

Seleção de subconjuntos de pontos amostrais em modelos digitais de elevação

Sidnei João Siqueira Sant'Anna
Laércio Massaru Namikawa

Instituto Nacional de Pesquisas Espaciais - INPE
Caixa Postal 515 - 12227-010 - São José dos Campos - SP, Brasil
{sidnei, laercio}@dpi.inpe.br

Abstract. Two automatic feature selection methods were implemented and applied to digital terrain model (DTM) sampling points selection. These methods are based on exhaustive and selective (sequential backward feature selection) search techniques. The methods were evaluated using a subset of sampling points extracted from a set of points acquired by differential GPS and conventional techniques of topography (by using total station) in a slightly non-flat terrain. The mean error, the absolute mean error, and the root mean squared error measures were used as quality criterion of sampling point selection. The results showed that the method of feature selection can be employed to search and select DTM's sampling points presenting highly satisfactory outcomes.

Palavras-chave: geoprocessing, digital terrain model, feature selection, geoprocessamento, modelo digital de terreno, seleção de atributos.

1. Introdução

O estudo de fenômenos geográficos requer o uso de representações, sendo que para os estudos baseados em computador estas representações digitais são criadas a partir de amostras por algum processo de criação de modelos. O Modelo Digital de Terreno (MDT) é o modelo que representa o fenômeno geográfico elevação do terreno na forma digital. O procedimento de criação de MDTs consiste na aquisição, seleção e organização de amostras de elevação. Estas amostras podem ser provenientes de diversas fontes, tais como isolinhas de mapas topográficos e levantamentos realizados in-situ com equipamentos de posicionamento (por exemplo, os que se baseiam no *Global Positioning System* – GPS).

As amostras adquiridas são selecionadas e organizadas em estruturas que facilitem a manipulação, com a menor perda de informação possível. As estruturas comumente utilizadas são a grade regular retangular e a grade irregular triangular. O modelo de grade retangular regular pode ser manipulado mais facilmente do que a triangular, mas é mais ineficiente em termos de armazenamento. Isto se deve ao tamanho de cada elemento da grade retangular ser definido em função da variação de elevação necessária para uma boa representação e regiões de menor variação de elevação terão elementos redundantes.

A estrutura em grade irregular triangular, também conhecida pela sigla TIN, de *Triangulated Irregular Network*, representa a superfície por uma malha de triângulos. O procedimento de triangulação consiste em conectar os pontos amostrais por meio de segmentos de reta, que são as arestas dos triângulos, que não devem se intersectar. A eficiência no armazenamento em relação aos modelos de grade regular retangular surge devido a arestas de triângulos serem menores em regiões de grande variação e maiores onde o valor de elevação varia mais suavemente.

Para um conjunto de pontos amostrais, o número de diferentes triangulações que podem ser geradas é muito grande e algum critério deve ser aplicado para que a triangulação seja única. O critério utilizado comumente é que a soma dos comprimentos de todas as arestas de triangulação seja a menor possível, porque desta maneira a superfície interna a um triângulo depende dos pontos amostrais mais próximos a esta superfície. Apesar de a triangulação de Delaunay não satisfazer a este critério, esta triangulação está muito próxima de satisfazê-la e é aceita como a triangulação padrão para a representação de superfícies (Namikawa, 1995).

A triangulação de Delaunay é definida como sendo aquela que é a mais equilateral possível. No caso de MDTs, consideramos a triangulação planar, isto é, onde os pontos a serem conectados estão em um plano. Além disto, as arestas não devem se intersectar e as

arestas mais externas da triangulação devem formar a fronteira convexa (a fronteira do menor polígono convexo que contenha todos os pontos) dos pontos amostrais. Para este tipo de triangulação do conjunto de pontos amostrais S , a de Delaunay D é aquela onde, para todo triângulo T da triangulação, o círculo que passa pelos seus vértices V_1 , V_2 e V_3 pertencentes a S , não contém nenhum outro elemento de S .

Em geral, o MDT representado por um TIN considera que um plano está ajustado a cada triângulo, formando um poliedro de faces planas. No entanto, superfícies curvas podem ser ajustadas a cada face triangular. Para o caso de se requerer que as faces tenham continuidade C^1 entre elas (os valores da função, de direção e de magnitude da primeira derivada devem ser iguais nas duas faces ao longo das arestas, na direção perpendicular a estas arestas), a função polinomial de superfície de menor grau é uma de quinto grau (Akima, 1978).

2. Seleção de Subconjuntos de Pontos

A seleção de subconjuntos de pontos amostrais pode ser posta no mesmo contexto da seleção de atributos. Isto é, a seleção de subconjuntos de pontos amostrais pode ser interpretada como a seleção de um certo conjunto de atributos. Portanto, métodos automáticos de seleção de atributos serão empregados na seleção de subconjuntos de pontos amostrais de um MDT. Seleção automática de atributos é uma técnica de otimização, onde a partir de dado conjunto de atributos pretende-se selecionar um subconjunto. Esta seleção é realizada minimizando ou maximizando um critério de qualidade ou função objetivo.

Apesar da importância da técnica de seleção de atributos não há regras ou procedimentos definitivos em cada particular aplicação, principalmente quando o número de atributos a ser pesquisado é grande. Existem vários métodos de busca empregados para a seleção de atributos, dentre os quais encontram-se os métodos ótimos e os subótimos. O método de busca realmente ótimo é o conhecido por busca exaustiva. Isto porque, para se fazer a seleção de um subconjunto de atributos deve-se buscar em todo o espaço de atributos, ou seja em todos os subconjuntos possíveis de serem construídos a partir do conjunto inicial de atributos. Dado um conjunto inicial com N atributos o número total de subconjuntos que podem ser construídos é 2^N , o que torna inviável o procedimento de busca exaustiva em boa parte dos problemas práticos, principalmente para grandes valores de N . Este procedimento de busca, geralmente, é muito custoso computacionalmente. A fim de minimizar o custo computacional surgem os métodos de busca heurística que são métodos subótimos, pois não há garantia de que no processo de busca seja encontrado o subconjunto ótimo. Dentre os métodos subótimos, os de busca sequencial são os mais simples de serem implementados, possuem baixo custo computacional e são bastante utilizados na seleção de atributos. O método sequencial pode ser caracterizado por busca para frente (SFS – *Sequential Forward Feature Selection*) ou para trás (SBS – *Sequential Backward Feature Selection*).

Um exemplo ilustrativo do procedimento de busca no método SBS é mostrado na Figura 1, para um conjunto inicial contendo 4 atributos ($[a_1, a_2, a_3, a_4]$). Assim, dado um critério de qualidade C_Q o procedimento de busca SBS é iniciado pelo conjunto completo de atributos, a partir do qual o critério C_Q é calculado para todas as combinações de subconjuntos obtidas quando um atributo é eliminado ($[a_1, a_2, a_3,]$, $[a_1, a_2, a_4]$, $[a_1, a_3, a_4]$ e $[a_2, a_3, a_4]$). O subconjunto que apresentar o melhor valor de C_Q (valor mínimo no caso da Figura 1) é, então, selecionado e o processo continua da mesma forma até que o subconjunto com o número mínimo de atributos seja alcançado. Na Figura 1 as setas indicam o caminho seguido no procedimento de busca através do método SBS.

Uma desvantagem dos métodos sequenciais busca é que uma vez eliminado um atributo não há a possibilidade dele ser considerado em etapas posteriores do processamento (Theodoridis e Koutroumbas, 2009). Isto é conhecido como efeito *nesting*. Desta forma, se um atributo que fizer parte do subconjunto ótimo for eliminado em alguma etapa, o procedimento de busca não encontrará o subconjunto ótimo de atributos. Vale ressaltar que o número de combinações de subconjuntos a ser pesquisado é menor que na busca exaustiva.

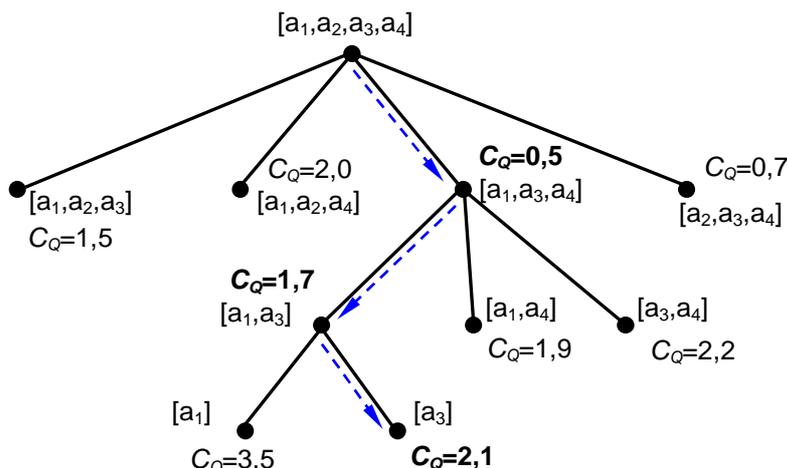


Figura 1. Gráfico ilustrativo do procedimento de busca SBS.

3. Conjunto de Pontos

O conjunto de pontos amostrais utilizado no presente estudo é proveniente do trabalho de Pierre (2005), no qual o autor avalia-se o modelo interferométrico gerado a partir das imagens nas bandas X e P de um SAR. Para tanto foram realizados 2 trabalhos de campo, nos anos de 2002 e de 2003, nos arredores da FLONA de Tapajós, situada na região sul da cidade de Santarém, Pará. Nos trabalhos de campo foram coletados um total de 55 conjuntos distintos de pontos amostrais, denominados de seções, a fim de que modelos digitais de elevação fossem gerados. Os pontos amostrais de cada seção foram levantados com o uso de GPS Geodésico de dupla frequência e por irradiação com estação total.

A seção En27, que contém 282 pontos amostrais dos quais 8 pertencem à fronteira convexa, foi escolhida para efetuar a análise dos resultados deste estudo. Pierre (2005) classifica como movimentado o tipo de relevo no qual os pontos amostrais desta seção foram levantados. A distribuição espacial dos pontos amostrais da seção En27 bem como a sua triangulação são ilustradas na Figura 2. Nesta figura também se observa a triangulação de um subconjunto de pontos amostrais extraído da seção En27, o qual é denominado por seção exEn27. A seção exEn27 contém 15 pontos amostrais dos quais 9 pontos pertencem à fronteira convexa e esta seção será utilizada para exemplificar o processo de seleção de subconjuntos de pontos amostrais. Os pontos não pertencentes à fronteira convexa serão denominados apenas por pontos internos e são numerados de 0 até 5. Esta nomenclatura de numeração tornará a formação dos subconjuntos mais evidente, quando da análise dos resultados.

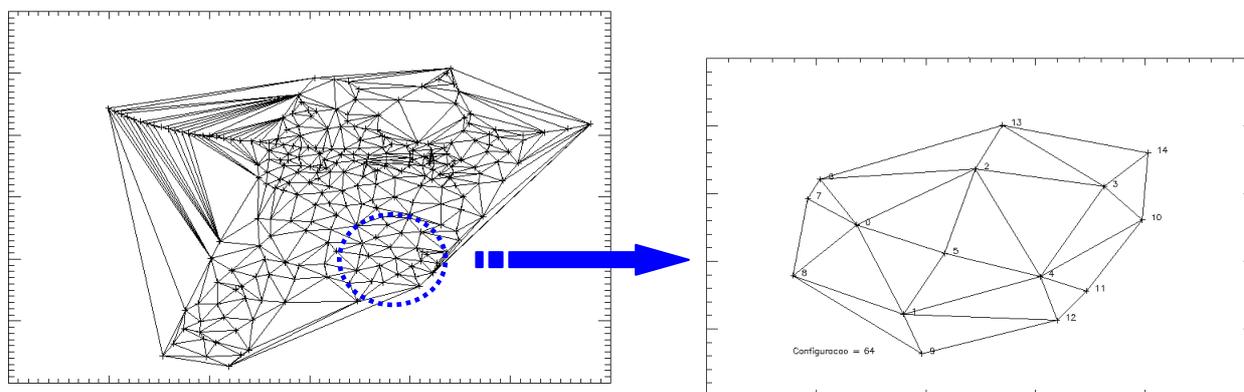


Figura 2. Triangulação dos modelos En27 e exEn27.

4. Seleção de Subconjuntos de Pontos Amostrais

A seleção de subconjuntos de pontos amostrais em MDT's neste trabalho é baseada nos métodos de seleção de atributos, como já mencionado na Seção 2. Para tanto, foram implementado 2 algoritmos de busca, um de busca exaustiva e outro de busca sequencial para trás (SBS). A primeira etapa destes algoritmos é a identificação, no conjunto dos pontos amostrais, dos pontos pertencentes à fronteira convexa. Estes pontos sempre estarão presentes em todos os subconjuntos de pontos amostrais pesquisados. O processo de busca se inicia com os pontos interiores, gerando-se um MDT para cada subconjunto a ser avaliado. Cada MDT obtido (modelo avaliado – MA) é comparado ao MDT gerado a partir de todos os pontos amostrais (modelo de referência – MR). Na geração do MDT usa-se a triangulação de Delaunay e a interpolação por suavização quártica.

A diferença entre MA e MR ($e = MA - MR$) define uma superfície de erro entre os modelos, da qual se originará 3 medidas de avaliação: o erro médio (EM), o erro absoluto médio (EAM) e o erro médio quadrático ($RMSE$). Estas medidas são utilizadas como critério de qualidade (C_Q) no processo de busca e, para uma superfície contendo M pontos, são dadas, respectivamente, por:

$$EM = \frac{1}{M} \sum_{m=1}^M e_m \quad (1)$$

$$EAM = \frac{1}{M} \sum_{m=1}^M |e_m| \quad (2)$$

$$RMSE = \sqrt{\frac{1}{M} \sum_{m=1}^M e_m^2} \quad (3)$$

Nos processos de busca implementados procura-se um subconjunto de pontos amostrais que minimize estes critérios de qualidade.

5. Análise dos Resultados

O modelo exEn27 será utilizado para análise dos resultados obtidos com os 2 procedimentos de busca implementados. Com este modelo que contém apenas 6 pontos amostrais interiores (elementos) à fronteira convexa, é possível formar 64 (2^6) subconjuntos de pontos amostrais interiores. Cada um destes subconjuntos juntamente com o conjunto de pontos da fronteira convexa forma um subconjunto de pontos amostrais com o qual é realizada uma triangulação para análise. Os subconjuntos avaliados são denominados de configuração, isto é, para o modelo exEn27 são avaliadas um total de 64 configurações distintas.

A partir do procedimento de busca exaustiva realizada com o modelo exEn27 determinaram-se os erros EM , EAM e $RMSE$, equações (1)-(3), de todas as 64 configurações possíveis de serem geradas. Estes erros são mostrados na Figura 3, onde se observam também barras erros nos valores dos erros EM e EAM medidos, as quais são representadas por um desvio padrão da medida. Cada gráfico da Figura 3 é dividido por linhas verticais tracejadas que delimitam as configurações de subconjuntos de pontos amostrais interiores contendo o mesmo número de elementos. Desta figura pode-se notar a existência de uma tendência geral dos erros diminuir com o aumento do número de pontos amostrais. No entanto, o maior número de pontos (associado ao maior número de triângulos da triangulação) não implica necessariamente num menor erro, demonstrando que há pontos melhores posicionados dentro do modelo, isto é, os pontos ótimos para que se tenham os menores erros no modelo a ser ajustado. Este fato pode ser verificado, por exemplo, pela configuração 27 que contém 3 elementos e possui erro maior que várias das configurações contendo apenas 2 elementos. Da Figura 3 observa-se também que o erro médio entre os modelos avaliado e de referência é

sempre negativo implicando que o modelo com menor número de pontos amostrais tende a subestimar as altitudes das superfícies estimadas.

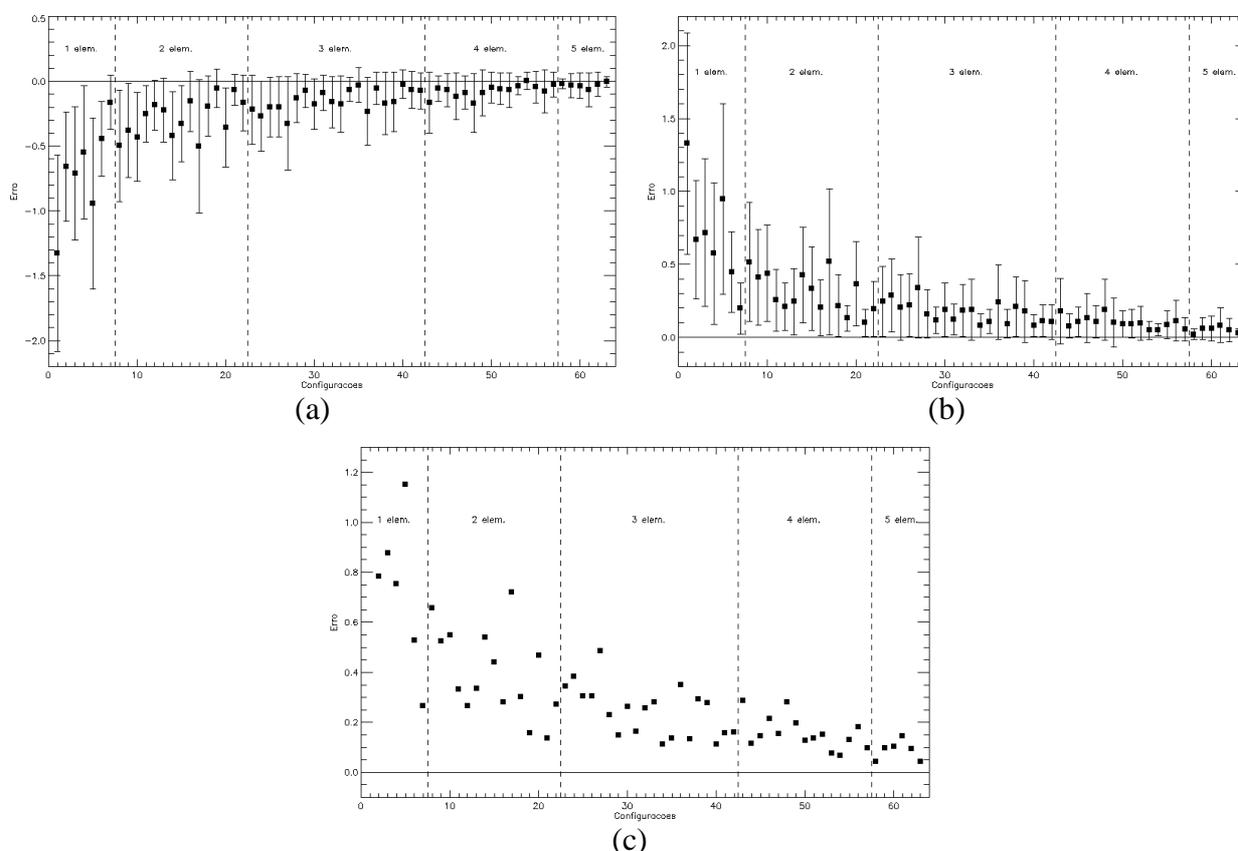


Figura 3. Erros dos modelos avaliados na busca exaustiva: (a) *EM*, (b) *EAM* e (c) *RMSE*.

Os subconjuntos de pontos interiores que apresentaram os menores erros na busca exaustiva, ou seja, os subconjuntos ótimos, podem ser vistos listados na Tabela 1, para cada medida de erro, bem como para o número de elementos contidos em cada subconjunto. Nesta tabela nota-se que os subconjuntos ótimos podem variar de critério de qualidade para critério, uma vez que o processo de busca é fortemente dependente do critério de qualidade adotado. Portanto, ressalta-se aqui a grande importância da escolha de um critério razoavelmente robusto para o processo de busca de um subconjunto ótimo. Desta forma, um critério de qualidade que unifique as 3 medidas de erros (*EM*, *EAM* e *RMSE*) pode ser visto como um critério mais robusto para ser usado no processo de busca. Este critério é denominado de unificado na Tabela 1 e é definido para selecionar o subconjunto ótimo de pontos que tenha a maior frequência de ocorrência considerando conjuntamente as 3 medidas de erro analisadas. Na Tabela 1 o número que aparece entre parênteses no critério unificado refere-se a configuração selecionada, isto quer dizer que o subconjunto de pontos [1,2,3,4,5] (configuração 63) é o subconjunto ótimo contendo 5 elementos selecionado pelo critério unificado. Na Figura 4 são ilustradas as triangulações das 5 configurações selecionadas com o critério unificado, bem com a triangulação da configuração com o pior resultado (configuração 5).

Tabela 1. Subconjuntos selecionados em cada critério de qualidade através da busca exaustiva.

	5	4	3	2	1
<i>EM</i>	[1,2,3,4,5]	[1,2,3,5]	[2,3,5]	[2,5]	[5]
<i>EAM</i>	[0,1,2,3,4]	[1,2,3,4]	[1,2,4]	[3,5]	[5]
<i>RMSE</i>	[1,2,3,4,5]	[1,2,3,5]	[1,2,4]	[3,5]	[5]
UNIFICADO	[1,2,3,4,5] ⁽⁶³⁾	[1,2,3,5] ⁽⁵⁴⁾	[1,2,4] ⁽³⁴⁾	[3,5] ⁽²¹⁾	[5] ⁽⁷⁾

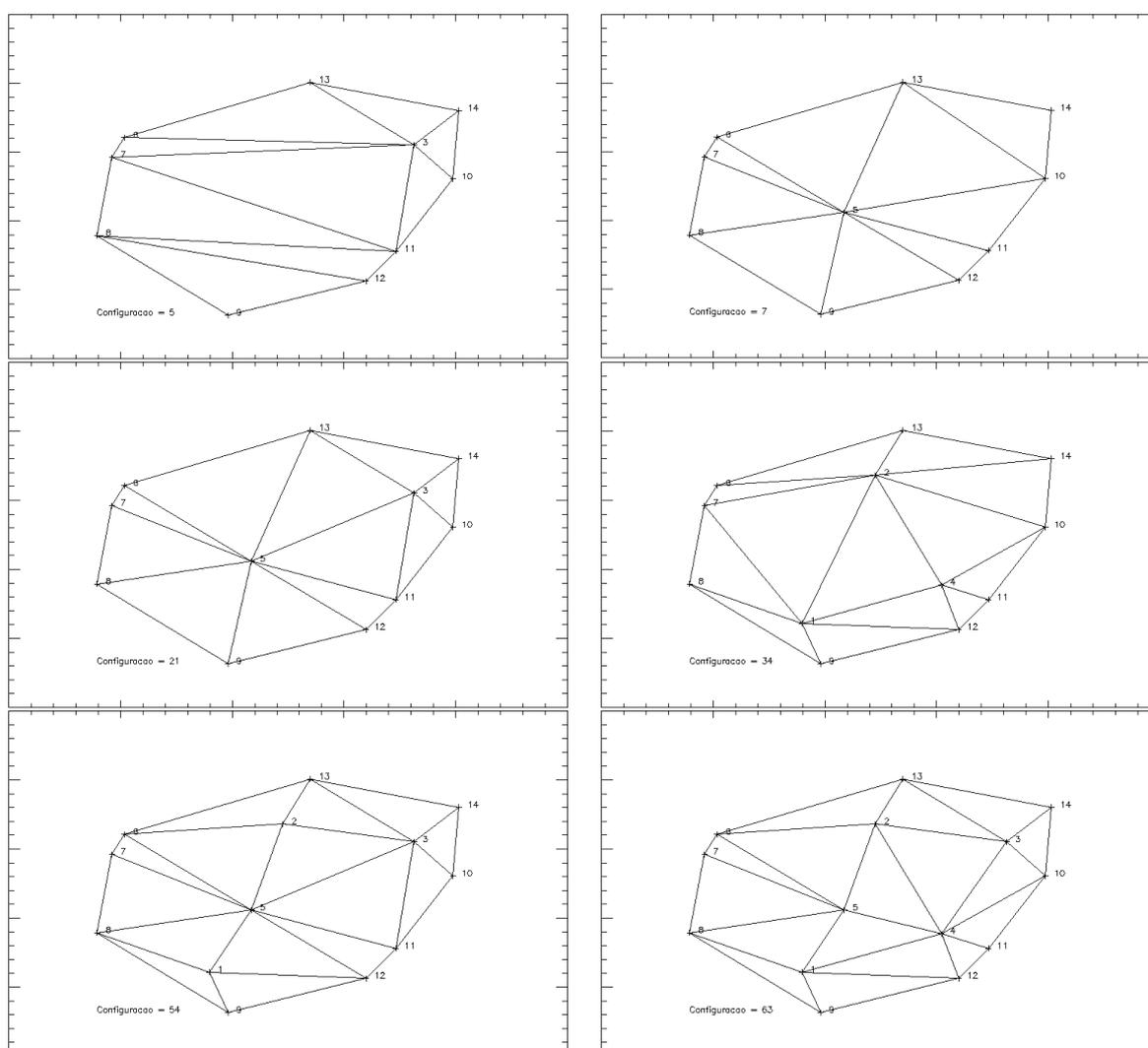


Figura 4. Triangulações dos subconjuntos de pontos selecionados.

O processo de busca exaustiva de subconjuntos de pontos em problemas encontrados na prática, onde o número de pontos amostrais é da ordem de milhares de pontos, pode tornar-se computacionalmente inviável. Portanto, implementou-se um processo de busca seletiva baseado no método de busca sequencial para trás. Os subconjuntos de pontos amostrais interiores do modelo exEn27 selecionados através da busca seletiva são mostrados na Tabela 2 para os 3 critérios de qualidade adotados, assim como para o critério unificado.

Tabela 2. Subconjuntos selecionados em cada critério de qualidade através da busca seletiva.

	5	4	3	2	1
EM	[1,2,3,4,5]	[1,2,3,5]	[2,3,5]	[2,5]	[5]
EAM	[0,1,2,3,4]	[1,2,3,4]	[1,2,4]	[2,4]*	[4]*
RMSE	[1,2,3,4,5]	[1,2,3,5]	[2,3,5]*	[3,5]	[5]
UNIFICADO	[1,2,3,4,5]	[1,2,3,5]	[2,3,5]	**	[5]

No processo de busca seletiva o subconjunto contendo $n-1$ elementos sempre será um subconjunto do conjunto contendo n elementos selecionado na etapa anterior. Fato este que não necessariamente ocorrerá com o método de busca exaustiva. Portanto, dependendo do critério de qualidade escolhido no processo de busca seletiva nem sempre irá se encontrar o

subconjunto ótimo de pontos, fato este já mencionado na Seção 2. Esta situação é ilustrada pelos subconjuntos marcados com um asterisco na Tabela 2.

No geral, os valores de cada critério de qualidade obtidos para os subconjuntos de pontos contendo o mesmo número de elementos não possui uma diferença muito grande. Alguns destes valores, em unidade de metros, podem ser vistos na Tabela 3, onde estão listados as configurações que obtiveram os mínimos (valores em negrito) e os máximos valores em cada critério de qualidade bem como as configurações que foram selecionadas pelo método de busca seletivo. Nesta tabela os desvios padrões das medidas de *EM* e *EAM* também são listados.

Tabela 3. Erros de algumas configurações selecionadas.

Config.	Pontos Interior	<i>EM</i>	σ_{EM}	<i>EAM</i>	σ_{EAM}	<i>RMSE</i>
1	-	-1,328	0,757	1,328	0,757	1,529
5	[3]	-0,944	0,659	0,948	0,653	1,151
6	[4]	-0,443	0,287	0,450	0,276	0,527
7	[5]	-0,163	0,209	0,199	0,175	0,265
17	[2,3]	-0,502	0,516	0,518	0,500	0,720
18	[2,4]	-0,192	0,234	0,216	0,211	0,302
19	[2,5]	-0,055	0,149	0,131	0,089	0,159
21	[3,5]	-0,067	0,119	0,099	0,093	0,136
27	[0,2,3]	-0,325	0,362	0,341	0,346	0,486
34	[1,2,4]	-0,063	0,094	0,081	0,079	0,113
40	[2,3,5]	-0,023	0,109	0,083	0,075	0,112
43	[0,1,2,3]	-0,164	0,236	0,179	0,225	0,288
48	[0,1,4,5]	-0,168	0,226	0,189	0,209	0,282
53	[1,2,3,4]	-0,036	0,069	0,049	0,060	0,077
54	[1,2,3,5]	0,004	0,067	0,053	0,042	0,067
58	[0,1,2,3,4]	-0,018	0,039	0,021	0,037	0,042
61	[0,1,3,4,5]	-0,064	0,130	0,082	0,120	0,145
63	[1,2,3,4,5]	-0,005	0,042	0,029	0,030	0,042

Por exemplo, as configurações 19 e 21 apresentam valores em cada critério de qualidade muito próximos entre si, suscitando a questão de quão diferentes estes valores são. Desta forma, um teste de diferença de média poderia ser utilizado para melhorar o processo de seleção. Ou ainda, um teste de diferença de médias poderia ajudar na seleção de um subconjunto de pontos numa situação similar aquela ilustrada pelos 2 asteriscos na Tabela 2, onde cada critério de qualidade fornece um subconjunto diferente de pontos.

O processo de busca seletiva também foi empregado no conjunto completo de pontos amostrais da seção En27 (282 pontos amostrais, sendo 8 pertencentes à fronteira convexa). As triangulações referentes a subconjuntos de pontos internos contendo 10, 50 e 100 elementos são mostradas na Figura 5, para cada uma das medidas de erro. Neste caso de seleção de busca seletiva foi encontrada uma configuração distinta para cada critério de qualidade, mostrando uma vez mais a necessidade de um critério único para a seleção de pontos amostrais. No entanto, da Figura 5 pode-se notar que as regiões interiores à fronteira convexa onde se encontram os pontos amostrais selecionados em cada critério são semelhantes. Pode-se observar ainda que as regiões em que se encontram os menores números de pontos selecionados estão presentes naquelas selecionadas com um número maior de pontos. Isto é, a região que contém os 10 pontos internos está presente na região que contém os 50 pontos internos, que está presente na região que contém os 100 pontos internos selecionados. Contudo, estas regiões são um pouco diferentes dependendo do critério de qualidade adotado, ou seja, o adensamento de pontos amostrais selecionados é distinto para cada critério.

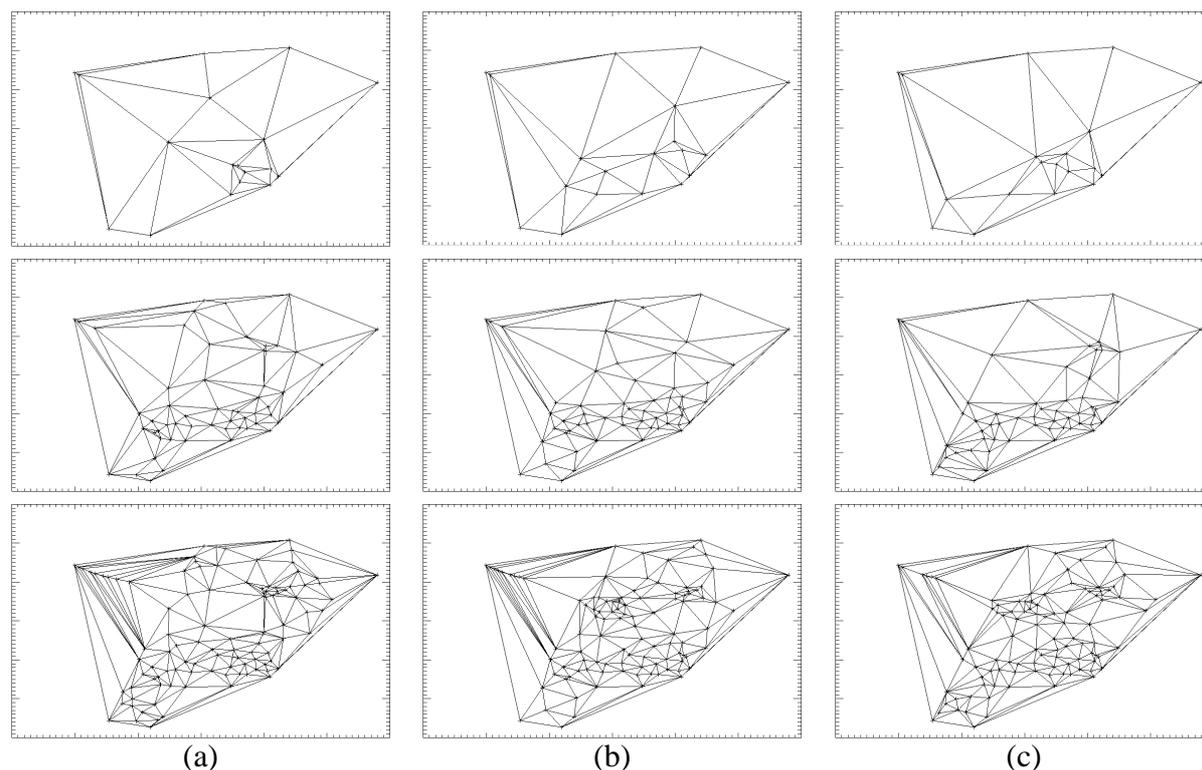


Figura 5. Triangulações da seção En27 contendo 10, 50 e 100 pontos internos obtidas com o método SBS e baseadas nos critérios: (a) *EM*, (b) *EAM* e (c) *RMSE*.

Para melhorar o processo de busca poder-se-ia utilizar um limiar de erro mínimo aceitável para que o subconjunto em questão fizesse parte da etapa de seleção. Isto é, somente seriam usados no processo de busca aqueles subconjuntos cujo erro fosse menor que um certo valor, por exemplo, 0,01m. Ou ainda, a utilização de outros critérios de qualidade, tais como as medidas de aspecto e/ou declividade em MDT's.

6. Conclusões

Neste trabalho foram implementados dois métodos de busca e seleção de subconjuntos de pontos amostrais em modelos digitais de terreno (MDT). Métodos estes baseados nas buscas exaustiva e seletiva (busca sequencial para trás – SBS). Os resultados foram avaliados em um extrato de um modelo de pontos adquiridos por GPS geodésico de dupla frequência e por irradiação com estação total, assim como o método SBS também foi aplicado ao conjunto completo de pontos amostrais. Os resultados mostraram que o processo de busca seletiva de atributos pode ser empregado como técnica de seleção de pontos amostrais em MDT's. Para a continuidade deste trabalho vislumbra-se a utilização de outros critérios de qualidade e também a implementação de outros métodos de busca e seleção de pontos amostrais.

Referências Bibliográficas

- Akima, H. A method of bivariate interpolation and smooth surface fitting for irregularly distributed data points. **ACM Transactions on Mathematical Software**, v. 4, n. 2, p. 148-159, 1978.
- Moura, P. Calibração dos modelos digitais de elevação gerados por interferometria SAR nas Bandas P e X. 2005. 238p. (INPE-15241-TDI/1328). Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2005.
- Namikawa, L.M. Um método de ajuste de superfície para grades triangulares considerando linhas características. 1995. 136p. Dissertação (Mestrado em Computação Aplicada) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 1995.
- Theodoridis, S.; Koutroubas, K. **Pattern Recognition**. London, Academic Press, 2009. 961 p.