

APLICAÇÃO DE ANÁLISE FATORIAL PARA DADOS HIDROQUÍMICOS DO GRUPO BAURU NO ESTADO DE SÃO PAULO

Ademar Tokio Ogawa
Heraldo Cavalheiro Navajas Sampaio Campos

Departamento de Águas e Energia Elétrica — DAEE

ABSTRACT - Factor Analysis was applied to results of chemical analysis of 427 water samples from wells in the Bauru Group in the State of São Paulo, Brazil. Seven variables were analyzed. The analysis showed that 70% of all variance were accounted for by only three factors. Factor 1 accounts for almost 40% of the total variance. Factor-Scores were computed and plotted to show the areas in which the variables were more significant.

1. INTRODUÇÃO

As análises químicas de água representam exemplo típico de dados multivariantes, ou seja, dados constituídos de observações de várias variáveis para cada elemento de uma amostra de indivíduos. Devido às diversas fontes de variação no conjunto de dados, sua análise se torna complexa e normalmente difícil de ser interpretada.

Uma das soluções para o problema é a técnica denominada Análise Fatorial, que trata das relações internas de um conjunto de variáveis correlacionadas por um conjunto menor de "Fatores" ou variáveis hipotéticas que podem ser correlacionadas ou não e que explicam a maior parte da variação do conjunto original.

As amostras foram tomadas como um todo, não tendo sido feita nenhuma separação entre as águas provenientes das diferentes unidades litológicas que compõem o Grupo Bauru.

O objetivo portanto foi o de se verificar se existiam fontes de variação nos dados que explicassem o conjunto total de amostras e não cada uma das unidades do Grupo Bauru em particular. Assim sendo trata-se de uma tentativa de exploração do método como ferramenta de auxílio na interpretação de análises químicas de águas subterrâneas.

2. CARACTERIZAÇÃO GEOLÓGICA E HIDROQUÍMICA DA ÁREA

A área objeto de estudo corresponde ao domínio de ocorrência do Grupo Bauru no Estado de São Paulo. (Figura 1).

A geologia de superfície da área de pesquisa é caracterizada por sedimentos cretáceos recobrendo discordantemente os basaltos da Formação Serra Geral. A sequência sedimentar constitui o sistema aquífero Bauru, principal fonte de exploração de água subterrânea no Estado, abrangendo uma área de 104.000 km².

Na base dessa sequência, tem-se a formação Caiuã, constituída essencialmente por arenitos de granulação fina a média, baixo teor de matriz, arredondamento moderado a bom, friáveis, com coloração característica roxo violeta e vermelho escuro. Sua área de ocorrência restringe-se ao extremo sudoeste da área.

Acima da Formação Caiuã ocorrem sedimentos da Formação Santo Anastácio, apresentando contato transicional com a unidade subjacente. É constituída por arenitos de granulação média e fina, dominando a fração fina, teor baixo de matriz de coloração vermelho-escuro, com raras lentes lamíticas intercaladas. Aflora principalmente a oeste da região, adentrando a leste através das partes baixas dos vales dos rios Aguapeí e Peixe.

A Formação Adamantina representa um conjunto de fácies cuja principal característica é a presença de bancos de arenitos de granulação fina a muito fina, cor róseo a castanho, portando estratificação cruzada, alternadas com bancos de lamitos, siltitos e arenitos lamíticos, de cor marron avermelhado e cinza castanho, maciços ou com acamamento plano paralelo grosseiro. É a unidade de maior distribuição em superfície no Estado de São Paulo, aflorando extensivamente na região norte (bacia do Rio Turvo); norte ocidental (bacia do Rio Turvo); norte ocidental (bacia do Rio São José dos Dourados e baixo Tietê) nos espigões entre os rios Tietê e Aguapeí; entre os rios Aguapeí e Peixe, a oeste de Tupã; entre os rios Peixe e Santo Anastácio na Região de Presidente Prudente - Presidente Venceslau; entre os rios Santo Anastácio e Paranapanema (região de Paranapanema); espigão entre o rio do Peixe e Paranapanema e médio e alto vale do Rio do Peixe e Vale do Rio do Turvo na região de São Pedro do Turvo - Ubirajara.

O contacto basal da Formação Adamantina, em toda a parte oeste do planalto é com a Formação Santo Anastácio e na parte superior com

a Formação Marília.

A Formação Marília é uma unidade caracterizada por arenitos grossos a conglomeráticos, imaturos textural e mineralogicamente, com abundantes nódulos calcíferos. Aflora na parte leste do Planalto Ocidental do Estado de São Paulo, prolongando-se para oeste nos espigões entre os principais rios.

Do ponto de vista hidroquímico, de modo geral, as águas que circulam no aquífero Bauru apresentam baixa concentração salina, com valores de resíduo seco raramente atingindo 300 mg/l. A distribuição dos teores salinos em área (figura 1) mostra que o intervalo de 100-200 mg/l recobre grande parte do domínio de ocorrência da formação Adamantina; valores acima de 200 mg/l são encontrados ao longo dos espigões de Pompéia-Adamantina e de Vaparaíso-Mirandópolis. As águas de concentração salina mais baixa, inferior a 100 mg/l, têm ocorrência dominante nos vales baixos, especialmente à jusante dos principais rios interiores.

3. METODOLOGIA

3.1. O MODELO FATORIAL

Dado em conjunto de p variáveis, com N observações para cada variável, que formam o arranjo de valores:

$$(X_{ik}), \quad i=1 \dots p \quad K=1 \dots N$$

denominado matriz de dados e supondo que as p variáveis são correlacionadas, pergunta-se se é possível reduzir a dimensão inicial (p) do problema através de novas variáveis hipotéticas que mesmo em menor número explicarão a maior parte da variação original. Assim, supõe-se que cada variável $X_i, i=1 \dots p$ é representada por uma função linear de um número menor ($m < p$) de variáveis hipotéticas denominadas fatores comuns (por serem comuns a várias variáveis) mais um fator residual ou único para cada variável.

O modelo fatorial pode ser expresso como:

$$X_i = \sum_{j=1}^m \lambda_{ij} f_j + e_i \quad i=1 \dots p$$

onde:

f_j = j -ésima fator comum.

e_i = resíduo representando fontes de variação que afetam apenas a variável X_i

λ_{ij} = cargas fatoriais (refletem a importância do j -ésima fator na

composição da i -ésima variável X_i .

A variância de e_i é denominada variância residual (ou especificidade) e é denotada por ψ_i .

A parte da variância de cada variável, devida só aos fatores comuns é dada por

$$h_i^2 = \sum_{j=1}^m \lambda_{ij}^2$$

e é denominada comunalidade da variável X_i .

Na prática os parâmetros do modelo a serem estimados, a partir dos dados experimentais são os λ_{ij} (cargas fatoriais) e os ψ_i (especificidades).

Após as cargas fatoriais terem sido obtidas o analista deve interpretar os fatores comuns da melhor maneira possível. Com este objetivo faz-se a rotação dos fatores.

Finalmente pode-se estimar os valores dos fatores comuns para cada indivíduo, como função das variáveis observadas. Esses valores denominam-se escores fatoriais.

3.2. A ANÁLISE EFETUADA

Neste trabalho foram utilizados os resultados de 427 análises químicas de águas de poços profundos distribuídos homogeneamente por todo o Grupo Bauru.

Os dados foram submetidos ao subprograma Factor do SPSS (Statistical Package for the Social Sciences) disponível no Centro de Computação Eletrônica da Universidade de São Paulo.

Foi feita uma análise preliminar com 28 variáveis sendo que 15 delas eram variáveis simples e as 13 restantes eram índices calculadas a partir das anteriores. Os resultados não foram claros devido à grande quantidade de fatores a serem analisados.

Após várias tentativas observou-se que um menor número de variáveis poderia fornecer resultados mais adequados para interpretação. Assim sendo, foram submetidas ao programa apenas as concentrações, em meq/l dos seguintes ions: Na, K, Ca, Mg, HCO_3 , SO_4 e Cl.

3.3. SEQUÊNCIA DE CÁLCULOS

Inicialmente o programa calcula os coeficientes de correlação - produto momento de Pearson entre as variáveis. Os resultados obtidos encontram-se na tabela 1.

Tabela 1: COEFICIENTE DE CORRELAÇÃO

	Na	K	Ca	Mg	HCO ₃	SO ₄	Cl
Na	1.000	-	-	-	-	-	-
K	-0.050	1.000	-	-	-	-	-
Ca	-0.026	<u>0.452</u>	1.000	-	-	-	-
Mg	0.029	0.279	0.511	1.000	-	-	-
HCO ₃	0.038	0.310	<u>0.845</u>	<u>0.558</u>	1.000	-	-
SO ₄	0.2123	-0.049	-0.29	-0.010	0.051	1.000	-
Cl	0.057	<u>0.341</u>	<u>0.39</u>	0.363	0.123	0.014	1.000

Como se pode observar, de um modo geral os coeficientes de correlação são baixos. Os maiores valores encontrados encontram-se sublinhados. Não foram observadas correlações negativas com valores altos. (em módulo)

A seguir o programa efetua os cálculos dos Auto valores e fornece a porcentagem de variância devida àquele fator. Neste estágio são extraídos tantos fatores quanto o número de variáveis. Utilizou-se o método onde a diagonal principal da matriz de correlação é substituída por estimativas de comunalidades.

O resultado desta operação é dado na tabela 2.

Tabela 2:

Fator 1	Auto Valor	% da variância	% Variância Acumulada
1	2.73	39.1	39.1
2	1.23	17.6	56.7
3	0.96	13.8	70.5
4	0.79	11.3	81.8
5	0.69	9.9	91.7
6	0.49	7.0	98.7
7	0.09	1.3	100.0

Pode-se observar que apenas 3 fatores explicam cerca de 70% de toda a variância contida nos dados; quatro fatores explicariam mais que 80% da variância.

Após o cálculo dos autovalores o programa permite que se decida quantos fatores devem ser retidos para explicar a maior variância

possível. A decisão, mais do que baseada em critérios metodológicos é uma decisão pessoal. O programa possibilita a exclusão automática de todos os fatores cujos auto-valores sejam menores que 1.0. Foram feitas algumas tentativas e devido à dificuldade na interpretação dos fatores, optou-se por reter apenas 3 deles.

Uma vez dividido o nº de fatores que controlam o sistema, o programa calcula a matriz de cargas fatoriais, que é apresentada na tabela 3.

Tabela 3:

	Fator 1	Fator 2	Fator 3	Comunalidade
Na	0.01343	-0.01722	0.466507	<u>0.21677</u>
K	0.47217	0.23146	-0.10369	<u>0.28866</u>
Ca	0.92088	-0.03277	-0.08522	<u>0.85635</u>
Mg	0.61416	0.07381	0.05208	<u>0.38535</u>
HCO ₃	0.89060	-0.44270	0.03842	<u>0.99063</u>
SO ₄	-0.00092	-0.06837	0.43120	<u>0.19061</u>
Cl	0.45397	0.56164	0.12199	<u>0.57100</u>

Com o intuito de se facilitar a interpretação dos fatores o programa realiza uma rotação dos mesmos. Diversos são os processos disponíveis podendo a rotação ser ortogonal ou oblíqua. Escolheu-se a rotação ortogonal realizada pelo processo Varimax. Os dados obtidos são apresentados na tabela 4

Tabela 4:

	Fator 1	Fator 2	Fator 3
Na	-0.00098	0.02718	0.46479
K	0.31310	0.41904	-0.12264
Ca	0.83327	0.39472	-0.07879
Mg	0.50636	0.35596	0.04736
HCO ₃	0.99158	0.02911	0.08095
SO ₄	0.01165	-0.02720	0.43559
Cl	0.12080	0.74263	0.07009

A próxima etapa do programa é a plotagem das cargas fatoriais resultantes da rotação em gráficos bidimensionais.

As observações feitas através destes gráficos podem ser obtidas através do exame das cargas fatoriais após a rotação.

Como último passo o programa calcula os coeficientes dos escores fatoriais e obtém os escores fatoriais através da multiplicação da matriz de coeficientes pela matriz dos dados na sua forma standartizada.

4. DISCUSSÃO DOS RESULTADOS

A observação dos valores obtidos através do programa indica que cerca de 70% de toda a variabilidade dos dados pode ser explicada - por apenas 3 fatores, sendo que o fator 1 por si só explica quase que 70% da variação se tomarmos apenas esses 3 fatores. O restante, da variação é explicada pelos fatores 2 e 3 que contribuem com 17.5 e 12.6% respectivamente.

As correlações entre as variáveis e os fatores em ordem decrescente pode ser observada na tabela 5.

Tabela 5:

	Fator 1	Fator 2	Fator 3
HCO ₃	= 0.99158	Cl = 0.74263	Na = 0.46479
Ca	= 0.83327	K = 0.41904	SO ₄ = 0.43559
Mg	= 0.50636	Ca = 0.39472	HCO ₃ = 0.08095
K	= 0.3131	Mg = 0.35596	Cl = 0.07009
Cl	= 0.1208	HCO ₃ = 0.02911	Mg = 0.04736
SO ₄	= 0.01165	Na = 0.02718	Ca = -0.07879
Na	= -0.00098	SO ₄ = 0.02720	K = -0.12264

Observa-se que o fator 1 apresenta correlações relativamente altas com as variáveis HCO₃, Ca e Mg. Não são observadas correlações negativas altas (em módulo).

A distribuição das amostras com valores de escores fatoriais - maiores que 0.5 é apresentado na figura 1. Observa-se em termos gerais que o zoneamento obtido corresponde aos domínios geomorfológicos dos planaltos e espigões que são via de regra áreas de ocorrência da parte superior da Formação Adamantina e da Formação Marília onde os fenômenos de evapotranspiração adquirem maior importância. As concentrações salinas nestes domínios apresentam valores de resí

duo seco no intervalo de 100 a 200 mg/l.

Desta forma, o fator 1 sugere representar o teor salino, onde as variáveis HCO_3 , Ca e Mg são as de maior importância.

O fator 2 apresenta correlação pronunciada apenas com o ion Cl. Apresenta também correlações menores com o K e Ca. Não são observadas correlações negativas altas (em módulo) com nenhuma outra variável.

A distribuição das amostras com valores de escores fatoriais positivos corresponde em termos gerais às áreas geomorfologicamente mais elevadas do Grupo Bauru. O zoneamento obtido (não apresentado) corresponde aproximadamente àquele obtido para o fator 1. As baixas correlações entre as variáveis e o fator 2 não permitiram uma interpretação adequada deste fator.

Como o fator 2, o fator 3 apresenta dificuldades de interpretação devido às baixas correlações obtidas.

5. CONSIDERAÇÕES FINAIS

O método permitiu distinguir fontes de variação para o conjunto de amostras, embora não tenha sido possível analisar as mesmas com clareza. Isto se deve possivelmente à dificuldade na escolha das variáveis.

Estudos mais pormenorizados onde se inclua algum critério para a separação do grupo de amostras, poderão possibilitar uma melhor compreensão dos fatores.

6. REFERÊNCIAS BIBLIOGRÁFICAS

1. DAULTREY, S. Principal components analysis. s.n.t. (Concepts and Techniques in Modern geography, 8)
2. DAWDY, D.R. & FETH, J.H. Applications of factor analysis in study of chemistry of groundwater: quality, Mojave River Valley, California. Water Resources Research, V.3. n.2, 1967.
3. FACHER, J.M.G. Análise Fatorial. São Paulo, 1976 (Dissertação de Mestrado - IME/USP).
4. GODDARD, J. & KIRBY, A. An introduction to factor analysis s.n.t. (Concepts and Techniques in Modern Geography, 7)

5. LANDIM, P.M.B.; SOARES, P.C. & GAMA JR., E.G.
Estratigrafia do nordeste da bacia sedimentar do Paraná. Rio Claro, Instituto de Geociências e Ciências Exatas, 1980.
(Curso de Especialização - Convênio IPT-UNESP)
6. ROCHA, G.A.; BERTACHINI, A.C.; CAMPOS, H.C.N.S.; CAIXETA, J.B. Tentativa de zoneamento das características hidráulicas e hidroquímicas do aquífero Bauru. In: SBG & ABAS. ENCONTRO DE GEOLOGIA & HIDROGEOLOGIA, São Paulo, novembro de 1981.
7. SÃO PAULO, Secretaria da Indústria, Comércio, Ciência e Tecnologia Mapa geológico do Estado de São Paulo 1981. 2v.

Figura 1 : Distribuição do grupo Bauru no Estado de São Paulo



Fonte : Mapa Geológico do Estado de São Paulo / Secretaria da Indústria, Comércio, Ciência e Tecnologia, 1981.