

**Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística**

**Seleção de Perfis
Granulométricos para a
Determinação do Padrão de
Transporte de Sedimentos
Via Método Seqüencial de
Cadeias de Markov e
Associação Cruzada**

P. S. Lucio, E. C. Bodevan

Relatório Técnico RTP-01/99

**Relatório Técnico
Série Pesquisa**

SELEÇÃO DE PERFIS GRANULOMÉTRICOS PARA A DETERMINAÇÃO DO PADRÃO DE TRANSPORTE DE SEDIMENTOS VIA MÉTODO SEQUENCIAL DE CADEIAS DE MARKOV E ASSOCIAÇÃO CRUZADA.

*O CASO DE WESTERSCHELDE (HOLANDA/BÉLGICA)
ENTRE BAARLAND E RUPELMONDE.*

P. S. Lucio, E. C. Bodevan
Departamento de Estatística – ICEX - UFMG
H. Dupont, L. V. Ribeiro
Departamento de Geologia – IGC – UFMG

RESUMO

Sabe-se que a granulometria de um sedimento é uma característica bastante complexa que requer estudos e definições de parâmetros, que vão além de uma simples categorização ou classificação. O estudo granulométrico define a quantidade de sedimento situado em faixas de tamanhos de grãos, ou distribuição de frequências por faixa granulométrica. Esta distribuição de frequências do tamanho dos grãos, deve ser analisada sob um enfoque estatístico, para que se possa estabelecer os parâmetros que permitam inferir sobre um padrão de informação do transporte de sedimentos.

Os estudos, em geral, são realizados com sedimentos não coesos, que podem ser separados em faixas granulométricas por peneiramento. Assim, uma amostra é peneirada e cada fração deve ser pesada, com certa precisão, e estes pesos são transformados em percentagens de um sistema fechado. Um problema, freqüentemente, abordado é aquele que se refere aos limites das classes granulométricas e ao número destas classes. Na prática, determina-se uma escala, com intervalos de classes, crescente com o tamanho do grão (escala logarítmica); o número mínimo de classes deve ser tal que não apresente variabilidade exagerada em cada uma delas. Já o número máximo destas classes, deve ser aquele que torne o experimento economicamente viável, uma vez que um número muito grande, iria onerar as operações de peneiramento e de representação gráfica.

Parâmetros granulométricos, como: média, variância e assimetria, têm sido amplamente utilizados para identificar ambientes de erosão ou deposição sedimentar. Recentemente, estes parâmetros têm apresentado um papel de extrema importância na identificação do padrão de transporte de sedimentos (McLaren, 1981; McLaren & Bowles, 1985; Gao & Collins 1991). Alguns métodos até então desenvolvidos, têm por base que, na interpretação das mudanças espaciais na textura do sedimento associada ao processo sedimentar, deve-se levar em consideração, entre outras, três hipóteses físicas: o material depositado deve ser produto de uma única fonte; os grãos menores e mais leves, quase certamente, serão transportados com maior facilidade que os grãos maiores e mais pesados e a probabilidade para deposição é diretamente proporcional ao diâmetro do grão. Não obstante, acredita-se que uma análise exploratória, através de um estudo estocástico, deve preceder a análise do padrão de informação da direção de transporte, cujo objetivo deve convergir à confiabilidade dos resultados obtidos, que nos fornecerá informações quanto à estabilidade dos métodos utilizados para a determinação do processo de deposição/erosão. Os métodos *de cadeias markovianas e associação cruzada*, mostram-se importantes ferramentas no estudo da definição do padrão granulométrico da área sob investigação, otimizando o custo computacional das análises realizadas através da técnica estrutural e do teste exato aqui abordados.

Palavras-chave: análise de conglomerados; granulometria; processos estocásticos; associograma.

1. INTRODUÇÃO

A análise de dados em seqüência ou em cadeia que incluam levantamentos mineralógicos ou geoquímicos ao longo de um linha definida por uma propriedade geomorfológica, são muito comuns em Ciências da Terra. As técnicas usuais para examiná-los devem fazer parte dos conceitos fundamentais de Processos Estocásticos. Dentro deste contexto propõe-se, neste ensaio, o estudo de alguns parâmetros associados à granulometria [na escala Φ , que corresponde a $-\log_2$ do diâmetro do grão em milímetros] ao longo de perfis longitudinais (seqüencialmente selecionados) com vias à identificação de um padrão de transporte de sedimentos. Desta forma, através do enfoque de Cadeias de Markov, sugere-se uma metodologia estrutural para a caracterização de depósitos em agregação e/ou erosão.

A região sob estudo se estende de Baarland na Holanda, a Rupelmonde na Bélgica (Fig. 1). Concernente à aquisição dos dados (*confidenciais*), as 867 amostras de sedimentos tratadas e analisadas granulometricamente, nos foram gentilmente cedidas pelo *Dr. Patrick McLaren* da *GeoSea Consulting* (UK), em arquivo ASCII, especialmente para este estudo. Estas amostras foram coletadas com um espaçamento regular de 500 metros, salvo em determinadas áreas de batimetria complexa, onde esta malha foi localmente reduzida a um intervalo de 250 metros.

As linhas selecionadas foram determinadas uma ao centro (L2), uma à margem superior (L1) e outra à margem inferior (L3), sem levar em consideração que suas batimetrias pudessem apresentar padrões não similares (Fig. 2a, Fig. 2b e Fig. 2c), para que não tivéssemos que considerar o efeito de suporte proveniente de uma amostragem não regular.

2. ANÁLISE DE CONGLOMERADOS

A análise de conglomerados é freqüentemente denominada classificação automática, onde as similaridades (ou dissimilaridade) entre observações ou variáveis são avaliadas com o objetivo de formar agrupamentos com base em características semelhantes através de partições. Propondo-se investigar o conjunto de variáveis granulométricas deste estudo, a partição e categorização foram fundamentadas sob uma discretização nominal utilizando a função de ligação “*completa*” [Johnson & Wichern, 1998], que leva em consideração a máxima distância entre grupos, baseando-se em seus “*centróides*” – centro de gravidade da massa de dados.

Deve-se observar a diferença entre análise discriminante, onde há uma partição particular de interesse conhecida e previamente definida, e análise de conglomerados, onde partições são geradas pela própria gênese da análise. Assim, como a discriminação pode ser considerada uma forma discreta de regressão, a análise de conglomerados pode ser considerada uma forma discreta de escalonamento multidimensional, no seguinte sentido: técnicas de escalonamento geram conjuntos de valores discretos que alocam as observações dentro dos grupos [Johnson & Wichern, 1998].

Executando-se o *dendograma* (sob métrica “*completa*”), realizado para a região, sob um enfoque global, percebe-se a nítida formação de 6 (seis) conglomerados (Fig. 3) granulométricos. Note ainda, que detectado através desta técnica, o grau de similaridade, em fase, entre os grupos de variáveis é muito baixo, em torno de 16%. No que tange à região, as quatro primeiras colunas devem ter seu comportamento bem controlado, para se fazer a análise de conglomerados, visto que são praticamente constantes, o que também é observado para as cinco últimas classes.

Desta forma, obtém-se para a partição final global:

GRUPO (núm. de elementos)	ELEMENTOS
1 (4)	1 - 4
2 (4)	5 - 8
3 (2)	9 - 10
4 (2)	11 - 12
5 (2)	13 - 14
6 (12)	15 - 26

Através do *dendograma* (sob métrica “*completa*”), realizado para cada uma das três linhas selecionadas (Fig. 3a, Fig. 3b e Fig. 3c) observamos com bastante nitidez, a formação de seis (6) grupos granulométricos, onde pode-se notar o alto grau de similaridade entre as variáveis. Observe ainda, que detectado através desta técnica, o grau de similaridade, em fase, entre os grupos de variáveis é muito baixo, em torno de 12% referente à linha L1, 16% à linha L2 e 13% à linha L3.

Assim, obtém-se para a partição final local

- linha L1:

GRUPO (núm. de elementos)	ELEMENTOS
1 (4)	1 - 4
2 (4)	5 - 8
3 (2)	9 - 10
4 (2)	11 - 12
5 (3)	13 - 15
6 (11)	16 - 26

- linha L2:

GRUPO (núm. de elementos)	ELEMENTOS
1 (4)	1 - 4
2 (3)	5 - 7
3 (2)	8 - 9
4 (2)	10 - 11
5 (4)	12 - 15
6 (11)	16 - 26

- linha L3:

GRUPO (núm. de elementos)	ELEMENTOS
1 (4)	1 - 4
2 (5)	5 - 9
3 (2)	10 - 11

4 (1)	12
5 (9)	13 - 21
6 (5)	22 - 26

Desta forma, neste ensaio, estar-se-á estudando sob um enfoque probabilístico, classes de variáveis granulométricas definidas por meio da determinação de categorias identificadas através de intervalos empíricos por tamanho dos grãos $[\Phi]$, obtidos através da técnica de análise de conglomerados. Assim, definiu-se da seguinte forma a classificação, sem perda de generalidades e da informação geológica:

1 – Grânulos e Areia Muito Grossa (GG) $[-2 < \Phi \leq 0]$;

2 – Areia Grossa e Média (GMG) $[0 < \Phi \leq 2]$;

3 – Areia Fina e Muito Fina (GMF) $[2 < \Phi \leq 3,5]$;

4 – Silte Grosso (GF) $[3,5 < \Phi \leq 5]$;

5 – Silte Fino (S) $[5 < \Phi \leq 7]$;

6 – Argila (A) $[\Phi > 7]$.

Não obstante, foi levado em consideração o estudo feito anteriormente e o conhecimento geomorfológico da área sob estudo e discriminou-se 6 (seis) grupos de variáveis com [4; 4; 3; 3; 4; 8] elementos, respectivamente.

É importante observar que estas variáveis devem formar um sistema fechado por amostra, ou seja a soma do conteúdo amostrado por unidade deve ser constante e igual a 100%. Sendo assim, concernente à correlação linear, o aumento de um grupo acarretará naturalmente em uma diminuição do seu complementar.

A aplicação de modelos matemáticos em geologia geralmente é dificultada pelo pressuposto de casualidade em modelos que invocam eventos totalmente independentes ou, em outro extremo, pela exigência nos modelos determinísticos clássicos de que os eventos estejam condicionados a variáveis específicas, predeterminadas e perfeitamente conhecidas. O modelo de Cadeias de Markov, ocupando justamente uma posição intermediária nesse espectro de modelos dinâmicos, presta-se com mais propriedade a ser utilizado na tentativa de explicação

de processos geológicos. Além disso, se aos valores obtidos por meio das Cadeias de Markov forem aplicados métodos estatísticos multivariados classificatórios, como a análise de agrupamentos ou a análise discriminante, os resultados, quando comparados, permitirão com grande facilidade verificar se os processos estudados apresentam alguma equivalência.

3. PROCESSO ESTOCÁSTICO

Um processo estocástico unidimensional, é uma coleção de variáveis aleatórias, todas definidas em um mesmo espaço de probabilidades. Uma função aleatória ou uma realização de um processo estocástico unidimensional $Z = \{Z(\mathbf{x}), \mathbf{x} \in \mathfrak{R}\}$, é uma determinação, para cada $\mathbf{x} \in \mathfrak{R}$, de um valor possível de $Z(\mathbf{x})$, onde \mathfrak{R} é chamado conjunto de índices. O conjunto sobre o qual a variável (Z) assume valores é denominado espaço de estados.

3.1 PROCESSOS DE MARKOV DE PRIMEIRA ORDEM

Um processo estocástico unidimensional $Z = \{Z(\mathbf{x}), \mathbf{x} \in \mathfrak{R}\}$, é um Processo de Markov de primeira ordem se para qualquer escolha de $\mathbf{x} \in \mathfrak{R}$, tem-se que:

$$P(a < Z(\mathbf{x}) \leq b \mid Z(\mathbf{x}_1) = z_1, Z(\mathbf{x}_2) = z_2, Z(\mathbf{x}_3) = z_3, \dots, Z(\mathbf{x}_n) = z_n) = P(a < Z(\mathbf{x}) \leq b \mid Z(\mathbf{x}_n) = z_n),$$

$$\text{onde } 0 < \mathbf{x}_1 < \mathbf{x}_2 < \dots < \mathbf{x}_n < \mathbf{x}.$$

Um processo de Markov com espaço de estados discretos é dito Cadeia de Markov. O método das Cadeias de Markov foi selecionado como ferramenta para análise dos dados de natureza seqüencial, provenientes deste estudo, por julgá-lo de grande aplicação para o entendimento dos fenômenos geológicos de sedimentação.

3.2 CADEIAS DE MARKOV

Uma Cadeia de Markov com parâmetro espacial discreto $\{Z_n\}$ é um processo estocástico markoviano cujo espaço de estados é um conjunto finito ou enumerável, e para o qual $\mathbf{x} = (0, 1, 2, \dots)$. Assim,

$$P_{ij}^{n, n+1} = P(Z_{n+1} = j \mid Z_n = i).$$

A quantidade $P_{ij}^{n,n+1}$ é dita probabilidade de transição do estado (i) na posição (n) para o estado (j) na posição ($n+1$). Quando $P_{ij}^{n,n+1}$ não depende de (n), ou seja, não depende da posição, a cadeia possui a propriedade de probabilidade de transição estacionária. Neste caso, sem perda de generalidades, utiliza-se:

$$P_{ij} = P(Z_{n+1} = j | Z_n = i).$$

As Cadeias de Markov consideram seqüências ou cadeias de estados em que a transição de um estado para outro, dentro de um intervalo no espaço, depende de estados prévios, de tal forma que as probabilidades associadas às transições entre os estados sejam estacionárias, isto é, não dependem da posição. Neste sentido, pode-se afirmar que muitos processos geológicos de sedimentação (deposição ou erosão) possuem a propriedade markoviana, que se traduz pela influência de eventos prévios em eventos subseqüentes .

O grau de dependência de um estado em relação à estados prévios nos informa a ordem da cadeia. Desta forma, se existe um padrão de dependência apenas em relação ao estado imediatamente anterior a cadeia é classificada como de *primeira ordem*, se dois estados imediatamente anteriores estão associados a cadeia é dita de *segunda ordem* e envolve uma unidade espacial, no entanto se os dois estados, que estão associados, não são imediatamente precedentes, a cadeia possui *ordem superior* à segunda e os intervalos espaciais envolvidos podem ser maiores que uma unidade espacial.

Para aplicarmos o método das Cadeias de Markov à nossa pesquisa, são necessárias algumas definições prévias, adequando a teoria explicitada acima ao nosso estudo. Assim, doravante, um *processo estocástico* será traduzido como o processo de transporte de sedimentos, caracterizado pela predominância dos materiais granulométricos ao longo de cada uma das linhas amostrais e os *estados da cadeia* serão os seis (6) estágios de predominância granulométrica. O *espaço de estados* será discreto e finito, devendo ser classificado segundo o teor granulométrico.

As cadeias de Markov de primeira ordem e dependência simples são as mais utilizadas na abordagem de processos sedimentares, tendo em vista sua relativa facilidade de aplicação. No entanto, é comum seqüências sedimentares exibirem superposição de ciclos maiores sobre

ciclos menores, situações estas, que podem ser representadas por matrizes markovianas de alta ordem, cuja estruturação e abordagem matemática são tanto mais complexas quanto maior o grau de dependência revelado pelo processo markoviano.

3.3 MATRIZ DE FREQUÊNCIAS DE TRANSIÇÕES

As matrizes de probabilidade de transições constituem uma forma simples e prática de apresentação das Cadeias de Markov. O cálculo das probabilidades de transições baseia-se na frequência de ocorrência de transições de um estágio para outro, observadas na representação do processo considerado, isto é, na matriz de registros ou de frequências de transições.

A verificação da propriedade markoviana no processo de transporte de sedimentos é feita geralmente por meio da matriz de transições dos parâmetros granulométricos onde os estágios envolvidos no processo correspondem à granulometria que compõe a seqüência estudada. O método utilizado na estruturação de informações geológicas das seqüências granulométricas em matrizes de transições paramétricas, são registros feitos a intervalos fixos a partir da base da seqüência, caso em que o tempo é considerado uma variável discreta. Para tanto considera-se que os parâmetros tempo e localização estejam intimamente associados.

Assim, para determinar a matriz de registro de transições é necessário que observemos antes a ordem seqüencial dos estágios granulométricos predominantes.

O diferencial da metodologia estrutural, apresentada neste estudo, em relação à aplicação clássica é devido ao fato de que nosso interesse está voltado para a região no que tange à influência entre vizinhos mais próximos, e não para uma única linha (1D) como um todo. Desta forma, sem perda de generalidades ou informação, caracterizamos a região segundo a análise de dados seqüenciais. É conveniente, ressaltarmos que, não estamos interessados em verificar a propriedade markoviana no processo sedimentar, e sim verificar a hipótese do transporte de sedimentos em determinada localidade..

4. CONSTRUÇÃO DA SÍNTESE NUMÉRICA GRANULOMÉTRICA

Na amostragem foram feitas as observações por linhas ao longo de uma determinada direção. Em cada um destes perfis foram coletados os dados granulométricos de interesse para este estudo.

A seqüência global foi então tabulada em uma matriz de frequências (M_f), composta pelas frequências de transições (n_{ij}) de um estágio (i) para seu sucessor (j). Assim, é obtido o número de vezes em que ocorre transição de um estágio para outro.

No caso de registros feitos a intervalos fixos, os elementos da diagonal principal exprimem as taxas relativas dos diferentes estágios granulométricos.

4.1 MATRIZ DE PROBABILIDADES DE TRANSIÇÕES

A partir das matrizes de transições granulométricas, objetiva-se representar os diferentes estágios por meio da sucessão de tamanho de grãos categorizados que foram depositadas em determinada seqüência sedimentar, buscando caracterizar seu ambiente (malha) ou sub-ambiente (perfis) específico. Desta forma, as diversas composições ou associações granulométricas de diferentes perfis de um ou diferentes sistemas deposicionais podem ser representadas por diferentes matrizes markovianas. Os problemas advindos de tal visão simplificada do processo sedimentar, ocorrida a deposição ou não em seqüência, são óbvios. Não são registrados períodos de não - deposição ou mesmo erosivos, pois as frequências são tabuladas considerado-se unicamente o registro granulométrico local. Problemas similares são evidentes no caso das cadeias, em que intervalos fixos de tamanho da amostra são diretamente relacionados a intervalos discretos na escala espacial de deposição, desprezando-se fatores tais como: diferentes graus de compactação de diferentes locais, hiatos deposicionais ou erosivos, diferentes taxas de sedimentação, entre outros.

Tais situações podem ser em parte resolvidas pela diminuição dos intervalos dos registros de transições, ou mesmo tomando cuidado para que interrupções de expressão no processo deposicional, isto é, mudanças significativas no ambiente sedimentar, não sejam incluídas numa mesma matriz, mas constituam pontos de separação entre matrizes. Este último caso é de grande importância referente à estacionariedade da cadeia.

Considerando as observações feitas acima e a partir dos resultados obtidos na matriz de frequências de transições (M_f) passamos então ao cálculo da matriz de probabilidades de

transição (M_p), que deve ser obtida a partir da correspondente matriz de registros, dividindo-se cada elemento (n_{ij}) da i -ésima linha pelo total desta linha, ou seja,

$$\hat{p}_{ij} = \frac{n_{ij}}{n_i}$$

onde \hat{p}_{ij} é a probabilidade de transição do estado (i) para o estado (j), (n_{ij}) é o número de vezes em que ocorre transição do estado (i) para o estado (j), (n_i) é a frequência total de transições do estado (i) para um qualquer dos outros estados.

Após o cálculo da matriz de probabilidades de transições granulométricas (M_p) o próximo passo será um teste estatístico para a verificação da propriedade markoviana relativa ao processo dinâmico de erosão/deposição - transporte.

A estruturação de uma sucessão de eventos em uma matriz de probabilidades de transições, não significa, necessariamente, que o processo físico original seja um processo markoviano. É evidente, pois, que a aplicação do modelo markoviano a seqüências sedimentares prescinde da verificação da propriedade markoviana. Para tanto, as suposições do teste são:

- a matriz de probabilidade de transições deve ser estacionária;
- as matrizes devem ser tabuladas segundo intervalos fixos.

4.2 A ESTATÍSTICA DE TESTE

O objetivo do teste estatístico de hipóteses é fornecer ferramentas que nos permitam validar ou rejeitar uma hipótese através dos resultados da amostra. O teste, aqui apresentado, é uma aproximação do teste qui-quadrado (χ^2) onde estaremos investigando a hipótese nula de que as observações, em pontos sucessivos no espaço unidimensional, segundo uma certa orientação determinada sob um linha, sejam estatisticamente independentes *versus* a hipótese alternativa de que as observações provenham de uma Cadeia de Markov de *primeira ordem*.

Desta forma, a propriedade markoviana da seqüência é verificada mediante o uso de matrizes tabuladas segundo mudanças do perfil granulométrico. Deve-se enfatizar que a estatística de teste aqui apresentada aplica-se somente a matrizes nas quais todos os elementos são positivos e diferentes de zero. As limitações estatísticas impostas por elementos nulos podem ser superadas pela omissão de um grau de liberdade para cada zero.

Assim, as hipóteses são definidas como se segue:

H_0 : A probabilidade do (i)-ésimo grupo granulométrico ser seguido pelo (j)-ésimo é a mesma de ser seguido por outra granulometria qualquer, ou seja, na deposição/erosão, as várias granulometrias independem umas das outras.

H_1 : A probabilidade do (i)-ésimo grupo granulométrico ser seguido pelo (j)-ésimo é diferente de ser seguido por uma outra granulometria qualquer, ou seja, no processo de deposição/erosão em determinado local depende daquele imediatamente anterior.

A estatística de teste é assim definida, segundo (Anderson & Goodman, 1956):

$$-2 \log \lambda = 2 \sum_{i,j=1}^m n_{ij} \log \left[\frac{\hat{p}_{ij}}{\hat{p}_j} \right].$$

Esta estatística de teste para averiguar a propriedade markoviana, segue uma distribuição *qui-quadrada* com $(m-1)^2$ graus de liberdade, onde m é o número de estágios granulométricos, n_{ij} é a freqüência de transições na (ij)-ésima cela, \hat{p}_{ij} é a probabilidade de transições na (ij)-ésima cela e \hat{p}_j é a probabilidade total para a (j)-ésima coluna.

Desta forma, rejeita-se a hipótese de que os eventos sejam independentes se o valor calculado para $-2 \log \lambda$ exceder ao valor tabelado da χ^2 para $(m-1)^2$ graus de liberdade a um nível de significância α preestabelecido.

Em nosso estudo, obtemos empiricamente, através da Análise de Conglomerados, 6 estágios granulométricos categorizados. Neste caso, n_{ij} é obtido através da matriz M_f . As tabelas (Tab. 1a, Tab. 2a e Tab. 3a) mostram as freqüências absolutas de transições de primeira ordem para cada um destes estágios. Nas tabelas (Tab. 1b, Tab. 2b e Tab. 3b) estão determinadas as respectivas probabilidades de transição estacionária.

O valor tabelado é obtido a partir da *qui-quadrado* com 25 graus de liberdade e nível de significância de 5%. De acordo com a região de rejeição especificada através do nível de significância, não se rejeita a hipótese nula (H_0) se nos baseamos nos dados apresentados. Isto

é, a probabilidade da (*i*)-ésima granulometria ser seguida pela (*j*)-ésima é a mesma de ser seguida por outra predominância granulométrica qualquer, ou seja, no processo de deposição e/ou erosão, as várias granulometrias locais independem umas das outras. Referentes as três linhas sob estudo, estas estatísticas de teste estão referenciadas na Tab. 4.

Embora os processos markovianos possam abranger desde as relativamente simples e regulares Cadeias de Markov de primeira ordem até modelos complexos com memórias que se estendem por dois ou mais intervalos no espaço, o teste aqui abordado encontra aplicação apenas para o caso de verificação da propriedade markoviana com memória de um passo.

Anderson e Goodman (1956) apresentam outros testes que são aplicáveis às cadeias de ordem mais elevadas. Assim, foram também implementados testes onde investigou-se a hipótese de que as observações, em pontos sucessivos no espaço unidimensional, segundo uma certa orientação determinada sob um linha, fossem provenientes de uma Cadeia de Markov de *segunda ordem*. Sem perda de informação, as hipóteses são definidas como se segue:

H₀: A probabilidade do (i)-ésimo grupo granulométrico ser seguido pelo (j)-ésimo e pelo (j+1)-ésimo grupo é a mesma de ser seguido por duas outras granulometrias quaisquer, ou seja, no processo de deposição/erosão em determinado local independe daqueles dois imediatamente anteriores.

H₁: A probabilidade do (i)-ésimo grupo granulométrico ser seguido pelo (j)-ésimo e pelo (j+1)-ésimo é diferente de ser seguido por duas outras granulometrias quaisquer, ou seja, no processo de deposição/erosão em determinado local depende dos dois imediatamente anteriores.

Resultou-se deste teste, p-valores muito altos, rejeitando-se, desta forma, a hipótese de que o processo são seja somente de gênese de primeira ordem.

5. ANÁLISE CLÁSSICA

McLaren e Bowles (1985), propõe um modelo que demonstra como as distribuições granulométricas dos depósitos sedimentares mudam na direção de transporte utilizando-se os parâmetros de granulometria (média, variância e assimetria). Nos métodos clássicos (McLaren, 1981 ; McLaren e Bowles, 1985) não são levadas em consideração o grau de influência ou interação entre vizinhos próximos.

McLaren (1981) propôs um método dedutivo que utiliza as mudanças espaciais dos parâmetros de granulometria com o objetivo de identificar fontes de sedimentação, dinâmica de transporte e ambientes de deposição, segundo as seguintes suposições:

1. um depósito é o produto de uma única fonte;
2. a probabilidade de transportar grãos finos é maior que a de grãos mais grosseiros (desde que os sedimentos finos não sejam coesivos);
3. a probabilidade de deposição de grãos mais grossos é maior que a de grãos mais finos.

Desta forma chega-se às seguintes conclusões:

1. os sedimentos em transporte devem ser mais finos, com menor variância e mais negativamente assimétrico que os sedimentos na fonte (F, B, -);
2. um depósito residual deve ser mais grosso, com menor variância e mais positivamente assimétrico que a fonte (C, B, +).

McLaren e Bowles (1985) utilizam uma aproximação estatística para determinar a direção de transporte:

- n amostras;
- $\binom{n}{2}$ vetores que podem exibir uma tendência sugerindo transporte numa direção. Estes são obtidos comparando-se duas amostras de cada vez, levando-se em consideração uma influência dinâmica global interativa;
- oito possíveis tendências (equiprováveis), mas somente duas dão indicativos de transporte:

$$\{F, B, -\} \quad \{C, B, +\};$$

o que significa grãos mais finos, melhor selecionado e negativamente assimétricos ou grãos mais grossos, melhor selecionado e positivamente assimétricos.

Realiza-se um teste Z (ao nível de significância de $\alpha\%$) para

$$H_0: p \leq 0.125 \text{ (não existe direção preferencial)}$$

$$H_1: p > 0.125 \text{ (o transporte está ocorrendo em uma direção preferencial).}$$

A hipótese nula, H_0 é rejeitada se $Z = \frac{X - N \cdot p}{\sqrt{N \cdot p \cdot q}} > z_{\alpha}$, onde, X é o número observado de pares representando um caso particular em uma das duas direções opostas e N é o número total de pares unidirecionais possíveis $N = \binom{n}{2}$.

Gao e Collins (1991) confirmaram as tendências anteriores de McLaren & Bowles (1985) e sugeriram mais dois indicativos de transporte:

$$\{F, B, +\} \quad \{C, B, -\};$$

o que significa grãos mais finos, melhor selecionado e positivamente assimétricos ou grãos mais grossos, melhor selecionado e negativamente assimétricos.

No âmbito de verificar tais informações, implementamos algoritmos em SPlus com vistas ao cálculo destas medidas sugeridas, para estudar a dinâmica granulométrica na determinação do transporte de sedimentos. Assim, seguem nas tabelas Tab. 5a, Tab. 5b, Tab. 6a, Tab. 6b, Tab. 7a e Tab. 7b, alguns resultados, onde os testes foram feitos sob significância de $\alpha = 0.05$ ($Z_{tab} = 1.645$). Sugerindo um ambiente granulométrico C,B,+ (grão mais grossos, melhor selecionado e mais positivamente assimétrico) para todas as linhas em ambas as direções (indicativo de um ambiente caótico de deposição/erosão), exceto para a linha L3 direção NW-SE, onde nenhum teste foi significativo.

6. ANÁLISE ESTRUTURAL

Todavia, propõe-se, nesta pesquisa, levar-se em conta a gênese do processo, isto é estudá-lo sob a condição de que o fenômeno se determina localmente como uma Cadeia de Markov de primeira ordem. Assim:

- n amostras,
- $N = (n-1)$ vetores que podem exibir uma tendência sugerindo transporte numa direção preestabelecida. Estes são obtidos comparando-se duas amostras vizinhas, na seqüência, a cada vez. Portanto, N é o número total de pares unidirecionais possíveis,
- Vinte sete casos, dos quais identifica-se oito possíveis tendências equiprováveis, mas apenas quatro destas fornecem indicativos de transporte (Gao e Collins, 1981), isto é:
 1. F, B, - (grão mais finos, melhor selecionado e mais negativamente assimétrico);
 2. C, B, + (grão mais grossos, melhor selecionado e mais positivamente assimétrico);

3. F, B, + (grão mais finos, melhor selecionado e mais positivamente assimétrico) e
4. C, B, - (grão mais grossos, melhor selecionado e mais negativamente assimétrico).

Um teste exato para proporções é aplicado, ao nível de significância de α , baseado em (Dudewicz e Mishra, 1988), para as seguintes hipóteses:

$$H_0: p_i = p_0 = 0.125 \text{ (não existe direção preferencial concernente ao caso "i"), } i=1,2,3,4;$$

$$H_1: p_i > p_0 = 0.125 \text{ (o transporte está ocorrendo em pelo menos uma direção preferencial).}$$

Desta forma, determina-se um teste onde rejeita-se H_0 , ao nível α , se e somente se $P(X_i \geq j) \leq \alpha$ para um determinado j . Neste caso a variável aleatória X_i é o número observado de pares representando um caso particular (apenas uma das quatro configurações acima nos determinariam um padrão de informação condicional) em uma das direções preestabelecida. Assim, $X_i \sim B(n, p_i)$ sob H_0 , uma vez que as restrições de dicotomia, probabilidade constante e independência são facilmente verificadas, através da própria gênese da variável.

No âmbito de verificar tais informações, implementamos algoritmos em SPlus com vias ao cálculo destas medidas sugeridas para estudar a dinâmica granulométrica na determinação do transporte de sedimentos. Assim, seguem nas tabelas Tab. 8a, Tab. 8b, Tab. 9a, Tab. 9b, Tab. 10a e Tab. 10b, alguns resultados, onde os testes foram feitos sob significância de $\alpha = 0.05$ com a obtenção do valor crítico ($j=21$). Sugerindo um ambiente granulométrico C,B,+ (grão mais grossos, melhor selecionado e mais positivamente assimétrico) para todas as linhas em ambas as direções (indicativo de um ambiente caótico de deposição/erosão).

Desta forma, obtém-se um teste estrutural exato, robusto e eficiente (tempo de cálculo aproximadamente dez vezes menor) que nos fornece resultados análogos ao teste clássico proposto por McLaren e Bowles (1985).

7. ASSOCIAÇÃO CRUZADA

O Método da Associação Cruzada é aqui apresentado com o intuito de se atingir o segundo objetivo do estudo, isto é, comparar os perfis quanto à seqüência granulométrica identificada em cada um, individualmente. Este método é usado para comparação de duas seqüências que contêm

somente dados nominais, onde o uso do método convencional da correlação cruzada não é apropriado.

O método proposto, sugere que a cada posição correspondente, conservada a forma da seqüência, o número de correspondências e o número total de comparações sejam considerados. Assim, a razão simples:

$$R_s = \frac{\text{número de correspondências}}{\text{número de comparações}}$$

é calculada e pode ser usada como um índice de similaridade de duas seqüências nas posições sobrepostas.

Se representamos graficamente as posições de comparação *versus* a razão simples (R_s) obtemos um gráfico que pode ser chamado de “Associograma” (Fig. 4a, Fig. 4b e Fig. 4c).

A significância da razão simples (R_s) será determinada através de um teste qui-quadrado (χ^2), assim estaremos testando o nível de similaridade entre as duas seqüências em estudo. Para realizar tal teste é necessário determinar o número de correspondências e o número de não correspondências nas seqüências sobrepostas. Precisamos, também, da probabilidade de ocorrer um número dado de correspondências entre duas seqüências totalmente aleatórias.

Existem m possíveis categorias (estágios granulométricos) nas quais as observações podem ser classificadas. Se denotamos por X_{1k} o número de observações no (k)-ésimo estado (estágio) da seqüência 1 (linha L1), então o número total de observações nesta seqüência será dado por:

$$n_1 = \sum_{k=1}^m X_{1k} .$$

De forma análoga, o número total de observações da seqüência 2 (linha L2) será dado por:

$$n_2 = \sum_{k=1}^m X_{2k} ,$$

onde X_{2k} é o número de observações no (k)-ésimo estado da seqüência 2.

Assim, para se determinar a probabilidade de que um número dado de observações de duas seqüências sejam correspondentes, deve-se encontrar a soma dos produtos de cada uma das m categorias nas duas séries e determinar a razão pelo produto dos totais, isto é,

$$p = \frac{\sum_{k=1}^m X_{1k} X_{2k}}{n_1 n_2}.$$

As hipóteses a serem testadas consistem em:

H_0 : o índice de similaridade entre as duas seqüências não é significativo, ou seja, o número de correspondências entre as duas sucessões é menor que o esperado em duas seqüências aleatórias de composições equivalentes.

H_1 : o índice de similaridade entre as duas seqüências é significativo, ou seja, o número de correspondências entre as duas sucessões é maior que o esperado em duas seqüências aleatórias de composições equivalentes.

A estatística de teste é dada pela qui-quadrado de Pearson:

$$X^2 = \sum_{i=1}^2 \frac{(O_i - E_i)^2}{E_i} \sim \chi_1^2$$

onde O_1 = Número observado de correspondências,

O_2 = Número observado de não – correspondências,

E_1 = Número esperado de correspondências e

E_2 = Número esperado de não – correspondências.

Desta forma, obtém-se os Associogramas e os Testes de Similaridade, apresentados nas figuras Fig. 4a, Fig. 4b, Fig. 4c e na tabela Tab. 11, respectivamente.

8. CONCLUSÕES

De acordo com os resultados apresentados, podemos concluir que as linhas de seqüências granulométricas apresentam a característica markoviana de dependência de primeira ordem, dando consistência às suposições do modelo estrutural.

Segundo o que nos fornecem os associogramas, podemos concluir que as linhas sob estudo são similares. Desta forma, bastaria um estudo exaustivo de uma destas (1D) para poder-se concluir sobre o fenômeno regionalizado (2D). Deve-se ressaltar que na prática o processo não se apresenta tão simples assim, dever-se-ia estudar linhas segundo a sua batimetria e ambientes de agregação ou erosão, isto é, selecionar linhas menores que pudessem ser comparadas dentro de seu ambiente geomorfológico – pesquisa em andamento.

Vale lembrar que os resultados obtidos podem ter sido influenciados pela classificação granulométrica determinada. Esta deve levar em conta o conhecimento geológico do fenômeno, assim como satisfazer determinadas suposições teóricas estatísticas (como a suposição de normalidade).

Os métodos *de cadeias markovianas e associação cruzada*, mostram-se importantes ferramentas no auxílio ao estudo da definição do padrão granulométrico de determinada área sob investigação, para que se possa desta forma estudar a variabilidade do processo de *deposição/erosão*.

Quanto ao teste exato que foi implementado, sabe-se que abordagens devem ser realizadas no que concerne ao estudo de seu tamanho e poder, para que se possa efetivamente comparar o método clássico ao estrutural – estudo a ser realizado utilizando-se a técnica de “*bootstrap*”.

9. BIBLIOGRAFIA

Anderson, T. W. & Goodman, L. A. (1956) “*Statistical Inference About Markov Chains*”, *Annals Mathematical Statistics*, Vol. 18, p. 89 – 110.

Davis, J. C. (1986) *Statistics and Data Analysis in Geology*. 2nd.ed. John Wiley & Sons.

Dudewicz, E. J. & Mishra, S. N. (1988) “*Modern Mathematical Statistics*”. John Willey & Sons, New York.

Gao, S. & Collins, M. (1991) “*A critique of the ‘ McLaren method’ for defining sediment transport paths – discussion*”, *Journal of Sedimentary Petrology*, Vol. 61, p. 143 – 146.

Johnson, R. A. & Wichern, D. W. (1998) “*Applied multivariate statistical analysis*”. 2nd.ed, Prentice Hall, New Jersey.

McLaren, P. (1981) “*An Interpretation of Trends in Grain Size Measures*”, *Journal of Sedimentary Petrology*, Vol. 51, p. 611 – 624.

McLaren, P. & Bowles, D. (1985) “*The Effects of Sediment Transport on Grain-Size Distributions*”, *Journal of Sedimentary Petrology*, Vol. 55, p. 457 – 470.

LISTA DE FIGURAS E TABELAS

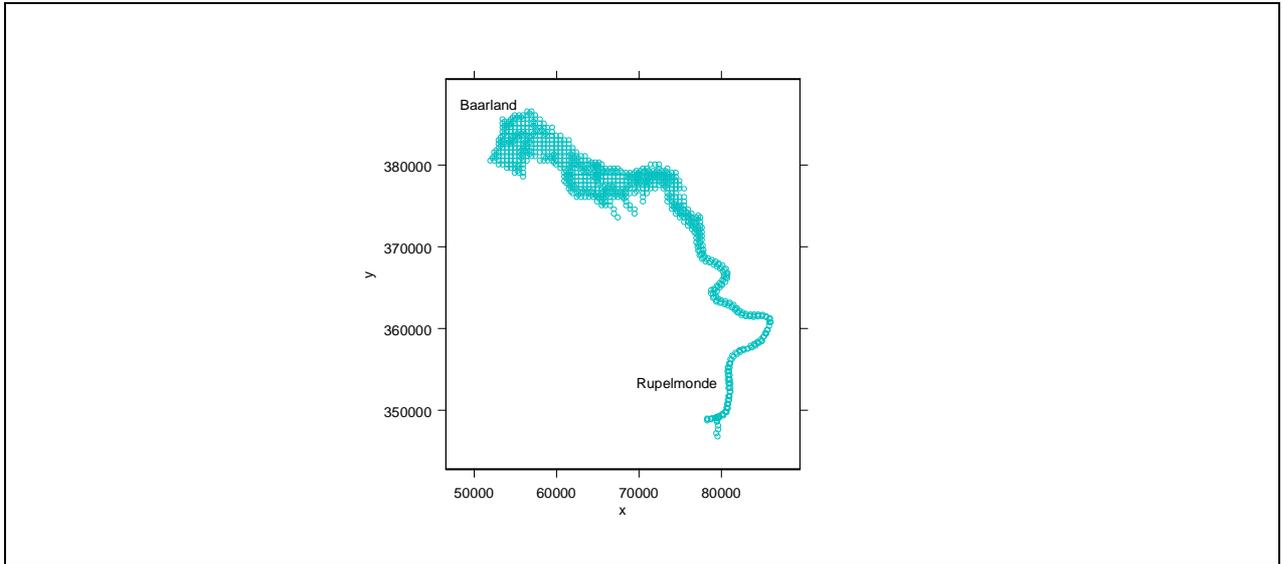


Fig. 1: Área global amostrada (867 amostras).

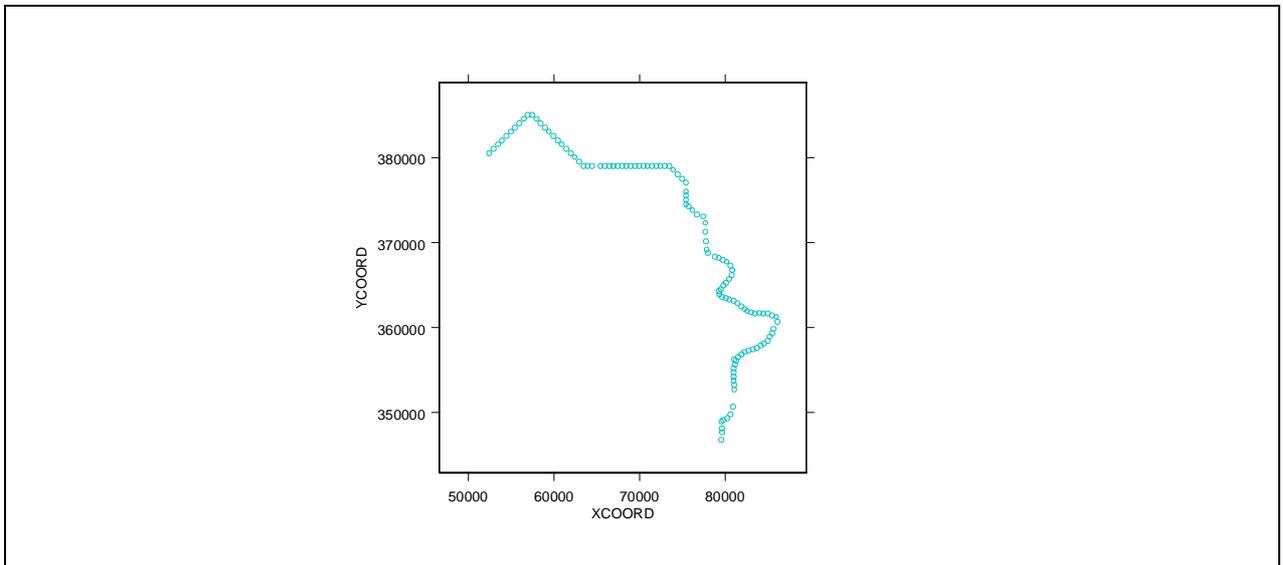


Fig. 2a: Linha L1, 117 amostras selecionadas de NW a SE.

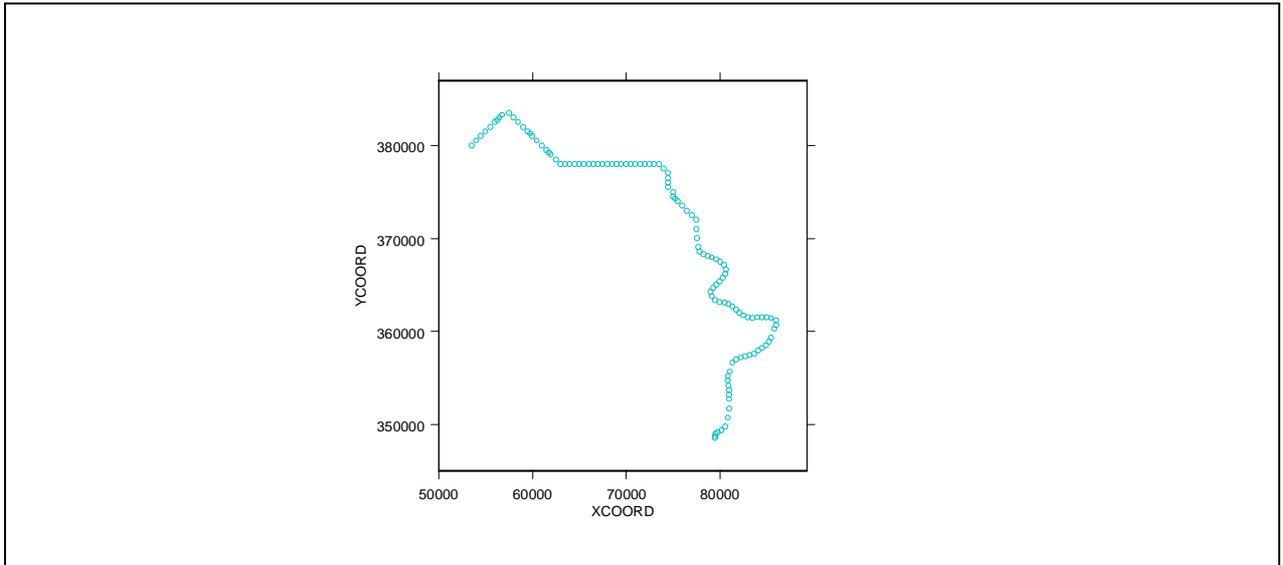


Fig. 2b: Linha L2, 118 amostras selecionadas de NW a SE.

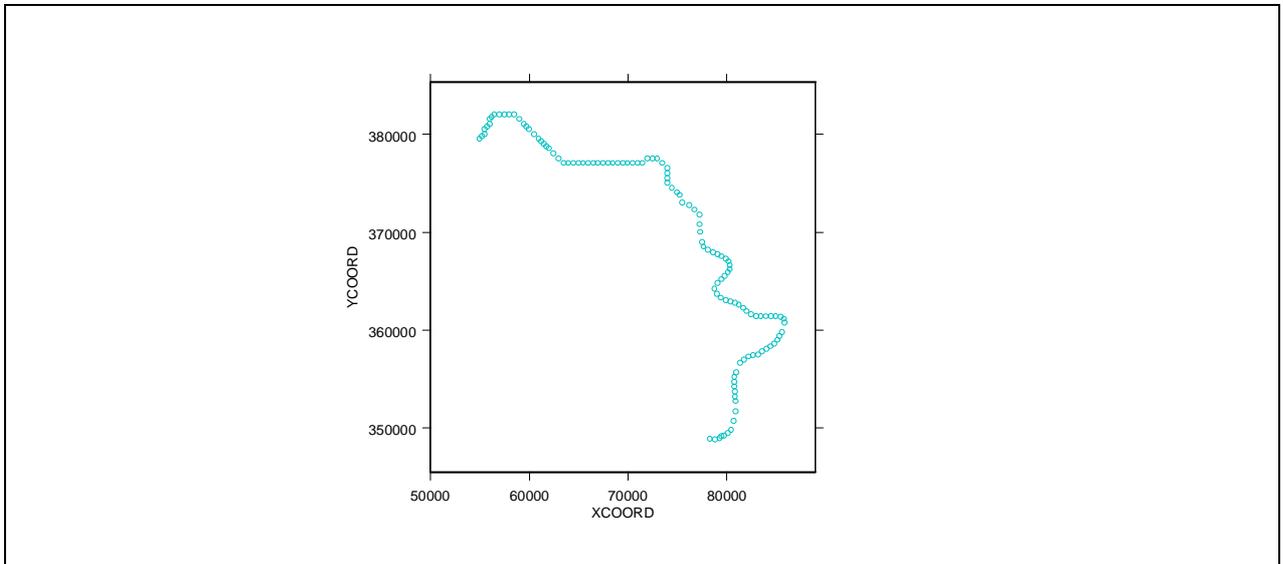


Fig. 2c: Linha L3, 119 amostras selecionadas de NW a SE.

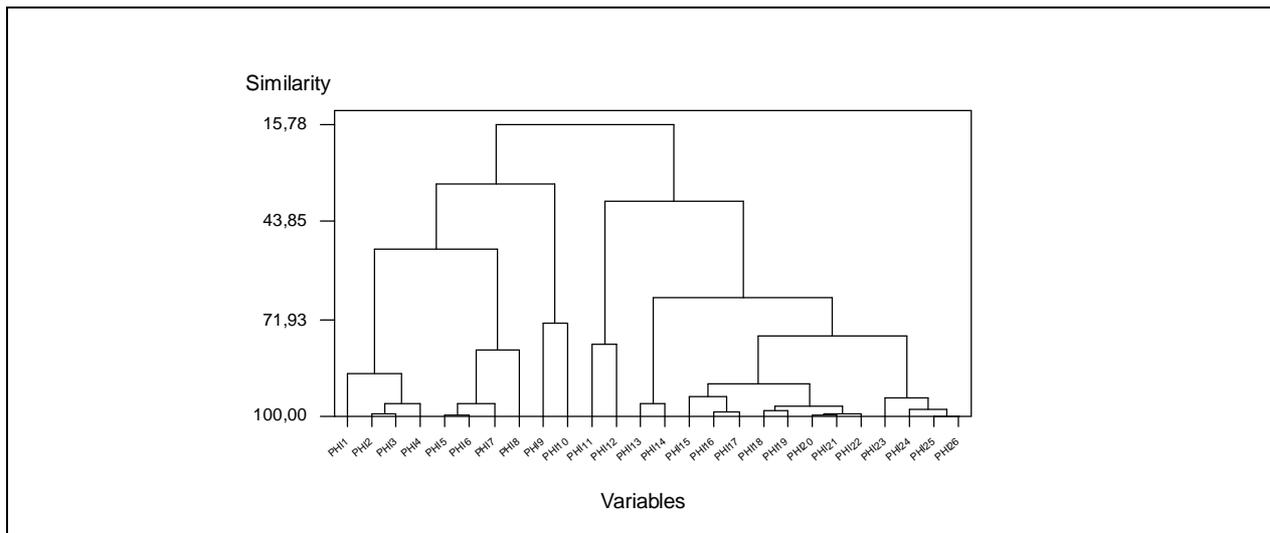


Fig. 3: Dendograma das 26 classificações granulométricas originais concernentes à região global.

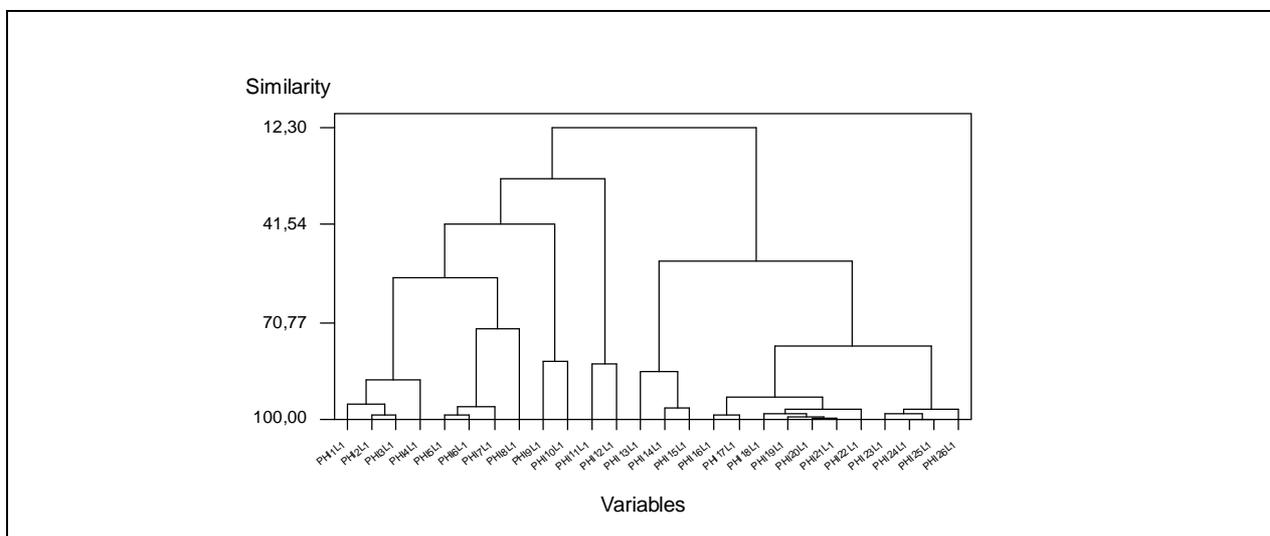


Fig. 3a: Dendograma das 26 classificações granulométricas originais concernentes à linha L1.

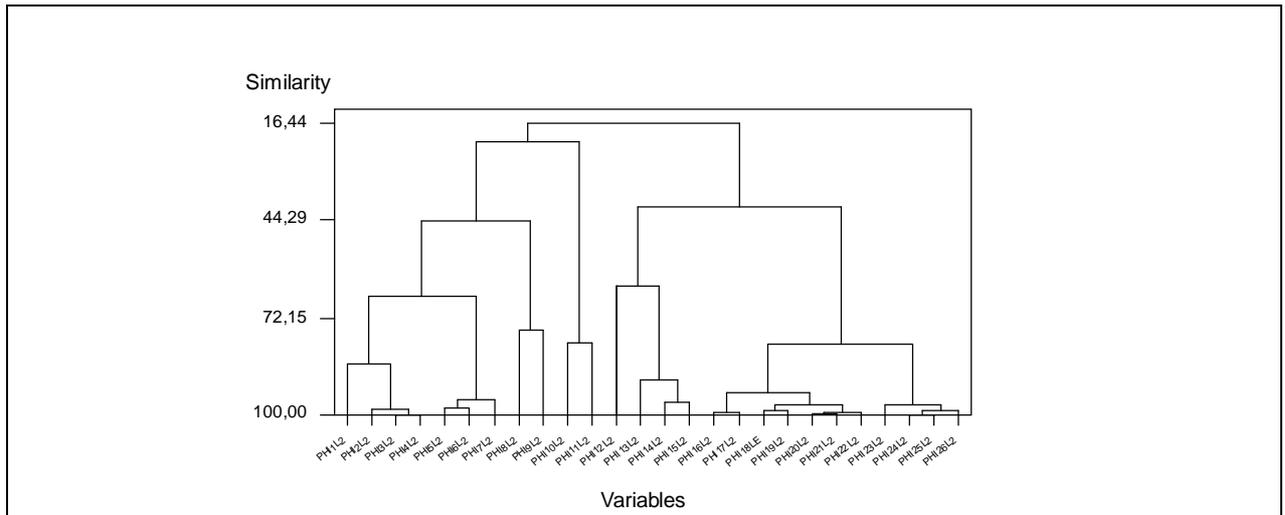


Fig. 3b: Dendrograma das 26 classificações granulométricas originais concernentes à linha L2.

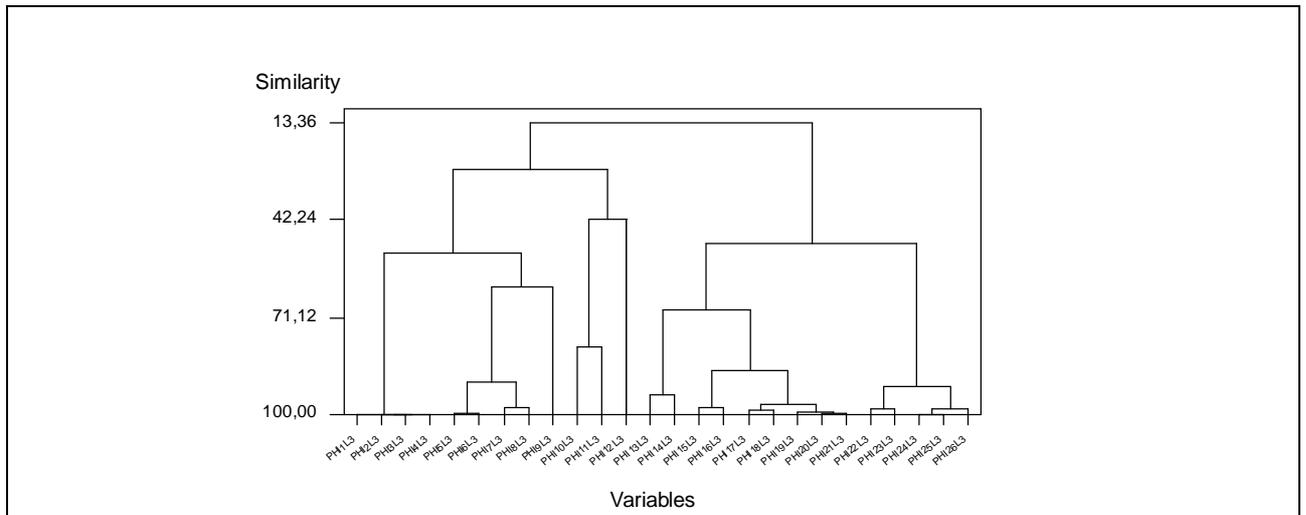


Fig. 3c: Dendrograma das 26 classificações granulométricas originais concernentes à linha L3.

	GMG	GMM	GMF	GF	S	A
GMG	6	1	4	1	0	1
GMM	1	0	5	0	1	0
GMF	5	6	30	5	1	4

GF	0	0	5	3	1	3
S	0	0	1	2	3	2
A	2	0	5	1	2	15

Tab. 1a: Matriz de transição referente à linha L1.

	GMG	GMM	GMF	GF	S	A
GMG	0.46	0.08	0.31	0.08	0.00	0.08
GMM	0.14	0.00	0.71	0.00	0.14	0.00
GMF	0.10	0.12	0.59	0.10	1.02	0.08
GF	0.00	0.00	0.42	0.25	0.08	0.25
S	0.00	0.00	0.125	0.25	0.375	0.25
A	0.08	0.00	0.20	0.04	0.08	0.60

Tab. 1b: Matriz de probabilidade de transição referente à linha L1.

	GMG	GMM	GMF	GF	S	A
GMG	12	6	1	0	1	2
GMM	9	37	7	0	0	3
GMF	1	9	8	0	0	2
GF	0	0	0	1	0	0
S	0	0	1	0	0	2
A	1	3	3	0	2	7

Tab. 2a: Matriz de transição referente à linha L2.

	GMG	GMM	GMF	GF	S	A
GMG	0.55	0.27	0.05	0.00	0.05	0.09
GMM	0.16	0.66	0.125	0.00	0.00	0.05
GMF	0.05	0.45	0.40	0.00	0.00	0.10
GF	0.00	0.00	0.00	1.00	0.00	0.00
S	0.00	0.00	0.33	0.00	0.00	0.67
A	0.06	0.19	0.19	0.00	0.125	0.44

Tab. 2b: Matriz de probabilidade de transição referente à linha L2.

	GMG	GMM	GMF	GF	S	A
GMG	9	2	3	0	1	3
GMM	1	1	3	0	0	0
GMF	5	2	37	2	5	4
GF	0	0	3	2	1	4
S	3	0	3	3	0	1
A	1	0	5	3	3	8

Tab. 3a: Matriz de transição referente à linha L3.

	GMG	GMM	GMF	GF	S	A
GMG	0.50	0.11	0.17	0.00	0.05	0.17

GMM	0.20	0.20	0.60	0.00	0.00	0.00
GMF	0.09	0.04	0.67	0.04	0.09	0.07
GF	0.00	0.00	0.30	0.20	0.10	0.40
S	0.30	0.00	0.30	0.30	0.00	0.10
A	0.05	0.00	0.25	0.15	0.15	0.40

Tab. 3b: Matriz de probabilidade de transição referente à linha L3.

	χ^2_{Calc}	χ^2_{Tab}	χ^2_{Tab}
L1	69.645	37.653 (25gl)	26.29 (16gl)
L2	58.245	37.653 (25gl)	19.65 (11gl)
L3	63.177	37.653 (25gl)	26.29 (16gl)

Tab. 4: Teste estatístico para a verificação da propriedade de dependência de primeira ordem

	F B -	C B +	F B +	C B -
N	6786	6786	6786	6786
X	776	1407	346	267
Z	-2.652	20.509	-18.435	-21.335

Tab. 5a: Método Clássico (linha L1) na direção NW-SE.

	F B -	C B +	F B +	C B -
N	6786	6786	6786	6786

X	473	2859	440	218
Z	-13.774	73.806	-14.985	-23.134

Tab. 5b: Método Clássico (linha L1) na direção SE-NW.

	F B -	C B +	F B +	C B -
N	6903	6903	6903	6903
X	364	1335	349	311
Z	-18.156	17.182	-18.702	-20.085

Tab. 6a: Método Clássico (linha L2) na direção NW-SE.

	F B -	C B +	F B +	C B -
N	6903	6903	6903	6903
X	589	2487	1031	437
Z	-9.968	59.107	6.119	-15.499

Tab. 6b: Método Clássico (linha L2) na direção SE-NW.

	F B -	C B +	F B +	C B -
N	7021	7021	7021	7021
X	730	872	470	236
Z	-5.327	-0.203	-14.709	-23.154

Tab. 7a: Método Clássico (linha L3) na direção NW-SE.

Direção: SE-NW

	F B -	C B +	F B +	C B -
N	7021	7021	7021	7021
X	499	3265	328	621
Z	-13.663	86.151	-19.834	-9.261

Tab. 7b: Método Clássico (linha L3) na direção SE-NW.

	F B -	C B +	F B +	C B -
n	117	117	117	117
x	11	28	13	9

Tab. 8a: Método Estrutural (linha L1) na direção NW-SE.

	F B -	C B +	F B +	C B -
n	117	117	117	117
x	16	28	7	4

Tab. 8b: Método Estrutural (linha L1) na direção SE-NW.

	F B -	C B +	F B +	C B -
n	118	118	118	118
x	7	29	9	7

Tab. 9a: Método Estrutural (linha L2) na direção NW-SE.

	F B -	C B +	F B +	C B -
n	118	118	118	118
x	12	29	16	8

Tab. 9b: Método Estrutural (linha L2) na direção SE-NW.

	F B -	C B +	F B +	C B -
n	119	119	119	119
x	13	23	12	10

Tab. 10a: Método Estrutural (linha L3) na direção NW-SE.

	F B -	C B +	F B +	C B -
n	119	119	119	119
x	15	25	10	10

Tab. 10b: Método Estrutural (linha L3) na direção SE-NW.

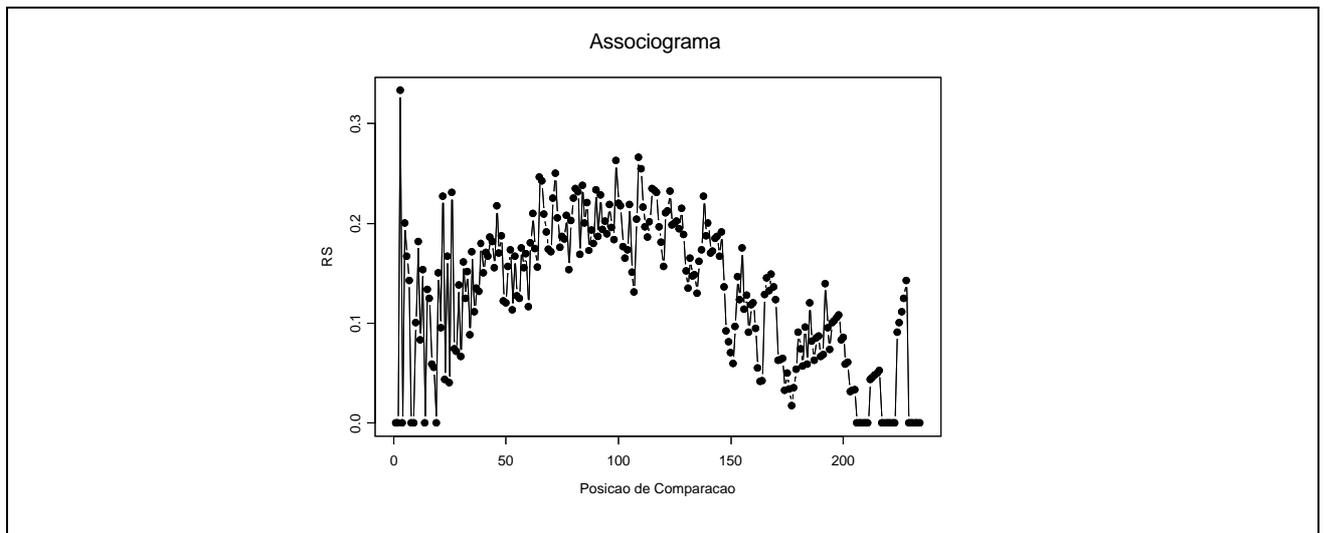


Fig. 4a: Associograma entre as linhas L1 e L2

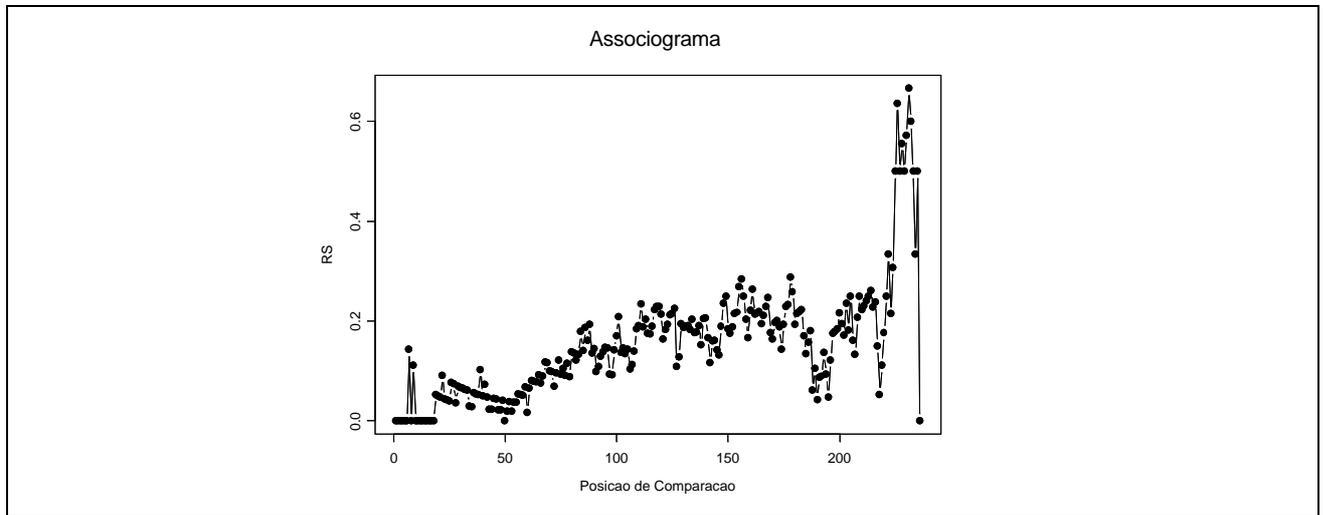


Fig. 4b: Associograma entre as linhas L2 e L3

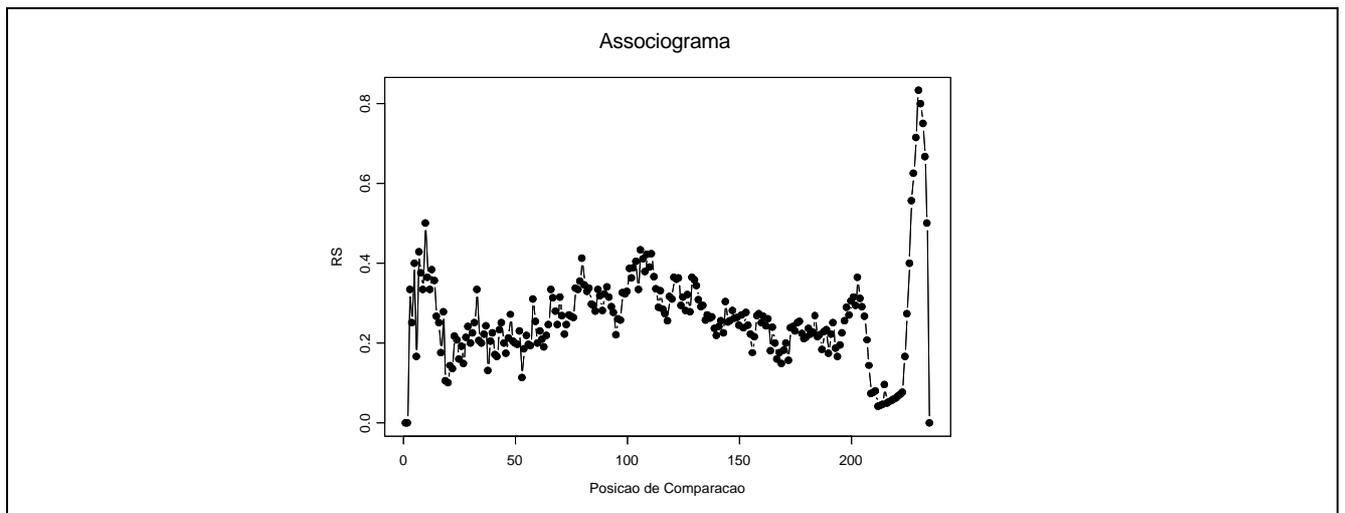


Fig. 4c: Associograma entre as linhas L1 e L3.

	Posição	χ^2_{Calc}	χ^2_{Tab}
L1 e L2	110	8.05	3.84

L2 e L3	161	6.89	3.84
L1 e L3	110	7.65	3.84

Tab. 11: Teste de similaridade associado à posição de maior similaridade entre as linhas. O número de pontos testados deve ser representativo.