



Revista Brasileira de Cartografia (2014) N<sup>o</sup> 66/5: 1039-1050  
Sociedade Brasileira de Cartografia, Geodésia, Fotogrametria e Sensoriamento Remoto  
ISSN: 1808-0936

## **ANÁLISE DE AGRUPAMENTOS DE BACIAS HIDROGRÁFICAS COMO PROCEDIMENTO-GUIA PARA O ESTUDO E A AMOSTRAGEM DE COBERTURAS SUPERFICIAIS**

*Cluster Analysis of Watersheds as a Guide-Procedure for Study and Sampling of Regoliths*

**Chrystiann Lavarini, Antônio Pereira Magalhães Júnior & Fábio Soares de Oliveira**

**Universidade Federal de Minas Gerais – UFMG**  
**Instituto de Geociências, Programa de Pós-Graduação em Geografia**  
Antônio Carlos, 6.627, Pampulha - CEP 31270-901. Belo Horizonte, MG, Brasil.  
lavarini.c@gmail.com; magalhaesufmg@yahoo.com.br; fabiosolos@gmail.com

*Recebido em 26 de Dezembro, 2013/ Aceito em 27 de Março, 2014*  
*Received on December 26, 2013/ Accepted on March 27, 2014*

### **RESUMO**

Conhecendo-se a priori as variáveis, ou fatores, que influenciam a gênese das coberturas superficiais, em conjunto à distribuição espacial destes elementos, é possível deduzir informações valiosas para as etapas posteriores de campo via análises estatísticas e geoprocessamento. Assim, insere-se este trabalho que objetiva, em termos gerais, demonstrar a possibilidade de agrupamentos de bacias hidrográficas a partir dos fatores de formação dos solos e sua subsequente relevância no diagnóstico de campo. Para este fim, foram extraídas no software ArcGIS 10 as variáveis morfométricas amplitude altimétrica, curvatura horizontal, curvatura vertical, declividade e integral hipsométrica referentes às bacias de cabeceira do Alto-Médio do Curso do Ribeirão Mata Porcos, Quadrilátero Ferrífero, MG, associadas a técnicas de agrupamentos amplamente conhecidas, como a Análise de Agrupamentos Hierárquica e Não-Hierárquica. Em suma, constatou-se que os Agrupamentos Hierárquicos foram relevantes, sobretudo pela oportunidade de análises exploratórias dos grupos, elementos e variáveis. Através do dendograma gerado foi possível observar oito grupos que melhor descreveriam os dados, no entanto, estes grupos se reduziram, através de inúmeros testes de análise não-hierárquica, para apenas sete estatisticamente consistentes. Em termos empíricos, foram encontradas em campo evidências morfológicas e/ou pedoestratigráficas que registram a complexidade regional. Ademais, por serem dados agrupados naturalmente, com menor intencionalidade e subjetividade humana, os resultados oriundos das análises de agrupamento mencionadas tendem a representar a complexidade existente de forma mais fidedigna do que a simples análise de produtos cartográficos.

**Palavras chaves:** Análise de Agrupamentos, Diagnóstico de Campo, Contribuição Metodológica, Ribeirão Mata Porcos, Quadrilátero Ferrífero.

### **ABSTRACT**

Knowing a priori the variables or factors which influence the genesis of regoliths, in association with them spatial distribution, it is possible to deduce relevant information for the next steps on field via statistical analysis and geoprocessing. Thus, this paper aims to demonstrate the possibility of watersheds clustering from the factors of soil formation as well as its subsequent relevance in exploratory fields. For these purpose, we extracted in ArcGIS 10 morphometric

variables such as relief, horizontal curvature, vertical curvature, slope and hypsometric integral from the Upper-Middle Course of Ribeirão Mata Porcos headwaters, Quadrilátero Ferrífero, Minas Gerais state, Brazil. These variables were joined by techniques clustering widely known as Hierarchical Cluster and Non-Hierarchical Analysis. The results showed that the Hierarchical Cluster were relevant, especially by exploratory group, elements and variables analyzes. According to the dendrogram we observed eight groups which best represent the data, however, these groups were reduced to seven statistically consistent groups based on several non-hierarchical analysis tests. Empirically, we found morphological evidence and/or pedostratigraphical records reflecting the regional complexity. Moreover, since the results of cluster analysis described above are naturally grouped data, with less intentionality and human subjectivity, they tend to represent the complexity existing much more reliably than the simple analysis of cartographic products.

**Keywords:** Cluster Analysis, Field Analysis, Methodological Contribution, Mata Porcos Stream Channel, Iron Quadrangle.

## 1. INTRODUÇÃO

O trabalho do geocientista frequentemente inicia-se muito antes dos trabalhos de campo. No caso dos pesquisadores em pedologia e geomorfologia são, sobretudo, nas etapas realizadas em gabinete que se adquirem concepções prévias sobre a diversidade de coberturas superficiais que poderão ser observadas em campo (SCHAETZL e ANDERSON, 2005). Tradicionalmente, a interpretação de gabinete ocorre por meio de análises de cartas topográficas, geológicas e de cobertura vegetal (OLLIER e PAIN, 1996), uma vez que, correspondendo ao relevo, material de origem e organismos, tais mapeamentos representam os fatores de formação do solo dominantes em meso-escalas de análise (JENNY, 1944).

Os recentes avanços tecnológicos das ciências têm permitido a extração de dados em diversas escalas, assim como uma melhoria significativa da qualidade dessas informações (SHEPPARD, 2001). Nesta perspectiva, desde a década de 1960, favorecidas pelos avanços na escala de análise microscópica, assim como pelos dados de sensoriamento remoto, as ciências da Terra puderam compreender os fenômenos de seus objetos em uma perspectiva multi-escalar e, conseqüentemente, mais abrangente (VIDAL-TORRADO *et al.*, 2005).

Por outro lado, ao reconhecer a complexidade multi-escalar geossistêmica, esta nova abordagem, ao mesmo tempo em que suscita a importância de um maior aporte de dados, revela, também, a impossibilidade de quantificá-los em sua totalidade, originando a necessidade de agrupá-los segundo uma ordem (HARVEY, 1969). Visando estes objetivos, uma das técnicas estatísticas mais robustas é a análise

de agrupamentos (*cluster analysis*), que permite, entre outras coisas, organizar um grande número de dados em grupos com maior homogeneidade, assim como testar hipóteses sobre a qualidade dos agrupamentos realizados (EVERITT *et al.*, 2001).

Desta forma, conhecendo-se *a priori* as variáveis, ou fatores, que influenciam a gênese das coberturas superficiais, em conjunto à distribuição espacial destes elementos, é possível deduzir informações valiosas para as etapas posteriores de campo via análises estatísticas e geoprocessamento.

Considerando-se estas relevantes informações, insere-se este trabalho que objetiva, em termos gerais, analisar a possibilidade de agrupamentos de bacias hidrográficas a partir dos fatores de formação dos solos e sua subsequente relevância no diagnóstico de campo. Como objetivos específicos, pretende-se (i) testar a análise de agrupamentos hierárquica e de k-médias em uma bacia de drenagem fortemente heterogênea em seus aspectos litológicos e geomorfológicos; (ii) comparar e discutir os dados obtidos a partir dos diferentes métodos de agrupamento utilizados; e, em último nível, (iii) verificar em campo a relevância dos agrupamentos produzidos em gabinete.

## 2. ÁREA DE ESTUDOS

Como pressuposto elementar das classificações e agrupamentos estatísticos reconhece-se que quanto maior a quantidade de elementos e variáveis mensuradas, mais difícil e complexa se torna a ordenação lógica dos fenômenos (ROMESBURG, 1984). Especificamente no caso desta pesquisa, se poderia deduzir, portanto, que bacias hidrográficas

com maior diversidade morfológica e litológica tendem a tornar as tentativas de agrupamento de coberturas ainda mais complexas.

Neste sentido, insere-se o Alto-Médio Curso do Ribeirão Mata Porcos, que pode ser concebido como uma área de estudos privilegiada, dada a sua grande heterogeneidade. Estas características permitem a formação de agrupamentos de suas sub-bacias e, ao mesmo tempo, verificar e testar a robustez deste procedimento em uma área ambientalmente complexa.

Em síntese, o Alto-Médio Curso do Ribeirão Mata Porcos, apresenta 163 km<sup>2</sup> de área drenada e está localizado no Sinclinal Moeda, uma megaestrutura que compõe a borda oeste

do Quadrilátero Ferrífero, na região centro-sul de Minas Gerais (Fig.1).

Este Sinclinal é composto por rochas supracrustais Paleoproterozóicas do Supergrupo Minas, essencialmente itabiritos e quartzitos nas bordas W e E, com registros de lateritas também nas bordas (CPRM, 2010), e filitos na porção central (REIS *et al.*, 2002). Em termos geoestruturais, superpõem-se desde estruturas anteriores aos planos de metamorfismo dúctil com vergência W, até o truncamento destas últimas por um grande número de falhas transcorrentes E-W do Ciclo Brasileiro (CHEMALE JR. *et al.*, 1994).

Morfologicamente, predominam dois grandes compartimentos, um referente às frentes

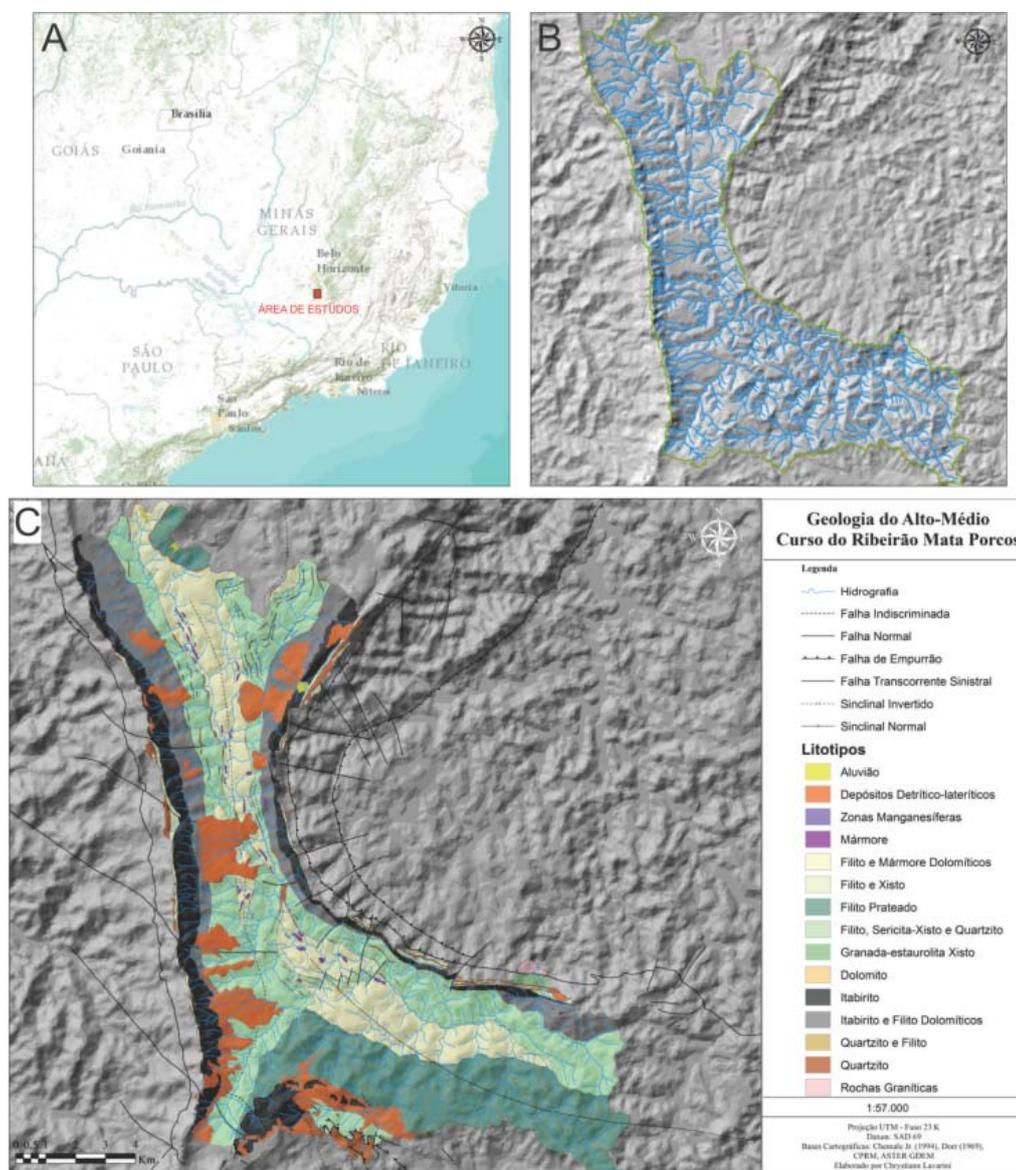


Fig. 1 - (A) Localização da bacia do Ribeirão Mata Porcos (em vermelho); (B) configuração do relevo regional; (C) compartimentação litoestrutural da bacia investigada. Bases cartográficas fornecidas pelo Instituto Mineiro de Gestão das Águas (IGAM) e Serviço Geológico do Brasil (CPRM).

de dissecação declivosas das bordas leste e oeste, e outro de colinas convexas alongadas, na porção central, com valores de declividade menores ao sul do Sinclinal e à montante da bacia do Ribeirão Mata Porcos.

Segundo a classificação climática de Köppen, a região apresenta clima mesotérmico com temperaturas moderadas, verão quente e chuvoso, e inverno frio e seco (Cwb).

### 3. METODOLOGIA

Conforme Milligan (1996), para a construção de agrupamentos consistentes é necessária a seguinte sequência de passos: (i) identificar os elementos a serem agrupados, (ii) selecionar as variáveis destes elementos; (iii) padronizar as variáveis; (iv) estabelecer uma medida de similaridade ou dissimilaridade entre os grupos; (v) selecionar o método de agrupamento; (vi) determinar o número ótimo de grupos; e, por fim, (vii) interpretar e testar os resultados.

Para facilitar o reconhecimento de possíveis padrões da cobertura superficial em uma etapa pré-campo, foram considerados os fatores de formação do solo relevo e litologia como as variáveis dos elementos a serem agrupados estatisticamente. Os fatores organismos e clima não foram considerados na análise, uma vez que a heterogeneidade no local aparentemente não é expressiva.

Na caracterização do relevo, foram utilizados produtos orbitais do SRTM interpolados para 30 m de resolução espacial, provenientes do projeto Topodata do INPE. O processamento inicial dos dados geoespaciais foi feito em ambiente computacional através do software *ArcGIS 10*, onde foram extraídas as variáveis morfométricas amplitude altimétrica, curvatura horizontal, curvatura vertical e declividade. Além destas, calculou-se também a Integral Hipsométrica, que possibilita deduzir o grau de dissecação e evolução geomorfológica de uma determinada bacia (SINGH *et al.*, 2008).

Em relação à litologia, foram consideradas como variáveis as cinco litologias principais do Ribeirão Mata Porcos, representando: filitos, quartzitos, itabiritos, cangas e mármore, conforme mapeamento realizado pelo CPRM (2010). Estas variáveis categóricas foram substituídas por algarismos entre 1 e 5, onde cada

litologia corresponderia a um determinado valor.

Os elementos detentores das variáveis anteriormente descritas são as bacias de primeira e segunda ordem, também denominadas bacias de cabeceira, ou *headwaters* (STRAHLER, 1952). No software *ArcGIS 10*, as bacias de cabeceira foram extraídas da base hidrográfica do IGAM (escala 1:50.000) através da seleção automática, em banco de dados, de bacias iguais ou inferiores à segunda ordem.

Esta escolha se justifica pela facilidade e possibilidade de, em uma posterior etapa de campo, utilizar tais bacias como elementos-chave para amostragens representativas de ambientes que se repetem no espaço. Além disto, os cálculos iniciais demonstraram que aproximadamente 80% da bacia do Ribeirão Mata Porcos é composta por bacias de cabeceira. Consequentemente, este procedimento possibilita, dentro de certos limites, a extrapolação dos dados em nível de detalhe, típicos das análises de coberturas, para escalas de análise em contexto regional, gerando uma investigação multi-escalar.

Desta forma, o total de bacias de cabeceira analisadas perfizeram 247 elementos detentores de 11 variáveis cada (Fig. 2).

As variáveis do relevo e da litologia, extraídas para as bacias de cabeceira, foram posteriormente trabalhadas em planilhas do *Excel 2010*, sendo padronizadas ( $Z$ ) segundo a equação:

$$Z=(X-\mu)/\sigma \quad (1)$$

onde,  $X$  é o valor a ser padronizado;  $\mu$  é a média aritmética da variável; e  $\sigma$  é o desvio padrão.

A padronização se caracteriza como um

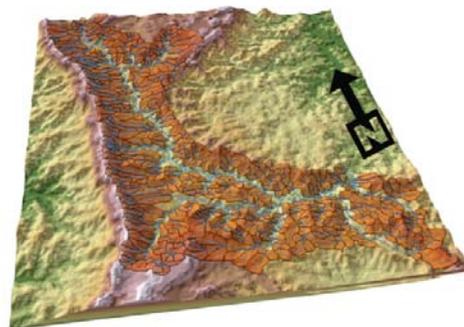


Fig. 2 – Localização das 247 bacias de cabeceira analisadas (em laranja). Bases cartográficas fornecidas pelo INPE e IGAM.

elemento obrigatório para anular o efeito de escala, tendo em vista que variáveis em escalas de mensuração diferentes podem gerar resultados falsos (EVERITT *et al.*, 2001; ROMESBURG, 1984).

Em seguida, os dados padronizados foram exportados para tratamento estatístico no *software SPSS 20*, sendo feitas as análises de agrupamentos hierárquica e não-hierárquica.

Em função do desconhecimento do número de grupos existentes, foi utilizada primeiramente a Análise de Agrupamentos Hierárquica (*Hierarchical Cluster Analysis*). Neste método cada elemento representa, inicialmente, um grupo por si só, nos quais, em etapas posteriores, são agrupados em função da dissimilaridade, ou distância, entre suas variáveis. A medida de dissimilaridade entre os grupos foi realizada com base no Método de Ward, reconhecido por ser bastante eficiente, sobretudo, por utilizar análises de variância e produzir os menores aumentos possíveis de erro (MILLIGAN, 1996).

Após a utilização do Agrupamento Hierárquico, e do reconhecimento da quantidade de possíveis grupos, realizou-se a Análise de Agrupamentos Não-Hierárquica a partir do método de K-Médias (*K-means*). Este método permite formar grupos, e posteriormente testá-los a partir de Testes-F da Análise de Variância, que demonstram a adequação ou não dos agrupamentos gerados (ALDENDERFER e BLASHFIELD, 1984; TIMM, 2002).

Finalizados os processamentos estatísticos, os agrupamentos gerados foram exportados para o *software ArcGIS 10*, sendo construídos os produtos finais (mapas) necessários aos trabalhos de reconhecimento iniciais em campo.

Desta forma, como última etapa deste trabalho, foi conduzido um trabalho de campo inicial para reconhecimento de semelhanças da cobertura superficial nos grupos gerados, assim como de constatação da heterogeneidade entre grupos diferentes.

#### 4. RESULTADOS

De uma maneira geral, os resultados obtidos a partir das duas análises de agrupamento realizadas – hierárquica e não-hierárquica – permitiram explorar as características dos dados e, conseqüentemente, contribuir para a seleção dos grupos de amostragem. Através

da espacialização em SIG e da verificação de campo, constatou-se que as coberturas pedológicas possuem localizações condizentes às variáveis iniciais – relevo, litologia, etc – e, portanto, uma distribuição espacial semelhante aos agrupamentos gerados, como se segue.

#### 4.1 Análise de Agrupamentos Hierárquica

O primeiro *output* do método hierárquico foi a matriz de dissimilaridade (ou proximidade), baseada no método de Ward, com valores oscilando entre 0,002 e 0,937, indicando a coexistência de bacias de cabeceira pouco diferentes (0,002) e outras extremamente dissemelhantes (0,937). A dissimilaridade média de todas as bacias foi de 0,095, com desvio padrão médio de 0,088.

Ressalta-se que a matriz de dissimilaridade não foi apresentada neste trabalho devido à sua elevada dimensão (61.009 células), entretanto, no Quadro 1 é possível observar uma amostra da matriz gerada, bem como algumas de suas variações.

Com os dados da matriz de dissimilaridade foi gerada a tabela de aglomeração (*agglomeration schedule*), que representa as soluções de agrupamento encontradas iterativamente para os 247 elementos analisados.

De um modo geral, as distâncias entre os elementos variaram de 0,052 a 2.460, com média 200,4 e desvio padrão de 321,8. Inicialmente as distâncias (coeficientes) entre os elementos que compunham os grupos foram baixas, contudo à medida que o processo iterativo, responsável pela aglomeração, continuou, as distâncias aumentaram, tendo em vista que elementos formadores de grupos maiores eram mais dissemelhantes (Quadro 2).

É importante mencionar que na tabela

Quadro 1: Amostra dos 8 primeiros elementos da matriz de dissimilaridade construída pelo Método de Ward. Nesta matriz os valores elevados apresentam maior dissimilaridade

Casos	1	2	3	4	5	6	7	8
1	0,000	,009	,066	,124	,105	,124	,069	,107
2	,009	0,000	,043	,144	,114	,116	,064	,130
3	,066	,043	0,000	,166	,141	,129	,021	,164
4	,124	,144	,166	0,000	,004	,023	,140	,019
5	,105	,114	,141	,004	0,000	,011	,130	,022
6	,124	,116	,129	,023	,011	0,000	,122	,047
7	,069	,064	,021	,140	,130	,122	0,000	,143
8	,107	,130	,164	,019	,022	,047	,143	0,000

supracitada constam as diversas fases de agrupamento, tais como: (i) elementos e grupos de elementos agrupados (grupo combinado); (ii) a utilização ou não destes elementos em outras etapas (fase de aparição); e (iii) qual a próxima união a ser realizada após a fase anterior (próxima fase).

De forma subsequente, no *software SPSS*, a leitura do gráfico, denominado *Icicle Plot*, possibilitou a observação visual de quais elementos foram reunidos na tabela de agregação acima, assim como a ordem de formação dos respectivos grupos. Constatou-se que os três primeiros elementos que compuseram grupos foram 247 e 237, seguidos por 84 e 77, 136 e 20. As uniões mais próximas da raiz, ou seja, mais diferentes entre si, foram 16 e 85, 95 e 223, 4 e 158. Novamente, o gráfico encontra-se representado neste trabalho apenas em formato reduzido, devido à sua grande dimensão (Fig. 3).

Como último estágio da análise de agrupamento hierárquica realizada neste trabalho, foi gerado o dendograma de ligação de Ward, que possibilitou a estimativa visual da quantidade de grupos existentes (Fig. 4).

Além disso, utilizando-se a distância de corte entre 2,5 e 5 – adotada como a mais visualmente coerente em relação à diversidade de agrupamentos – é possível distinguir a existência de 8 diferentes grupos referentes às bacias de cabeceira do Ribeirão Mata Porcos (grupos A, B, C, D, E, F, G e H).

#### 4.2 Análise de Agrupamentos Não-Hierárquica

Quadro 2: Amostra da tabela de aglomeração. Contêm os agrupamentos iniciais (1-4), os intermediários (122-125) e os últimos (243-246) a serem formados

Fase	Grupos Combinados		Coeficientes	Fase de Aparição		Próxima fase
	Grupo 1	Grupo 2		Grupo 1	Grupo 2	
1	237	244	,052	0	0	49
2	77	84	,125	0	0	86
3	20	136	,227	0	0	46
4	2	206	,349	0	0	16
122	44	225	79,637	66	0	170
123	27	153	81,074	0	88	186
124	34	50	82,576	89	0	180
125	163	204	84,086	0	0	185
243	4	29	1532,737	242	218	246
244	16	95	1727,167	237	232	245
245	1	16	1950,623	241	244	246
246	1	4	2460,000	245	243	0

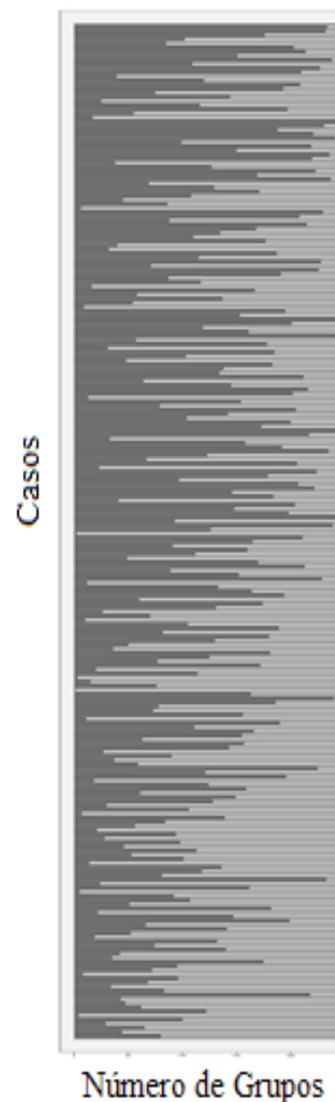


Fig. 3 – Representação Gráfica (Icicle Plot), que resume o processo realizado na tabela de agregação.

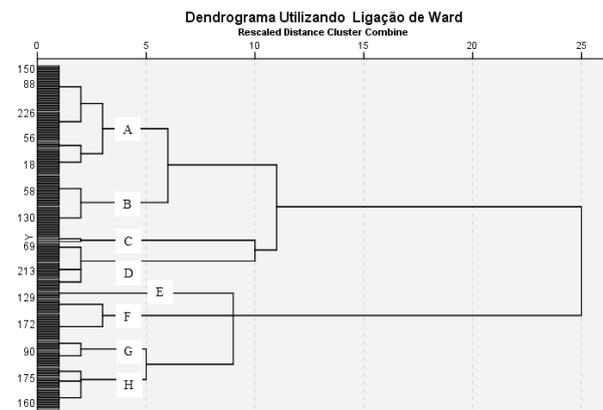


Fig. 4 – Dendrograma gerado como output do processo de agrupamento do método de Ward. As letras (A, B, C, D, E, F, G e H) indicadas no dendrograma referem-se aos grupos visualizados que aparentemente descrevem os dados com melhor exatidão.

Com base na estimativa visual de 8 grupos, resultantes da análise de agrupamento hierárquica, foram feitas uma sequencia de testes para verificar qual a quantidade de grupos melhor representa o universo amostral.

Inicialmente, foi testado o agrupamento de oito classes construído pelo método anterior. Conforme descrito no Quadro 3, todos os níveis de significância (Sig.) estão próximos de zero, indicando que as variáveis adotadas para esta quantidade de agrupamentos possuem grande influência na diferenciação entre grupos. Com efeito, admitindo-se somente a significância estatística das variáveis, o total de oito grupos poderia ser perfeitamente aplicável a este estudo. Entretanto, considerando-se a relevância de grupos reduzidos em estudos ambientais, e igualmente nas fases de campo, foram realizados outros testes, que possibilitaram uma redução sem perda de representatividade.

Após a realização dos testes de significância a partir da quantidade de grupos, constata-se que a representação em sete grupos se revela como a mais adequada em termos de significância estatística, quando comparada à construção de seis grupos (Quadros 4 e 5). Este fato encontra-se justificado pelo valor de 0,183 encontrado para a variável declividade. Tal valor ultrapassa o limite usualmente aceito em estatística de 0,05, que representa o limiar de 5% de probabilidade dos grupos serem iguais e não diferentes.

Desta forma, admitindo-se o agrupamento em sete classes como o mais coerente, verifica-se a existência de importantes resultados além dos níveis de significância, sobretudo nos *outputs* (i) história da iteração, (ii) centros dos grupos finais e (iii) as distâncias entre os centros dos grupos finais.

Quadro 3: Análise de variância das variáveis utilizadas na geração dos oito grupos

ANOVA						
	Grupo		Erro		F	Sig.
	Quadrado Médio	df	Quadrado Médio	df		
Filitos	29,940	7	,152	239	196,492	,000
Itabiritos	27,997	7	,209	239	133,759	,000
Quartzitos	25,698	7	,277	239	92,896	,000
Cangas	27,626	7	,220	239	125,473	,000
Mármore	30,861	7	,125	239	246,080	,000
Amplitude_Alt	14,620	7	,601	239	24,322	,000
Curva_Horizontal	17,186	7	,526	239	32,677	,000
Curva_Vertical	17,139	7	,527	239	32,501	,000
Declividade	11,556	7	,691	239	16,728	,000
Int_Hipsome	11,321	7	,698	239	16,226	,000

Quadro 4: Análise de variância das variáveis utilizadas na geração de sete grupos

ANOVA						
	Grupo		Erro		F	Sig.
	Quadrado Médio	df	Quadrado Médio	df		
Filitos	35,459	6	,139	240	255,976	,000
Itabiritos	33,430	6	,189	240	176,637	,000
Cangas	32,217	6	,220	240	146,732	,000
Amplitude_Alt	16,481	6	,613	240	26,886	,000
Declividade	13,006	6	,700	240	18,584	,000
Quartzitos	29,977	6	,276	240	108,777	,000
Mármore	32,467	6	,213	240	152,192	,000
Curva_Horizontal	20,668	6	,508	240	40,662	,000
Curva_Vertical	19,035	6	,549	240	34,665	,000
Int_Hipsome	12,770	6	,706	240	18,095	,000

Quadro 5: Análise de variância das variáveis utilizadas na geração de seis grupos

ANOVA						
	Grupo		Erro		F	Sig.
	Quadrado Médio	df	Quadrado Médio	df		
Filitos	42,692	5	,135	241	316,184	,000
Itabiritos	40,158	5	,188	241	214,056	,000
Quartzitos	35,939	5	,275	241	130,631	,000
Cangas	38,787	5	,216	241	179,535	,000
Mármore	38,923	5	,213	241	182,560	,000
Amplitude_Alt	19,086	5	,625	241	30,549	,000
Curva_Horizontal	20,937	5	,586	241	35,707	,000
Curva_Vertical	22,487	5	,554	241	40,573	,000
Declividade	1,506	5	,989	241	1,522	,183
Int_Hipsome	11,550	5	,781	241	14,787	,000

Os grupos 1 e 3 apresentam as maiores quantidade de iterações, possibilitando inferir, inclusive, que o processo tenha cessado somente quando atingido o limite máximo pré-estabelecido de 13 iterações por grupo (Quadro 6).

Por outro lado, com características opostas, os grupos 4 e 6 necessitaram de apenas 1 iteração para atingir o valor central de seus grupos, enquanto os agrupamentos 2, 5 e 7 apresentaram um padrão intermediário, oscilando entre 3 e 7 iterações.

Os centros calculados para os grupos finais apresentam valores que demonstram, em termos quantitativos, a grande heterogeneidade existente entre os agrupamentos (Quadro 7).

No grupo 1 verifica-se que a área representada por filitos é maior que média, bem como as variáveis curvatura vertical e integral hipsométrica. O grupo 2, por sua vez, registra o maiores valores de todos os grupos nas variáveis itabirito e amplitude altimétrica, assim como valor de declividade superior à média. Por outro lado, o grupo 3 detém valores acima da média para os filitos e a declividade. O grupo 4 apresentou valores acima da média para as variáveis mármore, curvatura horizontal, vertical e integral hipsométrica. O grupo 5 registra os

Quadro 6: História da iteração dos sete grupos

História da Iteração							
Iteração	Mudanças no Centro dos Grupos						
	1	2	3	4	5	6	7
1	3,370	2,803	3,494	3,952	3,792	2,437	3,058
2	,610	,771	,350	0,000	,535	0,000	1,037
3	,338	,231	,276	0,000	,316	0,000	,242
4	,168	0,000	,148	0,000	,074	0,000	0,000
5	,095	0,000	,065	0,000	,092	0,000	0,000
6	,064	0,000	,061	0,000	,046	0,000	0,000
7	,116	0,000	,085	0,000	0,000	0,000	,130
8	,083	0,000	,071	0,000	0,000	0,000	0,000
9	,023	0,000	,019	0,000	0,000	0,000	0,000
10	,030	0,000	,025	0,000	0,000	0,000	0,000
11	,051	0,000	,016	0,000	,056	0,000	0,000
12	,026	0,000	,020	0,000	0,000	0,000	0,000
13	0,000	0,000	0,000	0,000	0,000	0,000	0,000

maiores valores de todo o rol analisado nas variáveis filito, declividade, curvatura horizontal, vertical e integral hipsométrica. O grupo 6 detém valores acima da média para cangas, quartzitos, amplitude altimétrica e declividade. O grupo 7, por sua vez, registrou valores das variáveis itabirito, canga e amplitude acima da média geral dos grupos.

O último *output* do método de K-Médias, antes da análise de variância inicialmente abordada, foi o quadro relativo às distâncias entre os centros dos grupos finais (Quadro 8). Em síntese, é possível verificar que os grupos 1 e 3, e 1 e 5, com 2,124 e 2,578 respectivamente, são os mais semelhantes, enquanto os grupos 4 e 6, e 4 e 7, detentores de 9,863 e 8,898 respectivamente, são os mais diferentes.

Os valores dos demais grupos apresentam caráter intermediário entre estes extremos, com evidente heterogeneidade entre si.

No total os sete grupos apresentaram a seguinte distribuição: grupo 1 com 59 elementos; grupo 2 com 44; grupo 3 com 73; grupo 4 com 6 membros; grupo 5 com 36; grupo 6 com 7; e grupo 7 com 25 bacias de cabeceira.

### 4.3 Espacialização dos Grupos em SIG

De um modo geral, os grupos 2 e 7 ocorrem predominantemente nas bordas leste e oeste, referentes às frentes de dissecação do vale, sob os itabiritos e cangas (Fig. 5).

Em outro domínio, os demais grupos de bacias apresentam-se restritos às porções norte e sudeste, morfologicamente condizentes às colinas alongadas sob os filitos.

Ao analisar esta espacialização, constata-se a existência de correlação entre a localização dos grupos estimados e os compartimentos geológico-geomorfológicos, que compuseram as variáveis de entrada (*input*) do modelo.

Quadro 7: Valores dos centros dos grupos finais

Centros dos Grupos Finais							
	Grupos						
	1	2	3	4	5	6	7
Filitos	,61	-1,48	,62	,01	,66	-1,05	-1,32
Itabiritos	-,53	1,89	-,47	-,26	-,51	-,19	,14
Cangas	-,31	-,15	-,38	-,45	-,40	,58	2,58
Amplitude Alt	-,66	1,04	-,06	-,67	-,54	,19	,71
Declividade	-,84	,15	,28	-,19	,84	,33	-,39
Quartzitos	-,10	,03	-,20	-,08	-,26	4,96	-,24
Mármore	-,07	-,11	-,13	8,00	-,05	-,18	-,09
Curva Horizontal	,00	-,12	-,64	,47	1,59	-,55	-,10
Curva Vertical	,25	-,03	-,79	,12	1,33	-,39	-,04
Int Hipsome	,51	-,29	-,68	,54	,80	-,14	,11

Quadro 8: Matriz de distâncias entre os centros dos grupos finais

Distâncias Entre os Centros dos Grupos Finais							
Grupo	1	2	3	4	5	6	7
1	0	3,859	2,124	8,146	2,578	5,706	3,861
2	3,859	0	3,511	8,772	4,410	5,517	3,342
3	2,124	3,511	0	8,410	3,490	5,564	3,930
4	8,146	8,772	8,410	0	8,319	9,863	8,898
5	2,578	4,410	3,490	8,319	0	6,368	4,644
6	5,706	5,517	5,564	9,863	6,368	0	5,692
7	3,861	3,342	3,930	8,898	4,644	5,692	0

### 4.4 Verificação em Campo

Nesta etapa de análise empírica, diversos ambientes foram analisados e, juntamente com os dados de gabinete, conduzem a um entendimento mais abrangente, assim como de locais que seriam representativos para amostragens.

As porções referentes aos grupos 2 e 7, situadas sobre as frentes de dissecação E e W, apresentam, em sua ampla maioria, um desenvolvimento incipiente da cobertura superficial, marcado pela presença de material coluvionar arenoso pouco espesso (~ 4cm) nos fundos de vale. Estes compartimentos são seguidos por neossolos litólicos na média-alta encosta, e afloramentos esparsos de quartzitos e itabiritos, sobretudo próximos às linhas de cumeada (Fig. 6).

Nas áreas dos grupos 1,3, 4, 5 e 6 ocorrem mantos de alteração bastante espessos, muitos dos quais detêm registros coluviais variáveis, podendo atingir até 1 m. Próximos aos eixos fluviais coexistem tanto feições de caráter aluvial, quanto depósitos detríticos de origem coluvionar aparentemente associados ao desmonte de materiais do terço superior de suas encostas (Fig. 7).

Os segmentos referentes a tais grupos estão localizados exclusivamente nas porções menos declivosas, de N e SE, onde os materiais de origem estão diretamente associados aos filitos do Supergrupo Minas, correspondendo, grosso

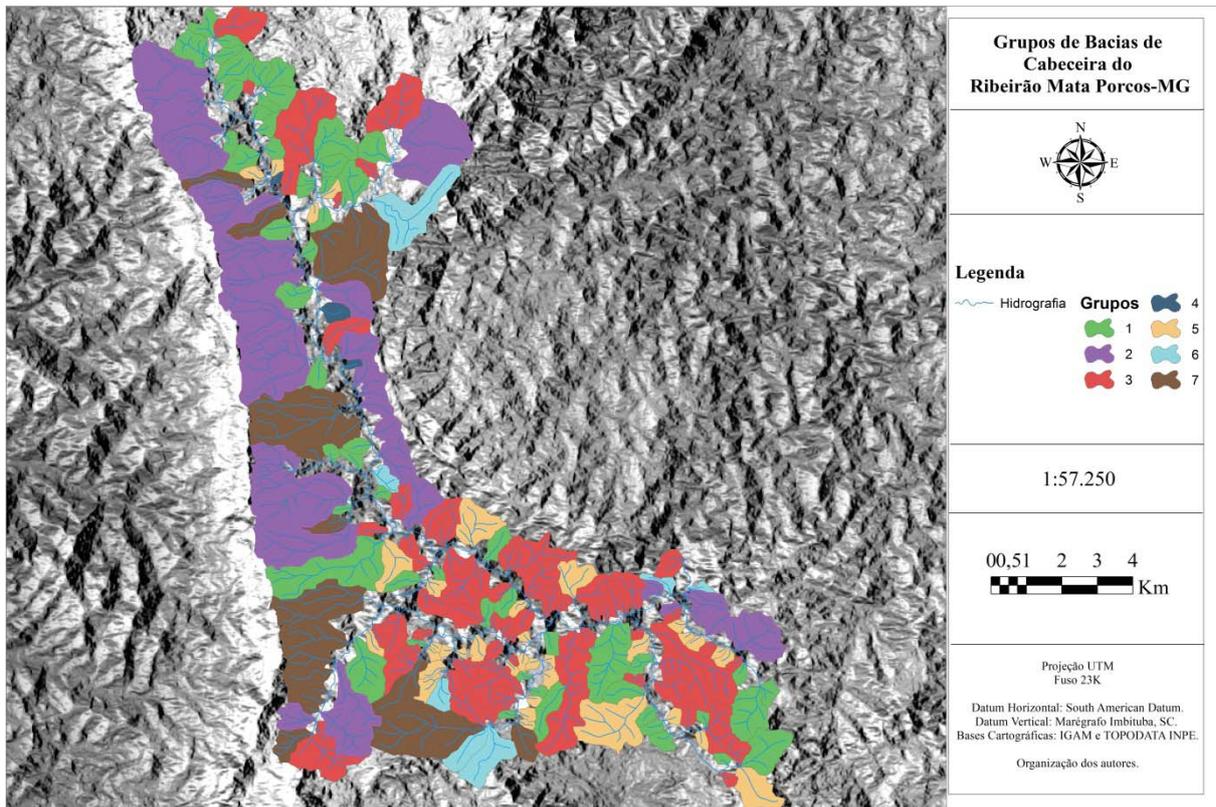


Fig. 5 – Espacialização dos grupos de bacias de cabeceiras gerados pelo Método de K-Médias. Bases cartográficas fornecidas pelo INPE e IGAM.



Fig. 6 – Compartimento típico das bacias de cabeceira dos grupos 2 e 7. Detalhe: Em primeiro plano, o fundo de vale; as baixas e médias encostas nas laterais da imagem; e afloramentos de quartzito ao fundo, próximos ao interflúvio.



Fig. 7 – Discordância erosiva entre o saprolito (acinzentado) e material coluvionar sobreposto bastante espesso. Na porção superior do perfil observa-se também a existência de horizonte A retrabalhado.

modo, às colinas alongadas mais próximas à calha do Ribeirão Mata Porcos (Fig. 8).

Ademais, coexistem diversas morfologias e padrões de cobertura, sobretudo nas porções fora das frentes de dissecação, que refletem a complexidade apresentada no mapeamento. Nos grupos 1, 3 e 5, em que predominam filitos, o manto de intemperismo, em especial o

saprolito, chega a atingir dezenas de metros. No grupo 6, também inserido neste compartimento geomorfológico, ocorre grau de desenvolvimento pedogenético similar, embora o material de origem seja mármore.

## 5. DISCUSSÃO

Os dados obtidos a partir da análise de



Fig. 8 – Aspecto morfológico regional. No centro, o domínio das colinas alongadas, e à direita, as frentes de dissecação bastante declivosas.

agrupamento hierárquica demonstram vantagens e desvantagens bastante semelhantes às descritas na literatura científica. Em suma, constata-se que tanto a matriz de dissimilaridade, quanto a tabela de aglomeração e sua representação gráfica (*Icicle Plot*) foram relevantes, sobretudo pela oportunidade de análises exploratórias do comportamento dos grupos, elementos e variáveis. Estes procedimentos tornaram-se robustos nesta pesquisa, fundamentalmente, pela determinação quantitativa das diferenças entre os elementos e os agrupamentos realizados pelo método de Ward, assim como pela validação baseada em níveis de significância, de forma concordante às considerações de Rao e Srinivas (2008).

Entretanto, o produto final da análise de agrupamento hierárquica, o dendograma, por ser uma representação visual, com uma interpretação puramente subjetiva (PALUMBO e D'ENZA, 2010), apresentou grande incerteza, justificando novamente, uma observação mais quantitativa através do método de K-Médias. Contudo, apesar do alto grau de subjetividade pode-se considerar que o dendograma também foi satisfatório, sobretudo pelo erro de apenas um grupo, na observação visual, em relação ao valor ótimo estimado pelo método de K-Médias.

No que concerne à análise de agrupamento não-hierárquica, os níveis de significância ressaltam a adequabilidade das variáveis adotadas na definição dos grupos. Igualmente, o histórico de iterações, e as distâncias entre os centros dos grupos finais evidenciam a homogeneidade de certos grupos em comparação

a outros mais heterogêneos.

Os grupos 1 e 3, que apresentaram as maiores iterações e, portanto, mais heterogêneos internamente (ALDENDERFER e BLASHFIELD, 1984) são, quando comparados, os grupos com maior semelhança seguidos pelos grupos 1 e 5. Especialmente, é importante notar que estes grupos são vizinhos uns aos outros, enquanto os grupos 4, 6 e 7, mais homogêneos internamente e com as menores semelhanças entre grupos, localizam-se sem nenhuma relação de vizinhança. Uma hipótese plausível para esta relação espacial diferenciada pode estar relacionada às características que definem os agrupamentos, como litologia e relevo, que, obviamente, apresentam continuidade no espaço.

Em relação à morfologia da paisagem, a compartimentação em (i) frentes de dissecação e (ii) colinas alongadas também parece ser bastante aplicável. Em termos estatísticos, os grupos 4 e 6 representam menos de 5% de todas as bacias estudadas, o que as torna uma exceção (*outlier*), principalmente em relação às suas vizinhas. Com efeito, estes grupos (4 e 6) não comprometem a generalização referente ao compartimento das colinas.

Em termos empíricos, foram encontradas em campo evidências morfológicas e/ou pedoestratigráficas que registram a complexidade local e sua possibilidade de regionalização geomorfológica nos dois grupos supracitados. Contudo, por serem dados agrupados naturalmente, com menor intencionalidade e subjetividade humana, os resultados oriundos das análises de agrupamento mencionadas, tendem a representar a complexidade existente de forma mais fidedigna (TIMM, 2002).

Igualmente, em etapas essenciais de campo, sobretudo em coletas pedoestratigráficas, as análises de agrupamentos realizadas neste trabalho parecem ser tanto quantitativa quanto qualitativamente melhor fundamentadas.

## CONSIDERAÇÕES FINAIS

As análises de agrupamentos realizadas apresentaram resultados robustos e consistentes em termos estatísticos, conforme constatado nos diversos resultados (*outputs*) gerados. Em termos empíricos, o Alto-Médio Curso do Ribeirão Mata Porcos apresentou uma heterogeneidade de coberturas superficiais coincidente com

os agrupamentos gerados, assim como em relação aos dois compartimentos geológico-geomorfológicos delineados.

Por conseguinte, a adoção das bacias de cabeceira, como unidades de estudo que se repetem no espaço, associada a caracterizações por análises de agrupamentos, parece ser bastante relevante. Não somente por permitir o detalhamento das características da área, mas, de forma semelhante, por auxiliar na identificação de prováveis pontos de amostragem de coberturas superficiais.

Desta forma, utilizando-se apenas produtos cartográficos de distribuição gratuita, este trabalho demonstrou a adequabilidade e possibilidade de extração de informações espaciais e estatísticas em etapas pré-campo, transformando-a em uma ferramenta-guia.

Por fim, espera-se que esta pesquisa tenha contribuído não somente na elaboração de uma metodologia-guia na caracterização das coberturas superficiais do Ribeirão Mata Porcos, mas igualmente, possa representar um instrumental metodológico para outras áreas. Dessa forma, acredita-se que este trabalho tenha contribuído positivamente para a construção de uma metodologia simples, com potencial de contribuição para diversos pesquisadores envolvidos em ciências da Terra.

#### **AGRADECIMENTOS**

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de Mestrado de um dos autores; ao Grupo de Pesquisas Geomorfologia e Recursos Hídricos (CNPq); ao Centro de Estudos em Geomorfologia (CEGEO) do IGC/UFMG; e ao MSc Felipe Chagas do DEGEO/UFOP pelo auxílio em campo.

#### **REFERÊNCIAS BIBLIOGRÁFICAS**

ALDENDERFER, M. S.; BLASHFIELD, R. K. **Cluster Analysis**. Flórida: Sage University Paper, 1984. 88 p.

BERTALANFFY, L. V. **Teoria Geral dos Sistemas**. Petrópolis: Vozes, 1973. 351 p.

CHEMALE JR. F., ROSIÈRE, C. A., ENDO, I. The Tectonic Evolution of the Quadrilátero Ferrífero, Minas Gerais, Brazil. **Prec. Res.** 65, 25-64, 1994.

CPRM – Serviço Geológico do Brasil. Folhas Belo Horizonte (1:100.000) e Itabirito (1:50.000). **Geobank**. Acessado em 29/06/2010.

EVERITT, B. S.; LANDAU, S.; LEESE, M. **Cluster Analysis**. Londres: Arnold, 2001. 346 p.

HARVEY, D. **Explanation in Geography**. Londres: Edward Arnold, 1969. 521 p.

JENNY, H. **Factors of Soil Formation: A System of Quantitative Pedology**. Nova York: Dover Publications, 1944. 191 p.

MILLIGAN, G. W. Clustering Validation: Results and Implications for Applied Analysis. In: ARABIE, P.; HUBERT, L. J.; DE SOETE, G. **Clustering and Classification**. New Jersey: World Scientific Publication, p.341-375, 1996.

OLLIER, C.; PAIN, C. **Regolith, Soils and Landforms**. Canberra: John Wiley and Sons, 1996. 316 p.

PALUMBO, F.; D'ENZA, A. I. A Two-Step Iterative Procedure for Clustering of Binary Sequences. In: PALUMBO, F.; LAURO, C. N.; GREENACRE, M. J. **Data Analysis and Classification**. Berlin: Springer, p. 33-40, 2010.

PHILLIPS, J. D. Deterministic Complexity, Explanation and Predictability in Geomorphic Systems. In: RHOADS, B. L.; THORN, C. E. **The Scientific Nature of Geomorphology**. Binghamton: John Wiley and Sons, p. 315-336, 1996.

RAO, A. R.; SIRINIVAS, V. V. **Regionalization of Watersheds: an approach based on cluster analysis**. West Lafayette: Water Science and Technology Library, 2008. 256 p.

ROMESBURG, H. C. **Cluster Analysis for Researchers**. Carolina do Norte: Lulu Press, 1984. 334 p.

SCHAETZL, R. J.; ANDERSON, S. **Soils: Genesis and Geomorphology**. Cambridge: Cambridge University Press, 2005. 817 p.

SHEPPARD, E. Quantitative Geography: representations, practices and possibilities. **Environment and Planning D: Society and Space**, v. 19, pags. 535-554, 2001.

SINGH, O.; SARANGI, A.; SHARMA, M. C. Hypsometric Integral Estimation Methods and its Relevance on Erosion Status of North-Western

Lesser Himalayan Watersheds. **Water Resour Manage**, 22, 1545–1560, 2008.

STRAHLER, A. N. Hypsometric (area–altitude) analysis of erosional topography. **Geol Soc Amer Bull** 63, 1117–1141, 1952.

TIMM, N. H. **Applied Multivariate Analysis**.

New York: Springer, 2002. 693 p.

VIDAL-TORRADO, P.; LEPSCH, I. F.; CASTRO, S. S. Conceitos e Aplicações das Relações Pedologia e Geomorfologia em Regiões Tropicais Úmidas. **Tópicos Ci. Solo**, v. 4, pgs. 145-192, 2005.