

Um estudo Monte Carlo na avaliação de erro em modelos digitais de elevação

Sidnei João Siqueira Sant'Anna
Corina da Costa Freitas

Instituto Nacional de Pesquisas Espaciais - INPE
Caixa Postal 515 - 12245-970 - São José dos Campos - SP, Brasil
{sidnei,corina}@dpi.inpe.br

Abstract. A Monte Carlo study was conducted to assess the errors present in digital elevation model (DEM) estimation when the number of sample points increases. The influence of spatial location and the correlation of these sample points on the DEM errors was also investigated. The errors were quantified by computing the root mean squared error and mean absolute deviation measures, both calculated global and locally. The results showed that for an area of 100m x 100m with DEM resolution of 1.25m the increasing of sample points tend to decrease the DEM errors. The effect of increasing the sample points on the errors seems to be higher over surfaces with correlated points. The spatial locations of the sample points are highly related to the triangles formed on the triangular irregular net grid and consequently have great influence on the DEM errors, thus grids having more acute triangles tend to present greater errors than those having more equilateral triangles.

Palavras-chave: modelo digital de elevação, estudo Monte Carlo, MAE, RMSE.

1. Introdução

Um Modelo Digital de Elevação (MDE) pode ser definido com uma representação computacional (discreta) de uma superfície contínua. Os modelos digitais de elevação têm aplicação em várias áreas de conhecimento, tais como, engenharia civil, planejamento de enlaces de radiocomunicação, modelagem hidrológica, produção de mapas topográficos entre outras. Através de um MDE procura-se estimar as elevações de uma superfície terrestre real a partir de um conjunto de pontos amostrais com elevações conhecidas. Quanto mais próximo da superfície real estas estimativas estiverem melhor será o MDE. Entretanto, como em todo e qualquer modelo existem erros associados aos métodos de modelagem.

De acordo com El-Sheimy (1999) não existe um MDE ideal já que as técnicas utilizadas para a geração de MDE's não conseguem representar toda a complexidade de uma superfície terrestre, sempre havendo problemas tanto na sua amostragem quanto na sua representação. A quantificação dos erros associados a um MDE ainda é alvo de estudo, evidenciado pelo grande número de trabalhos publicados na literatura, dentre os quais pode-se citar: Wood e Fisher (1993), Rees (2000), Wechsler (2003) e Widayati et al. (2004). A qualidade de um MDE está diretamente relacionada à própria superfície que se quer representar, incluindo sua extensão e o tipo de relevo (plano ou movimentado), assim como à densidade e à distribuição dos pontos amostrais, à grade (regular ou irregular) usada na sua geração, ao método de interpolação na estimativa dos outros pontos da grade e à resolução espacial do modelo final. Portanto, cada elevação estimada em um MDE contém erros (provenientes de várias fontes) devido à técnica e aos dados (conjunto de pontos amostrais) utilizados na sua geração.

É neste contexto, então, que se insere este trabalho, cujo objetivo central é a avaliação da influência do número de pontos amostrais (densidade) e de sua distribuição espacial (posicionamento) na geração de MDE's. Para tanto, realizou-se um estudo Monte Carlo da geração de MDE's a partir de grades irregulares triangular e interpolador linear, para dois tipos de superfície. A densidade e a posição dos pontos amostrais são dois parâmetros importantes na quantidade e forma dos triângulos que compõem a grade irregular utilizada na interpolação dos outros pontos da superfície. Em Câmara et al. (2001) podem ser encontradas descrições mais detalhadas sobre as definições dos tipos de grades e de intepoladores.

2. Dados e Metodologia

Neste trabalho foram idealizadas duas regiões distintas possuindo relevo plano correspondendo a uma área de 100m x 100m, com resolução espacial de 1,25m para representar a superfície terrestre. As regiões foram simuladas considerando que as elevações da superfície do relevo possuam uma distribuição Gaussiana, com função de correlação espacial bidimensional, possuindo também a forma Gaussiana.

O método de simulação de processos aleatórios Gaussianos correlacionados utilizado na geração das regiões é baseado na decomposição da matriz de covariância através dos seus autovalores e autovetores (Yanasse, 1991). Neste método admite-se que o processo bidimensional seja separável, ou seja, a matriz de covariância V deste processo pode ser escrita na forma de um produto de Kronecker entre as matrizes de covariância nas direções ortogonais x e y ($V = V_x \otimes V_y$). Assim, a realização de um processo bidimensional Gaussiano correlacionado a pode ser descrita por um produto de matrizes dado por:

$$a = P\Lambda^{1/2}\varepsilon_{ij} \text{ com } P = P_x \otimes P_y \text{ e } \Lambda = \Lambda_x^{1/2} \otimes \Lambda_y^{1/2}$$

onde ε_{ij} ($i = 1, \dots, n_x$ e $j = 1, \dots, n_y$) é uma realização de um processo Gaussiano com média zero e variância unitária, Λ_x e Λ_y são as matrizes diagonais dos autovalores de V_x e V_y , respectivamente, e P_x e P_y são matrizes de dimensão $n_x \times n_x$ e $n_y \times n_y$ de autovetores ortonormais de V_x e V_y , respectivamente.

A função da correlação espacial bidimensional adotada é expressa por:

$$\rho(x, y) = e^{-\alpha(x-y)^2}$$

onde x e y são as distâncias nas direções horizontal e vertical, respectivamente, e o parâmetro α controla o grau de correlação. Note que para pequenos valores de α a função de correlação possui uma variância maior, caracterizando uma alta entre as elevações. Assim, para distinguir os dois tipos regiões foram usados diferentes graus de correlação espacial um alto ($\alpha = 0,07$) e um baixo ($\alpha = 1,7$).

Na **Figura 1** é mostrado o gráfico de um corte da correlação espacial na direção vertical, que é igual à correlação na direção horizontal, para os dois graus de correlação e uma imagem de uma possível superfície (uma realização) para cada uma das regiões. Cabe ressaltar que a elevação média e sua variância foram mantidas as mesmas para as duas regiões, somente mudou-se o grau de correlação em cada região. Os valores de 161,76m e 4,16m foram

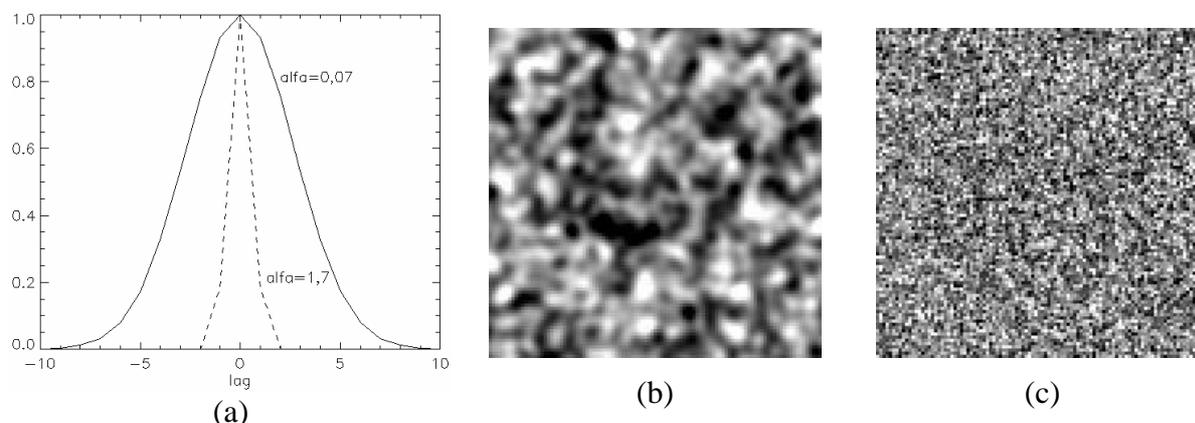


Figura 1 - (a) Gráfico da correlação espacial na direção vertical para alta (linha sólida - $\alpha = 0,07$) e baixa (linha tracejada - $\alpha = 1,7$) correlação e uma imagem de uma realização de cada superfície: (b) com alta correlação e (c) baixa correlação.

estipulados, respectivamente, para a elevação média e sua variância. Estes valores que foram calculados a partir de um arquivo de 165 pontos cotados obtidos em levantamento planialtimétrico, realizado no ano de 2003, usando GPS diferencial e técnica de topografia convencional em uma área de relevo plano na região de Tapajós (Pará). Este procedimento tem por objetivo ter valores de elevações nas simulações mais próximas de valores reais.

Na avaliação do erro de um MDE, o qual é definido como a diferença entre o valor da superfície simulada e o estimado através do MDE, realizou-se um estudo Monte Carlo baseado em 5000 realizações (replicações) de cada região. Neste estudo foram calculados dois tipos de erro, denominados por local e global médio. O erro local é descrito pela média dos erros de todas as replicações em cada elevação da superfície (ponto amostral), neste caso gera-se uma imagem de erro. O erro global médio é obtido primeiro calculando-se um erro global para cada superfície em cada replicação, o qual leva em conta todos os pontos da superfície, e em seguida faz-se uma média destes 5000 erros globais, gerando um único valor.

Os erros foram quantificados através das medidas do Erro Médio Quadrático (*Root Mean Squared Error* - RMSE) e do Erro Absoluto Médio (*Mean Absolute Error* - MAE). As medidas RMSE e MAE para o erro local são expressas por:

$$RMSE_j = \sqrt{\frac{1}{N} \sum_{i=1}^N (z_{ij} - \hat{z}_{ij})^2} \text{ e } MAE_j = \frac{1}{N} \sum_{i=1}^N |z_{ij} - \hat{z}_{ij}|$$

onde z_{ij} e \hat{z}_{ij} representam, respectivamente, as elevações geradas estimadas do j -ésimo ponto amostral ($j = 1, \dots, M$) na i -ésima replicação ($i = 1, \dots, N$), M é o número total de elevações da superfície e N é o número de replicações. No caso do erro global médio estas medidas são dadas por:

$$RMSE = \frac{1}{N} \sum_{i=1}^N \left(\sqrt{\frac{1}{M} \sum_{j=1}^M (z_{ij} - \hat{z}_{ij})^2} \right) \text{ e } MAE = \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{M} \sum_{j=1}^M |z_{ij} - \hat{z}_{ij}| \right)$$

Na avaliação da influência do número de pontos amostrais e de sua distribuição na geração de MDE's, estes modelos foram construídos a partir de grades irregulares triangular e interpoladores lineares. A versão 6.0 *software* IDL (*Interactive Data Language*) foi utilizada na geração dos modelos e das superfícies simuladas, assim como para o cômputo das medidas de erro.

Primeiramente, foram construídos 11 modelos variando-se o número de pontos amostrais (de 4 até 25 pontos) onde a posição de cada ponto era conhecida, isto é, a distribuição espacial dos pontos amostrais foi controlada. Nesta distribuição dos pontos amostrais foi imposta a restrição de que os pontos referentes aos quatro cantos da superfície estivessem presentes em todas as grades. Esta restrição garante que todos os modelos gerados possuam a mesma área. As grades triangulares, obtidas através do método de Delaunay, para estes modelos são apresentadas na **Figura 2**. Já numa análise preliminar pode-se observar da **Tabela 1** e da **Figura 2** que para um mesmo número de pontos amostrais (grupo de modelos C, E e F) o número e a formas dos triângulos que compõem a grade são diferentes e isto pode acarretar em estimações distintas das elevações.

Tabela 1 – Número de pontos amostrais e de triângulos em cada modelo.

Modelo	A	B	C1	C2	C3	C4	D	E1	E2	F1	F2
# pontos	4	5	9	9	9	9	13	17	17	25	25
# triângulos	2	4	8	12	12	7	16	24	24	32	40

Após esta primeira avaliação, uma outra avaliação foi efetuada gerando-se 50 outros modelos com número de pontos amostrais variando de 5 até 495 em passos de 10 pontos. Nesta análise manteve-se a restrição anterior, contudo a distribuição dos pontos amostrais

restantes foi realizada de maneira aleatória no interior da superfície. As grades triangulares obtidas para os modelos com 5 e 495 pontos amostrais são mostradas na **Figura 3**, as quais possuem 4 e 943 triângulos, respectivamente. Para esta avaliação somente se calculou o erro global médio dos modelos.

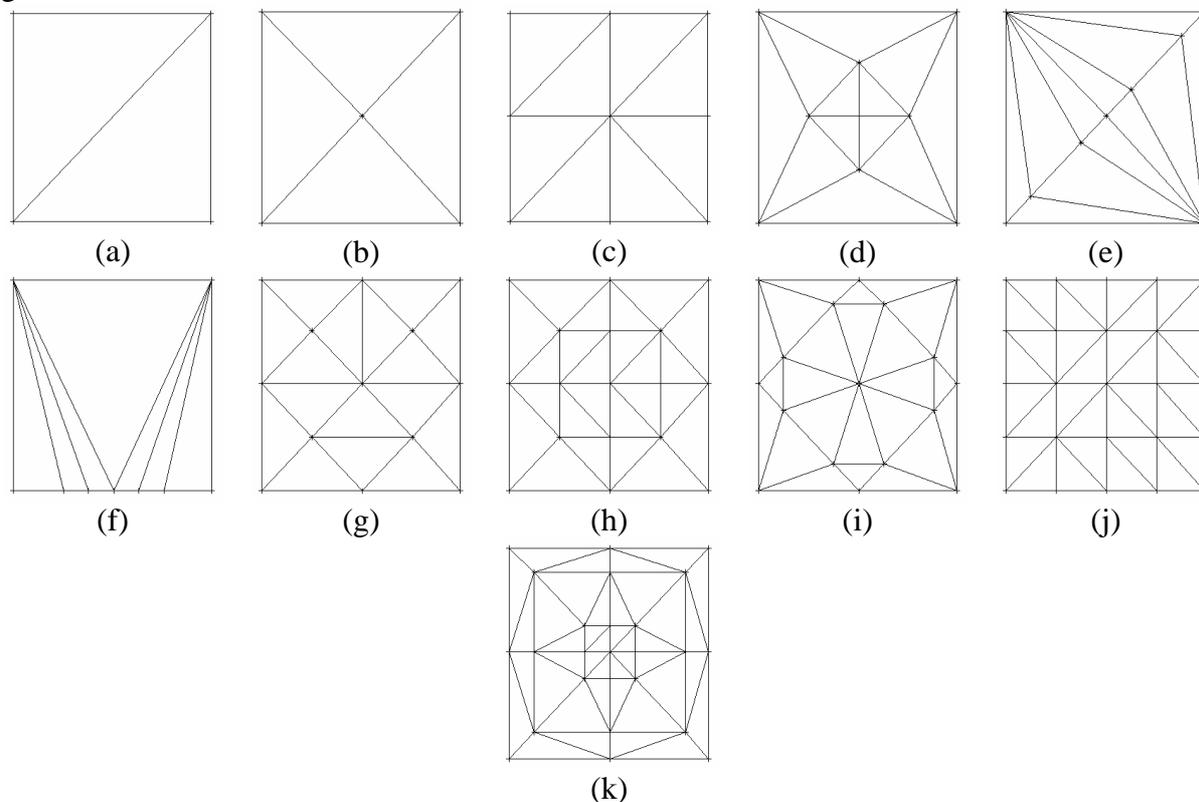


Figura 2 – Grades irregulares triangulares dos modelos: (a) A, (b) B, (c) C1, (d) C2, (e) C3, (f) C4, (g) D, (h) E1, (i) E2, (j) F1e (k) F2.



Figura 3 – Grades triangulares dos modelos possuindo pontos amostrais com posicionamento aleatório: (a) 5 e (b) 495.

3. Resultados e Discussão

Os erros globais médios para os 11 primeiros modelos foram calculados, baseados nas 5000 replicações das superfícies. As medidas do MAE e do RMSE médio e de suas respectivas variâncias são mostrados na **Figura 4**, onde a barra de erro é referente a um desvio padrão.

Pode-se observar, para o caso das superfícies com alto grau de correlação, uma ligeira tendência ao decréscimo dos valores médios das medidas com o aumento do número de pontos amostrais, assim como uma diminuição de sua variância (**Figuras 4a e 4c**). Este fato indica que o aumento do número de pontos amostrais acarretará em estimativas das elevações da superfície mais precisas e acuradas. Por outro lado, as medidas obtidas para as superfícies com baixo grau de correlação apresentam valores médios praticamente iguais, porém há uma

diminuição de sua variância (**Figuras 4b e 4d**), indicando que as estimativas das elevações da superfície são mais precisas, mas não necessariamente mais acuradas.

A partir dos gráficos da **Figura 4** nota-se que embora haja uma diminuição na variância das medidas com o aumento do número de pontos amostrais, estas medidas podem ser consideradas iguais. Isto é confirmado através de um teste Z de diferença de médias, o qual mostrou que os valores das médias do RMSE e MAE para ambos tipos de superfície não são estatisticamente diferentes.

Observando as medidas MAE e RMSE calculadas para os modelos C1 a C4 sobre as superfícies com alto grau de correlação (**Figuras 4a e 4c**) pode-se afirmar que a forma e o número dos triângulos obtidos na triangulação podem influir nas medidas de erro, uma vez que o número de pontos amostrais é o mesmo para estes modelos. O modelo C4, por exemplo, que possui apenas 7 triângulos (sendo 6 agudos) em sua grade apresenta as medidas com maior média e variância, isto é, o maior erro global e as medidas mais imprecisas do erro. As grades dos modelos C2 e C3 possuem 12 triângulos cada, contudo há uma pequena diferença entre os valores de média e variância das medidas de RMSE e MAE. O modelo C3 apresenta valores de média e variância maiores que os obtidos para o modelo C2, indicando maior erro. Note que a grade triangular do modelo C2 possui triângulos mais próximos de um triângulo equilátero que os triângulos da grade do modelo C3, este fato conduz à afirmação de que a forma dos triângulos da grade irregular pode vir a influenciar na medida de erro do modelo.

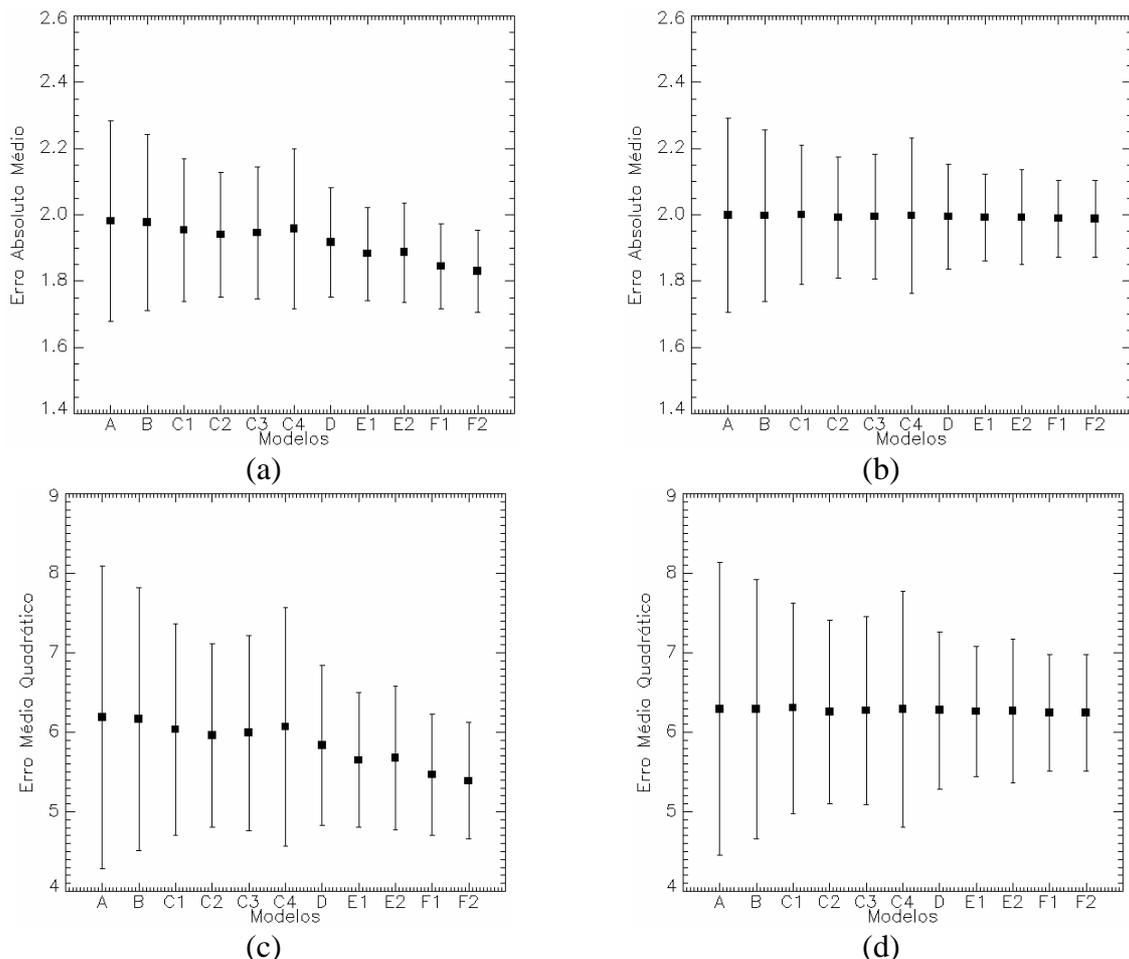


Figura 4 - Erro Médio Absoluto Global: (a) $\alpha = 0,07$ e (b) $\alpha = 1,7$ e Erro Médio Quadrático Global (c) $\alpha = 0,07$ e (d) $\alpha = 1,7$.

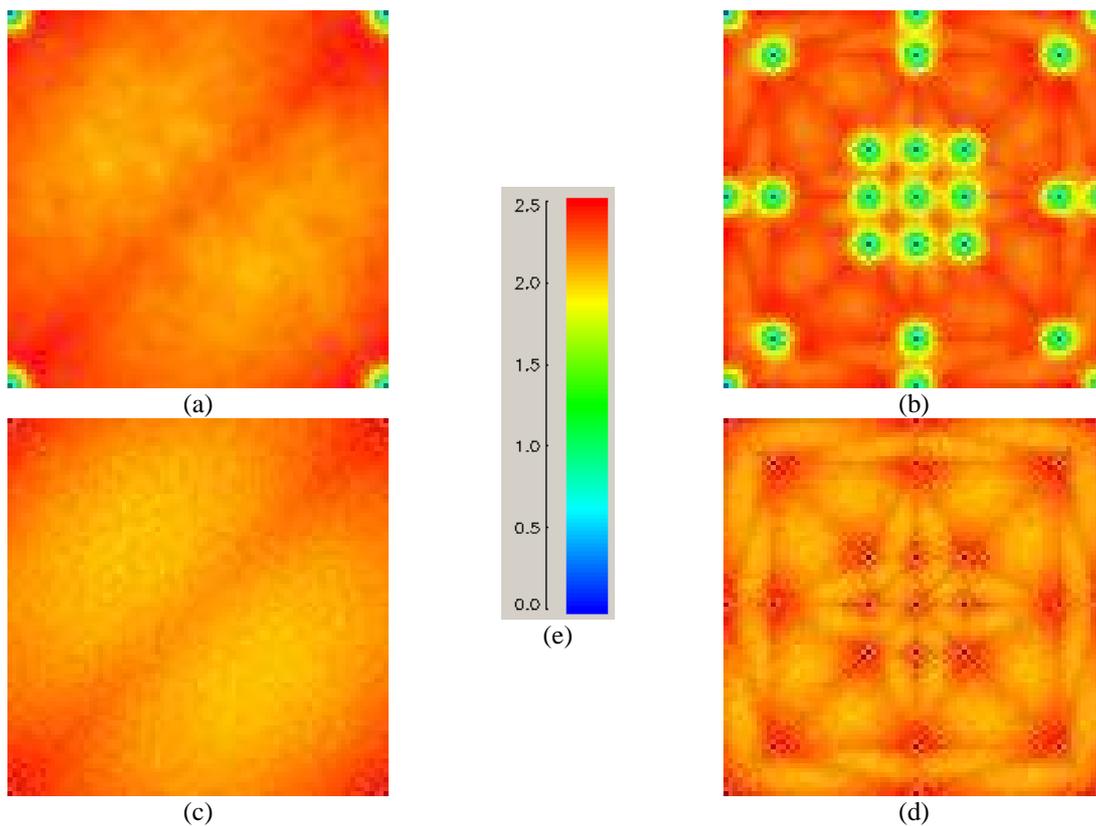


Figura 5 - Erro Absoluto Médio Local:(a) Modelo A e $\alpha = 0,07$, (b) Modelo F2 e $\alpha = 0,07$, (c) Modelo A e $\alpha = 1,7$, (d) Modelo F2 e $\alpha = 1,7$ e (e) Escala.

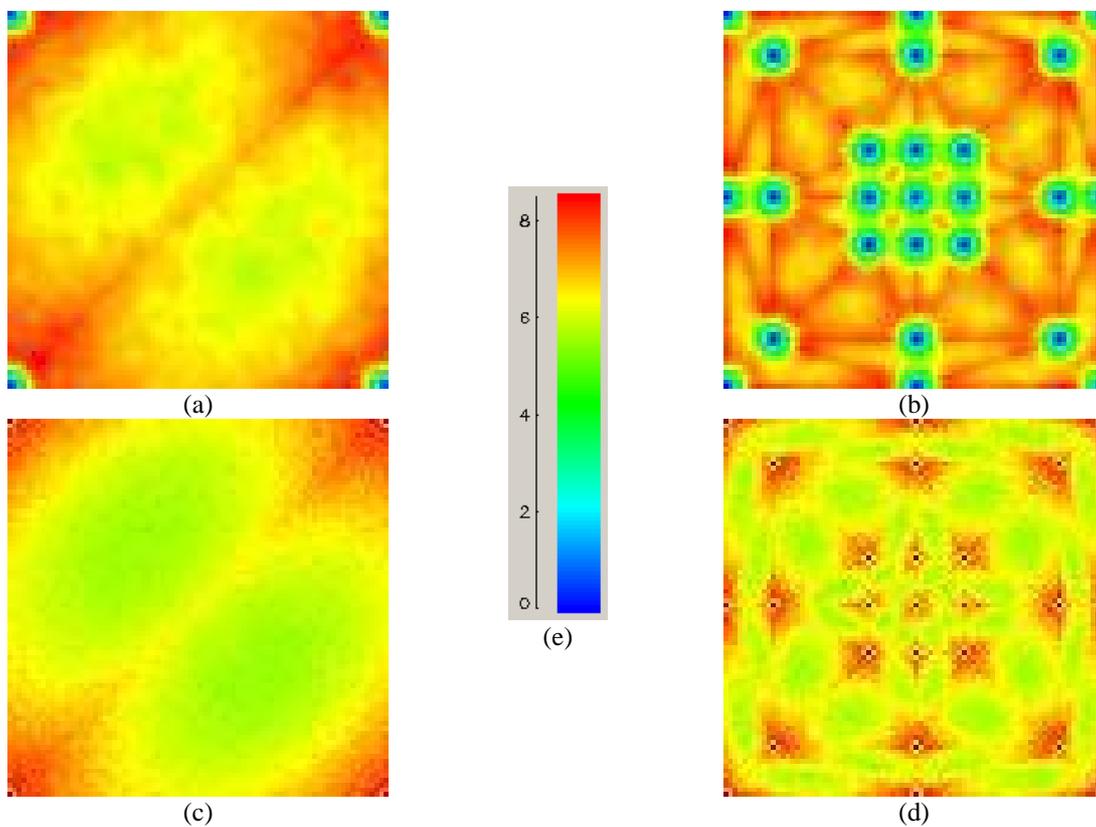


Figura 6 - Erro Médio Quadrático Local:(a) Modelo A e $\alpha = 0,07$, (b) Modelo F2 e $\alpha = 0,07$, (c) Modelo A e $\alpha = 1,7$, (d) Modelo F2 e $\alpha = 1,7$ e (e) Escala.

O comportamento das medidas RMSE e MAE para os erros locais é muito similar, a diferença básica entre elas reside na faixa de variação de cada uma. A medida MAE do erro varia entre 0 e 2,5 enquanto que a RMSE varia entre 0 e 8,5. As imagens representativas dos erros locais dos modelos A e F2 para as medidas MAE e RMSE são ilustradas, respectivamente, na **Figura 5** e na **Figura 6**. Os modelos A e F2 são os modelos que apresentaram, respectivamente, os maiores e os menores erros locais, evidenciando outra vez que com o aumento do número de pontos amostrais tende-se a diminuir os erros.

Para as superfícies com alto grau de correlação os menores erros locais ocorrem em torno dos pontos amostrais aumentando gradativamente na direção do interior dos triângulos até alcançar os seus valores máximos que ocorrem nas arestas destes triângulos. Um padrão inverso pode ser observado no caso das superfícies geradas com baixo grau de correlação, isto é, os maiores erros locais são encontrados ao redor dos pontos amostrais e vão diminuindo na direção do interior dos triângulos.

Para o outro tipo de avaliação, em que 50 modelos foram gerados com número de pontos amostrais variando de 5 a 495, os erros globais médios foram calculados, baseados também nas 5000 replicações das superfícies, a partir das medidas do MAE e do RMSE médio e de suas respectivas variâncias. Na **Figura 7** são mostrados os valores destas medidas médias com suas respectivas barras de erro referentes a um desvio padrão. Nota-se, mais uma vez, que o aumento dos pontos amostrais diminui o erro das elevações estimadas a partir dos modelos. Esta diminuição do erro é mais lenta para as superfícies geradas com baixo grau de correlação. É necessário um estudo mais detalhado para verificar se esta diminuição está relacionada à problemas de estimativas em dados correlacionados.²

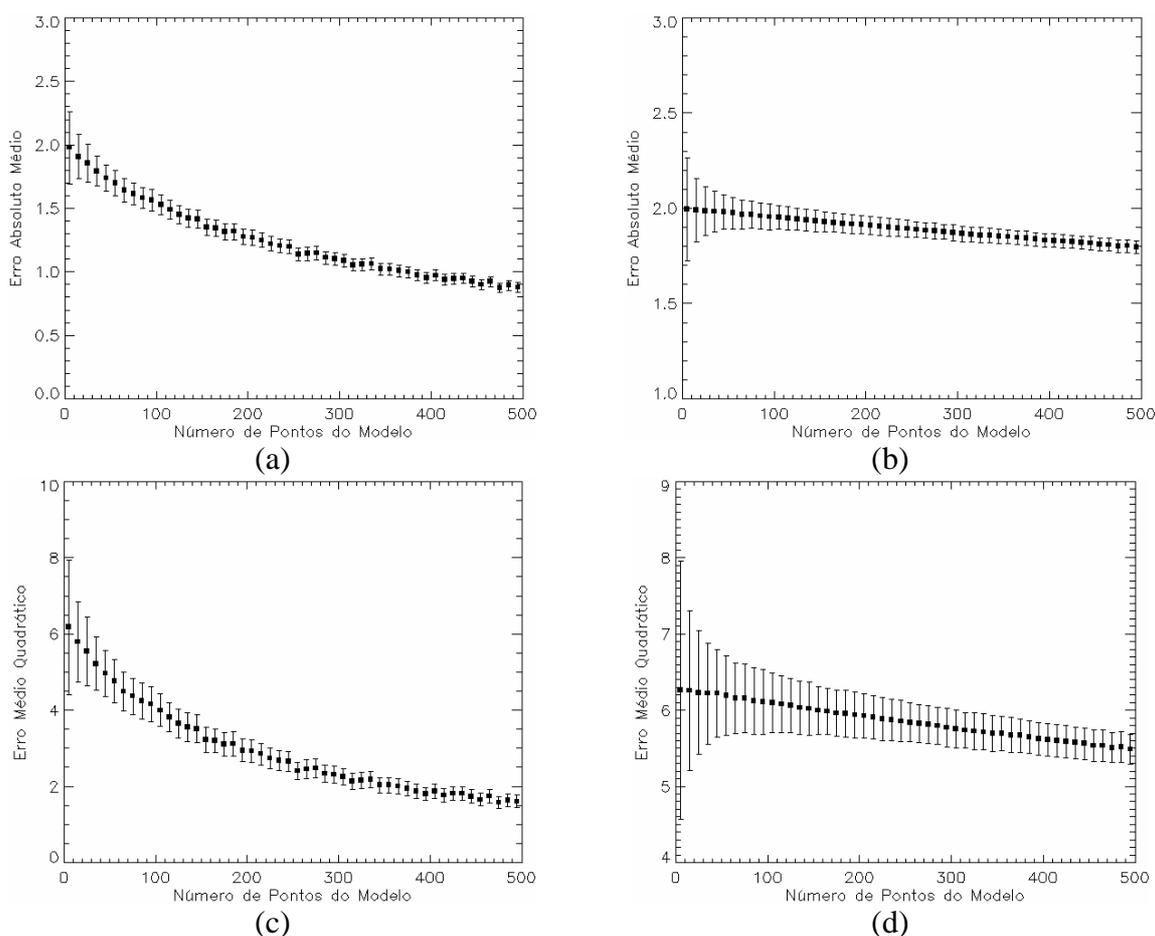


Figura 7 - Erro Médio Absoluto Global: (a) $\alpha = 0,07$ e (b) $\alpha = 1,7$ e Erro Médio Quadrático

Global (c) $\alpha = 0,07$ e (d) $\alpha = 1,7$.

Das **Figuras 7a e 7c** observa-se que a partir do valor em torno de 400 pontos amostrais existe uma estabilização dos erros, tendendo a um valor constante o que demonstra que o aumento dos pontos amostrais não influencia mais no erro do modelo. Este valor constante expressa o erro intrínseco do método adotado. Neste caso em consideração inclui a triangulação de Delaunay, a interpolação linear dos dados, área da superfície e a resolução do MDE.

4. Conclusões

De acordo com os resultados alcançados a partir deste estudo Monte Carlo notou-se que número de pontos amostrais, sua distribuição espacial e a correlação entre as elevações da superfície possuem influência sobre os erros de um modelo digital de elevação.

Nas grades irregulares triangular o aumento do número de pontos amostrais, em geral, tende a diminuir os erros oriundos dos modelos digitais de elevação. Entretanto, este aumento não pode ser indiscriminado, pois sempre haverá um erro intrínseco do modelo que será o erro mínimo alcançado. Já a distribuição destes pontos amostrais está diretamente relacionada com a forma dos triângulos produzidos na triangulação e por consequência com erro do MDE também. No caso de triangulações que originem triângulos mais próximos possíveis de triângulos equiláteros (triangulação de Delaunay) os erros do MDE são menores quando comparados com os erros calculados por triangulação que produza um número de triângulos semelhantes, porém mais agudos. A diminuição do erro é mais lenta para as superfícies geradas com baixo grau de correlação, sendo necessários mais estudos para se verificar se esta diminuição está relacionada à problemas de estimativas em dados correlacionados.

A área da superfície a ser modelada, a resolução espacial dos dados, assim como o tipo de interpolador usado nas estimativas das elevações da superfície também são parâmetros a serem levados em conta na avaliação dos erros em um MDE. A avaliação da influência destes parâmetros no erro de um MDE será alvo de trabalhos futuros.

Referências

- Câmara, G.; Davis, C.; Monteiro A.M.V. **Introdução à Ciência da Geoinformação**. 2001. Repositório da URLib: <<http://www.dpi.inpe.br/gilberto/livro/introd/>>. Acesso em: 13 de jan. de 2003.
- El-Sheimy, N. **Digital Terrain Modelling**, The University of Calgary (ENGO 573), 1999.
- Rees, W.G. The accuracy of digital elevation models interpolated to higher resolutions. **International Journal of Remote Sensing**, v.21, n.1, p. 7-20, 2000.
- Wechsler, S.P. Perceptions of Digital Elevation Model Uncertainty by DEM Users, **Urban and Regional Information Systems Association Journal**, v. 15, n. 2, 2003.
- Widayati, A.; Lusiana, B.; Suyanto, D.; Berbist Uncertainty and effects of resolution of digital elevation model and its derived features: case study of Sumberjaya, Sumatera, Indonésia In: International Symposium of Photogrammetry and Remote Sensing, 2004, Istanbul. **Proceedings...**
- Wood, J.D.; Fisher, P.F. Assessing interpolation accuracy in elevation models. **IEEE Computer Graphics and Applications**, v. 13, n. 2, p. 48-56, 1993.
- Yanasse, C.C.F. **Statistical Analysis of Synthetic Aperture Radar Images and its Applications to System Analysis and Change Detection**. 1991. 318 p. Phd. Thesis, University of Sheffield. Sheffield. 1991.