



V CONGRESSO DE GEOQUÍMICA DOS
PAÍSES DE LÍNGUA PORTUGUESA

VII CONGRESSO BRASILEIRO
DE GEOQUÍMICA

ANNAIS



SOCIEDADE BRASILEIRA DE GEOQUÍMICA - SBGq

17 a 21 de outubro de 1999
PORTO SEGURO - BAHIA - BRASIL



PROPOSTA METODOLÓGICA DE UTILIZAÇÃO DA ANÁLISE DE COMPONENTES PRINCIPAIS EM PROSPECÇÃO GEOQUÍMICA

J.L.S. ANDRIOTTI

andri@portoweb.com.br - CPRM - Serviço Geológico do Brasil

INTRODUÇÃO

Uma das características mais comumente observadas nos estágios iniciais das campanhas de prospecção geoquímica é a de serem baseadas no estudo de diversas variáveis simultaneamente. Isso se deve ao fato de que, na conformação de uma paisagem geoquímica atual, diversas variáveis interagiram no tempo e no espaço para o seu estabelecimento. Quando o prospector decide analisar diversas variáveis ele já está fazendo um reconhecimento da importância de todas elas no resultado final (a paisagem geoquímica presente); se, de posse da totalidade dos dados, for tomada a decisão de utilizar a Estatística como técnica auxiliar de interpretação, a Estatística Multivariada passa a ter um papel muito importante na eficácia dos resultados obtidos na interpretação. Em palavras mais simples e diretas, a utilização da Estatística não se constitui em condição *sine qua non* para a interpretação de dados oriundos de prospecção geoquímica, mas, ao se optar por utilizá-la, se torna imperativo que se o faça do modo mais adequado; ou seja, na presença de diversas variáveis deve-se usar, também, a Estatística Multivariada.

Diversas são as técnicas disponíveis que podem trazer ganhos para a interpretação desse tipo de dados (multivariados), como a Análise de Agrupamentos, a Análise Discriminante, a Análise de Variância Multivariada, a Análise Fatorial e a Análise de Componentes Principais. Dessas, apenas a última é abordada no presente trabalho.

ANÁLISE DE COMPONENTES PRINCIPAIS

A Análise de Componentes Principais (ACP) é uma técnica analítica que obtém transformações lineares de um grupo de variáveis correlacionadas de modo que certas condições ótimas são obtidas, como a de que as novas variáveis obtidas pela transformação sejam não correlacionadas. Além disso, a dimensionalidade do problema é reduzida, pois as primeiras variáveis hipotéticas geradas - as primeiras componentes principais - respondem por grande parte da variabilidade total dos dados. Muito embora sejam geradas tantas componentes principais quantas são as variáveis originais, idealmente apenas as primeiras são analisadas, por representarem grande parte das variações totais presentes nos dados disponíveis.

Após a geração das componentes principais a partir dos dados originais deve-se tentar dar uma interpretação às mesmas, sendo usual escolher, para a composição das componentes principais, as variáveis para as quais os coeficientes na componente principal sejam relativamente grandes, positivos ou negativos. Se uma grande parte (70% ou 80%, por exemplo) da variância total da população puder ser atribuída às primeiras poucas componentes, então elas podem substituir as variáveis originais sem muita perda de informação.

Existem várias técnicas que orientam a decisão sobre a retenção de componentes principais para interpretação, sendo a mais aceita e recomendada a que propõe a retenção das componentes principais cujos percentuais de variabilidade total expliquem uma proporção importante das variações totais presentes no conjunto original de dados; a não utilização de componentes principais a partir de um certo nível (a partir da quarta, ou da quinta componente principal, por exemplo) não causa nenhum problema sobre as utilizadas, por não afetar sua interpretação. O registro é pertinente porque tal não ocorre, por exemplo, na aplicação da Análise Fatorial.

Um outro procedimento muito utilizado e recomendado, e que está presente na grande maioria dos softwares disponíveis, é o de reter apenas as componentes principais cujos autovalores tenham valor superior a um.

METODOLOGIA PROPOSTA

Andriotti (1999) detalha a presente proposta metodológica de aplicação da Análise de Componentes Principais a dados obtidos em campanhas de amostragem geoquímica, com exemplos práticos de aplicação.

A figura 1 mostra o fluxograma da metodologia proposta, cujos passos principais são descritos resumidamente a seguir.

1 - Compilação de todos os dados disponíveis para a área de estudo, separando-os por grupos conforme o método de amostragem adotado, o método analítico utilizado, a época de coleta, a entidade ou equipe responsável pela coleta ou pela análise dos materiais disponíveis.

2 - Verificar, por meio de métodos estatísticos (univariados e multivariados), se os diferentes grupos separados no item anterior apresentam, entre si, diferenças significativas a um nível de significância pré-estabelecido; se não houver diferenças significativas, reuni-los formando um novo grupo, e, se houver, escolher qual dos grupos de dados melhor representa a área estudada.

3 - Identificação da presença ou não de outliers em cada grupo selecionado para estudo. Outliers devem ser re-amostrados e re-analisados ou, em caso de impossibilidade, eliminados dos grupos criados.

4 - Criação de matrizes para cada grupo criado. No presente caso, cada grupo corresponde a um meio de amostragem: sedimentos de corrente, concentrados de minerais pesados, rochas ou outro. As matrizes devem conter os dados em mesmas unidades; se as unidades de medida não forem originalmente as mesmas (percentagens, partes por milhão, etc.) deve-se padronizar os dados.

5 - Aplicação da técnica de Análise de Componentes Principais aos dados disponíveis, por meio

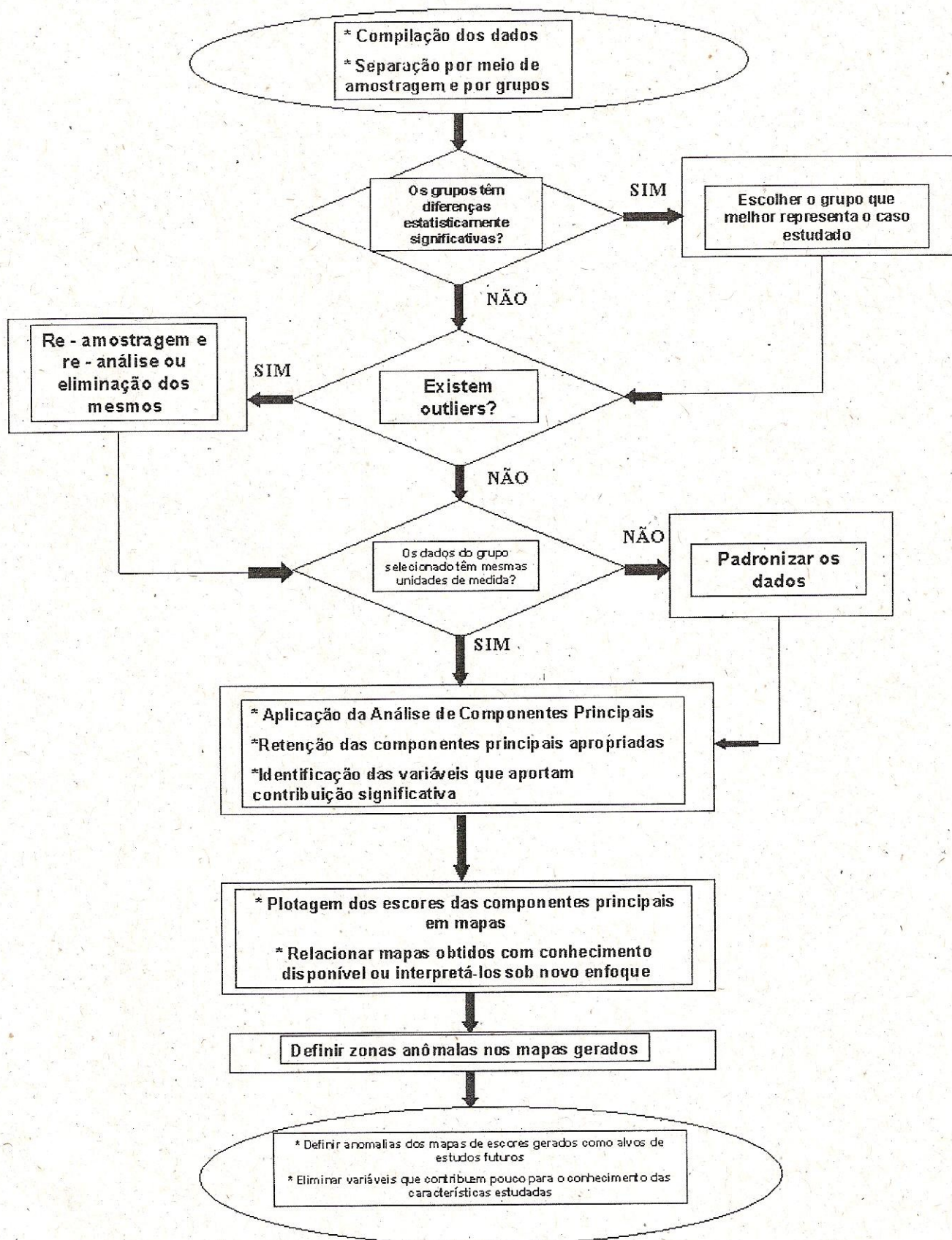


Figura 1 - Fluxograma da Metodologia Proposta



de *softwares* específicos. Se a aplicação fornecer resultados interpretados como respondendo satisfatoriamente, sob o aspecto estatístico, aos dados disponíveis, reter apenas as componentes principais adequadas. Pelos resultados obtidos neste item se identifica, para cada caso, quais variáveis contribuem significativamente e quais contribuem pouco para o avanço do conhecimento das características estudadas com este tipo de abordagem (Análise de Componentes Principais).

6 - Plotar em mapa os escores de cada componente principal retida, e relacionar os mapas obtidos com o conhecimento geológico da área estudada (mapa geológico, mineralizações, estruturas, anomalias, grupos litológicos, mapa metalogenético, etc.). As componentes principais não retidas não são interpretadas.

7 - Com a associação de cada componente principal com um processo geológico atuante na área (ou com um objetivo de interesse), definir as anomalias dos mapas gerados no item 6 como alvos de futuras etapas de trabalho, identificando as áreas prioritárias. Nas etapas seguintes apenas as variáveis que aportam contribuição significativa (item 5) devem ser analisadas, gerando redução de custos na etapa de análises químicas.

CONCLUSÕES

Em se tratando de variáveis regionalizadas, caso da Geologia, em que a representação em mapa é de fundamental importância, a Análise de Componentes Principais representa ganhos em termos de associações de variáveis e sua ligação com os processos geológicos atuantes, na simplificando a interpretação pela redução da dimensionalidade do problema.

A ACP enseja diminuição de custos em etapas posteriores do trabalho e, em especial, permite a visualização, em mapas, da distribuição de valores correspondentes a determinados eventos geológicos (as componentes principais), o que raramente é possível em mapas de uma única variável.

Os mapas das componentes principais retidas para estudo são de grande utilidade na interpretação dos processos atuantes na área de trabalho, devendo ser acrescentados aos das variáveis individuais na interpretação dos resultados de campanhas de prospecção geoquímica em que diversas variáveis estejam disponíveis para estudo.

REFERÊNCIA BIBLIOGRÁFICA

ANDRIOTTI, J.L.S. (1999) Interpretação e modelamento de dados geoquímicos e de sensoriamento remoto por meio da Análise de Componentes Principais: implicações na exploração de depósitos de metais no Escudo Sul-riograndense. Tese de Doutorado, UFRGS, 227 p.