



DETECÇÃO DE "OUTLIERS" MULTIVARIADAS EM PROSPECÇÃO GEOQUÍMICA. ESTUDO DE CASO NO PRÉ-CAMBRIANO URUGUAIO

FILIPPINI, J.⁽¹⁾, BJÖRKLUND, Å.⁽²⁾, GUSTAVSSON, N.⁽³⁾, OLIVEIRA, S.M.B. DE⁽⁴⁾

⁽¹⁾ Pós-graduando Curso de Geoquímica e Geotectônica. Departamento de Geologia Geral. Instituto de Geociências. Universidade de São Paulo. Caixa Postal 11348. CEP 05422-970. São Paulo - SP. e-mail: jmfilipi@usp.br

⁽²⁾ Department Geology. Åbo Academic University. 20500 Åbo - Finland. e-mail: alf.bjorklund@abo.fi

⁽³⁾ Geological Survey of Finland. fin02150. Espoo, Finland. e-mail: nils.gustavsson@gsf.fi

⁽⁴⁾ Departamento de Geologia Geral. Instituto de Geociências. Universidade de São Paulo. Caixa Postal 11348. CEP 05422-970. São Paulo - SP. e-mail: soniaoli@usp.br

RESUMO

Foram aplicados dois métodos multivariados de detecção de "outliers", para dados geoquímicos multielementares de amostras de regolitos em terrenos precambrianos do Uruguai. O comportamento das "outliers" demonstrou estar controlado pela ocorrência de anomalias regionais e locais, que foram discriminadas. Ambos os métodos resultaram eficientes na detecção das ocorrências mineralizadas conhecidas.

Palavras-Chave: prospecção geoquímica, "outliers" estatísticas, métodos multivariados.

INTRODUÇÃO

No tratamento de dados de prospecção geoquímica, as "outliers" são amostras inconsistentes com o comportamento geral da população. Do ponto de vista univariado podem ser definidas como "as amostras nos extremos da distribuição", porém sua concepção é diferente no caso multivariado. Uma "outlier" multivariada pode ser gerada por uma sucessão de desvios moderados, que não são visualizados ao se considerar cada variável em forma independente.

Neste trabalho foram aplicados dois procedimentos de detecção de "outliers" multivariadas sobre dados geoquímicos de sedimentos de corrente em terrenos pré-cambrianos do Uruguai. Um deles baseia-se na Análise das Componentes Principais-ACP (Jolliffe, 1986; Alecrim & Royer, 1991) e o outro considera as Distâncias de Mahalanobis-DM (Garrett, 1989). Esses procedimentos foram discutidos e comparados visando avaliar sua eficácia na detecção e discriminação de anomalias locais e regionais. Algumas ocorrências minerais conhecidas permitiram checar a validade do método do ponto de vista prospectivo.

A ÁREA DE PESQUISA

A região considerada possui forma poligonal e está integrada por quatro folhas 1/50.000. Trata-se de uma região com menos de 5.000 habitantes, clima temperado, com hidrografia pouco densa e homogênea. A paisagem típica é de pradarias, mas também existem ervas altas nos setores de relevo baixo (25-50m) e florestas autóctones que concentram-se nas margens dos rios e arroios.

Uma das principais feições geológicas da região é a ocorrência de duas grandes zonas de cisalhamento (Fesefeldt, 1988; Bossi et al., 1993), que marcam os limites de uma faixa de rochas sedimentares e vulcânicas de idade Brasileira, de 10km de largura, onde ocorrem rochas argilosas, carbonatadas e quartzosas afetadas por metamorfismo de baixo grau, lavas intermediárias máficas, anfíbolitos, leptinitos, e granitóides sin e tarditectônicos. A maior das falhas localiza-se no setor oriental da região, sendo representada por uma faixa de blastomilonitos de orientação N20°E. A outra, que atravessa a porção central da área, determina a separação entre o ciclo Brasileiro e o ciclo mais antigo, representado por um complexo granito-gnaissico, onde alternam-se termos félsicos e máficos.



MÉTODOS DE COLETA E TRATAMENTO DA INFORMAÇÃO

Amostragem e análises químicas

Foram coletadas 2172 amostras no total, em etapas de um mês, cobrindo cerca de um quarto da área cada uma: (i) 640 amostras-setembro/1982, (ii) 601 amostras-fevereiro/1985, (iii) 412 amostras-dezembro/1985 e (iv) 519 amostras -dezembro/1986. Alguns setores cobertos por sedimentos Fanerozoicos não foram prospectados. A densidade média de amostragem foi de uma amostra/km². Em 89% dos casos foram amostrados sedimentos em flat, que correspondem aos "overbank sediments" na definição do IGCP 259 (Darnley et al., 1995). Em forma alternativa foram considerados sedimentos em leito ou solos. Também foram coletadas 29 duplicatas de amostragem e 16 duplicatas analíticas para avaliar a precisão dos dados.

As amostras foram secadas a 50°C, peneiradas a 80 mesh, atacadas com ácidos (HClO₄-HCl-HF) e analisadas por espectrometria de emissão por plasma de arco de corrente contínua (Dutra, 1984). Os limites de detecção e os erros de amostragem e analíticos foram apresentados e discutidos em Filippini et al. (1997). Dez dos 22 elementos analisados (Ag, As, B, Be, Cd, Mo, Nb, Sb, Sn e W) apresentaram sérios problemas analíticos e foram descartados. Fe, Mn, P, Ba, Co, Cr, Cu, Ni, Pb, V, Y e Zn apresentaram erros analíticos inferiores a 10% e foram os elementos considerados para os fins deste trabalho.

Tratamento estatístico dos dados

Foram aplicados métodos não paramétricos e gráficos para o estudo das distribuições, análise de variância-ANVA e a análise das dependências. A ANVA foi útil para avaliar a influência do tipo de material amostrado e do litotipo da rocha de procedência sobre a resposta geoquímica. Isso significa investigar, por exemplo, se as amostras de solos apresentaram assinatura diferente da dos sedimentos. No outro caso, foi investigado se a resposta das sub-populações de amostras classificadas segundo o litotipo predominante mantém uma assinatura semelhante ou se pelo menos uma sub-população mostra resposta diferente. Também foram elaborados mapas de círculos e de interpolação de variáveis reais e teóricas.

A ACP foi utilizada para reduzir o número de variáveis e para definir as "outliers". Foram calculadas as estatísticas definidas pela soma dos quadrados das CPs divididas pelos seus desvios padrões e foram analisados os diagramas de dispersão das CPs. As primeiras e as últimas CPs discriminariam as "outliers nos extremos das distribuições univariadas" das geradas por uma sucessão de desvios moderados (Jolliffe, 1986). Efetuando uma pequena modificação do método, neste trabalho foram calculadas duas estatísticas da soma de quadrados das CPs, uma considerando as CPs 1-5 (D^2_{CP1-5}) e a outra as CPs 6-12 (D^2_{CP6-12}). Para otimizar a resposta da ACP, foram realizados vários tratamentos, sendo comparados os autovalores, as cargas e os escores.

A DM de uma amostra é um número escalar obtido pelo produto vetorial entre o "vetor amostra fila", a matriz de correlação e o "vetor amostra coluna". As DM foram calculadas pelo Dr. Nils Gustavsson do Serviço Geológico da Finlândia com um programa próprio, elaborado segundo a definição de Afifi & Azen (1972), considerando as 12 variáveis citadas.

RESULTADOS

O conjunto mínimo-média-máximo para cada variável apresentou o seguinte resultado: Fe=1-3-8%, Mn=74-998-15610ppm, P=99-292-2528ppm, Ba=174-557-2272ppm, Co=1-12-78ppm, Cr=8-42-1980ppm, Cu=2-19-60ppm, Ni=0-22-940ppm, Pb=0-14-116ppm, V=18-65-240ppm, Y=8-24-266ppm e Zn=17-59-176ppm. Todos os elementos ajustaram-se melhor ao comportamento log-normal. Assim, quando foram aplicados métodos paramétricos, os dados foram log-transformados.

Com exceção do P, a ANVA em função do tipo do material amostrado, indicou empobrecimento das amostras de solos em relação aos sedimentos. Para uniformizar a resposta, os dados foram expressos em função da mediana dos grupos definidos pelo material amostrado. Ao considerar a classificação segundo o litotipo da rocha de procedência, a ANVA mostrou uma forte influência desse fator, sugerindo contrastes entre domínios geológicos em diferentes escalas. Os coeficientes de correlação e os diagramas de correlação mostraram correlações positivas entre Fe, Co, Cr, Cu, Ni, V e Zn.

As DM foram calculadas com os dados transformados em logaritmos neperianos e divididos segundo as medianas do material amostrado. Os diagramas probabilidade teórica-probabilidade empírica mostraram um grande afastamento do comportamento 12-normal ou 12-log-normal. Isto poderia significar que, considerando o espaço das 12 variáveis, existiriam várias sub-populações. O mapa de círculos das DM mostrou uma grande anomalia no setor NW, relacionada ao complexo granito-gnaissico. Outra anomalia regional ocorreu no N da área. O resto foram anomalias locais relacionadas à faixa de rochas entre as duas falhas de cisalhamento.

A ACP ótima considerou os dados log-transformados e divididos pelas medianas segundo o material amostrado. Foram eliminadas as 44 "outliers" superiores definidas pelas DM. No quadro abaixo são expostos os resultados das cinco primeiras CPs (os signos negativos indicam oposição do elemento em relação à CP e os parênteses dúvidas):

CP	Variância	Elementos	Interpretação
1	47%	Fe,(Mn),(P),(Ba),Co,Cr,Cu,Ni, -Pb,V,Y,Zn	Ocorrência de ferromagnesianos e/ou fenômenos supérgenos de adsorção em óxidos de Fe-Mn
2	13%	Mn,(Co),Pb,-Y	Adsorção em óxidos de Mn
3	11%	P,Ba,-Cr,-Ni	Rochas félsicas fosforadas
4	7%	-Mn,Ba,Pb	Granitóides com K-feldspato
5	5%	Y,-Cu,-Zn	Duvidosa

A CP5 não mostrou significado claro e foi interpretada como o limite entre as primeiras e as últimas CPs. Os mapas de círculos mostraram distribuições diferentes para as estatísticas D^2_{CP1-5} e D^2_{CP6-12} .

CONSIDERAÇÕES FINAIS

O tratamento estatístico evidenciou a existência de várias sub-populações. O mapa das DM mostrou uma anomalia regional importante no setor NW da área. Trata-se de uma anomalia de mais de 100km², com associação entre elementos femáfilos (Fe, Cr, Cu, Ni...) e boa sobreposição com os anfíbolitos que afloram no setor, estando provavelmente relacionada aos ferromagnesianos que ocorrem nessas rochas. Os critérios para discriminar anomalias regionais e locais foram: (i) o tamanho da anomalia (Darnley et al., 1995), (ii) a existência de sobreposição com litotipos ou unidades geológicas e (iii) as características da associação geoquímica.

As "outliers" definidas pelas primeiras CPs também evidenciaram a anomalia regional do setor NW, mas isso não aconteceu com as "outliers" definidas pelas últimas CPs, que estariam relacionadas às anomalias locais Cu-Pb-Zn geradas pela faixa de rochas localizada entre as duas zonas de cisalhamento. As distribuições diferentes das "outliers" definidas pelas primeiras e pelas últimas CPs pareceram confirmar o conceito intuitivo de que as anomalias regionais estariam associadas às primeiras CPs e as anomalias locais às últimas.

As "outliers" definidas pelas DM e pelas últimas CPs detectaram as quatro ocorrências minerais conhecidas com precisão quilométrica (0-2 km), sendo as últimas algo mais sensíveis. A comparação das "outliers" definidas pelas primeiras e as últimas CPs poderia ajudar a discriminar anomalias regionais e locais.

Finalmente, uma característica importante de ambos os métodos é a possibilidade de detectar "outliers" multivariadas, definidas por combinações diferentes de teores e associações, que corresponderiam a eventos específicos, em concordância com a diversidade de aspectos que dão origem e controlam os fenômenos naturais.

AGRADECIMENTOS: À Universidade Acadêmica de Åbo-ÅAU e ao Serviço Geológico da Finlândia-GSF por permitir, através de um estágio de seis meses na Finlândia, efetuar o processamento dos dados deste trabalho. Ao Lic. Eero Lampio e ao Lic. Heimo Savolainen, ambos do Departamento de Geoquímica do GSF, pela sua colaboração na produção de mapas de alta qualidade.

REFERÊNCIAS BIBLIOGRÁFICAS

- AFIFI, A. & AZEN, S. (1972) Statistical Analysis-A Computer Oriented Approach. Academic Press, p.228-230.
- ALECRIM, J. & ROGER, J. (1991) A detecção das "outliers" na análise das componentes principais e sua aplicação ao tratamento de dados em prospecção. In: III Congresso Brasileiro de Geoquímica e I Congresso de Geoquímica dos Países de Língua Portuguesa. Resumos, 2:447-451.
- BOSSI, J., PRECIOZZI, F.; CAMPAL, N. (1993) Predevoniano en el Uruguay. Tomo I: Terreno Piedra Alta. MIE-DINAMIGE, 50p.
- DARNLEY, A.; BJÖRKLUND, A.; BØLBIKEN, B.; GUSTAVSSON, N.; KOVAL, P.; PLANT, J.; STEENFELT, A.; TAUCHID, M.; XIE XUEJING (1995) A global geochemical database for environmental and resource management. Final Report of IGCP 259. UNESCO Pub., 122p.
- DUTRA, C. (1984) Espectrografia óptica de emissão. In: C.B. Gomes (Coord.). Técnicas analíticas instrumentais aplicadas à geologia. Ed. Edgard Blücher Ltda., p.121-158.
- FESEFELDT, K. (1988) Asesoriamiento geológico y yacimentológico a la DINAMIGE. BGR, 189p.
- FILIPPINI, J. (1992) Aplicação dos métodos estatísticos em prospecção geoquímica de sedimentos de correntes e rochas na região de "Isla Patrulla", Distrito de "Treinta y Tres", Uruguai. Dissertação de Mestrado, IG-USP, 186p.
- FILIPPINI, J., DE OLIVEIRA, S.M. B. & BJÖRKLUND, A. (1997) Avaliação de erros em prospecção geoquímica. III Congresso de Geoquímica dos Países de Língua Portuguesa. Universidade do Minho. Actas, p.179-182.
- GARRETT, R. (1989) The chi-square plot: a tool for multivariate outlier recognition. Journal of Geochemical Exploration, 32: 319-341.
- JOLLIFFE, I. (1986) Principal component analysis. Springer-Verlag, 271p.