \mathbf{u}_i é $z(\mathbf{u}_i)$) e vetores são realçados em **negrito** (exemplo: $\{z(\mathbf{u}_i), i = 1, \dots, n\}$, onde \mathbf{u}_i identifica uma posição em duas dimensões representada pelos pares de coordenadas (x_i, y_i)).

assumir diferentes valores para o atributo, cada valor com uma probabilidade de ocorrência associada. Nas n posições amostradas, \mathbf{u}_α com $\alpha = 1, 2, \dots, n$, os valores $z(\mathbf{u}_\alpha)$ são considerados determinísticos, ou ainda, podem ser considerados variáveis aleatórias cujo valor medido tem uma probabilidade de 100% de ocorrência. Uma variável aleatória é caracterizada pela sua função de distribuição acumulada (fda) univariada, $F(\mathbf{u}; z)$.

Segundo Landim (1988) a Krigeagem⁴ é um processo estimativo de valores de variáveis espacialmente distribuídas a partir de valores adjacentes considerados como interdependentes pelo semivariograma. Dessa maneira, a krigeagem é considerada um método de médias móveis.

A krigeagem é a âncora dos procedimentos geoestatísticos. Ela possibilita inferências de valores, a partir de amostras pontuais de um atributo espacial. Estas inferências são usadas, geralmente, para finalidades de mapeamento digital de atributos espaciais. Supondo-se a existência de um conjunto de n amostras pontuais, $z(\mathbf{u}_\alpha)$ de valores de um atributo, dentro de uma região da superfície terrestre, o procedimento da krigeagem permite a estimação de valores do atributo em localizações $\mathbf{u} \neq \mathbf{u}_\alpha$. Além disso, este procedimento possibilita a inferência da variância de krigeagem, que, em certas condições, pode ser usada como uma medida de incerteza do atributo em \mathbf{u} (Felgueiras, 1999).

Segundo INPE (2002), o Variograma é a ferramenta básica de suporte às técnicas de krigeagem, permitindo representar quantitativamente a variação de um fenômeno regionalizado no espaço. Considerando duas variáveis regionalizadas, X e Y onde $X = Z(\mathbf{x})$ e $Y = Z(\mathbf{x} + \mathbf{h})$ referem-se ao mesmo atributo medido em duas posições diferentes, medidas através do vetor \mathbf{h} , mostrados na Fig. 2. O nível de dependência entre as duas variáveis aleatórias é representado pelo variograma $2\gamma(\mathbf{h}) = E\{[Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})]^2\} = \text{Var}[Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})]$, definido como a esperança matemática do quadrado da diferença entre os valores de pontos no espaço, separados pelo vetor distância \mathbf{h} , que por sua vez é igual a variância do atributo medido nas duas posições consideradas acima.

⁴ *Krigeagem*, do francês "krigeage" e do inglês "kriging", é um termo cunhado pela escola francesa de geoestatística em homenagem ao engenheiro de minas sul-africano e pioneiro na aplicação de técnicas estatísticas em avaliação de minas D.G. Krige.

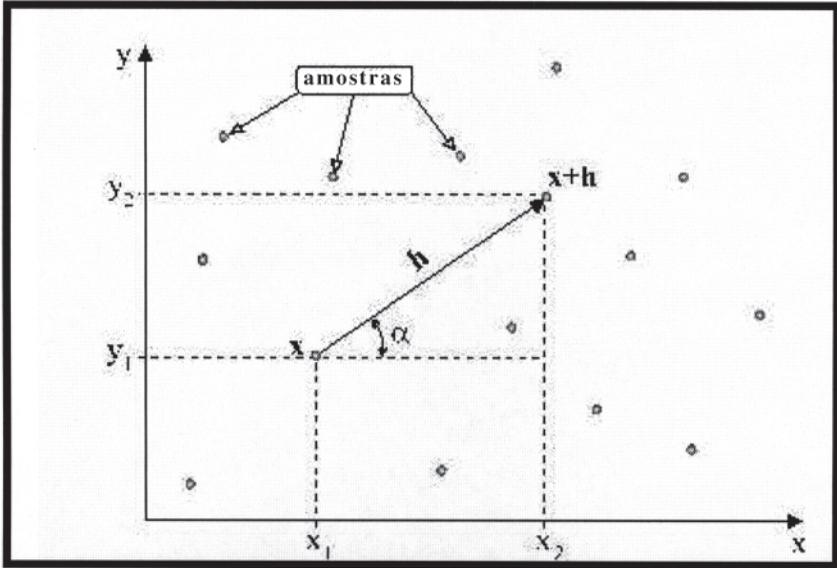


Fig. 2. Esquema amostral bidimensional explicativo do variograma como ferramenta básica de suporte às técnicas de krigagem.

Segundo Braga (1990), o variograma é a chave para a inferência e operação de modelos estatísticos, está relacionado com a covariância, cobrindo entretanto um domínio de casos mais amplo que esta.

Alguns autores definem o variograma de forma distinta, considerando o que comumente se refere como semivariograma, $\gamma(\mathbf{h}) = E\{[Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})]^2\} = 1/2 \text{Var}[Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})]$.

Com características muito próximas do ideal, o seu padrão representa o que intuitivamente se espera de dados de campo, que as diferenças $\{Z(\mathbf{x}_i) - Z(\mathbf{x}_i + \mathbf{h})\}$ decresçam à medida que \mathbf{h} , a distância que os separa decresce. É esperado que observações mais próximas geograficamente tenham um comportamento mais semelhante entre si do que aquelas separadas por maiores distâncias. Desta maneira, é esperado que $\gamma(\mathbf{h})$ aumente com a distância \mathbf{h} (INPE, 2002).

Os parâmetros de um semivariograma são representados na Fig. 3 e definidos a seguir:

1. Alcance (a): distância dentro da qual as amostras apresentam-se correlacionadas espacialmente. Na Fig. 2, o alcance ocorre próximo de 25 m.
2. Patamar (C): é o valor do semivariograma correspondente a seu alcance (a). Deste ponto em diante, considera-se que não existe mais dependência espacial entre as amostras, porque a variância da diferença entre pares de amostras ($\text{Var}[Z(\mathbf{x}) - Z(\mathbf{x}+\mathbf{h})]$) torna-se invariante com a distância.
3. Efeito Pepita (C_0): por definição, $\gamma(\mathbf{0}) = 0$. Entretanto, na prática, a medida que \mathbf{h} tende para 0 (zero), $\gamma(\mathbf{h})$ se aproxima de um valor positivo chamado Efeito Pepita (C_0). O valor de C_0 revela a descontinuidade do semivariograma para distâncias menores do que a menor distância entre as amostras. Parte desta descontinuidade pode ser também devida a erros de medição, mas é impossível quantificar se a maior contribuição provém dos erros de medição ou da variabilidade de pequena escala não captada pela amostragem.
4. Contribuição (C_1): é a diferença entre o patamar (C) e o Efeito Pepita (C_0).

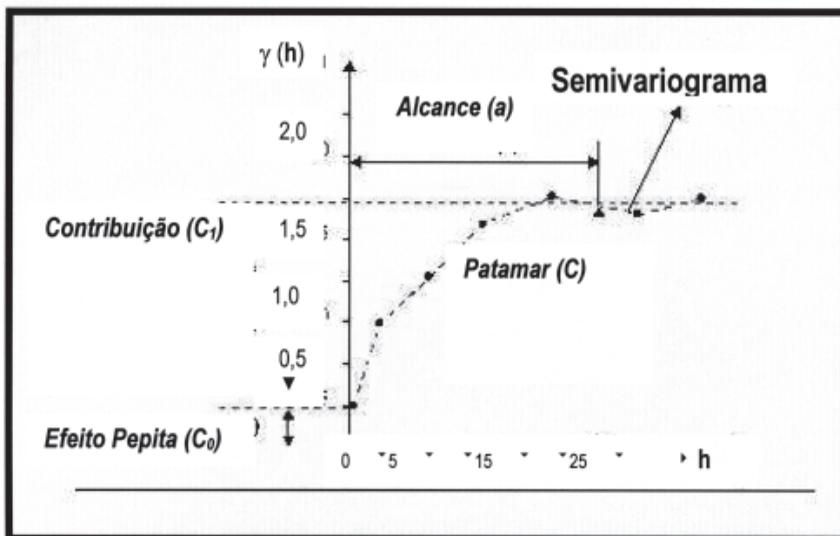


Fig. 3. Modelo de semivariograma, destacando seus parâmetros Alcance, Patamar Efeito Pepita e Contribuição.

Material

No presente trabalho utilizou-se procedimentos para avaliar algumas ferramentas de análise espacial, executadas pelo SPRING, apresentadas anteriormente.

Utilizou-se um contorno do mapa municipal do estado de São Paulo obtido de arquivos, disponíveis para download em <http://www.dpi.inpe.br/spring/portugues/mapaspr.html#gerais>, (Fig. 1) contendo polígonos identificadores e tabelas, juntamente com a versão 3.6.02 do SPRING disponível em <http://www.dpi.inpe.br/spring/>.

Empregou-se também um conjunto de planilhas contendo dados coletados, de variáveis de clima do ano 2000, de aproximadamente 114 estações climatológicas do Estado de São Paulo, obtidos de série histórica de trabalho de levantamento de campo executado pelo Instituto Agrônomo de Campinas (IAC), constantes do acervo da base de dados que a Embrapa Informática Agropecuária está montando contido no Projeto "Desenvolvimento e evolução de um sistema de monitoramento agroclimatológico para o Estado de São Paulo" dentro do Subprojeto Desenvolvimento de modelos operacionais para monitoramento e alerta agroclimatológico no Estado de São Paulo.

Resultados

São apresentados a seguir alguns resultados, numa seqüência de procedimentos do SPRING, mostrando as vantagens de se utilizar o Módulo Análise do SPRING na sua atual versão. Organizou-se um roteiro apresentando suas ferramentas tendo, como exemplo, utilização do atributo numérico Temperatura Máxima, mostrando as operacionalidades do módulo.

1. Análise Exploratória

- 1.1. Estatística Descritiva: fornece uma tabela contendo algumas informações da temperatura máxima (tx) no inverno do ano 2000 obtidas através dos dados de 105 estações do estado de São Paulo, apresentados na Fig. 4, dados esses obtidos através de estatísticas univariadas. São fornecidos o número de amostras totais

e as amostras que entraram no processo, como também média, variância, desvio padrão e outros mostrados a seguir. O valor mínimo e valor máximo são muito usados em processo de modelagem determinística no módulo Modelagem Numérica de Terreno (MNT).

ESTATÍSTICAS DESCRITIVA

Plano de Informação tx

Número de amostras	105
Número de amostras válidas	105
Média	25.12552390
Variância	10.45546896
Desvio Padrão	3.23349176
Coeficiente de Variação	0.12869351
Coeficiente de Assimetria	-0.49743946
Coeficiente de Curtose	2.90084223
Valor Mínimo	15.09000015
Quartil Inferior	23.13999939
Mediana	25.34000015
Quartil Superior	27.56999969
Valor Máximo	31.20999908

Fig. 4. Tabela contendo um resumo da estatística descritiva, de dados de Temperatura máxima de uma semana do Inverno no estado de São Paulo, obtida através de procedimento de análise exploratória do SPRING.

- 1.2. Histograma: fornece um gráfico, da temperatura máxima (t_x) no inverno do ano 2000, mostrado na Fig. 5, no formato de histograma dividido em 20 classes (amarelo) e a distribuição Gaussiana (vermelho), obtidos também através de estatística univariada.
- 1.3. Gráfico de Probabilidade Normal: obtido através de estatística univariada, esse gráfico, apresentado na Fig. 6, mostra na linha azul a distribuição dos dados de temperatura máxima no inverno do ano 2000, obtido das 105 estações meteorológicas do estado de São Paulo e na linha vermelha a distribuição Gaussiana, mostrando uma tendência de distribuição normal da variável.

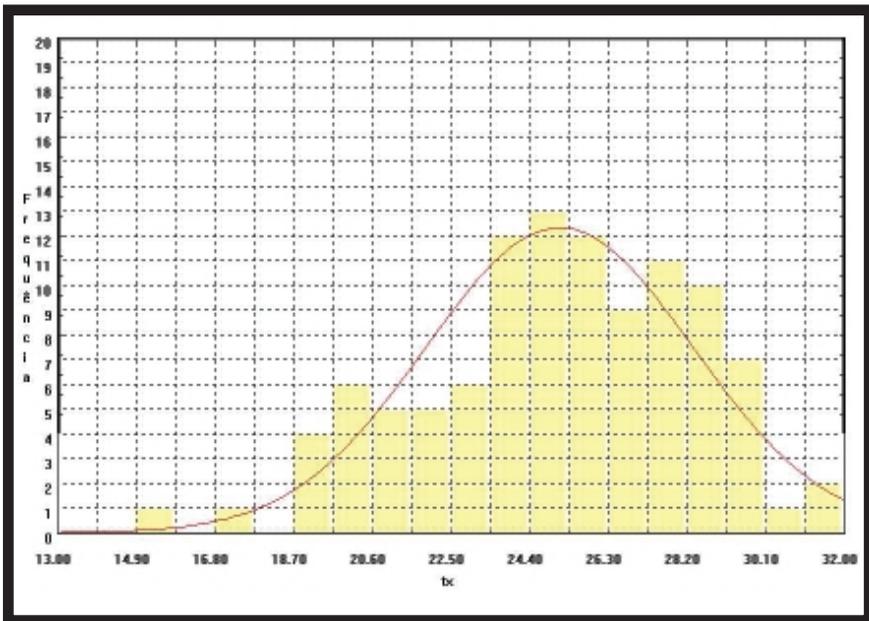


Fig. 5. Histograma da Temperatura máxima " t_x " no inverno do ano 2000.

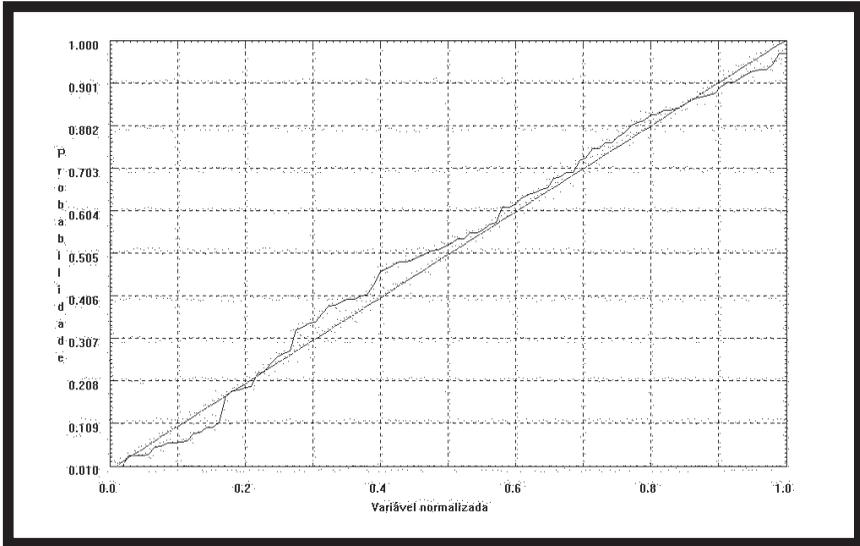


Fig. 6. Gráfico da Probabilidade Normal, apresentando em linha azul os dados amostrais e, em vermelho, refere-se a distribuição Gaussiana.

2. Espacialização⁵

A espacialização de amostras pontuais, na sua maneira mais simples, dos atributos de natureza numérica aqui estudado (Temperatura máxima, mínima e chuva, empregando no exemplo dados de temperatura máxima do inverno do ano 2000) através da Modelagem Numérica de Terrenos (MNT), gerando uma grade retangular, utilizando um interpolador Média Ponderada dos valores de temperatura máxima do ano 2000, cujo valor de cada ponto da grade é calculado a partir da média ponderada dos valores dos oito vizinhos mais próximos a este, atribuindo-se pesos variados para cada ponto amostrado através de uma função que considera a distância do ponto cotado ao ponto da grade INPE (2002), obtendo-se como produto uma imagem em níveis de cinza onde cada pixel assume um nível de cinza proporcional ao valor da variável em questão.

⁵ Espacialização é atribuir valores de um atributo para se obter uma representação computacional no espaço. Os SIGs representam atributos com distribuições espaciais na forma matricial, num processo de espacialização segundo uma representação de grade regular.

A geração de imagem em níveis de cinza estabelece que todos os valores absolutos da variável em questão (tx) serão representados em uma escala de níveis de cinza variando do zero (0) preto ao 255 (branco) totalizando 256 níveis, visto que os computadores trabalham com um número fixo de campos ou bytes para armazenar valores numéricos $2^8 = 256$. Dessa maneira os locais mais escuros terão valores absolutos temperaturas máximas de invernos menores e vice-versa (Fig. 7).

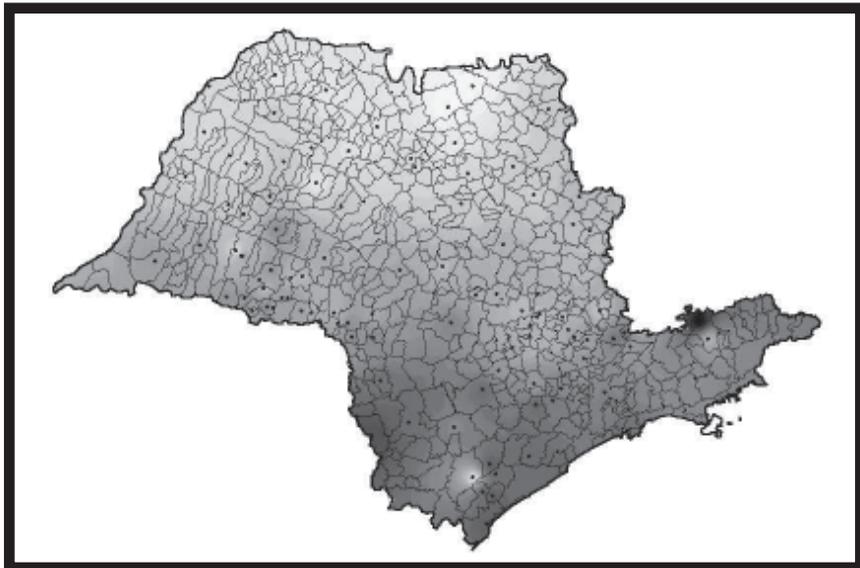


Fig. 7. Imagem em níveis de cinza, obtida pela geração de grade retangular para espacialização de dados pontuais de temperatura máxima tx no inverno do ano 2000. Na geração de grade utilizou-se o interpolador Média Ponderada contido no módulo.

3. Análise Espacial com geração de semivariograma

Concluída a análise exploratória passamos, na seqüência, para análise espacial com geração de um semivariograma experimental, constituindo-se uma etapa importante, pois a partir do modelo de variograma escolhido será efetuada a interpretação da estrutura de correlação espacial a ser utilizada nas inferências de krigagem. Nessa etapa o Spring executará diversos estágios da análise espacial, ajustando e validando os parâmetros do semivariograma experimental.

Utilizando dados de temperatura máxima gerou-se um semivariograma unidirecional, adotando-se um Lag, que vem a ser uma distância pré-definida utilizada no cálculo do semivariograma, como também uma direção já previamente preenchidos, podendo, em alguns casos, rever esses valores de forma a melhorá-lo. A opção de parâmetros de direção é de livre escolha e com cores correspondendo as do gráfico (Fig. 8). Um arquivo .txt, contendo os resultados numéricos poderá ser salvo para ser exibido.

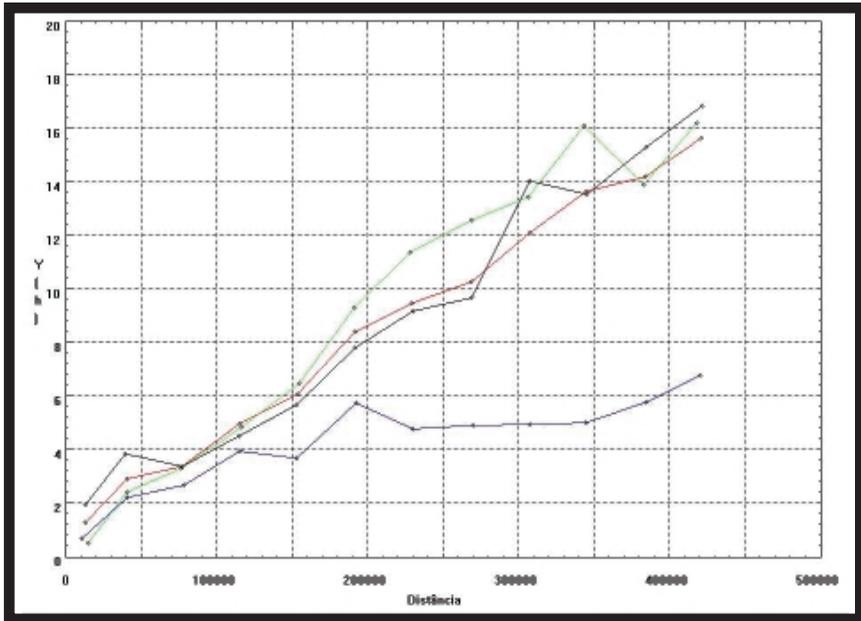


Fig. 8. Semivariograma unidirecional sobre amostras irregulares em 4 direções com cores individualizadas: 0° vermelho, 45° verde, 90° azul e 135° preta.

4. Ajuste de semivariograma

O semivariograma experimental, construído anteriormente, é calculado através de uma equação e formado por uma série de valores sobre os quais se objetiva ajustar uma função. É importante que o modelo seja ajustado representando sua tendência de se enquadrar em algum modelo teórico já existente. Deste modo, as estimativas obtidas a partir da krigeagem serão mais exatas e, portanto mais confiáveis.

O procedimento de ajuste não é direto e automático, mas sim interativo, pois nesse processo o intérprete faz um primeiro ajuste e verifica a adequação do modelo teórico. Dependendo do ajuste obtido, pode ou não redefinir o modelo, até obter um que seja considerado satisfatório.

Os modelos teóricos aqui apresentados na Fig. 9, são modelos básicos e divididos em dois tipos: a) com patamar ou transitivos sendo os mais utilizados: Esférico, a inclinação da tangente junto a origem, é o mais comum, equívulendo a função de distribuição normal da estatística clássica; Exponência, a inclinação junto a origem, o modelo atinge o patamar assintoticamente; Gaussiano, semelhante ao Exponencial, atinge o patamar tangenciando-o, sua principal característica é o comportamento parabólico próximo a origem, todos mostrados na Fig. 9 A; e b) sem patamar, continuam aumentando enquanto a distância aumenta, mostrado aqui o Modelo Potência, serve para modelar fenômenos com capacidade infinita de dispersão, Fig. 9 B.

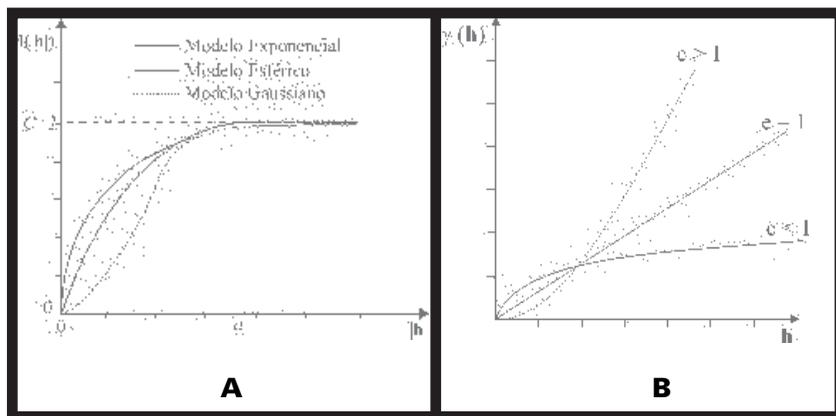


Fig. 9. Representação gráfica dos modelos teóricos: A: com patamar ou transitivos, representando os modelos Exponencial, Esférico e Gaussiano e B: sem patamar, representando o modelo Potência.

Segundo INPE (2002) essa etapa pode ser feita de modo automático ou visual. No primeiro caso, utiliza-se um algoritmo baseado no método dos mínimos quadrados, fornecendo também uma medida quantitativa, denominada informação de Akaike, indicando para qual modelo o ajuste é mais preciso.

Essa etapa, como o próprio nome sugere, ajusta o semivariograma de acordo com a combinação entre o número de estruturas necessárias e as quatro opções de modelos disponíveis.

Para cada escolha executada, ficam habilitados botões permitindo definir e verificar os ajustes, por tentativa executados, visualizando um Relatório de Dados contendo informações dos parâmetros componentes do modelo e da informação de Akaike, indicando o ajuste desejado através do seu menor valor, que irá ajustar também os outros componentes, definindo gradativamente o modelo teórico de semivariograma a ser utilizado nas etapas seguintes (Validação e Krigeagem).

Procedeu-se ao ajuste automático do semivariograma da Fig. 8, utilizando as três estruturas disponíveis ao ajuste, definindo-se, nas três opções, o modelo teórico esférico, verificando ser satisfatório o modelo de ajuste para 0° , conforme mostra a Fig. 10. O ajuste do modelo teórico escolhido (em vermelho) sobre o semivariograma experimental (em preto). Visualmente é possível dizer que o ajuste é satisfatório.

5. Validação do Modelo

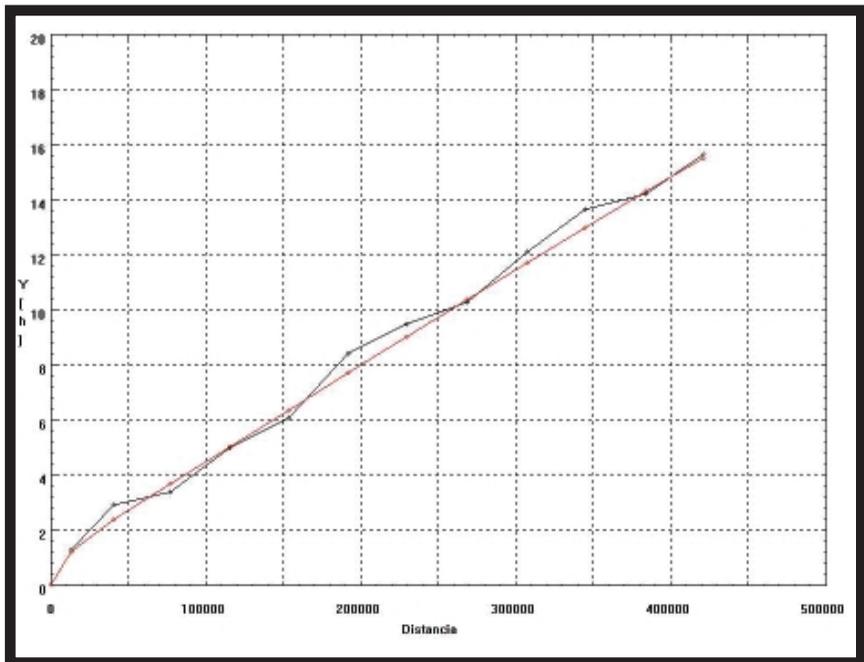


Fig. 10. Semivariograma experimental (em preto), ajustado por três modelos teóricos esféricos (em vermelho).

Até essa etapa, tem sido executados processos tentativos e preliminares, criando um certo grau de incerteza sobre os parâmetros ajustados aos modelos. Esta incerteza, obtida através do processo de validação, constitui-se o erro da estimativa. Será feita uma nova estimativa dos valores já conhecidos através dos parâmetros ajustados ao modelo do semivariograma (INPE, 2002). Utilizam-se os parâmetros, Efeito Pepita, Contribuição e Alcance, fornecidos no relatório de dados, no ajuste do semivariograma experimental pelos modelos esféricos selecionados para preencher os parâmetros estruturais do modelo a ser validado.

Para a validação do modelo, o Spring utiliza uma sub-rotina do programa kt3d.exe da Geostatistical Software Library (GSLIB) (<http://ekofisk.stanford.edu/SCRF.html>), uma biblioteca pública de funções geostatísticas, fornecendo um diagrama espacial do erro apresentado no Fig. 11.

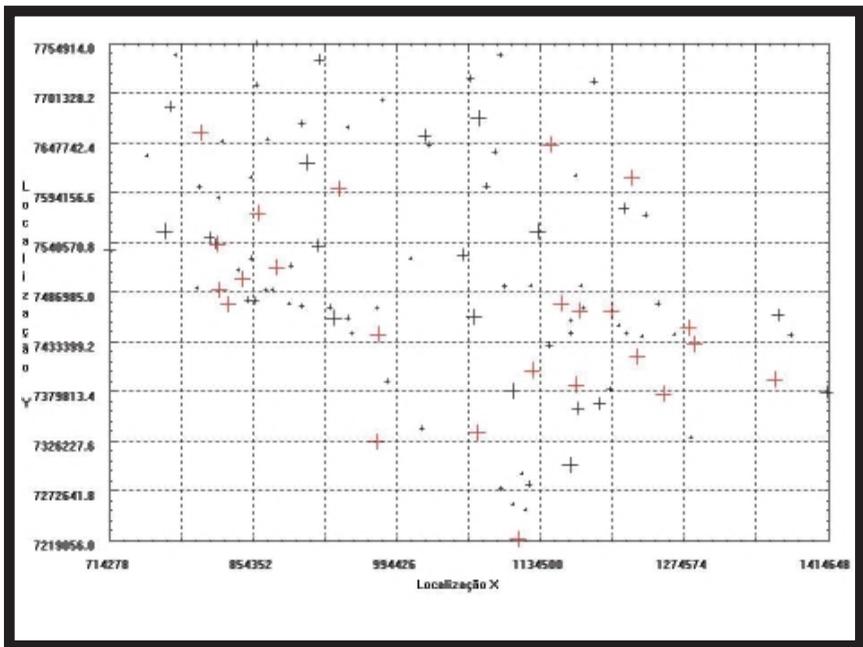


Fig. 11. Distribuição Espacial do Erro. Os símbolos tipo cruz indicam a localização geográfica das amostragens e a magnitude do erro (para os símbolos pequenos o erro é menor e vice-versa).

Esse diagrama exibe símbolos em forma de cruz em diversos tamanhos indicando a localização da amostragem, como também a magnitude do erro (símbolos pequenos demonstram erro menor e vice-versa). Outras informações adicionais e mais detalhadas poderão ser fornecidas ao acionar o mouse com o botão direito apertado em cima do sinal onde você deseja a informação, cujo resultado será informado no rodapé da janela em questão.

Nota-se que embora os ajustes tenham sido feitos, ainda existe uma certa porcentagem de erros, demonstrando o grau de complexidade da análise e seus ajustes para obter-se um processo de Krigeagem, satisfatório.

6. Krigeagem

É o uso de médias móveis para evitar a superestimação. Ela difere de outros métodos de interpolação pela maneira como os pesos são atribuídos às diferentes amostras. Na krigeagem, o procedimento é semelhante ao de interpolação por média móvel ponderada, exceto que os pesos são determinados a partir de uma análise espacial, baseada no semivariograma experimental. Além disso, a krigeagem fornece, em média, estimativas não tendenciosas e com variância mínima (Felgueiras, 1999).

A modelagem estocástica linear se caracteriza pela modelagem dos atributos espaciais segundo o paradigma geoestatístico linear. Para cada posição do dado o seu valor é modelado como uma variável que assume diferentes valores de atributos a cada posição com uma probabilidade de ocorrência associada, sendo tratada como uma função numérica com distribuição espacial, variando de um ponto a outro, com uma continuidade aparente, mas cujas variações não podem ser representadas por uma função matemática simples (Felgueiras, 1999).

Segundo Oliver e Webster (1990) citados em INPE (2002), a krigeagem linear engloba um conjunto de métodos de estimação, dos quais iremos demonstrar aqui somente a krigeagem simples e a krigeagem ordinária.

É a etapa final do nosso estudo com a obtenção de uma grade regular de valores a partir de dados amostrados. Esse módulo foi implementado pelo SPRING, baseado na sub-rotina do programa kt3d.exe da Geostatistical Software Library (GSLIB) (<http://ekofisk.stanford.edu/SCRF.html>), uma biblioteca pública de funções geoestatísticas englobando krigeagem simples, krigeagem ordinária.

6.1. Krigagem simples, ou simplesmente krigagem, foi inicialmente utilizada como um estimador de valores de atributos numéricos, em posições não observadas, para mapeamentos por médias ponderadas dos valores existentes das amostras locais. Atualmente vem sendo utilizada, com maior frequência, na construção de modelos probabilísticos de incerteza sobre os valores dos atributos. Pode ser usada para inferência dos parâmetros, média e variância, de um modelo probabilísticos gaussiano. A krigagem possibilita a inferência de um valor médio do atributo, em uma posição espacial \mathbf{u} , a partir de um quantidade $n(\mathbf{u})$ de amostras vizinhas $z(\mathbf{u}_a)$, $a = 1, \dots, n(\mathbf{u})$. (Fig. 12 e 13)

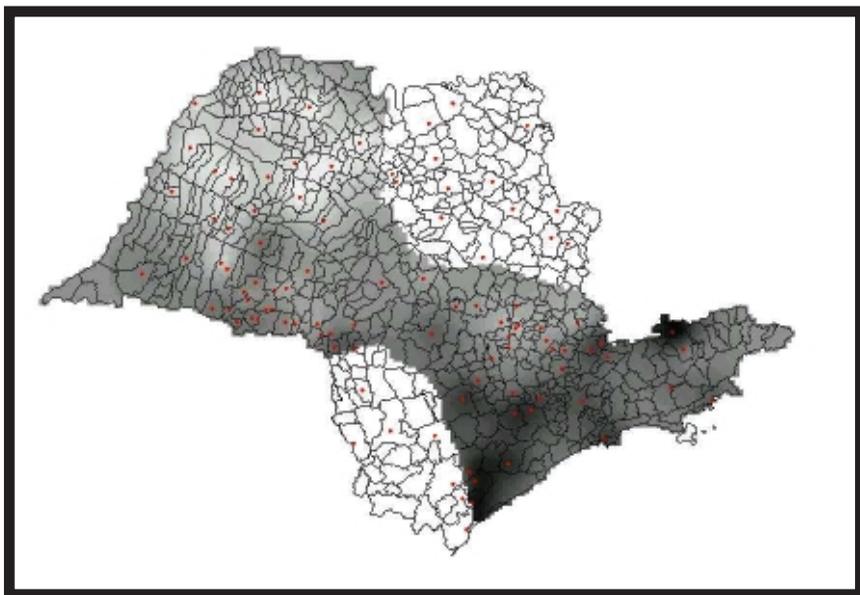


Fig. 12. Plano de Informação contendo a estimativa da variável temperatura máxima t_x , obtido através de krigagem simples.

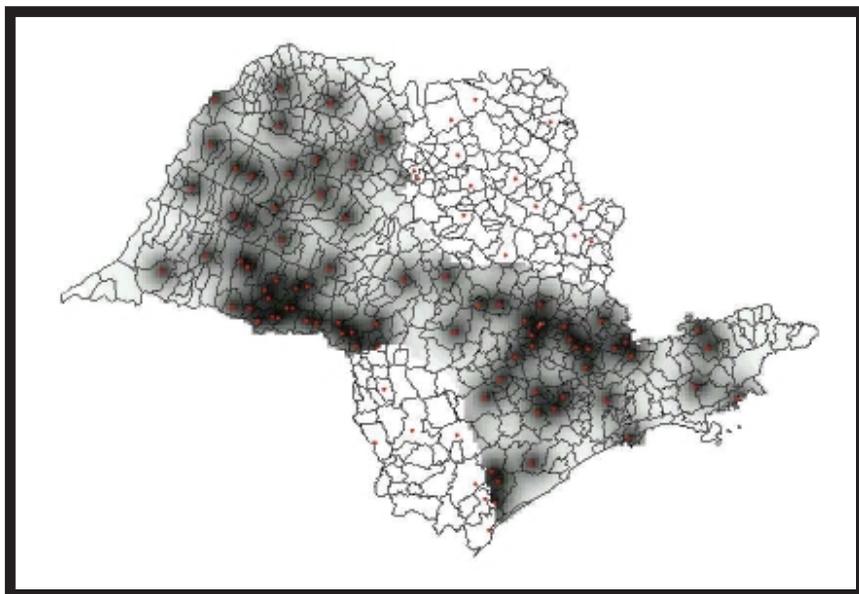


Fig. 13. Plano de Informação contendo variância da estimativa ou variância de krigeagem da temperatura máxima t_x , obtido através de krigeagem simples.

- 6.2. Krigeagem ordinária possibilita a inferência do atributo, numa posição u , sem a necessidade de se conhecer a média estacionária. Equivale a uma krigeagem simples com médias estimadas localmente a partir das amostras vizinhas. A substituição de uma única média estacionária por médias locais, ou tendências locais, explica a extrema robustez do algoritmo de krigeagem ordinária (Deutsch & Journel, 1998, citados por Felgueiras (1999) (Fig. 14 e 15).

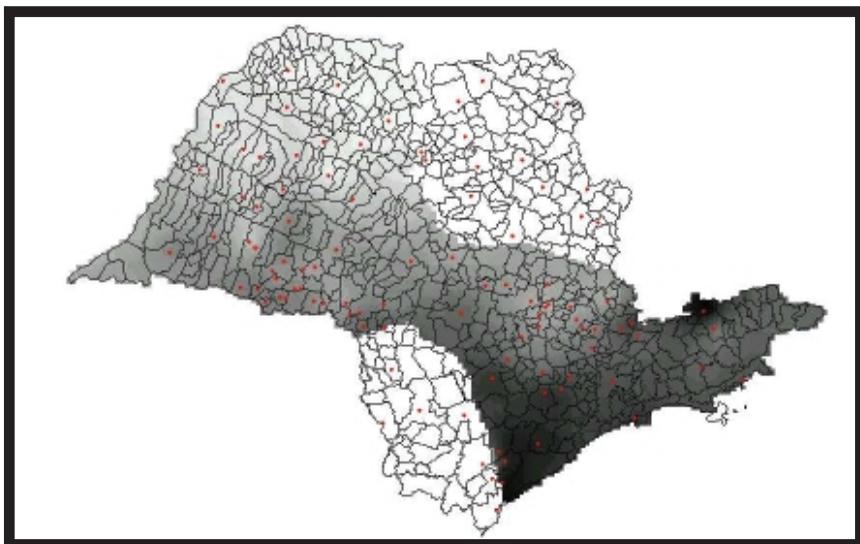


Fig. 14. Plano de Informação contendo a estimativa da variável temperatura máxima tx, obtido através de krigagem ordinária.

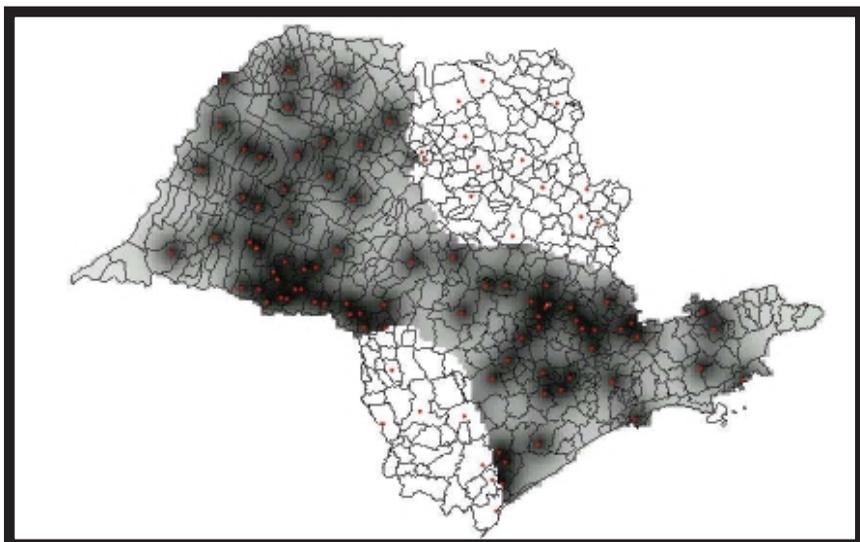


Fig. 15. Plano de Informação contendo a variância da estimativa ou variância de krigagem da temperatura máxima tx, obtido através de krigagem ordinária.

Considerações Finais

Os trabalhos aqui desenvolvidos, utilizando o módulo Análise do SPRING, tiveram o caráter puramente exploratório e foram, em grande parte, voltados para Análise Geográfica, utilizando-se algumas de suas ferramentas, em particular a geoestatística com o intuito de explorar esse paradigma de análise.

As experiências aqui adquiridas, serviram e servirão ainda mais para a motivação de novos estudos para melhoria dos processos de modelagem atualmente desenvolvidos.

O SPRING, dentro desse módulo, ainda carece de muita contribuição dos usuários para que ele possa ser aperfeiçoado. Atualmente existe um número muito restrito de usuários e estudiosos contribuindo para tal.

A Linguagem Espacial para Geoprocessamento Algébrico (LEGAL) e a análise para suporte a decisão, são estágios complementares que necessitam ser explorados para melhoria dos diversos processos para que o SPRING se propõe.

O processo de análise aqui desenvolvido, mostrou-se eficiente, prático e interativo, preenchendo os requisitos básicos para o enquadramento na categoria pretendida. Evidentemente que assuntos dessa natureza necessitam de exercício constante e contínuo para que sejam incorporadas todas suas rotinas.

Sugere-se a capacitação de um número maior de estudiosos no assunto e a adoção dessas práticas nos atuais processo de modelagem.

Referências Bibliográficas

BRAGA, L. P. V. **Geoestatística e aplicações**. São Paulo: ABE: IME: USP, 1990. 36 p. Trabalho apresentado no IX Simpósio Brasileiro de Probabilidade e Estatística, São Paulo, SP, 1990.

FELGUEIRAS, C. A. **Modelagem ambiental com tratamento de incertezas em sistemas de informação geográfica**: o paradigma geoestatístico por indicação. 1999. 212 f. Tese (Doutorado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos.

INPE. **Spring**: utilizando o Spring. Disponível em: <<http://www.dpi.inpe.br/spring/usuario/entrada.htm#acoes>>. Acesso em: 29 nov. 2002.

LANDIM, P. M. B. **Introdução a geoestatística**. Rio Claro: Unesp, 1988. 144 p. (Publicação Didática, 3).

Embrapa

Informática Agropecuária