# O uso de métodos de interpolação espacial de dados nas análises sociodemográficas[?]

**Alberto Augusto Eichman Jakob**[†]
**Andrea Ferraz Young**[‡]

**Palavras-Chave:** Dinâmica Demográfica; Geoestatística; Segregação Espacial; Interpolação Espacial

**Resumo**

A utilização de algum método de interpolação espacial de dados está se tornando cada vez mais freqüente nas análises sociodemográficas, em função de que, atualmente, diversos *softwares* já contêm vários destes métodos, permitindo análises bem mais detalhadas do que há algum tempo atrás. A interpolação de dados é importante para eliminar o chamado "efeito mosaico" ou "efeito xadrez" presentes em geral na visualização de mapas temáticos e para chamar a atenção para as principais concentrações espaciais de determinado atributo, suavizando suas diferenças. Mas como escolher qual é o método mais adequado aos nossos dados? Para tentar responder a esta pergunta, neste trabalho é feita inicialmente uma análise de componentes principais com as principais variáveis de estudo ao nível de setor censitário. Posteriormente, com os fatores resultantes da combinação linear destas variáveis é feita uma comparação entre os mais conhecidos métodos de interpolação espacial. Este estudo de caso foi realizado com as informações sociodemográficas para os municípios da Região Metropolitana da Baixada Santista (RMBS): Bertioga, Cubatão, Guarujá, Itanhaém, Mongaguá, Peruíbe, Praia Grande, Santos e São Vicente, uma região com perto de 1,5 milhão de pessoas segundo o Censo Demográfico de 2000. O resultado da aplicação destes métodos de interpolação para as variáveis sociodemográficas na região de estudo são mapas mostrando as principais concentrações espaciais da população segundo seu atributo, uma aproximação da segregação espacial da população segundo as variáveis analisadas.

---

[?] Trabalho apresentado no XV Encontro Nacional de Estudos Populacionais, ABEP, realizado em Caxambu – MG – Brasil, de 18 a 22 de setembro de 2006.

[†] Doutor em Demografia, Pesquisador Colaborador do Núcleo de Estudos de População/ UNICAMP e pós-doutorando da FAPESP no Centro de Estudos da Metrópole/ CEBRAP.

[‡] Doutora em Engenharia Agrícola, Pesquisadora Colaboradora do Núcleo de Estudos de População/ UNICAMP.

# O uso de métodos de interpolação espacial de dados nas análises sociodemográficas[?]

**Alberto Augusto Eichman Jakob**[†]
**Andrea Ferraz Young**[‡]

## Introdução

O termo "Geodemografia" vem sendo cada vez mais difundido na área acadêmica. Tem sido muito usada na área de marketing. Uma rápida pesquisa na internet resultou em 630 páginas com este termo e 14.200 com o termo "Geodemography"[1].

Segundo a enciclopédia Wikipedia, da internet, *"Geodemografia é o estudo das populações em seus ambientes. Ela une as ciências da demografia, o estudo da dinâmica populacional humana, com a geografia, o estudo da variação espacial e locacional dos fenômenos físicos e humanos na Terra."*[2]

A Wikipedia coloca também que a geodemografia tem sido usada para proporcionar serviços de consumo às populações "ideais" baseado em seu estilo de vida e lugar. Combinando bases de dados geográficas com características populacionais, os dados da geodemografia podem ser usados por "marketeiros" para influenciar pessoas com descontos, serviços ou informações políticas.

Segundo a Wikipedia, a Associação de Marketing Americana define a geodemografia como: *"Uma disponibilidade de dados de estilo de vida e comportamento demográfico do consumidor segundo limites geográficos arbitrários que são geralmente muito pequenos."*, e aponta que diversos *softwares* têm sido criados para analisar a informação obtida com a união entre dados populacionais e dados geográficos e proporcionar um resultado para as pessoas que desejam utilizar tal informação.

No que se refere à América Latina, a geodemografia tem sido muito utilizada especialmente na Espanha e no Chile. A AIMC (Asociación para la investigación de medios de comunicación), com sede na Espanha, denota que a geodemografia se trata de aplicar o "*Diga-me onde moras que eu te direi quem és e como consomes*", é a análise das pessoas de acordo com a localização de sua residência. Seu princípio básico é que duas pessoas que vivem no mesmo bairro se parecem entre si mais do que outras duas selecionadas ao acaso (Lamas, 1994). Este princípio básico é uma pequena variação do princípio da Principal Lei de Geografia de Tobler, que diz que unidades de análise mais próximas entre si são mais parecidas do que unidades mais afastadas (Tobler, 1970).

---

[?] Trabalho apresentado no XV Encontro Nacional de Estudos Populacionais, ABEP, realizado em Caxambu – MG – Brasil, de 18 a 22 de setembro de 2006.

[†] Doutor em Demografia, Pesquisador Colaborador do Núcleo de Estudos de População/ UNICAMP e pós-doutorando da FAPESP no Centro de Estudos da Metrópole/ CEBRAP.

[‡] Doutora em Engenharia Agrícola, Pesquisadora Colaboradora do Núcleo de Estudos de População/ UNICAMP.

[1] Pesquisa realizada com o Google (http://www.google.com.br) em 17 de março de 2006.

[2] Fonte: http://en.wikipedia.org/wiki/Geodemography <acesso em 17 de março de 2006>

Lamas (1994) assinala três dos principais componentes de um sistema de classificação geodemográfico:

- Uma infra-estrutura de unidades geográficas suficientemente pequenas, como os setores censitários;
- Uma base de dados de informação estatística sobre estas unidades geográficas, em geral o Censo ou outra fonte adicional; e
- A criação de um sistema de grupos de unidades geográficas resultante de uma Análise de Cluster, que se trata de agrupar as unidades espaciais que sejam mais próximas entre si e diferenciá-las de outras menos similares. Como normalmente o conjunto de variáveis é muito numeroso e com certo grau de correlação entre si, pode-se reduzir seu número por meio da aplicação de uma Análise de Componentes Principais ou outra técnica estatística de síntese similar, antes de fazer a Análise de Clusters.

No caso da área acadêmica, em especial nos estudos demográficos, se distanciando um pouco do enfoque do marketing, após a determinação dos grupos de unidades geográficas, que em geral são agrupamentos de setores censitários, podem ser feitas análises sociodemográficas destes grupos. Outras técnicas de estatística espacial podem ser também utilizadas, como os métodos de interpolação espacial, que eliminam os limites dos setores transformando dados no formato vetorial em dados no formato matricial ou raster.

A utilização de algum método de interpolação espacial de dados está se tornando cada vez mais freqüente nas análises sociodemográficas, em função de que, atualmente, diversos *softwares* já contêm vários destes métodos, permitindo análises bem mais detalhadas do que há algum tempo atrás. A interpolação de dados é importante para eliminar o chamado "efeito mosaico" ou "efeito xadrez" presentes em geral na visualização de mapas temáticos e para chamar a atenção para as principais concentrações espaciais de determinado atributo, suavizando suas diferenças.

Mas como escolher qual é o método mais adequado aos nossos dados? Para tentar responder a esta pergunta, neste trabalho é feita inicialmente uma análise de componentes principais com as variáveis de estudo mais utilizadas ao nível de setor censitário, para responsáveis por domicílios particulares permanentes e para as características destes domicílios. Posteriormente, com os fatores resultantes da combinação linear destas variáveis é feita uma comparação entre os mais conhecidos métodos de interpolação espacial.

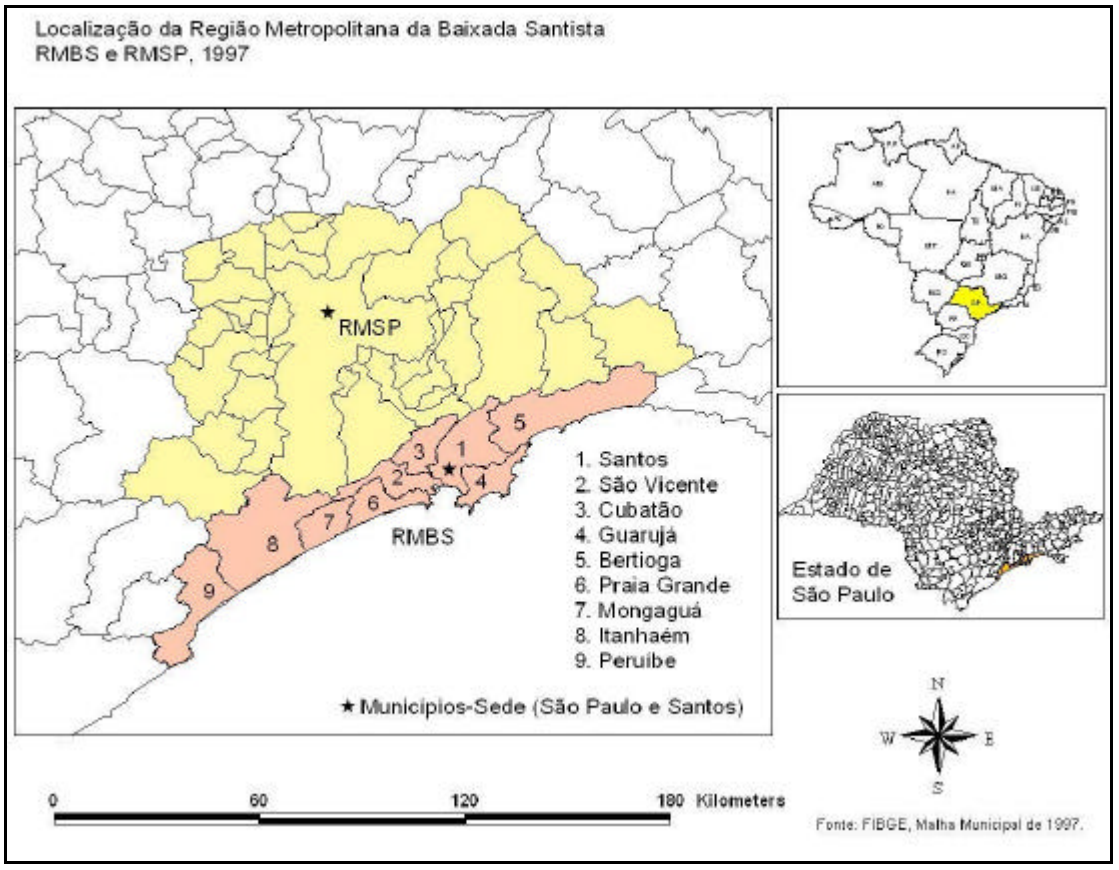
Este estudo de caso foi realizado com as informações sociodemográficas para os municípios da Região Metropolitana da Baixada Santista (RMBS): Bertioga, Cubatão, Guarujá, Itanhaém, Mongaguá, Peruíbe, Praia Grande, Santos e São Vicente, uma região com perto de 1,5 milhão de pessoas, e com quase 100% de sua população residindo em áreas urbanas, segundo o Censo Demográfico de 2000. O resultado da aplicação destes métodos de interpolação para as variáveis sociodemográficas na região de estudo são mapas mostrando as principais concentrações espaciais da população segundo seu atributo, uma aproximação da segregação espacial da população segundo as variáveis analisadas.

O próximo tópico mostra como foi criado o sistema de classificação geodemográfico com informações para os municípios da Região Metropolitana da Baixada Santista para o ano de 2000, o uso da análise de componentes principais, os agrupamentos, entre outras coisas.

## I. O Sistema de Classificação Geodemográfico para a Área de Estudo

A Figura 1 mostra a localização dos municípios pertencentes à Região Metropolitana da Baixada Santista (RMBS) em comparação com a Região Metropolitana de São Paulo e o resto do Estado de São Paulo.

**Figura 1**




Conforme dito anteriormente, Lamas (1994) sugeriu três dos mais importantes componentes do sistema de classificação geodemográfico: uma infra-estrutura de unidades espaciais suficientemente pequenas, que neste caso são os setores censitários dos municípios da RMBS em 2000; uma base de dados de informações estatísticas sobre estas unidades, o Censo Demográfico de 2000 neste caso, em especial o agregado de setores censitários disponibilizado pela Fundação IBGE; e a criação de um sistema de grupos das unidades espaciais, a partir das análises de clusters e de componentes principais, por exemplo.

Neste trabalho, consideraram-se as variáveis mais utilizadas em estudos para os responsáveis ou chefes de domicílios particulares permanentes, e as características mais usadas destes domicílios, em relação aos setores censitários, sendo estas:

- Idade média do responsável pelo domicílio (idmédia);
- Anos médios de estudo do responsável pelo domicílio (estmed);
- Renda média mensal (em Salários Mínimos) do responsável pelo domicílio (renmedsm);
- Porcentagem de domicílios sem rede geral de água (semágua);

- Porcentagem de domicílios sem rede geral de esgoto (semesg);
- Porcentagem de domicílios sem coleta de lixo (semlixo);
- Porcentagem de domicílios sem banheiro (sembanh).

Uma análise fatorial por componentes principais foi então realizada para reduzir o número de variáveis ou atributos, por meio do *software* estatístico SPSS (Statistical Package for Social Sciences). A Tabela 1 traz a correlação estatística entre estas variáveis.

**Tabela 1**
**Correlação estatística entre as variáveis selecionadas**
**Setores censitários dos municípios da RMBS, 2000**

| Variáveis | Idade Média do Chefe | Renda Média (em SM) | Anos Médios de Estudo | Sem Rede Geral de Esgoto | Sem Rede Geral de Água | Sem Lixo Coletado | Sem Banheiro |
|---|---|---|---|---|---|---|---|
| Idade Média do Chefe | 1,0000 | 0,5793 | 0,6763 | -0,5191 | -0,2264 | -0,1633 | -0,1661 |
| Renda Média (em SM) | 0,5793 | 1,0000 | 0,8701 | -0,3616 | -0,2023 | -0,1674 | -0,1411 |
| Anos Médios de Estudo | 0,6763 | 0,8701 | 1,0000 | -0,4891 | -0,2905 | -0,2378 | -0,2002 |
| Sem Rede Geral de Esgoto | -0,5191 | -0,3616 | -0,4891 | 1,0000 | 0,3028 | 0,2717 | 0,1546 |
| Sem Rede Geral de Água | -0,2264 | -0,2023 | -0,2905 | 0,3028 | 1,0000 | 0,5220 | 0,3037 |
| Sem Lixo Coletado | -0,1633 | -0,1674 | -0,2378 | 0,2717 | 0,5220 | 1,0000 | 0,4012 |
| Sem Banheiro | -0,1661 | -0,1411 | -0,2002 | 0,1546 | 0,3037 | 0,4012 | 1,0000 |

Fonte: FIBGE, Censo Demográfico de 2000. Análises estatísticas NEPO/UNICAMP

A Tabela 1 mostra que as correlações mais altas estavam em 2000 entre as variáveis de “renda média” e “anos médios de estudo” dos chefes de domicílio (0,87) e entre a “idade média” e os “anos médios de estudo” dos chefes (0,68). Assim, poderia se pensar, somente por esta análise, em retirar a variável “anos médios de estudo”, porque não se perderia muito dada sua alta correlação com outras variáveis. Mas vamos continuar com este conjunto de variáveis para ver como se comporta a análise fatorial.

A Tabela 2 traz a composição dos dois fatores extraídos da análise fatorial, assim como a variância dos dados explicada com os fatores.

**Tabela 2**
**Composição dos fatores extraídos dos dados**
**Setores censitários dos municípios da RMBS, 2000**

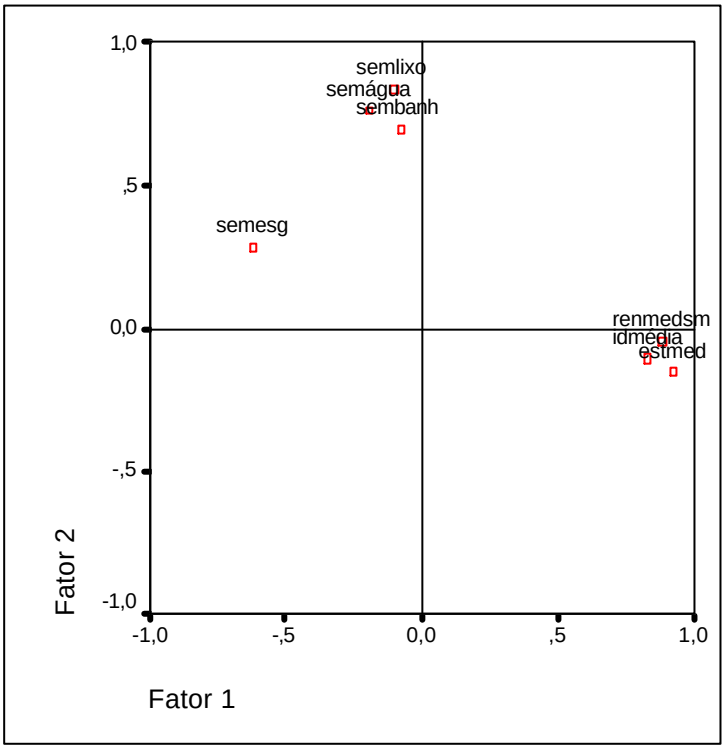
| Variáveis | Fator 1 | Fator 2 |
|---|---|---|
| Idade Média do Chefe | 0,8292 | -0,1040 |
| Renda Média (em SM) | 0,8797 | -0,0452 |
| Anos Médios de Estudo | 0,9225 | -0,1495 |
| Sem Rede Geral de Esgoto | -0,6209 | 0,2874 |
| Sem Rede Geral de Água | -0,1954 | 0,7605 |
| Sem Lixo Coletado | -0,1076 | 0,8370 |
| Sem Banheiro | -0,0774 | 0,6931 |
| % Variância Explicada | 39,34 | 66,16 |

Fonte: FIBGE, Censo Demográfico de 2000. Análises estatísticas NEPO/UNICAMP

A Tabela 2 mostra que o fator 1 pode ser considerado as características do chefe do domicílio, dado que o maior peso de sua composição foi dado às variáveis de atributos dos chefes, e o fator 2 pode então ser considerado com características de infra-estrutura do domicílio. O único senão é a variável "sem rede de esgoto", que por esta análise deve entrar no fator 1, mas com sinal contrário, ou seja, na medida em que melhoram as características do chefe, diminuem os valores de "sem rede de esgoto", melhorando também então as características de escoamento sanitário do domicílio, o que se era de esperar.

Esta indefinição do esgoto ocorre em vista de que sua maior correlação ocorre com a idade média e os anos médios de estudo dos chefes (Tabela 1). O Gráfico 1 mostra claramente esta condição do esgoto, que parece não se enquadrar em nenhum dos grupos formados.

**Gráfico 1**
**Variáveis selecionadas com relação aos dois fatores extraídos**
**Setores censitários dos municípios da RMBS, 2000**




Fonte: FIBGE, Censo Demográfico de 2000. Análises estatísticas NEPO/UNICAMP

É interessante destacar também que o primeiro fator sozinho explica 39,3% da variância dos dados, enquanto os dois fatores extraídos explicam 66,2% da variância (Tabela 2), o que não é ruim, uma vez que são dados bem distintos, em valores absolutos e em porcentagens, se referindo a características de pessoas e de domicílios, por exemplo.

Foi feita então uma combinação linear entre os valores observados (informações censitárias) e a composição dos fatores, resultando em duas novas variáveis (Fator 1 e Fator 2). Estas variáveis foram então padronizadas para facilitar sua análise, com a utilização da seguinte fórmula:

$$\frac{(Observação - Mínimo)}{(Máximo - Mínimo)}$$

Com esta padronização, os valores da combinação linear dos fatores ficaram entre 0 e 1.

Estas duas variáveis novas (Fator 1 e Fator 2) foram então selecionadas para se fazer a comparação entre os métodos de interpolação espacial, o que é feito no tópico a seguir.

## II. Os Métodos de Interpolação Espacial

Os avanços computacionais e o aprimoramento nas técnicas de mapeamento que temos vivenciado têm nos permitido uma avaliação cada vez mais precisa da qualidade dos atributos mapeados, assim como detectar os erros a eles associados, causados ao se determinar o modelo de representação espacial a ser utilizado, por exemplo, nas interpolações de dados. Com isto, surgiu a necessidade de se implantar, nos atuais Sistemas de Informação Geográfica (SIGs), formas mais sofisticadas de análise das informações espaciais, assim como a incorporação de procedimentos que permitam uma avaliação da confiabilidade e segurança dos resultados obtidos. No caso dos métodos de interpolação, a avaliação dos erros associados aos atributos mapeados seria um exemplo disto.

Lourenço (1998) aponta que os mapas de isovalores, que mostram a variabilidade dos dados, são resultados cada vez mais comuns que se espera dos SIGs, assim como as estimativas dos dados de pontos não amostrados, por meio de valores em pontos amostrados. O autor coloca também que nestas duas situações, os problemas de interpolação surgem, tornando necessário o uso de metodologias específicas, e as soluções deveriam vir com os erros associados às estimativas.

A interpolação é uma técnica utilizada para a estimativa do valor de um atributo em locais não amostrados, a partir de pontos amostrados na mesma área ou região. A interpolação espacial converte dados de observações pontuais em campos contínuos, produzindo padrões espaciais que podem ser comparados com outras entidades espaciais contínuas. O raciocínio que está na base da interpolação é que, em média, os valores do atributo tendem a ser similares em locais mais próximos do que em locais mais afastados. Esse conceito também fundamenta a base das relações espaciais entre fenômenos geográficos, utilizando a correlação espacial como meio de diferença dos atributos estimados (Câmara e Medeiros, 1998).

Os métodos de interpolação mais comuns dos SIGs em geral pertencem a duas categorias: globais e locais, sendo os globais mais utilizados em superfícies de tendência, e os locais podem ser polinômios de baixa ordem, funções *splin*e, poliedros, triangulação e médias móveis ponderadas. Porém, estes métodos não fornecem os erros associados às estimativas. Somente o método da krigagem o faz por meio de um "modelo contínuo de variação espacial". Os modelos de atributos quantitativos indeterminados no campo baseiam-se nas médias e nos erros médios aleatórios. As médias são armazenadas nos mapas criados e estes erros definidos por meio do cálculo dos desvios-padrão. Mas geralmente não é suficiente apenas o desvio-padrão para a identificação dos erros, mas também a inclusão das autocorrelações, as correlações espaciais dos atributos (Lourenço, 1998). Por isso a técnica da krigagem se torna importante.

Existe um conjunto particular de métodos determinísticos que não pretendem caracterizar completamente um fenômeno físico através do conjunto de fatores que estão na sua origem, mas têm simplesmente como objetivo a interpolação espacial dos valores observados. Trata-se de um dos problemas básicos da análise espacial, que a geoestatística propõe resolver através de uma metodologia probabilístico-estocástica. Por essa razão, nesse

estudo comparativo, são apresentados, resumidamente, métodos que tiveram e, em alguns casos, continuam a ter uma grande aplicação na cartografia de fenômenos espaciais, suas vantagens e desvantagens. Todos os métodos aqui apresentados – Ponderação do Inverso das Distâncias (IDW), Polinomial Global, Polinomial Local, Funções de Base Radial, Krigagem, Co-Krigagem, calculam um valor de uma dada grandeza no espaço entre as amostras ou observações a partir de uma combinação linear dos valores observados[3].

## 2.1. A Ponderação do Inverso das Distâncias (IDW)

A Ponderação do Inverso das Distâncias (*Inverse Distance Weighting*) implementa explicitamente o pressuposto de que as coisas mais próximas entre si são mais parecidas do que as mais distantes. Para predizer um valor para algum local não medido, o IDW usará os valores amostrados à sua volta, que terão um maior peso do que os valores mais distantes, ou seja, cada ponto possui uma influência no novo ponto, que diminui na medida em que a distância aumenta, daí seu nome.

Sua fórmula de cálculo é: $\hat{Z}(s_0) = \sum_{i=1}^{N} \lambda_i Z(s_i)$ ; onde:

$\hat{Z}(s_0)$ é o valor a ser predito para o local $s_0$ ;

N é o número de pontos observados a serem usados ao redor do valor a ser predito;

$\lambda_i$ são os pesos colocados para cada ponto observado a ser utilizado;

$Z(s_i)$ é o valor observado no local $s_i$ .

A fórmula para determinar os pesos é a seguinte: $\lambda_i = \frac{d_{i0}^{-p}}{\sum_{i=1}^{N} d_{i0}^{-p}}$ ; sendo $\sum_{i=1}^{N} \lambda_i = 1$

Na medida em que a distância aumenta, o peso é reduzido por um fator de “p”.

$d_{i0}$ é a distância entre o local predito, $s_0$, e cada um dos locais observados, $s_i$ .

Os pesos dos locais observados, a serem usados na predição, são ponderados, e sua soma é igual a 1.

O valor “p” é determinado minimizando o erro médio quadrático da predição (RMSPE), que é a estatística calculada por um procedimento de validação cruzada (*cross-validation*). Na validação cruzada, cada ponto observado é removido e comparado com o predito para aquele local. O RMSPE é a estatística resumo do erro desta superfície de predição. Pode-se tentar diferentes valores de “p” para identificar o que produz o menor RMSPE.

Existem dois componentes direcionais que podem afetar as predições na superfície de saída: as tendências globais e as influências direcionais (anisotropia). As tendências globais são, por exemplo, efeitos de ventos, poluição ou água escorrendo morro abaixo, etc. que podem ser retirados do modelo com processos de remoção de tendências (*detrending*). A anisotropia difere da tendência global em função de que esta última pode ser descrita como

[3] Os métodos apresentados nos próximos tópicos foram baseados principalmente em ESRI (2001).

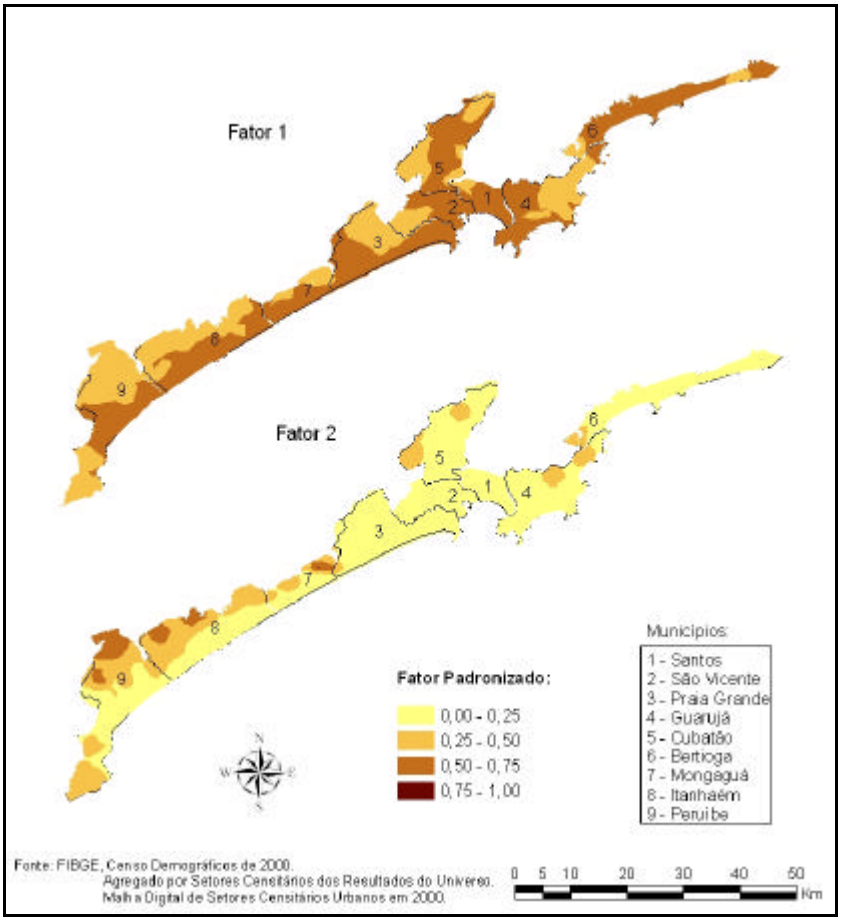
um processo físico, e modelada por fórmulas matemáticas. Já a causa da anisotropia não é conhecida, e é modelada como um erro aleatório, e não um processo determinístico que pode ser descrito com uma equação matemática. Ela pode ser tratada com a krigagem.

O IDW é um método interpolador que é exato. Poucas decisões são tomadas acerca dos parâmetros do modelo. Este método pode ser adequado para uma visualização ou interpretação preliminar da interpolação de uma superfície. Entretanto, não é realizada uma avaliação da predição de erros, que pode produzir um efeito "*bulls eyes*" ao redor da localização do dado, pequenas áreas que se diferenciam da suavização geral da variável.

Este método assume que a superfície possui uma variação local, e funciona melhor se os pontos amostrais estão igualmente distribuídos pela área, sem estarem concentrados em determinado local. Os parâmetros mais importantes a se detectar, então, são as especificações de vizinhança, o parâmetro de poder (*power*) "p" e o fator de anisotropia, se existir.

A Figura 2 mostra as variáveis Fator 1 e Fator 2 interpoladas com o IDW.

**Figura 2**
**Interpolação dos fatores com o método IDW**
**Setores censitários urbanos dos municípios da RMBS, 2000**



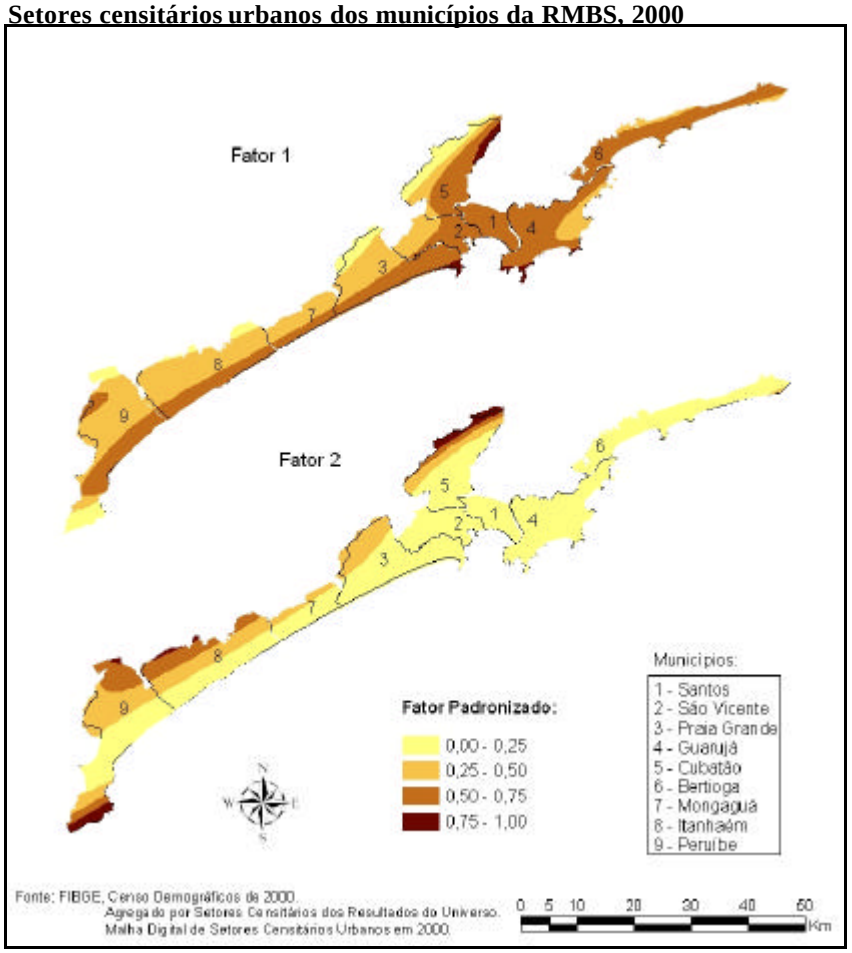

A Figura 2, assim como as demais figuras representando mapas, traz os mapas de setores censitários urbanos dos municípios pertencentes à RMBS, que concentravam em 2000 quase 100% da população residente na região. As categorias de legenda são os quartis dos fatores padronizados, interpolados segundo os diferentes métodos. Deve-se lembrar que o Fator 1 corresponde às características dos chefes dos domicílios e o Fator 2 às características da infra-estrutura domiciliar. As cores mais escuras para o Fator 1 representam os locais de

concentração de chefes mais idosos, com melhor renda e escolaridade (em geral as áreas próximas à orla marítima), e no Fator 2 representam as áreas com pior infra-estrutura domiciliar (áreas situadas mais no interior, distantes da orla marítima). Pode-se observar como serão os mapas resultantes com as interpolações realizadas com os outros métodos, o que é feito nos próximos tópicos.

## 2.2. A Interpolação Polinomial Global

A interpolação polinomial global ajusta uma superfície suavizada definida por uma função matemática (polinomial) aos pontos observados. Esta superfície gradualmente muda e captura o padrão de escala dos dados. Seria como ajustar um plano aos pontos observados, que pode ser linear (função polinomial de primeira ordem), de segunda ordem (quadrática), de terceira ordem (cúbica), até a décima ordem. O resultado é uma superfície matemática suavizada que representa as tendências graduais da superfície da área de interesse (Figura 3).

**Figura 3**
**Interpolação dos fatores com o método Polinomial Global**
**Setores censitários urbanos dos municípios da RMBS, 2000**




A Figura 3 mostra certa semelhança com a Figura 2, embora a Figura 3 traga uma maior suavização dos dados. Com o polinomial global pode-se escolher o valor de "p" que representa uma distribuição mais global (p=1) até a mais local (p=10). Neste caso foi

utilizado o valor de p=5 para se ter uma idéia geral. Mas deve-se lembrar de que o ideal é se obter o valor de "p" a partir do erro médio resultante, conforme apontado anteriormente.

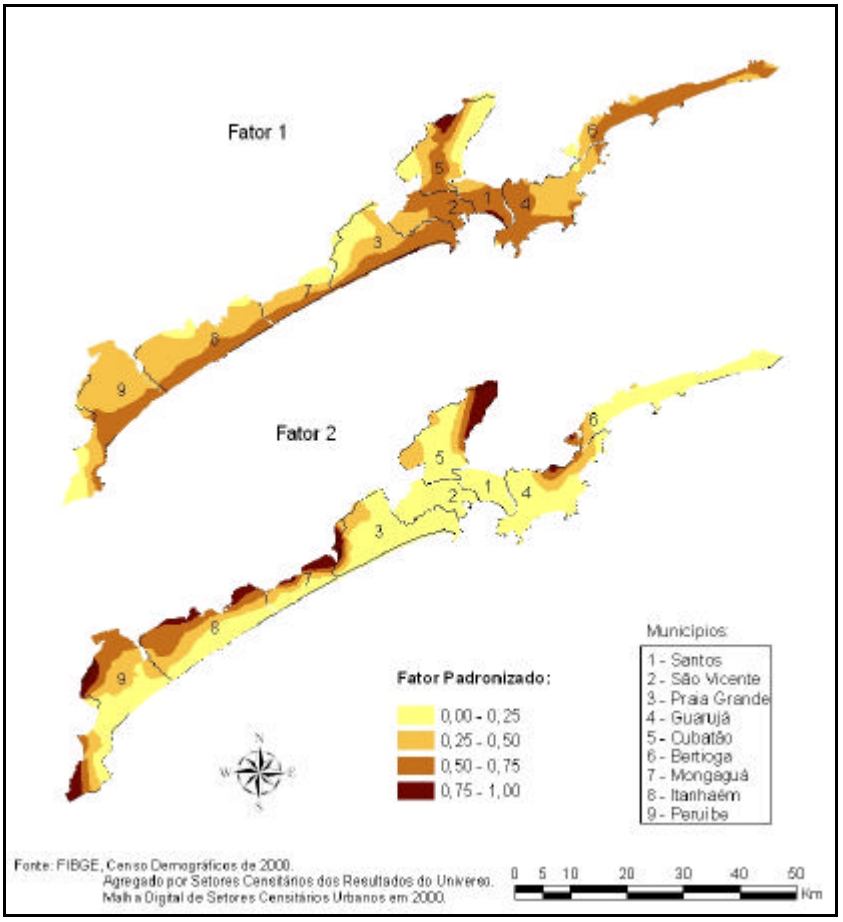
O método Polinomial Global é um interpolador determinístico rápido que não é exato (não apresenta certo grau exatidão). Pouquíssimas decisões são tomadas acerca dos parâmetros do modelo. É mais adequado para superfícies que apresentem mudanças graduais, ou seja, transformações gradativas. Este método não apresenta uma avaliação da predição de erros, podendo causar uma suavização geral da informação. A localização da média dos dados pode ter um amplo efeito (influência) sobre a superfície. Não é requerida uma hipótese inicial para os dados.

Este método também é utilizado para examinar e/ou remover os efeitos de tendências globais de longa duração, criando superfícies que descrevem processos físicos. Entretanto, quanto maior a ordem da função polinomial, maior a complexidade para a explicação. Estas superfícies também são muito suscetíveis aos pontos extremos (*outliers*).

## 2.3. A Interpolação Polinomial Local

A interpolação polinomial global ajusta um polinômio à superfície toda. A polinomial local pode ajustar muitos polinômios, cada um especificando sua vizinhança. Podem ser especificados a forma, o máximo e o mínimo número de pontos a serem usados, e a configuração do setor, assim como uma função de diminuição de pesos a partir da distância dos pontos, em conjunto com o poder do teste "p". A Figura 4 mostra esta interpolação.

**Figura 4: Interpolação dos fatores com o método Polinomial Local**
**Setores censitários urbanos dos municípios da RMBS, 2000**



O Polinomial Local é um método de interpolação determinístico moderadamente rápido que também não apresenta exatidão, pois suaviza os dados de maneira geral. Este método é mais flexível que o método Polinomial Global, entretanto, mais parâmetros de decisão são necessários. Neste caso também não é realizada uma avaliação da predição de erros. O método fornece apenas uma predição de superfícies que é comparável ao método de krigagem com medida de erros. Este método não permite que seja investigada a autocorrelação entre dados, tornando-o menos flexível e mais automático que o método de krigagem. Assim como no método polinomial global, hipóteses iniciais também não são requeridas para os dados.

Como este método produz superfícies com muito mais variação, podem ser utilizados polinomiais de diversas ordens (um único plano), ou então múltiplos planos polinomiais. A Figura 4 dá uma idéia desta variação maior das superfícies, em comparação com a Figura 3.

A interpolação polinomial global é boa para a identificação de grandes ou longas tendências nos dados. Entretanto, quando a variável apresenta uma curta variação da tendência, a interpolação polinomial local é a mais indicada. Este método é sensível à distância de vizinhança. Por isso, pode-se visualizar a superfície antes de ser criada. Como no IDW, pode-se também definir um modelo que leve em consideração a anisotropia.

## 2.4. Funções de Base Radial

As Funções de Base Radial são interpoladores determinísticos moderadamente rápidos e exatos. São mais flexíveis que o IDW, mas exigem mais parâmetros de decisão. Não é realizada uma avaliação da predição de erros. No entanto, fornece uma predição de superfícies que é comparável a forma exata do método de krigagem. Este método também não permite que seja investigada a autocorrelação entre dados, sendo do mesmo modo menos flexível e mais automático que o método de krigagem. Como nos métodos anteriores, hipóteses iniciais também não são requeridas para os dados.

Existem cinco funções básicas distintas: *thin-plate spline* (spline suave), *spline with tension* (spline com tensão), *completely regularized spline* (spline completamente regular), *multiquadratic function* (função multiquadrática) e *inverse multiquadratic spline* (spline multiquadrática inversa). Cada uma destas funções tem uma diferente forma e resulta em uma distinta superfície de interpolação. Estes métodos são um tipo de redes neurais artificiais.

Estas funções são semelhantes a ajustar uma membrana de borracha aos valores observados, minimizando a curvatura total da curva. Sendo interpoladores exatos, estes métodos diferem dos interpoladores globais e locais, que são inexatos e não requerem que a superfície passe sobre os pontos observados. Comparando as funções de base radial com o IDW, outro interpolador exato, o IDW nunca fará a predição de valores acima do valor máximo observado ou abaixo do valor mínimo observado, mas estas funções sim. O parâmetro otimizado é determinado usando variação cruzada de maneira parecida com o IDW e os interpoladores polinomiais locais.

As funções de base radial são usadas para se calcular superfícies suavizadas de um grande número de pontos. As funções produzem bons resultados para superfícies de pouca variação, como de elevação de terreno.

Estas técnicas são inapropriadas quando existem muitas mudanças nos valores em pouca distância ou quando existe a suspeita de que os dados amostrados estão propensos ao erro ou incerteza.

A Figura 5 mostra a interpolação por meio das funções de base radial, criada a partir da função spline completamente regularizada, e as figuras 6 a 10 mostram uma visualização inicial das superfícies a serem criadas com as funções disponíveis de base radial do Fator 1.

**Figura 5**
**Interpolação dos fatores com as Funções de Base Radial**
**Setores censitários urbanos dos municípios da RMBS, 2000**

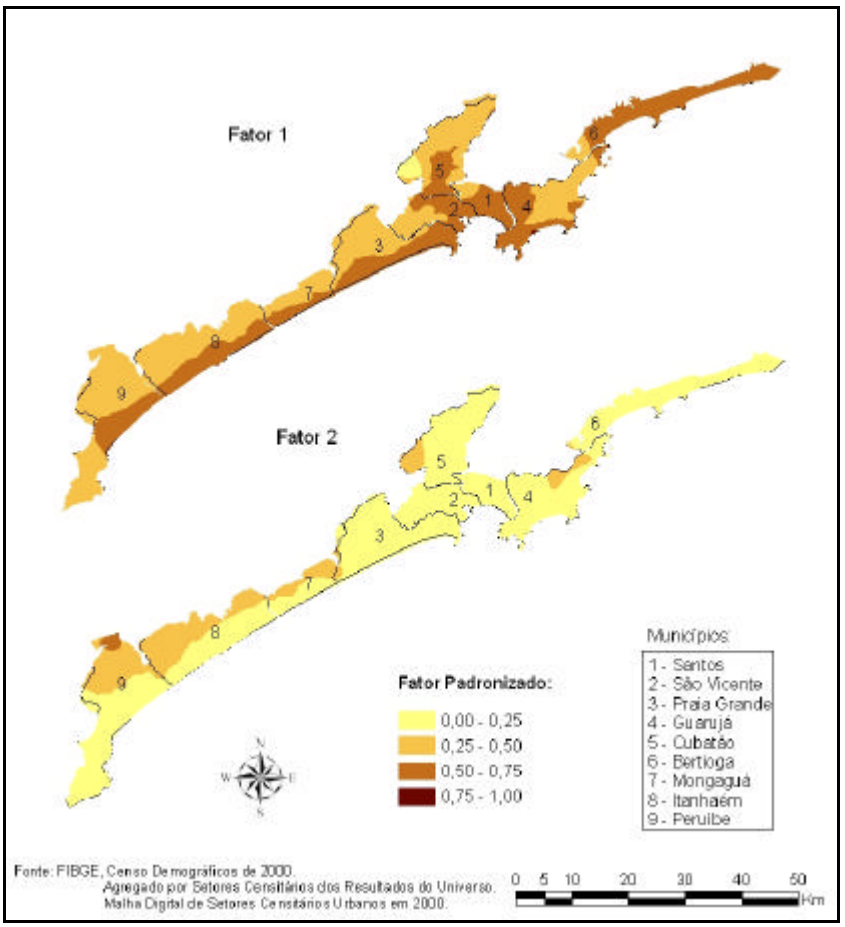



**Figura 6: Spline Regularizada**

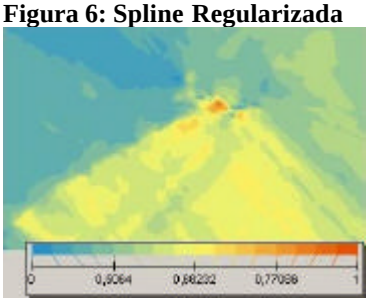


**Figura 7: Spline com Tensão**

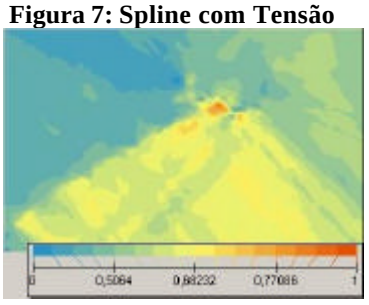


**Figura 8: Spline Suave**

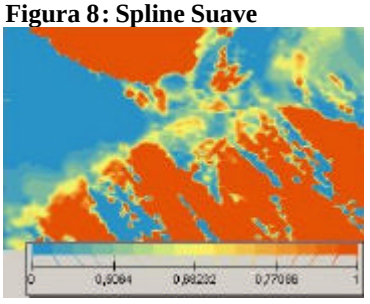


**Figura 9: Multiquádrico**

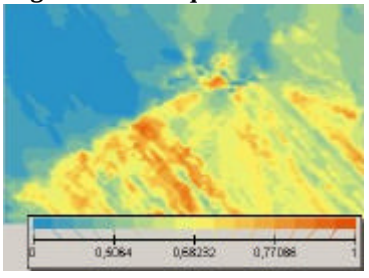
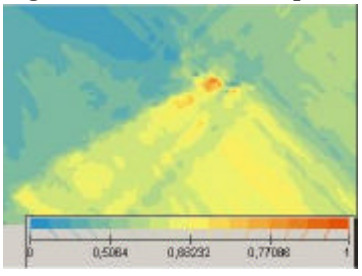


**Figura 10: Inverso Multiquádrico**



## 2.5. A Krigagem

Os métodos de krigagem dependem de modelos matemáticos e estatísticos, assim como da noção de autocorrelação. Na estatística clássica, assume-se que as observações são independentes, ou seja, não há correlação entre as observações. Na geoestatística, a informação dos locais espaciais permite o cálculo das distâncias entre as observações e modelar a autocorrelação como uma função da distância. Para isto, a função mais comum utilizada é o (semi)variograma.

O variograma é a descrição matemática do relacionamento entre a variância de pares de observações (pontos) e a distância separando estas observações (h). A autocorrelação espacial pode então ser usada para fazer melhores estimativas para pontos não amostrados (inferência = krigagem). A krigagem se baseia na idéia de que se pode fazer inferências a partir de uma função aleatória Z(x), originando os pontos $Z(x_1)$, $Z(x_2)$, ..., $Z(x_n)$.

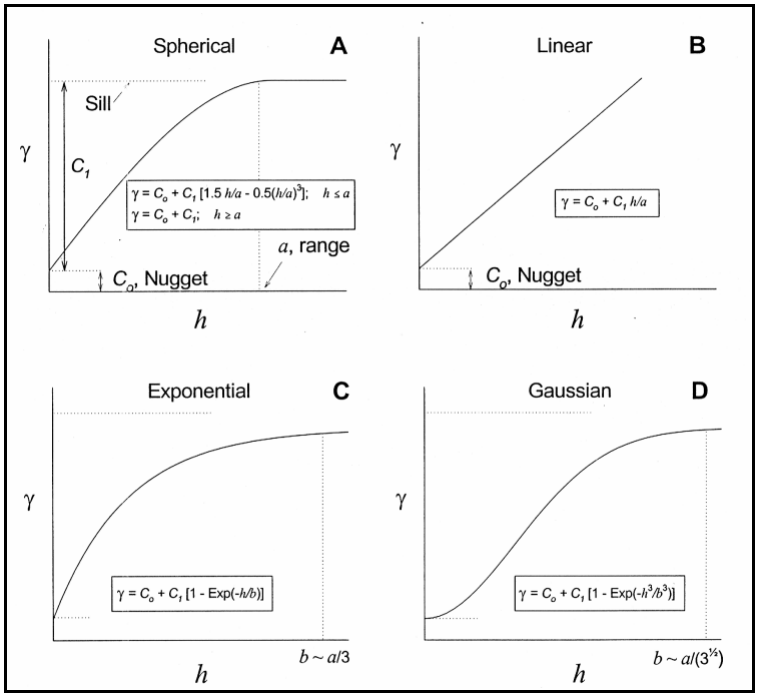
**(Equação 1):** A função Z(x) = m(x) + ?(h) + e apresenta a média, em geral constante, a correlação espacial e o erro residual. A correlação espacial é dada pelo variograma:

?(h) = ½ var [ Z(x) – Z(x+h) ] = ½ E [ $\{Z(x) – Z(x+h)\}^2$ ]; na prática:

?(h) = ½ N(h) $S_i[Z(x_i) – Z(x_i+h)]^2$, onde N(h) é o número total de pares de observações separadas pela distância h. E a curva ajustada minimiza a variância dos erros.

A Figura 11 mostra os componentes do variograma, e seus principais modelos. Dentre estes, os mais comuns são o esférico e o exponencial. O efeito pepita (nugget) é o ponto inicial da curva, onde a curva toca o eixo ?, quando h = 0. O patamar (sill) é o valor de ? máximo da curva, o ponto em que não existe mais nenhuma correlação entre as variáveis, sendo assim a variância do conjunto de dados. O alcance (range) é o ponto máximo onde existe autocorrelação espacial das variáveis.

**Figura 11: Componentes e Modelos do Variograma**




Fonte: Jakob (2003a)

É importante definir o patamar da curva para se analisar seu alcance. O alcance é uma medida importante de se obter, pois proporciona a autocorrelação espacial da variável em unidades conhecidas de distância, como metros. E se pode fazer um Proxy desta autocorrelação espacial como sendo entendida por segregação espacial. Com isto, se torna possível encontrar um valor aproximado (e mensurável) da segregação espacial de determinada variável de estudo. Já o efeito pepita traduz o quanto pequenas distâncias são parecidas ou diferentes. Um valor alto deste índice indica que se encontram grandes variações em curtas distâncias.

Outros fatores a serem estudados são a anisotropia e a linha de tendência. Existem funções específicas para o tratamento destes fatores. Pode-se dizer que a krigagem produz a melhor estimativa linear não-viciada dos dados de um atributo em um local não amostrado, com a modelagem do variograma.

A Krigagem usada para a predição não requer que os dados tenham distribuição normal. Entretanto, a normalidade é necessária para se obter mapas de quantis e de probabilidade na krigagem ordinária, simples e universal. Considerando apenas a predição criada por médias ponderadas, a krigagem é considerada o melhor estimador não viciado. No caso de uma distribuição normal dos dados, é o melhor estimador entre todos os estimadores não viciados, não apenas aqueles com médias ponderadas. Ela depende também do pressuposto de que os erros aleatórios são estacionários de segunda ordem, ou seja, têm média zero e a covariância entre dois erros aleatórios depende apenas de sua distância e direção que os separa, não de sua posição. Funções de transformação e remoção de tendências podem auxiliar quanto a estes pressupostos de normalidade e estacionaridade.

Existem diversos tipos de krigagem. As mais comuns são a krigagem simples, a ordinária, a universal, de indicadores, de probabilidade e a disjuntiva. Na krigagem de indicadores, a função aleatória resultante é uma variável binária, os dados observados seriam, por exemplo “0” ou “1”, o que não é o caso aqui. A krigagem de probabilidade produz um mapa de probabilidades ou erros-padrão de indicadores. Ela tenta fazer o mesmo que os indicadores, mas utiliza também a validação cruzada para melhorar os resultados, e requer uma normalidade dos dados, o que também não é o caso. Já a krigagem disjuntiva assume o modelo com uma função arbitrária da função Z(x), e com este método pode-se prever tanto o valor quanto o indicador, uma vez que a krigagem de indicadores é um caso especial da krigagem disjuntiva. Mas esta requer uma normalidade bivariada dos dados e aproximações para as funções arbitrárias, com pressupostos difíceis de se verificar e soluções complicadas ao nível matemático e computacional. Portanto, se concentrará aqui nas krigagens simples, ordinária e universal (figuras 12 a 14).

A Krigagem ordinária assume o modelo da Equação 1, onde a média m(x) é constante e desconhecida. Pode ser usada com dados que parece terem tendência, pois possui tratamentos específicos para retirar a tendência e anisotropia. Com a Krigagem Simples, a média m(x) deve ser conhecida e constante. E como m(x) é conhecida, pode-se conhecer o e também, o que melhora as estimativas. Mas geralmente é difícil se conhecer a média geral, em função de uma possível tendência nos dados. O método utiliza a krigagem nos resíduos, a diferença entre os valores preditos e os observados, assumindo que a tendência nos resíduos é zero. Já na krigagem universal, a média m(x) é uma função determinística, e portanto, não constante. É utilizada para dados que apresentam tendência , e uma função polinomial pode representar a média (e esta tendência). Comparando-se tal função com os dados observados se encontram os erros, assumidos como sendo aleatórios. A média destes é zero, e a autocorrelação é modelada destes erros aleatórios.

**Figura 12: Interpolação dos fatores com a krigagem ordinária**
**Setores censitários urbanos dos municípios da RMBS, 2000**

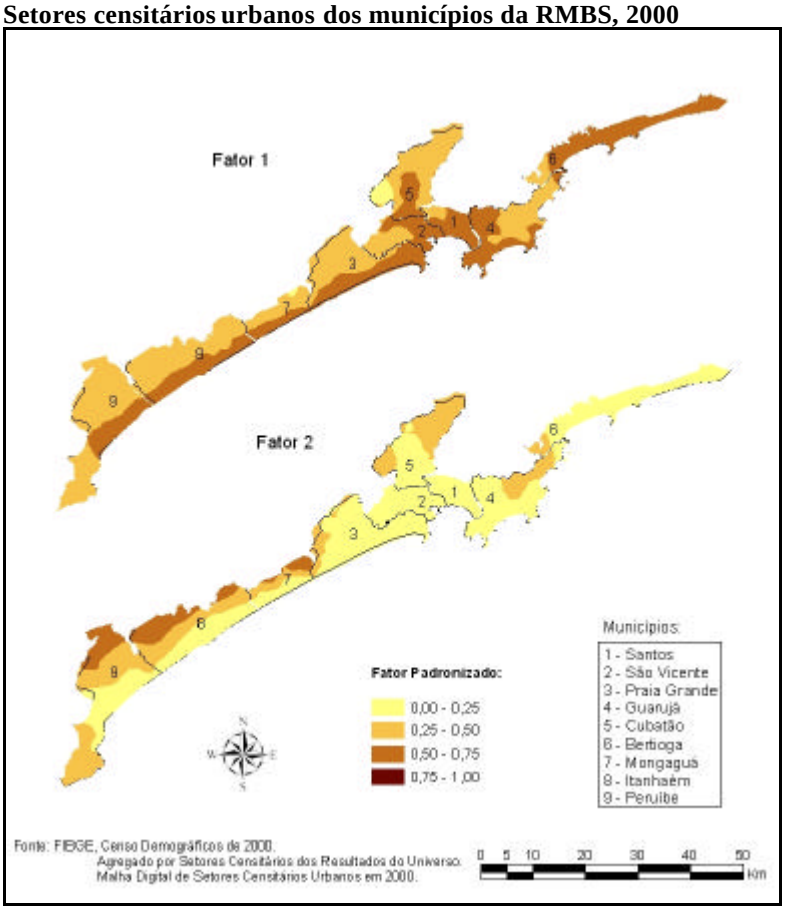



**Figura 13: Interpolação dos fatores com a krigagem simples**
**Setores censitários urbanos dos municípios da RMBS, 2000**

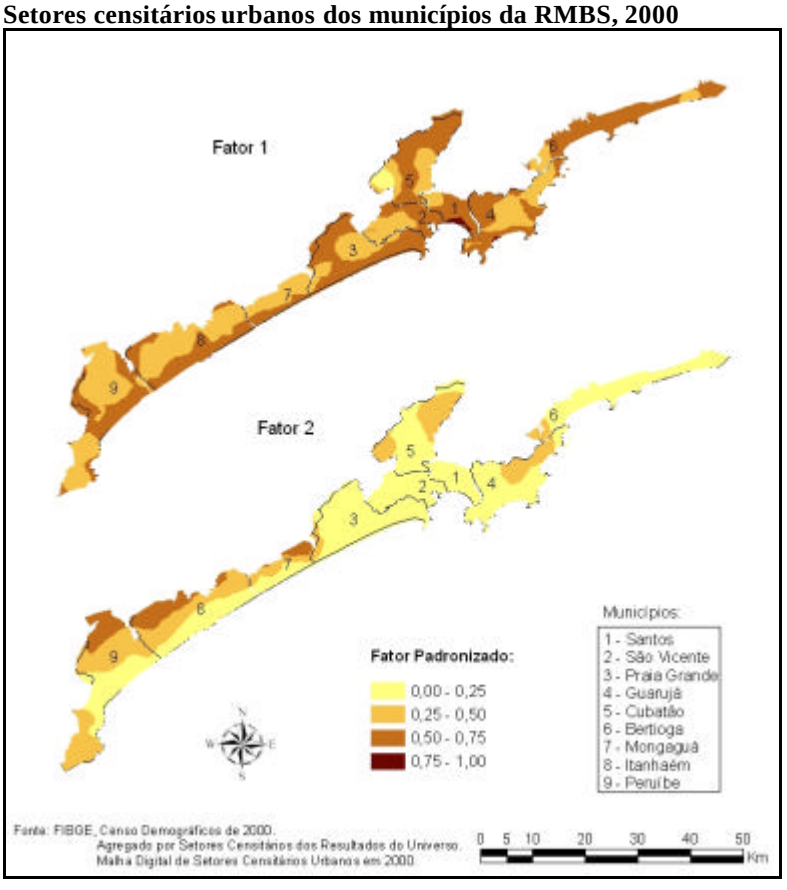
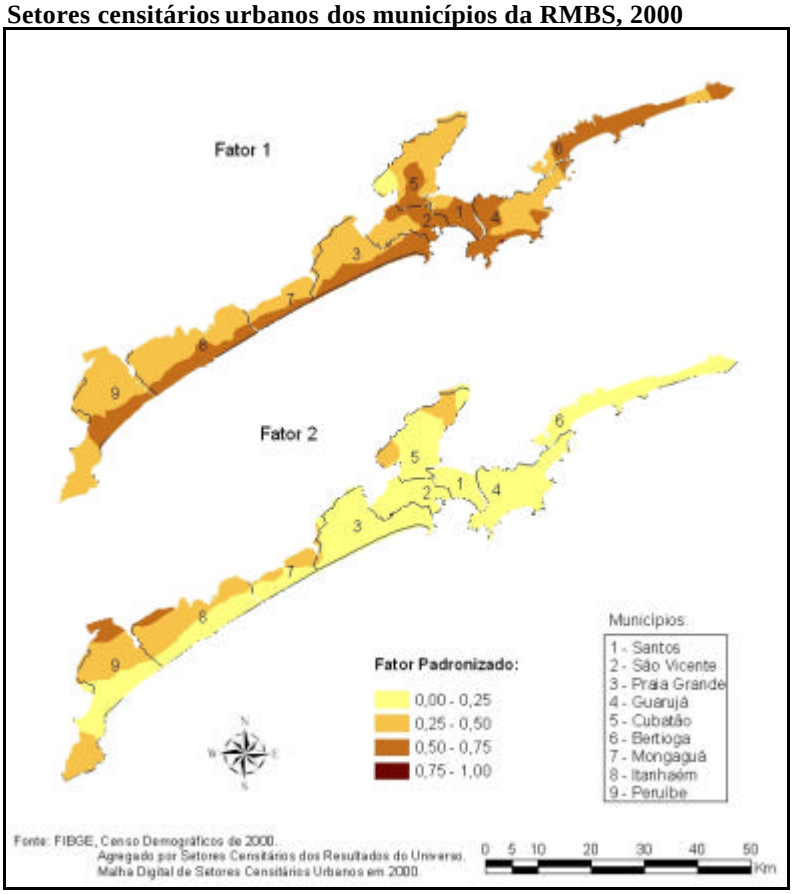


**Figura 14: Interpolação dos fatores com a krigagem universal**
**Setores censitários urbanos dos municípios da RMBS, 2000**




A krigagem universal faz um tipo de regressão com as coordenadas espaciais como sendo as variáveis explicativas, mas ao invés de assumir que os erros são independentes, eles são modelados para serem autocorrelacionados.

Do mesmo modo, a co-krigagem também é um método interpolador moderadamente rápido que pode ser exato se os dados apresentarem uma medida de erro, ou se possuir uma medida de erro para a suavização dos dados. Mas esse método utiliza a informação em múltiplos bancos de dados. Como a krigagem, é muito flexível e permite que se investigue a autocorrelação espacial entre dados, bem como se realize o cruzamento entre correlações. Por causa dos modelos estatísticos utilizados na co-krigagem, esse método também permite gerar uma variedade de mapas de saída, incluindo predições, predições de desvio-padrão, de probabilidade e de quantis. A flexibilidade desse método pode exigir muitas decisões para a maioria dos métodos ou que sejam definidos parâmetros padrões. Esse método também assume que os dados partem de um processo estocástico estacionário, sendo que, alguns métodos exigem que os dados se originem a partir de uma distribuição normal[4].

O próximo tópico tem como principal objetivo fazer uma comparação entre estes diversos métodos de interpolação explanados neste item, de forma a dar subsídios ao leitor para poder ter um conhecimento maior a respeito das formas de comparação entre eles, um suporte maior sobre como escolher qual é o melhor sem se preocupar apenas com a parte visual resultante.

[4] Para maiores detalhes sobre a krigagem e co-krigagem, consultar Matheron (1963); Journel e Huijbregts (1973); Burrough (1986); Cressie (1993); Isaaks e Srivastava (1989); Jakob (2003a); Jakob (2003b), entre outros.

## III. Comparação entre os métodos de interpolação

Os dados selecionados neste trabalho não deveriam ser utilizados para serem interpolados por todos estes métodos apresentados, dadas suas características em comparação com os pressupostos dos modelos. Isto foi feito apenas para efeito de visualização dos resultados, as superfícies de resposta. Se um estudioso deseja conhecer os locais com as maiores concentrações espaciais de determinado atributo, as figuras 2 em diante mostraram que teoricamente qualquer destes métodos poderia ser usado. O resultado seria basicamente o mesmo, apenas com áreas mais suavizadas ou com um maior detalhamento local. Eles mostraram que as maiores concentrações do Fator 1 ("melhores" características do chefe) se concentram junto à orla marítima da RMBS e do Fator 2 (inadequação dos domicílios) longe da orla, mais para o interior.

Se o resultado esperado é este, qualquer destes métodos poderia ser usado, embora os dados possam ser "forçados" a se adaptarem ao método. Mas se existe um desejo de uma análise mais precisa, com um maior embasamento estatístico da seleção do modelo final, deve-se fazer uma análise dos erros dos valores preditos. A krigagem permite que se faça uma validação cruzada para checagem dos dados, ou pelo menos uma comparação entre os erros. Por estes métodos, o ideal seria ter um erro médio padronizado (*Mean Standardized*) dos valores preditos próximo de "0", um erro quadrático médio (*Root-Mean-Square*) o mais baixo possível, um erro padrão médio (*Average Standard Error*) próximo do erro quadrático médio, e um erro quadrático médio padronizado (*Root-Mean-Square Standardized*) próximo de "1". No caso dos demais interpoladores determinísticos, que somente fornecem o erro quadrático médio, este tem que ser o mais baixo possível. A Tabela 3 traz os erros associados aos valores preditos destes métodos de interpolação utilizados neste trabalho.

**Tabela 3**
**Média e Erros dos Valores Preditos segundo os métodos de interpolação utilizados**
**Setores censitários urbanos dos municípios da RMBS, 2000**

| | | | Erros dos Valores Preditos | | | |
|---|---|---|---|---|---|---|
| Variável | Método | Média (*Mean*) | Erro Quadrático Médio (*Root-Mean-Square*) | Erro Padrão Médio (*Average Standard Error*) | Erro Médio Padronizado (*Mean Standardized*) | Erro Médio Quadrático Padronizado (*Root-Mean-Square Standardized*) |
| Fator 1 | IDW | 0,004505 | 0,09593 | - | - | - |
| | Polinomial Global (p=5) | -0,000335 | 0,1256 | - | - | - |
| | Polinomial Global (p=10) | 0,066 | 2,938 | - | - | - |
| | Polinomial Local | 0,0006595 | 0,1052 | - | - | - |
| | Funções de Base Radial | 0,0003672 | 0,09222 | - | - | - |
| | Krigagem Ordinária | 0,000363 | 0,09545 | 0,1121 | 0,0008113 | 0,841 |
| | Krigagem Simples | 0,002186 | 0,09124 | 0,07546 | 0,01226 | 1,171 |
| | Krigagem Universal | 0,00104 | 0,09261 | 0,07252 | 0,007942 | 1,251 |
| Fator 2 | IDW | -0,006525 | 0,06922 | - | - | - |
| | Polinomial Global (p=5) | 0,0000553 | 0,07704 | - | - | - |
| | Polinomial Global (p=10) | 0,03973 | 1,849 | - | - | - |
| | Polinomial Local | -0,0003936 | 0,06675 | - | - | - |
| | Funções de Base Radial | -0,005879 | 0,0676 | - | - | - |
| | Krigagem Ordinária | -0,0006433 | 0,06189 | 0,03885 | -0,0005871 | 1,477 |
| | Krigagem Simples | -0,0006226 | 0,06288 | 0,04503 | -0,0002366 | 1,301 |
| | Krigagem Universal | 0,0006003 | 0,06785 | 0,04303 | 0,01866 | 1,545 |

Fonte: FIBGE, Censo Demográfico de 2000. Análises estatísticas NEPO/UNICAMP

A Tabela 4 mostra os critérios para a comparação dos resultados dos métodos de interpolação, os menores erros, as diferenças entre um erro e outro, ou as diferenças entre um erro e algum valor, como “0” e “1”.

**Tabela 4**
**Critérios para comparação dos resultados dos métodos de interpolação**
**Setores censitários urbanos dos municípios da RMBS, 2000**

| Variável | Método | Menor Erro Quadrático Médio | Erro Médio Padronizado perto de "0" | Erro Padrão Médio próximo ao Erro Quadrático Médio | Erro Médio Quadrático Padronizado próximo a "1" |
|---|---|---|---|---|---|
| Fator 1 | IDW | 0,09593 | - | - | - |
| | Polinomial Global (p=5) | 0,12560 | - | - | - |
| | Polinomial Global (p=10) | 2,93800 | - | - | - |
| | Polinomial Local | 0,10520 | - | - | - |
| | Funções de Base Radial | 0,09222 | - | - | - |
| | Krigagem Ordinária | 0,09545 | -0,0008113 | 0,01665 | 0,159 |
| | Krigagem Simples | 0,09124 | -0,0122600 | -0,01578 | -0,171 |
| | Krigagem Universal | 0,09261 | -0,0079420 | -0,02009 | -0,251 |
| Fator 2 | IDW | 0,06922 | - | - | - |
| | Polinomial Global (p=5) | 0,07704 | - | - | - |
| | Polinomial Global (p=10) | 1,84900 | - | - | - |
| | Polinomial Local | 0,06675 | - | - | - |
| | Funções de Base Radial | 0,06760 | - | - | - |
| | Krigagem Ordinária | 0,06189 | 0,0005871 | -0,02304 | -0,477 |
| | Krigagem Simples | 0,06288 | 0,0002366 | -0,01785 | -0,301 |
| | Krigagem Universal | 0,06785 | -0,0186600 | -0,02482 | -0,545 |

Em Azul estão os Menores Valores Observados na Tabela!

Fonte: FIBGE, Censo Demográfico de 2000. Análises estatísticas NEPO/UNICAMP

Os dados da Tabela 4 mostram que os menores valores destes critérios (em azul) se encontram entre a krigagem simples e a krigagem universal. Assim, em termos somente dos erros, estes seriam os melhores métodos para se empregar as interpolações nas variáveis, mesmo estas variáveis sendo em forma de valores médios, como no Fator 1, como em termos de porcentagens, como no Fator 2. Resta então ao pesquisador a escolha do melhor método para utilização com seus dados. Em geral, como não se tem a média geral dos dados, o mais correto seria a aplicação da krigagem ordinária.

A Figura 15 mostra uma tela do *software* ArcGIS com toda a flexibilidade possível da krigagem ordinária, a escolha dos modelos para o semivariograma ou a covariância, uma vista prevista do semivariograma e da superfície de dados para busca de tendências e de uma direção para adotar, assim como valores de alcance, patamar, efeito pepita, tamanho do “*lag*”, número de “*lags*” e a possibilidade de tratar também a anisotropia, conforme abordado anteriormente. Neste caso a tela se refere ao Fator 1, e o modelo selecionado foi o esférico (a equação dele se encontra na parte inferior esquerda da tela, em azul. A curva final está em amarelo ajustada junto aos pontos do semivariograma. Deve-se buscar um modelo que ajuste os dados da melhor maneira possível.

Já a Figura 16 traz uma tela mostrando os erros associados aos valores preditos do Fator 2, criados por meio da krigagem simples, por exemplo. Espera-se que estes cheguem o mais próximo possível da linha azul do gráfico, que corresponde aos valores observados. Existem também os gráficos de valores preditos e dos erros padronizados, em comparação aos valores observados, assim como o gráfico QQPlot, que mostra os quantis dos valores preditos em comparação com os quantis dos valores observados, como forma de um maior subsídio às análises.

**Figura 15**
**Modelagem do Semivariograma do Fator 1 da Krigagem Ordinária**
**Setores censitários urbanos dos municípios da RMBS, 2000**

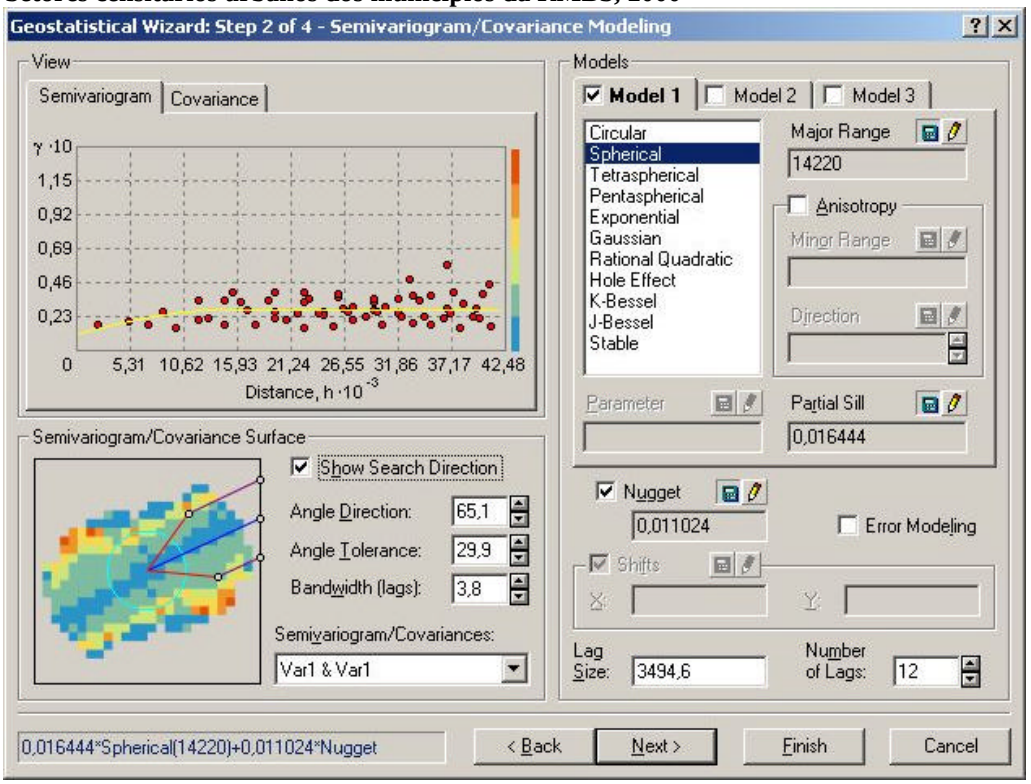
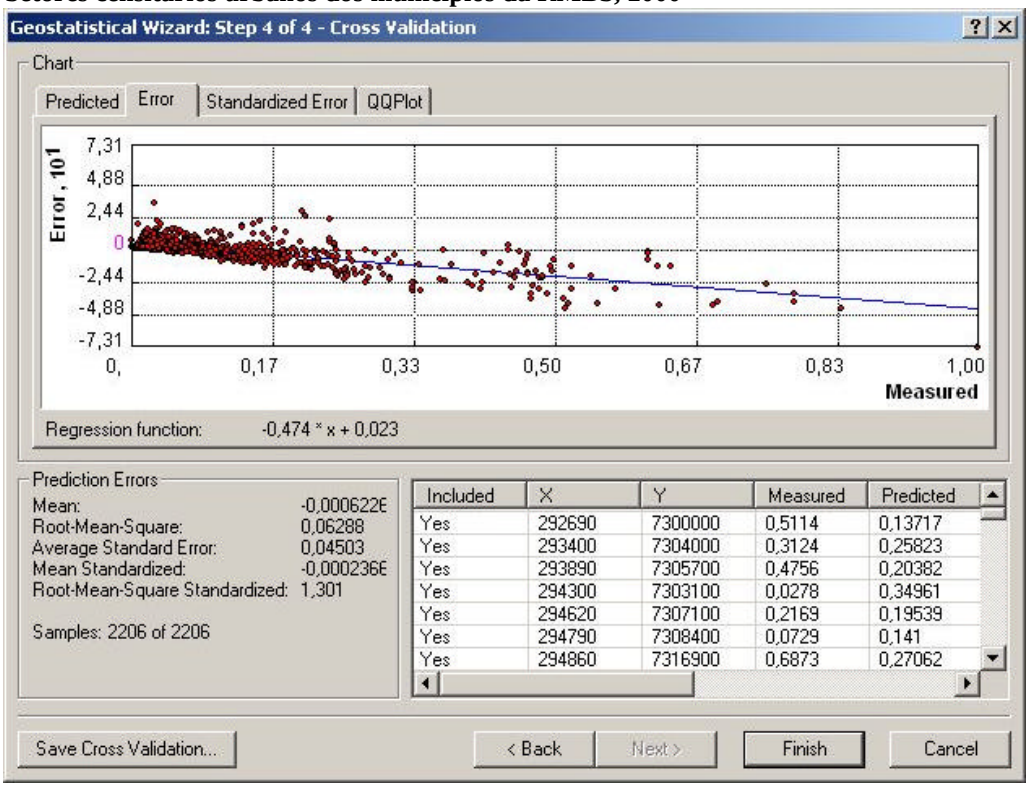



**Figura 16**
**Erros Associados ao Fator 2 da Krigagem Simples**
**Setores censitários urbanos dos municípios da RMBS, 2000**



## Conclusões

Este trabalho teve como principal objetivo mostrar a uma pessoa iniciante na área de geodemografia e de geoestatística como agir quanto ao número cada vez maior de métodos de interpolação de dados, presentes nos mais atuais Sistemas de Informação Geográfica (SIGs). A interpolação de dados seria então o segundo passo das análises nos SIGs, após a visualização simples destes.

Escrito na forma de um possível manual para utilização com métodos de interpolação, deve-se ressaltar também que os dados preditos a partir de tais métodos são suavizados, em maior ou menor grau, e assim, perde-se o valor do dado original. São usados então para a visualização de tendências e áreas de concentração espacial de determinada variável, uma possível estimativa de segregação espacial. Mas se o objetivo é mais específico e necessita de uma maior precisão quanto aos valores dos dados, outros métodos devem ser utilizados.

Estes outros métodos, talvez considerados como o terceiro passo das análises em SIGs, não envolveriam interpolação pura e simples, mas focariam mais na autocorrelação espacial, como a utilização dos índices de Moran global e local, ou até da lógica *fuzzy*, ou lógica nebulosa.

Como pôde ser comprovado neste trabalho, os resultados dos métodos de interpolação são muito bons, com alto grau de confiabilidade dos dados, uma vez que se têm os erros associados aos dados preditos, mas diversos deles requerem um razoável conhecimento prévio das técnicas utilizadas para os modelos matemáticos e seus pressupostos, como o conhecimento prévio da normalidade dos dados, estacionaridade, tendências, anisotropia, etc. Os melhores métodos de interpolação foram determinados neste caso como sendo a krigagem ordinária e a krigagem simples.

Mas deve-se lembrar também que as interpolações necessitam de um grande número de dados observados para resultar em uma maior precisão, no mínimo 100, mas o ideal é mais de 1000, dependendo da área de estudo. Neste estudo de caso, foram contabilizados mais de 2100 setores censitários da Baixada Santista e transformados em pontos (centróides) para as análises com os métodos de interpolação.

Por fim, espera-se também que este trabalho possa contribuir para um maior número de trabalhos relacionando a geoestatística com a geodemografia nas análises intra-urbanas, áreas estas relativamente novas nos círculos acadêmicos, e em especial na demografia.

## Referências Bibliográficas

BURROUGH, P.A. **Principles of Geographical Information Systems for Land Resources Assessment**. Oxford: Clarendon Press, 1986.

CÂMARA, G.; MEDEIROS, J. S. Princípios básicos em geoprocessamento. In: ASSAD, E. D.; SANO, E. E. (Ed.). **Sistemas de informações geográficas**: aplicações na agricultura. 2. ed. ver. ampl. Brasília, DF: Embrapa-SPI: Embrapa-CPAC, pp.3-11, 1998.

CRESSIE, N.A.C. **Statistics for Spatial Data**. New York: John Wiley, 1993.

ESRI. **Using ArcGIS Geostatistical Analyst** – GIS by ESRI. Redlands, CA: ESRI, 2001.

ISAAKS, E.H.; SRIVASTAVA, R.H. **An Introduction to Applied Geostatistics**. New York: Oxford University Press, 1989.

JAKOB, A.A.E. **Análise Sócio-Demográfica da Constituição do Espaço Urbano da Região Metropolitana da Baixada Santista no período 1960-2000**. Tese de Doutorado apresentada ao Programa de Doutorado em Demografia do Instituto de Filosofia e Ciências Humanas da Universidade Estadual de Campinas sob a orientação do Prof. Dr. José Marcos Pinto da Cunha. Campinas, SP: [s.n.], 2003a

_____. A Krigagem como método de análise de segregação espacial da população. IN: Encontro Nacional da ANPUR, 10. **Anais...**, Belo Horizonte (MG), 2003b

_____. A Krigagem como método de análise de dados demográficos. IN: Encontro Nacional de Estudos Populacionais, 13. **Anais...**, Ouro Preto (MG), 2002

JOURNEL, A.G.; HUIJBREGTS, C.J. **Mining Geostatistics**. New York: Academic Press, 1978.

LAMAS, C. **La Geodemografía Y la investigación de medios**. IV Seminario de AEDEMO sobre Medios Impresos, Radio y Publicidad Exterior. Bilbao, Noviembre de 1994. 12p.

LOURENÇO, R.W. **Comparação entre métodos de interpolação para Sistemas de Informações Geográficas**. Dissertação de mestrado elaborada junto ao curso de Pós-Graduação em Geociências – área de concentração em Geociências e Meio Ambiente do Instituto de Geociências e Ciências Exatas da Universidade Estadual Paulista, Campus de Rio Claro. Rio Claro, SP, 1998.

MATHERON, G. Principles of geoestatistics. **Economic Geology**, v. 58, p.1246-1266. El Paso, 1963.

TOBLER, W. A computer movie simulating urban growth in the Detroit region. **Economic Geography**, 46 (2), pp.234-240, 1970.