



CAPTAR
ciência e ambiente para todos

volume 2 • número 3 • p ...-...

Modelação espacial da temperatura na ilha de Santiago, Cabo Verde, com o *GeoStatistical Analyst*®

A utilização dos métodos de interpolação espacial é vez mais frequente nas análises espaciais de dados. Paralelamente, diversos softwares já contêm vários destes métodos, permitindo obter melhores estimativas como acontece com o *ArcGIS*® através da extensão *Geostatistical Analyst*®. Este artigo curto utiliza esta ferramenta informática para estimar a distribuição espacial da temperatura média da ilha de Santiago, Cabo Verde, numa visão educacional. Paralelamente, apresenta-se uma breve descrição dos procedimentos de interpolação espacial usados (*Kriging*, inverso da distância, *Voronoi* e tendência polinomial).

Palavras-chave

sistemas de informação geográficos
interpolação espacial
temperatura
Cabo Verde

Amândio Furtado¹

João Negreiros²•

¹ Universidade de Cabo Verde, Praia, Santiago, CP 379, Cabo Verde.

² Instituto Superior de Estatística e Gestão de Informação - Universidade Nova de Lisboa

• c8057@isegi.unl.pt

ISSN 1647-323X

NOTA INTRODUTÓRIA

Compreender a distribuição espacial de dados provenientes de fenómenos ocorridos no espaço constitui um grande desafio para a elucidação de questões centrais em diversas áreas do conhecimento como a saúde, ambiente, geologia, urbanismo, planeamento e climatologia. A tecnologia SIG (Sistemas de Informação Geográfica) constitui uma valiosa ferramenta em processos de integração, edição, análise e gestão de dados meteorológicos, por exemplo. É neste contexto que se pretende analisar a distribuição da temperatura para a ilha de Santiago, Cabo Verde, a partir das médias anuais obtidas para as leituras feitas nos 70 postos existentes na superfície da ilha, com a utilização do método de Krigagem, tendência polinomial (global e local), Voronoi e inverso da distância (do inglês Inverse Distance Weighted). Naturalmente, os diferentes métodos resultaram na geração de diferentes mapas.

A origem de Santiago é vulcânica enquanto o seu relevo é dominado por duas serras (ver Figura 1): Malagueta e Pico de Antónia (1394 metros). Na ilha existe uma diversidade de formas de relevo, marcados por declives acentuados e por uma grande diversidade de vertentes que, associado a diferentes graus de rocha e à variabilidade das precipitações, induz a uma forte erosão. A temperatura média anual ronda os 25°C e amplitude térmica anual é relativamente baixa (inferior a 10°C) ao passo que a precipitação é fortemente marcada pela altitude e pela exposição das íngremes vertentes. A média anual de precipitação varia aproximadamente entre 190 milímetros (mm) na região litoral (Chão Bom, a 20 m de altitude) e 472,6 mm (São Jorge dos Órgãos, a 319m de altitude) mas pode ultrapassar os 800 mm a partir dos 1000 m de altitude.

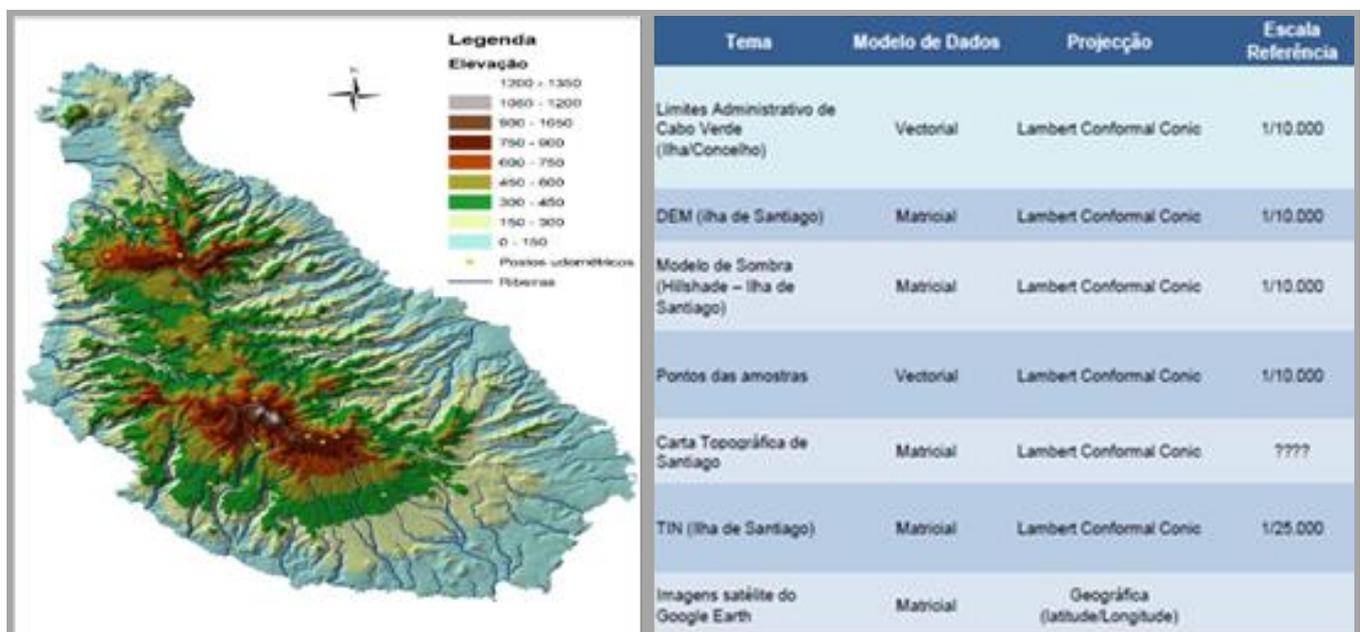


FIGURA 1: Mapa de elevação da ilha de Santiago (área total de 4033 Km²) e respectivos dados geográficos de *input* usados neste estudo.

ABORDAGEM GEOESTATÍSTICA

Inicialmente, a geoestatística foi utilizada na resolução de problemas de geologia e mineração. Tratando-se de técnicas específicas de interpolação espacial¹ e desenvolvido na Escola de Minas de Paris pelo Professor Matheron em 1963. O autor deste trabalho foi auxiliado pelos trabalhos estatísticos da escola Sul-africana já realizados nas minas de ouro e urânio. Estatisticamente, a geoestatística está baseada na teoria das variáveis regionalizadas que assume que os valores estão relacionados, de alguma forma, com a posição que ocupam no espaço.

Genericamente, a Krigagem é um método que usa esta dependência espacial entre as amostras vizinhas, expressa no variograma, para estimar os valores em qualquer posição dentro do campo de estudo. Por defeito, estes estimadores apresentam uma variância mínima e, paralelamente, consideram que a média é relativamente constante mas desconhecida (estacionaridade). Contudo, a modelação por Krigagem pode também levar em consideração as variações locais da variável, limitando o seu domínio de estacionaridade a uma vizinhança local. Assim, apenas as amostras da vizinhança do ponto a estimar são consideradas no cálculo. Como seria de esperar, os pesos da Krigagem dependem da localização do ponto a estimar e da configuração geométrica entre as restantes amostras.

ABORDAGEM DETERMINÍSTICA

No método de Voronoi, a decomposição do espaço métrico em regiões varia de acordo com a distância aos pontos da amostragem e onde se assume apenas variações nas fronteiras (ver Figura 2). Esta eliminação de diferenças sistemáticas antes de uma interpolação contínua e a avaliação rápida em termos computacionais são, portanto, duas características intrínsecas desta metodologia simples e eficaz. Obviamente, o mapa final depende do *layout* das amostras como mostra a Figura 2.

O método do inverso da distância (IDW) calcula os valores espaciais desconhecidos a partir do valor médio ponderado das amostras conhecidas em seu redor. Naturalmente, as amostras mais perto do ponto a interpolar contêm um peso maior no cálculo global que amostras mais distantes. No entanto e de forma similar a todas as médias ponderadas pela distância, a sua preocupação comum é o efeito da polarização condicional, isto é, subestimação dos valores altos e superestimação dos valores baixos (ver Figura 3).

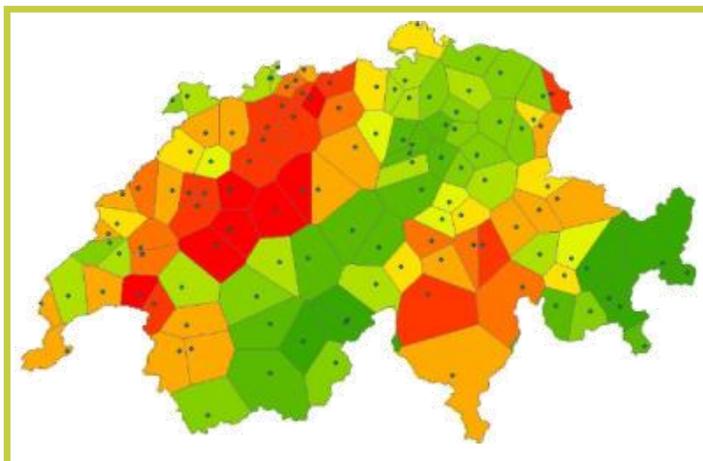


FIGURA 2: Mapa estimado de pontos genéricos na Suíça (Temperatura). Note-se que nunca existem regiões de valor nulo enquanto o tamanho de cada sub-região varia inversamente com a densidade dos pontos existentes. Seguindo as regras tradicionais das cores do arco-íris, a cor vermelha traduz valores altos enquanto o amarelo e o verde indicam valores mais baixos, respectivamente (<http://www.geo.unizh.ch/gis>, 2010).

¹Interpolação é um procedimento de estimação do valor de um atributo em locais não amostrados a partir de pontos amostrados na mesma área ou região.

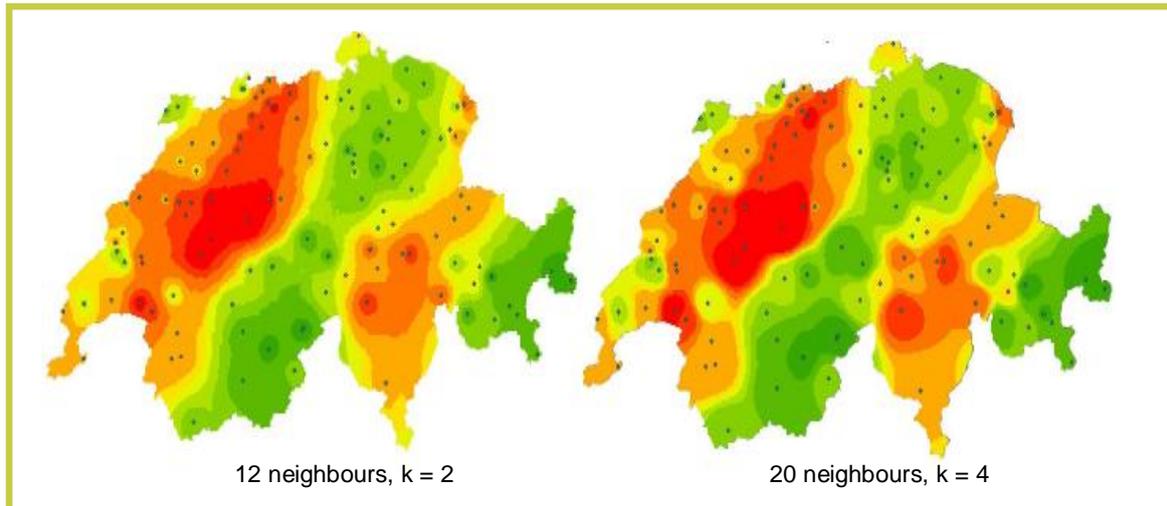


FIGURA 3: Mapa estimado pelo inverso da distância (IDW) em função do número de amostras a considerar e pesos de ponderação (www.geo.unizh.ch/gis, 2010). Note que as amostras iniciais são exactamente as mesmas que da figura anterior.

A superfície de tendência polinomial (ver Figura 4) utiliza um procedimento de regressão² normal mas baseado nas coordenadas espaciais das amostras e nos respectivos valores. Desenvolvidos por Whitten em 1957, estes coeficientes são encontrados pela resolução de um conjunto de equações simultâneas não-linear dos produtos cruzados das coordenadas (x,y) e os valores das respectivas amostras³.

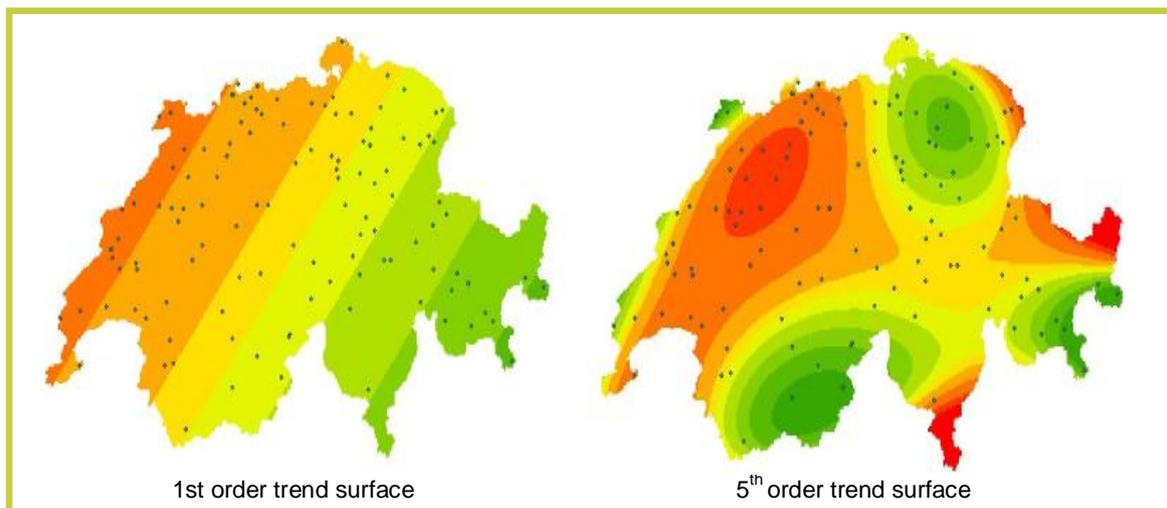


FIGURA 4: Estimação pela superfície de tendência polinomial do 1º e 5º grau (www.geo.unizh.ch/gis, 2010).

² Regressão é um método para se estimar a variável condicional (valor esperado) dados os valores de algumas outras variáveis.

³ Neste tipo de interpolação, a superfície é aproximada através do ajustamento de um polinómio aos dados pontuais (amostras) através de uma regressão múltipla dos valores do atributo em função da localização geográfica (coordenadas x e y). O polinómio é então usado para estimar valores de pontos numa grelha de estrutura raster ou num dado local. Por exemplo: Imagine que o atributo z num dado ponto (x1,y1) é dado pela equação linear do 1º grau $z=2+4x-8y$. Substituindo o x e o y pelos valores x1 e y1, respectivamente, obtem-se a estimacão z nesse determinado local.

INTERPOLAÇÃO ESPACIAL COM O *GEOSTATISTICAL ANALYST*®

Análise exploratória dos dados

Baseado nas setenta amostras de temperatura registada no último ano na ilha de Santiago, o coeficiente de *Skewness*⁴(0,07) mostra que o índice de temperatura apresenta uma assimetria quase nula onde a mediana (24°C) é ligeiramente superior à média (23,85°C). Numa análise do nível de achatamento da distribuição, a curva demonstra ser **platicurtica**⁵ (2,27). Em relação às medidas de tendência não central verifica-se que os valores estão concentrados em torno da média onde a amplitude interquartis (diferença entre o 3º e o 1º quartil) ronda apenas o valor 2,5. Dado que o limite máximo e mínimo são 27°C e 21°C (ver Figura 5), este parâmetro descritivo leva-nos a concluir as temperaturas em Santiago, Cabo Verde, é bastante homogênea ao longo do dia e do ano. Assim, não se verificam valores atípicos que possam indicar a existência de *outliers*⁶.

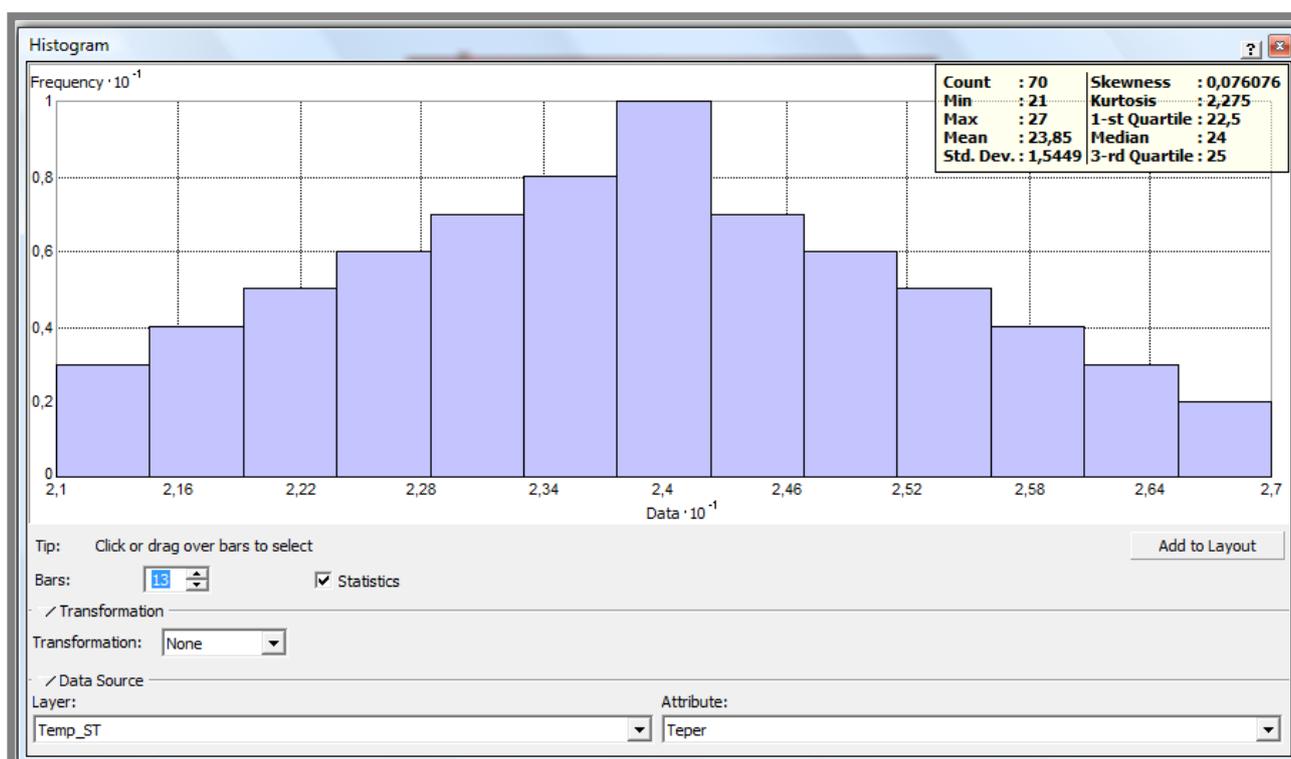


FIGURA 5: Histograma e medidas estatísticas dos valores médios de temperatura do ar em Santiago, Cabo Verde.

⁴ Numa distribuição estatística, a assimetria (*skewness*) é o quanto a sua curva de frequência se desvia ou afasta da posição simétrica. Se este valor for zero, então a distribuição dos valores é simétrica em relação à média. Se negativo, apresenta uma assimetria à esquerda. Se positivo, a distribuição é assimétrica à direita.

⁵ Em estatística, a Kurtosis é uma medida de dispersão que caracteriza o achatamento da curva da função de distribuição. Se este parâmetro for igual a zero então tem o mesmo achatamento que a distribuição Normal ou Gaussiana. Se o valor é maior que zero então a distribuição em questão é mais alta (afunilada) e concentrada que a distribuição Normal. Caso contrário, a função é mais achatada que a distribuição Normal [Universidade da Amazônia, 2010].

⁶ Um valor muito diferente do conjunto de dados existentes.

Outro elemento gráfico importante é a ferramenta de análise de tendência dos dados que permite identificar a presença/ausência de tendência global. No gráfico seguinte, esta análise apresenta os pontos projectados num plano perpendicular com direcção E-W e N-S. Verifica-se, na Figura 6, uma tendencial global da distribuição dos dados em forma de U (um polinomial do 2º grau parece ser o indicado para ser ajustado à tendência global), indicando uma tendência espacial de não estabilidade em ambas as direcções (N-S versus E-W).

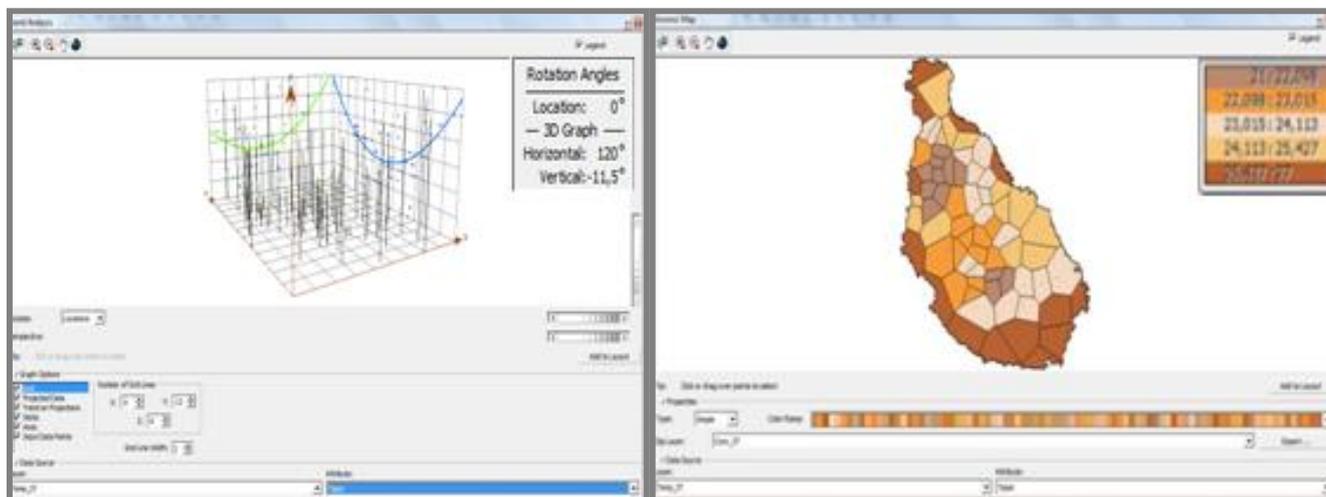


FIGURA 6: Análise de tendência global do ArcGIS® (esquerda) e mapa de Voronoi (direita). Note-se em ambas as figuras uma clara tendência de valores mais altos de temperatura na zona sul e litoral da ilha.

Modelação da Temperatura Espacial e Validação Cruzada

De acordo com a figura 7, constata-se uma alteração acentuada dos valores da temperatura. Assim, as temperaturas mais baixas ocorrem nas zonas elevadas (Serra Malagueta e de Pico de Antónia) enquanto as zonas de baixa altitude (litoral) apresentam altos valores (situação antagónica ao padrão da pluviosidade). O comportamento da variável em causa é, assim, anisotrópico⁷ com uma direcção principal de continuidade ou homogeneidade espacial na direcção de 325º (sentido directo ou sentido dos ponteiros do relógio onde 0º indica meio-dia).

Tendo em consideração os valores das amostras, métodos e parâmetros estabelecidos, diferentes superfícies foram interpoladas e, naturalmente, encontra-se associado um determinado erro. O processo de validação-cruzada (*cross-validation*) permite dar uma ideia do quão bem o modelo se adaptou, servindo assim como medida de incerteza da predição realizada. Genericamente, a validação cruzada omite uma amostra do conjunto de dados inicial e estima o seu valor usando os restantes dados. Comparando o valor estimado com o real, obtêm-se o erro associado. Realizando este processo sucessivamente com as restantes amostras, o objectivo é comparar modelos e ajudar o utilizador a decidir sobre qual o modelo que apresenta maior precisão de estimação. Assim, apresenta-se na Tabela 1 as duas estatísticas principais deste procedimento associado aos métodos de interpolação IDW, Polinomial e Kriging.

⁷Característica espacial que a substância em causa possui uma propriedade física de variação em função da direcção [Wikipedia, 2010].

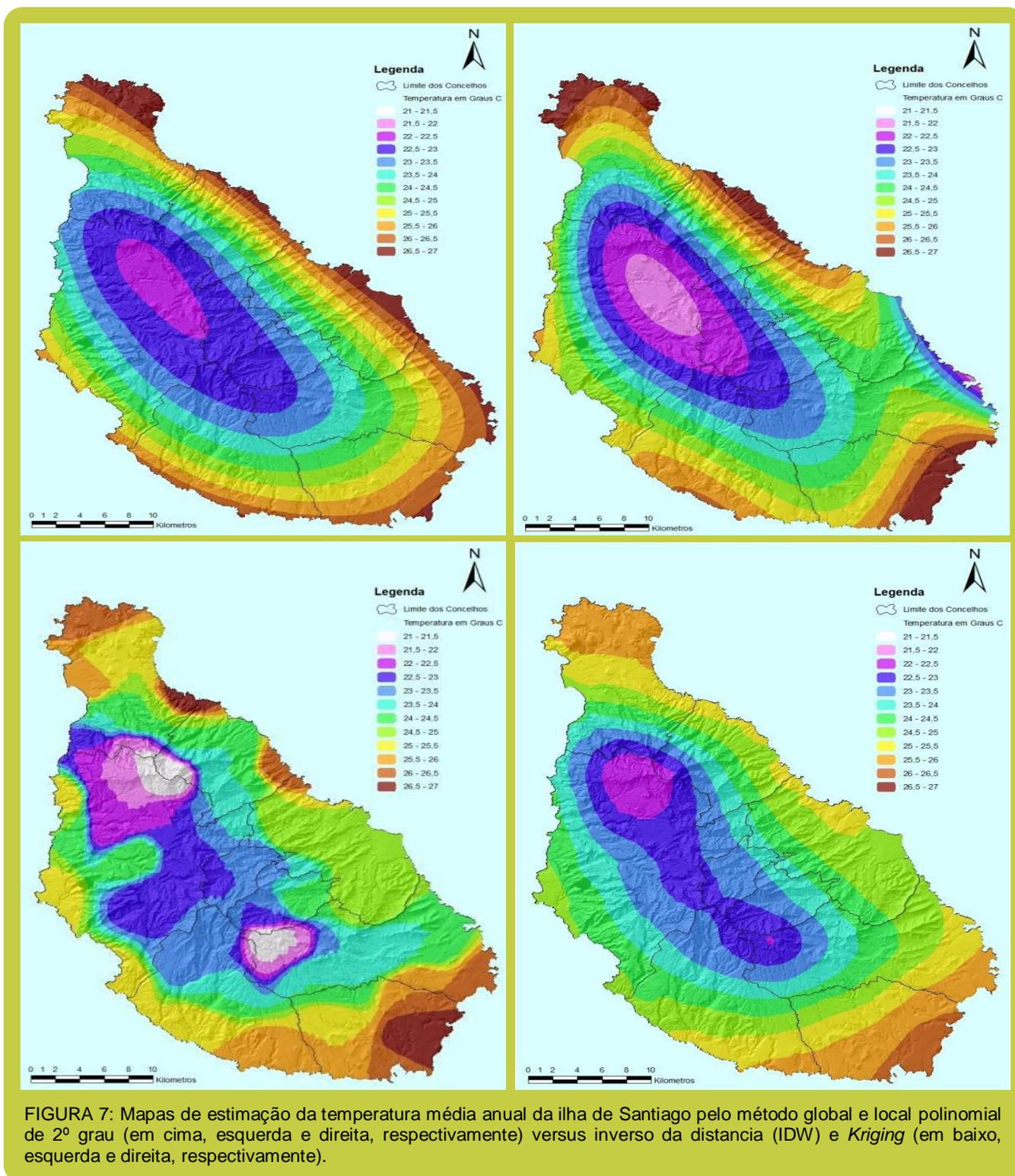


TABELA 1: Estatística dos erros estimados pelo *Geostatistical Analyst*® onde a média estimada deve ser o mais próximo de zero (*unbiased*) enquanto o erro quadrático médio deve ser mais baixo possível.

Método		Erro médio - EM	Erro Quadrático médio - EQM
IWD		-0,3166	0,8851
Polinomial	Local	-0,1164	1,46
	Global	0,05011	1,258
Kriging	Normal	0,03864	0,9331


NOTAS CONCLUSIVAS

Os Sistemas de Informação Geográficos (SIG) constituem numa importante ferramenta de análise espacial, facultando informações espaciais imprescindíveis para a tomada de decisão. A eficácia da interpolação, neste caso particular, acentua a necessidade de dominar a conceptualização da ferramenta, metodologia, grau de conhecimento que o modelador possui acerca do fenómeno em questão e do funcionamento dos algoritmos subjacentes ao SIG (ver Tabela 2). Estes factos reforçam a vantagem deste tipo de projectos serem desenvolvidos no seio de uma equipa multidisciplinar.

De acordo com os parâmetros da validação cruzada, pode ainda comprovar-se que a melhor escolha do método de interpolação não é única. Se IDW prima por um melhor valor do erro quadrático médio, o método de *Kriging* contrapõe um melhor erro médio. Contudo e neste caso particular, seria preferível a Krigagem pelo seu erro médio ser bastante próximo de zero (não enviesamento das interpolações, isto é, evitar constantemente a sobre e a subestimação dos valores desconhecidos).

TABELA 2: Comparação entre os diferentes métodos de interpolação [Johnston et al., 2001].

Método	Determinístico/ Estocástico	Típo de saída (superfície)	Tempo modelação	Exactidão do interpolador	Vantagens	Desvantagens	Requisitos
IWD (Inverse Distance Weighted)	Determinístico	Previsão	Rápido/Rá- pido	Sim	Poucos parâmetros de decisão	Sem acesso à previsão do erro; produz o efeito "olho-de-boi" em volta da locali- zação dos da- dos.	Nenhum
Polinomial global	Determinístico	Previsão	Rápido/Rá- pido	Não	Poucos parâmetros de decisão	Sem acesso à previsão do erro; pode ser dema- siado suave; pontos da borda têm uma grande influência	Nenhum
Polinomial local	Determinístico	Previsão	Moderada- mente rápido/ Moderado	Não	Alguns parâmetros de decisão	Sem acesso à previsão dos erros; pode ser demasiado automatizado	Nenhum
Krigagem	Estocástico	Previsão; Previsão do Erro Regularizado; Probabilidade	Moderada- mente rápido/Lento	Sim, sem medição de Erro; Não, com medição de Erro	Muito flexí- vel, permite estabelecer a auto corre- lação espa- cial, pode também estabelecer a previsão do erro re- gularizado; muitos parâmetros de decisão	Necessário to- mar algumas de- cisões nas transformações; tendências; modelos; Parâ- metros e Vizi- nhança	Pressupõe que os dados sejam prove- nientes de processos estocásticos estacionários e alguns mé- todos reque- rem que os dados tenham uma distri- buição normal

REFERÊNCIAS BIBLIOGRÁFICAS

Johnston, K., Hoef, J. M., Krivoruchko, K., Lucas, N (2001). Using ArcGIS Geostatistical Analyst. California – USA, ESRI Press, 306 pp.

Universidade da Amazônia, 2010 (http://arquivos.unama.br/nead/gol/gol_adm_2mod/estatistica/pdf/ESTA_impressoaula05.pdf).

Wikipedia, 2010 (<http://pt.wikipedia.org/wiki/Anisotropia>).