

Uso da Análise Estatística Multivariada na Caracterização de Heterogeneidades Hidrogeoquímicas dos Sistemas Aquíferos, uma Revisão Metodológica

Use of Multivariate Statistical Analysis in the Characterization of Hydrogeochemical Heterogeneity of Aquifer Systems, a Methodological Review

Deize Elle Ribeiro Moitinho¹; Natanael da Silva Barbosa²; Maria da Conceição Rabelo Gomes³; Sizenando Bispo-Silva⁴; Cristovaldo Bispo dos Santos⁵; Ludmila Amorim Gomes⁶

¹ Instituto de Geociências, Universidade Federal da Bahia (UFBA), Salvador, BA, Brasil. Companhia de Engenharia hídrica e saneamento da Bahia (CERB), Salvador, BA, 41745-005, Brasil. E-mail: deize.moitinho@cerb.ba.gov.br

ORCID: <https://orcid.org/0000-0002-7061-6914>

² Instituto de Geociências, Universidade Federal da Bahia (UFBA), Brasil. E-mail: ndbarbosa@ufba.br

ORCID: <https://orcid.org/0000-0003-0923-6266>

³ Universidade do Estado do Pará (UEPA), Brasil. E-mail: conceicaoabelo@yahoo.com.br

ORCID: <https://orcid.org/0000-0001-7841-4201>

⁴ Petrobrás, Rio de Janeiro, Brasil. E-mail: sizenando@petrobras.com.br

ORCID: <https://orcid.org/0000-0001-6049-7774>

⁵ Instituto de Geociências, Universidade Federal da Bahia (UFBA), Salvador, BA, 40170-290, Brasil. E-mail: bispo@ufba.br

ORCID: <https://orcid.org/0000-0002-1448-3643>

⁶ Instituto de Geociências, Universidade Federal da Bahia (UFBA), Salvador, BA, 40170-290, Brasil. E-mail: ludimilla.amorim@ufba.br

ORCID: <https://orcid.org/0000-0002-1651-3738>

Resumo: Cresce a preocupação em relação aos recursos hídricos subterrâneos, impulsionando o desenvolvimento de métodos mais refinados para avaliar as mudanças nesses recursos, sejam elas decorrentes de processos naturais, intervenções diretas ou indiretas no meio, de natureza física ou química. Este estudo aborda as principais técnicas do método de Análise Estatística Multivariada (AEM) aplicadas a esse propósito. A AEM tem sido amplamente empregada na caracterização de aquíferos e em estudos relacionados à contaminação proveniente de atividades mineiras, intervenções urbanas ou rurais (agropecuárias) e na identificação de processos de intrusão marinha resultando em salinização. Além disso, essa abordagem foi utilizada para segmentar o território em áreas com diferentes qualidades das águas subterrâneas. Foram analisados 40 artigos, sendo 80% destes provenientes de países da Ásia, África e América do Norte. Durante essa avaliação, foi possível também examinar a aplicabilidade relativa e a complementaridade da Análise de Cluster Hierárquica (HCA) e da Análise de Componentes Principais (PCA) para atingir o objetivo científico, ou seja, a investigação de processos naturais e/ou antropogênicos nas águas subterrâneas. Ambos os métodos, quando comparados ao agrupamento geoquímico convencional ou a outras técnicas da estatística multivariada, destacaram-se como os mais adequados e comuns na caracterização das heterogeneidades dos aquíferos. Concluímos que a aplicabilidade dessa abordagem estatística é universal, pois as técnicas empregadas independem do tipo de litologia, uma vez que estão vinculadas às nuances hidrogeoquímicas abordadas nos estudos. Com base nos artigos publicados entre 1965 e 2022, foi possível reconhecer nas ferramentas da AEM um instrumento crucial na gestão e desenvolvimento dos recursos hídricos subterrâneos, atendendo tanto às demandas presentes quanto às futuras.

Palavras-chave: Análise Estatística Multivariada; Recursos Hídricos; Revisão Bibliográfica.

Abstract: There is a growing concern regarding underground water resources, driving the development of more refined methods to assess changes in these resources, whether they arise from natural processes or direct and indirect interventions in the environment, of a physical or chemical nature. This study addresses the main techniques of the Multivariate Statistical Analysis (MSA) method applied to this purpose. MSA has been widely employed in characterizing aquifers and in studies related to contamination from mining activities, urban or rural interventions (agricultural and livestock), and in identifying processes of marine intrusion resulting in salinization. Additionally, this approach was used to segment the territory into areas with different qualities of groundwater. Forty articles were analyzed, with 80% of these originating from countries in Asia, Africa, and North America. During this evaluation, it was also possible to examine the relative applicability and complementarity of Hierarchical Cluster Analysis (HCA) and Principal Component Analysis (PCA) to achieve the scientific objective, i.e., the investigation of natural and/or anthropogenic processes in groundwater. Both methods, when compared to conventional geochemical grouping or other multivariate statistical techniques, stood out as the most suitable and common in characterizing aquifer heterogeneities. We conclude that the applicability of this statistical approach is universal, as the employed techniques are independent of lithology type, as they are linked to the hydrogeochemical nuances addressed in the studies. Based on articles published between 1965 and 2022, it was possible to recognize in MSA tools a crucial instrument in the management and development of underground water resources, addressing both current and future demands.

Keywords: Multivariate Statistical Analysis; Water Resources; methodological Review.

Recebido: 13/03/2024; Aceito: 21/02/2025; Publicado: 28/05/2025.

1. Introdução

Os recursos hídricos subterrâneos figuram como um dos principais elementos naturais explorados pela humanidade. A gestão sustentável desses recursos emerge como uma questão crucial para todos os países, especialmente os em desenvolvimento (Zghibi et al., 2014; Rakotondrabe et al., 2018). Globalmente, 97,5% da água é salgada, distribuída nos oceanos, restando apenas 2,5%. Desse montante, 68,9% encontra-se congelado, restando 31% de água doce, sendo 4% provenientes de fontes superficiais, com a maior parcela composta por águas subterrâneas (Revenga et al., 2000; MMA-SRHU, 2007). Os múltiplos usos dessas águas atendem a diversas finalidades: econômicas (indústria, agropecuária, mineração), geração de energia e subsistência de organismos vivos.

A notoriedade atribuída aos aquíferos deriva, em parte, do menor custo em comparação com as águas superficiais, que, sujeitas à poluição e esgotamento, contrastam com os aspectos de perenidade, pureza e proteção conferidos aos reservatórios subterrâneos. Essas características incitam a necessidade de uma política de gestão eficaz para o uso sustentável dos recursos hídricos subterrâneos (Wanda et al., 2011; Narany et al., 2014; Ahmed et al., 2017). Assim, os países requerem dados representativos para monitorar e antecipar as alterações nesses recursos.

Numerosos estudos hidrogeoquímicos recentes corroboram que a salinidade figura entre as principais causas da degradação da qualidade das águas em aquíferos costeiros submetidos a climas áridos (Zghibi et al., 2014; Slama et al., 2017; Ahmed et al., 2018; Sae-Ju et al., 2019). As variações na qualidade das águas subterrâneas frequentemente resultam da configuração regional e geológica do contexto subterrâneo (Melloul & Collin 1992; Shin et al., 2020).

Infelizmente, nas últimas décadas, observou-se uma considerável redução das capacidades nacionais para garantir o monitoramento hidrogeológico e a produção de dados de qualidade para embasar decisões (Keita & Zhonghua et al., 2017). O contexto atual demanda atenção cuidadosa e uma expansão do entendimento sobre a heterogeneidade dos sistemas aquíferos, dependentes de fatores litológico-estruturais (fissural, cárstico ou poroso) inerentes à gênese geológica e a fatores ambientais como clima, área de recarga, evaporação e interação rocha-água, com processos atuais ou históricos. Ademais, compreender o uso do solo, identificar contaminações por processos antropogênicos e conhecer características geoquímicas passadas ou futuras são de grande relevância (Narany et al., 2014; Ravikumar et al., 2015; Kumar Das et al., 2022).

Recentemente, hidrólogos têm voltado sua atenção cada vez mais para técnicas matemáticas e estatísticas, visando reduzir um grande número de variáveis para um mínimo de dimensões não correlacionadas (Seyan et al., 1985; Farnham et al., 2000). Logo, diante de uma infinidade de fatores associados, embora seja humanamente impossível analisá-los, existe uma valiosa ferramenta matemática com a qual é possível trabalhar simultaneamente com múltiplos elementos: a Análise Estatística Multivariada (AEM). Essa ferramenta é crucial na manipulação e interpretação de dados com muitas variáveis, auxiliando na identificação dos elementos mais relevantes e, portanto, merecedores de maior destaque e atenção.

O pioneiro na aplicação da AEM na hidrologia foi Wallis em 1965, em um trabalho comparativo entre técnicas comuns da época, como a Análise de Regressão, contrapostas à aplicação da AEM, empregando a Análise do Componente Principal (PCA) e análise de Cluster com rotação Varimax. Dois anos depois, Dawdy & Feth (1967) realizaram a primeira aplicação em dados hidrogeoquímicos, adotando o método em um estudo sobre a qualidade da água subterrânea em Mojave, Califórnia.

Desde então, essa ferramenta tem apresentado uma crescente utilização nos mais diversos campos. Sendo a pesquisa estado da arte uma eficiente ferramenta, por fornecer um prelúdio para entender o que tem sido estudado em determinado campo (Roma Romanowski & Ens, 2006), este trabalho corresponde a uma revisão baseada em uma pesquisa em escala global, com uma escala temporal abrangendo as últimas cinco décadas. O objetivo é alcançar os principais referenciais teóricos citados na literatura sobre a análise estatística multivariada, com o intuito de compreender como a AEM é aplicada e quais são as principais ferramentas para estudar as heterogeneidades geoquímicas (ou seja, evolução natural dos aquíferos, processos de salinização, contaminação antrópica, contaminação por mineralizações seja por atividades mineiras ou background natural) dos sistemas aquíferos.

2. Metodologia

2.1 Metodologia da Análise dos Componentes Principais (PCA)

A metodologia da Análise dos Componentes Principais (PCA), é um caso específico do método mais abrangente de análise fatorial. A PCA concentra-se na construção de novas variáveis, denominadas componentes principais, a partir de

um conjunto de variáveis originais existentes (Shin et al., 2020; Odat et al., 2015; Dawdy & Feth, 1967). Esses novos elementos são derivados por meio de uma combinação linear das variáveis originais.

A aplicação do método PCA visa a redução de dimensões de múltiplos parâmetros, categorizando aqueles mais relevantes que explicam a variância do fenômeno em estudo. Ao serem interpretados, esses componentes revelam estruturas de dados subjacentes. A abordagem básica consiste em agrupar as dimensões com base na correlação, ou seja, reunir variáveis altamente correlacionadas, enquanto aquelas com baixa correlação são separadas em classes distintas (Güler et al., 2002; PingHeng et al., 2010; Sae-Ju et al., 2019). O objetivo primordial é descrever os dados brutos, utilizando a soma de um fator especial, numa função linear do menor número possível de fatores comuns, simplificando assim as dimensões.

Com os componentes principais em mãos, realizam-se análises nos modos Q (quando se investigam similaridades entre indivíduos ou objetos), considerando simultaneamente todas as variáveis medidas em cada um deles, e no modo R (explorando semelhanças entre variáveis). A técnica fundamenta-se na análise dos autovalores de uma matriz de correlação ou covariância, sendo os resultados expressos em tabelas/matriz (Tabela 1) (Melloul & Collin, 1992; Güler et al., 2012; Shin et al., 2020).

Tabela 1 – Tabela com os Componentes Principais.

Variável	Componente Principal			
	CP1	CP2	CP3	CP4
pH	-0,182	0,036	0,952	0,045
EC	0,972	0,099	-0,144	-0,027
TDS	0,972	0,099	-0,144	-0,027
TH	0,991	0,045	-0,026	-0,027
CaH	0,967	0,076	0,080	-0,066
TA	0,994	0,004	-0,044	-0,042
Ca	0,967	0,076	0,256	-0,197
Mg	0,970	0,015	-0,123	0,010
Na	0,132	0,762	0,256	-0,197
K	0,937	0,044	-0,141	-0,007
HCO ₃	0,994	0,004	-0,044	-0,042
F	-0,035	0,659	-0,114	0,210
Cl	0,981	-0,007	-0,119	-0,066
SO ₄	0,962	0,118	-0,068	0,003
PO ₄	-0,055	-0,021	0,044	0,967
NO ₃	0,076	0,748	-0,023	-0,066
Autovalor	10,49	1,63	1,10	1,04
Proporção%	65,55	10,17	6,88	6,52
Acumulado %	65,55	75,71	82,59	89,11

Fonte: Ravikumar & Somashekar (2015).

No estudo conduzido por Ravikumar & Somashekar (2017) sobre a qualidade de um aquífero na bacia do rio Varahi, Karnataka, Índia, empregou-se a técnica PCA. A escolha da PCA para a avaliação baseou-se na necessidade de lidar com elementos variando em diferentes ordens de magnitude. A aplicação da PCA concentrou-se na matriz de correlação, onde cada variável foi normalizada para variância unitária, garantindo contribuição equitativa. Os resultados foram representados graficamente com base na tabela de componentes principais (Tabela 1) e representados na Figura 1.

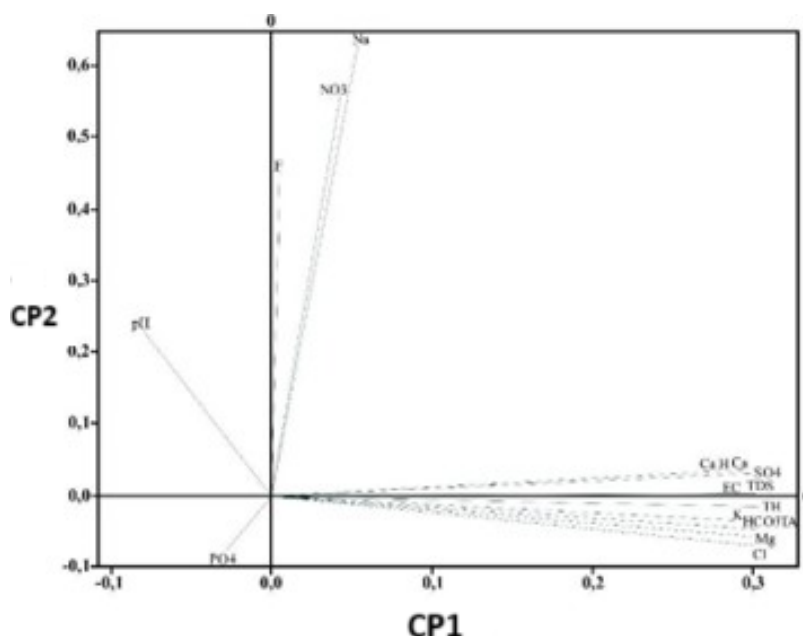


Figura 1 – Representação gráfica dos componentes principais elencados na tabela 1.

Fonte: Modificado de Ravikumar & Somashekar (2015).

As informações resumidas do gráfico PCA são cruciais para identificar relacionamentos, padrões, similaridades, composição química extrema e possibilitam a posterior modelagem estatística. Analisando o gráfico de carregamentos (Fig. 2) dos dois primeiros PCs (PC1 e PC2), observa-se a distribuição de todos os parâmetros físico-químicos nos quadrantes superior direito e inferior direito (canto superior e inferior direito, respectivamente). As linhas que conectam as variáveis e passam pela origem no gráfico das cargas fatoriais indicam a contribuição relativa de cada variável. Essa representação visual é crucial para uma compreensão mais aprofundada dos dados obtidos na avaliação da qualidade do aquífero. Essa seção sumariza a metodologia utilizada, elucidando o processo de aplicação da PCA na análise da qualidade do aquífero, preparando para a compreensão subsequente dos resultados obtidos, com esse método.

2.2 Metodologia da Análise Fatorial (AF)

Na análise fatorial, os fatores são rotacionados, utilizando a rotação varimax de Kaiser para obter interpretações claras. Esse método, proposto por Kaiser (1958), destaca cargas fatoriais mais significativas, fundamentais para a interpretação. A preparação dos dados inclui a padronização dos dados brutos, eliminando diferenças de unidade e valores. Em seguida, calcula-se a matriz de covariância (Farnham et al., 2003; Che et al., 2021; Tessema et al., 2012). Os autovalores e a proporção das variâncias associadas a cada fator são computados.

A determinação do número de fatores a serem extraídos é realizada somando as proporções e calculando os autovalores cumulativos. Diferentes métodos, como Scree plot e regra de Kaiser, são aplicados para orientar essa decisão, com a maioria dos estudos utilizando uma ou ambas as ferramentas.

2.2.1 Scree Plot

O *Scree Plot* é uma ferramenta eficaz para identificar fatores relevantes, destacando uma mudança abrupta nos valores próprios que resulta em uma alteração na inclinação da curva, passando de íngreme para uma angulação mais suave. Na Figura 3, observa-se esse padrão, onde a inclinação do gráfico *Scree Plot* muda notavelmente após os dois primeiros fatores. Os autovalores também diminuem, caindo abaixo de 1 ao passar do fator 4 para o 5. Essa observação sugere que uma solução com quatro componentes pode ser a escolha mais apropriada, abrangendo o melhor intervalo de variação (Ravikumar & Somashekar, 2015; Yang et al., 2015).

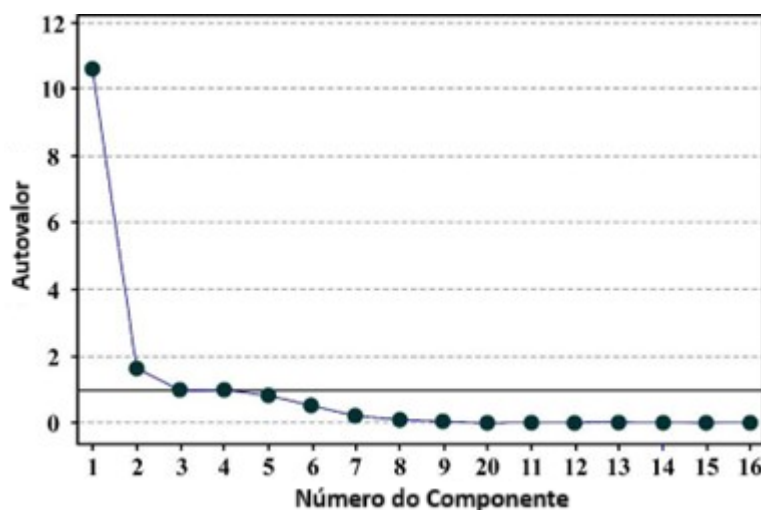


Figura 2 – Scree Plot gráfico de autovalores, usado para definir quantos componentes, sendo a linha contínua o marcador dos componentes mais representativos.

Fonte: Modificado de Ravikumar & Somashekar (2015).

2.2.2 Regra De Kaiser

Com base no método de Análise Fatorial, a escolha do número de fatores está vinculada aos elementos que explicam a maior parte da variância, geralmente em torno de 70% do conjunto de dados. A rotação varimax, frequentemente guiada pelos critérios de Kaiser (1958) usando autovalores superiores a 1, é aplicada para tornar os resultados interpretáveis. Os carregamentos fatoriais são categorizados como "fortes" ($>0,75$), "moderados" ($0,75-0,50$) e "fracos" ($0,50-0,30$) (Liu et al., 2003). A classificação das cargas fatoriais é realizada com base na matriz de correlação entre variáveis padronizadas, buscando simplificar a interpretação através da maximização da relação entre elas, utilizando a rotação varimax de Kaiser (1958). Nesse esquema, o primeiro componente (F1) geralmente reflete as variações mais significativas, enquanto o último contribui menos para a explicação dos dados (Usunoff e Guzman 1989; Helena 2000).

2.3 Metodologia - Análise de Cluster

A análise de cluster é um método estatístico de agrupamento, dividido em hierárquico e não hierárquico. Algoritmos hierárquicos geram dendrogramas, representando a similaridade entre amostras, enquanto algoritmos não hierárquicos formam grupos diretamente. O método hierárquico, amplamente usado, agrupa amostras com base em similaridades, ilustrando a semelhança geral das variáveis no conjunto de dados, muitas vezes utilizando a distância euclidiana (Davis, 1986). Entre as diversas técnicas, a ligação média e o método de Ward são destacados (Seyhan e Keet, 1981).

A Análise de Agrupamento Hierárquico (HCA) destaca-se como o principal método para identificar conjuntos homogêneos com base em características medidas. Inicia-se com cada caso como um cluster separado, unindo-os sequencialmente até restar apenas um grupo, fundamentado na teoria de que parâmetros semelhantes provêm da mesma fonte (Bhuiyan et al., 2016). Essa técnica, amplamente aplicada nas ciências da Terra (Davis, 1986), utiliza níveis de semelhança para construir um dendrograma (Güler et al., 2002; Cloutier et al., 2008; Güler et al., 2012; Moya et al., 2015).

O agrupamento hierárquico calcula a similaridade entre pares de objetos e forma clusters com base nessa similaridade. Mede a distância entre pontos no espaço multidimensional, onde uma menor distância indica maior semelhança e vice-versa (Tessema et al., 2012). Na Figura 3, as menores distâncias evidenciam semelhanças entre amostras, agrupadas com base em suas origens, como Cluster A (baixa poluição) subdividido em A-1 (águas subterrâneas urbanas) e A-2 (águas superficiais e nascente sujeitas à mineralização natural). O Cluster B representa águas a jusante de locais de mineração de ouro.

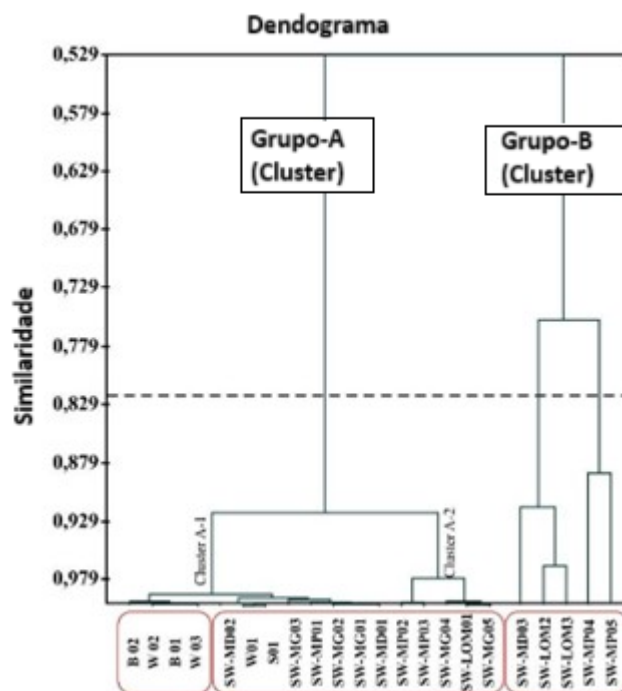


Figura 3 – Dendrograma desenvolvido na identificação de diferentes sítios de contaminação.

Fonte: Modificado de Rakotondrabe *et.al.* (2018).

2.4 Matriz De Correlação (MC)

A matriz de correlação é uma ferramenta essencial para identificar associações entre variáveis, especialmente em hidroquímica, onde parâmetros químicos individuais podem desempenhar diversos papéis (Orakwe e Chukwuma, 2015; Rakotondrabe *et al.*, 2018; Helena *et al.*, 2000).

Na análise de correlação, avaliamos as relações entre pares de variáveis usando o coeficiente de correlação de Pearson. Esse coeficiente, representando o nível teórico de relacionamento, reflete a significância estatística da relação entre duas variáveis. Os resultados são apresentados em uma matriz (Tabela 2), variando de -1 a +1, com valores superiores a |0,5| considerados relevantes (Schot e Vander Wal, 1992; Odat, 2015; Armed *et al.*, 2017; Zhang *et al.*, 2018).

Tabela 2 – Matriz de Correlação.

	<i>Ec</i>	<i>Tc</i>	<i>pH</i>	<i>TDS</i>	<i>Na</i>	<i>K</i>	<i>Ca</i>	<i>Mg</i>	<i>Cl</i>
<i>Ec</i>	1,00								
<i>Tc</i>	0,09	1,00							
<i>pH</i>	-0,25	-0,27	1,00						
<i>TDS</i>	0,99	0,09	-0,25	1,00					
<i>Na</i>	0,94	0,09	0,17	0,95	1,00				
<i>K</i>	0,70	0,05	-0,20	0,71	0,72	1,00			
<i>Ca</i>	0,96	0,06	0,29	0,96	0,86	0,69	1,00		
<i>Mg</i>	0,95	0,07	-0,24	0,94	0,82	0,66	0,92	1,00	
<i>Cl</i>	0,99	0,06	-0,23	0,99	0,94	0,73	0,95	0,94	1,00

Fonte: Modificado de Odat (2015).

3. Resultados e discussão

Ao lidar com uma variedade de parâmetros em amostras, a estatística multivariada emerge como uma ferramenta essencial para uma análise eficiente desses dados. A coleta de múltiplas medições proporciona uma compreensão abrangente, permitindo a consideração simultânea da variabilidade de diversas propriedades (Landim, 2011).

Dentre os métodos amplamente utilizados destacam-se: Análise dos Componentes Principais, Análise de Cluster, Análise Fatorial (incluindo Rotação e Scree Plot), Matriz de Correlação.

Foram examinados 40 artigos quanto às técnicas utilizadas e seus objetivos. A distribuição destes foi realizada com base nos temas para os quais a Análise Estatística Multivariada (AEM) foi aplicada e nas metodologias mais frequentes. A figura 4 ilustra quais técnicas foram empregadas nas pesquisas referenciadas, revelando o uso simultâneo de algumas delas. Destaca-se a predominância da ferramenta Análise dos Componentes Principais (PCA), com o tema mais explorado sendo a caracterização/evolução hidrogeoquímica. Os resultados para cada tema abordado pela AEM estão apresentados, destacando as técnicas empregadas, muitas vezes utilizadas concomitantemente.

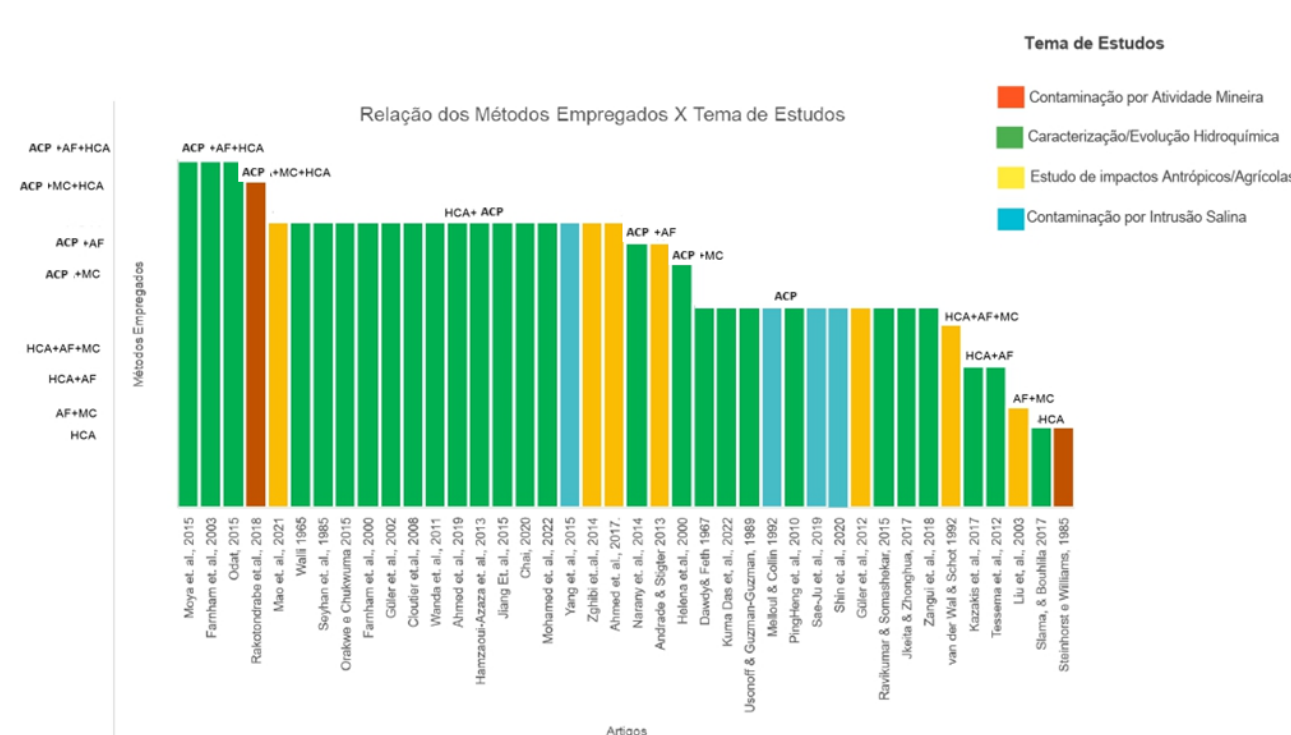


Figura 4 – Gráfico distribuição das técnicas empregadas.

Fonte: Autores (2024).

3.1 Caracterização e/ou Evolução Hidrogeoquímica

No estudo realizado no condado de Kangding, sudoeste da China, por Zhang et al. (2018), empregou-se a Análise dos Componentes Principais (PCA) juntamente com o gráfico Scree plot, identificando assim as três principais componentes, conforme representado na figura 5. Ao integrar essas informações com outras ferramentas, foi possível inferir que a água termal segue uma circulação profunda pela zona de falha de Xianshuihe, enquanto as águas subterrâneas fluem por fraturas, recarregando-se com água termal. Essas análises resultaram na construção de um modelo conceitual hidrológico, proporcionando uma compreensão mais aprofundada do sistema hídrico natural.

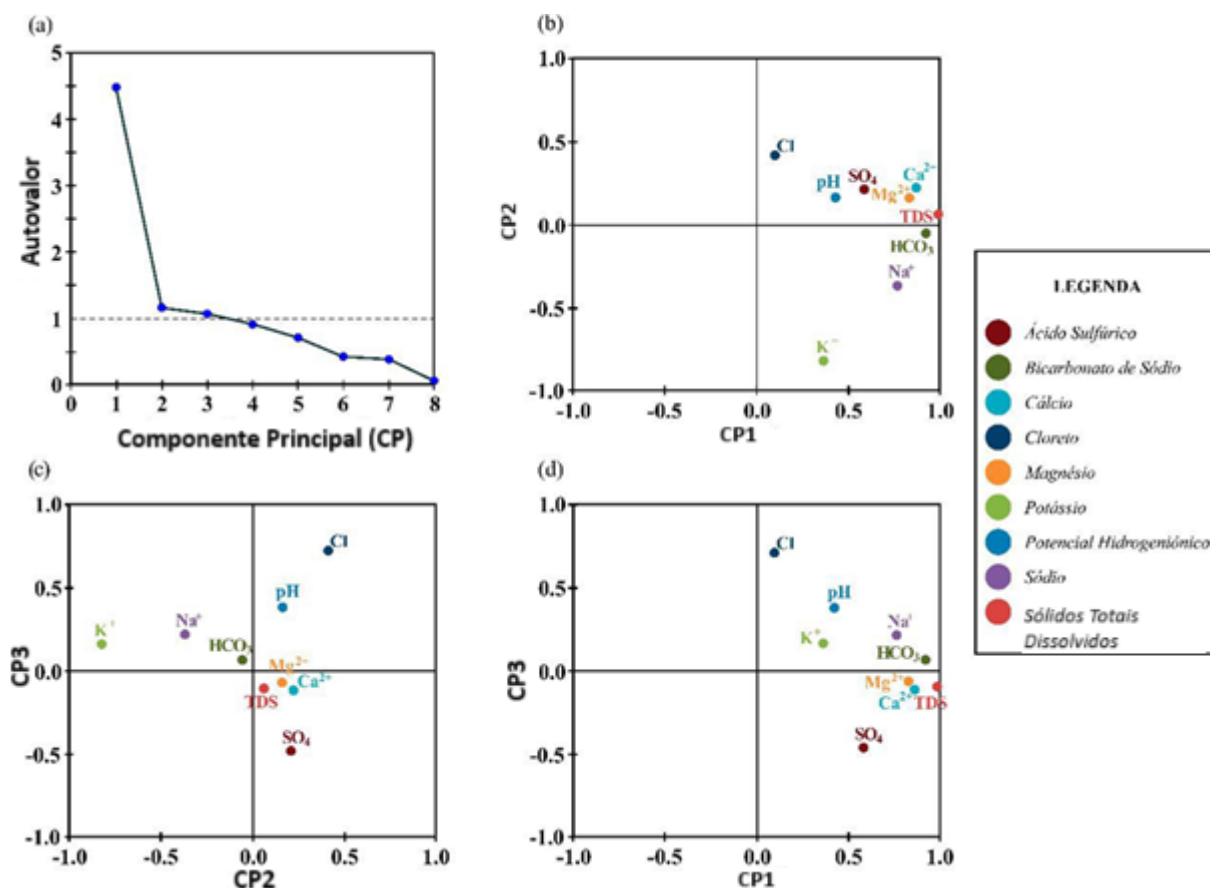


Figura 5 – Gráfico de Scree plot (gráfico de escoamento) e Diagramas PCA.

Fonte: Modificado de Zhang et al (2018).

3.2 Investigação de Impactos Antrópicos Urbanos/Agrícolas nas Águas Subterrâneas

No estudo da região de Sylhet, em Bangladesh, Ahmed et al. (2017) investigaram as características hidrogeoquímicas das águas subterrâneas, analisando possíveis influências das águas superficiais. Utilizando a Matriz de Correlação de Pearson (MC), PCA e HCA, avaliaram os fatores de controle atuantes. Correlações significativas foram observadas, como $K+SO_4^-$ (0,838) e $K+NO_3$ (0,543), indicando a influência de práticas agrícolas excessivas na área de estudo, associadas a alta precipitação ao longo da descarga.

A PCA revelou 6 componentes principais, abrangendo 76,16% da variância total. PC1, PC3, PC4 e PC6 foram indicativos de Fatores Geogênicos/Naturais, enquanto PC2 e PC5 destacaram Fatores Antropogênicos, relacionados ao uso de pesticidas e fertilizantes. A análise hierárquica de cluster (Figura 6) confirmou esses resultados, identificando 4 clusters distintos. Os clusters 1 e 3 refletem intemperismo e fatores naturais, o cluster 2 indica influências naturais e antrópicas, e o cluster 4 está associado exclusivamente a fatores antropogênicos.

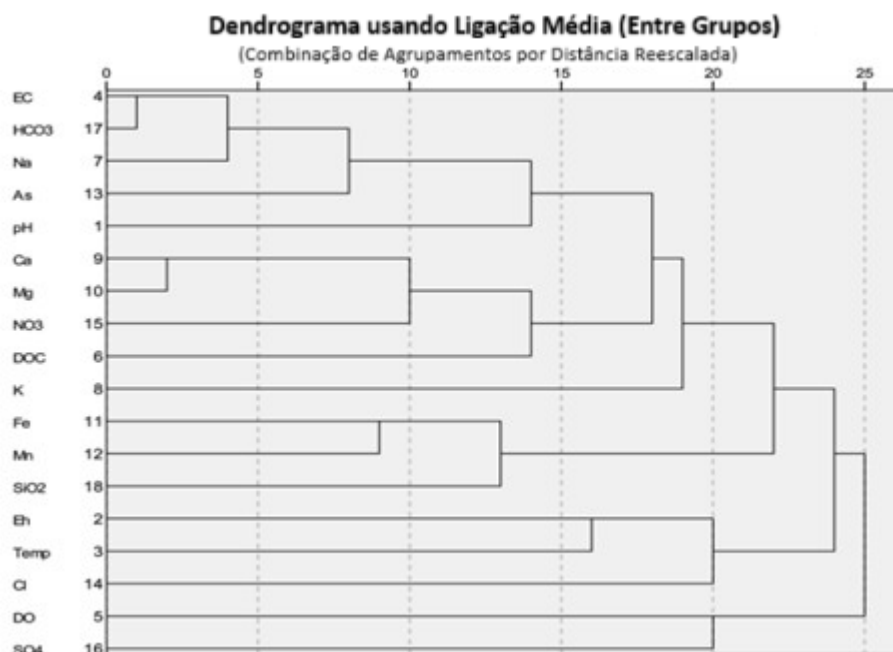


Figura 6 – Diagramas HCA.
Fonte: Ahmed et al. (2017).

3.3 Análise de Contaminação por Atividade Mineira

Na análise dirigida por Rakotondrab et al. (2018) analisou a influência das atividades de mineração de ouro na qualidade da água na bacia hidrográfica de Mari em Bétaré-Oya, situada no leste dos Camarões. Amostras de água subterrânea e superficial foram coletadas, e além dos métodos convencionais de análise e do Índice de Poluição por Metais Pesados (HPI), todos os dados foram submetidos a Análises Estatísticas Multivariadas, utilizando as técnicas de PCA, HCA e MC.

A matriz de correlação neste estudo permitiu identificar associações entre variáveis, avaliar a coerência do conjunto de dados e indicar a participação dos parâmetros químicos individuais em vários fatores de influência. Da mesma forma, a PCA foi empregada para distinguir e verificar (Figura 7 a-b), resultando na discriminação de cinco grupos: dois para elementos maiores e parâmetros físicos (Figura 7-a) e três para metais pesados (Figura 7-b).

No primeiro grupo (círculo vermelho), composto por turbidez e sólidos totais (TSS), sugere-se que a poluição física possa ser atribuída à erosão da mineração e aos resíduos. O segundo grupo (círculo verde), composto por CE, pH, alcalinidade e íons principais (K^+ , Ca^{2+} , Na^+ , Cl^- , HCO_3^- , NO_3^-), representa a mineralização da água, podendo ter origens naturais ou antropogênicas, como no caso dos íons nitrato.

Quanto aos grupos relacionados aos metais pesados, o primeiro grupo (círculo verde), composto por Fe e Cr, apresentou correlação significativa (0,728), sugerindo origens litogênicas para esses metais pesados. O segundo grupo (círculo vermelho), composto por Mn/Cd e Pb, pode ter uma combinação de fontes geogênicas e antropogênicas. O último grupo (Zn/Cu), no círculo roxo, foi atribuído a fontes antropogênicas, como mineração e atividades associadas, representando potenciais fontes de poluição na área de estudo.

A abordagem estatística multivariada permitiu a identificação de zonas altamente poluídas dentro da área de estudo, relacionadas tanto a poluentes físicos quanto químicos.

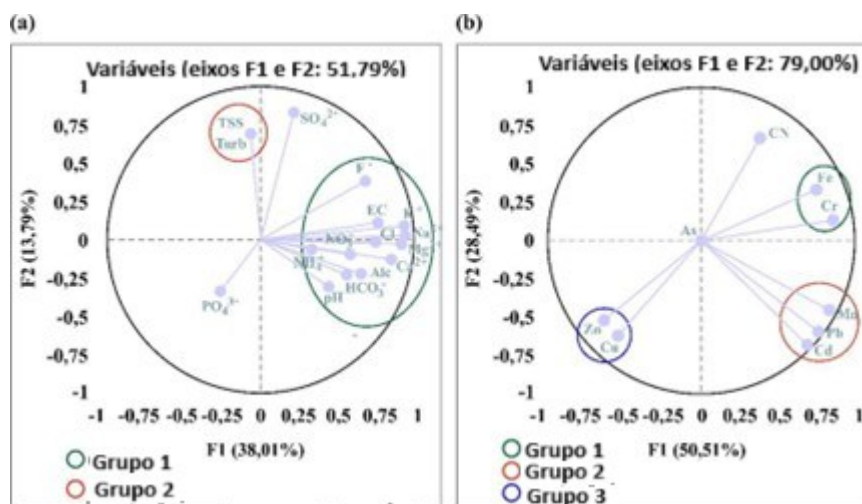


Figura 7 – Análise dos Componentes Principais dos elementos maiores e parâmetros físicos (a) e metais pesados (b).
Fonte: Modificado de Rakotondrab *et. al.* (2018).

3.4 Análises de Contaminação de Aquíferos por Intrusão Salina

Na pesquisa realizada por Shin *et al.* (2020) no arquipélago ao sudoeste do Mar da Coreia do Sul, adjacente à Península de Muanna, com o objetivo de compreender o impacto da água do mar nas águas subterrâneas, utilizou-se a ferramenta Análise de Componentes Principais (PCA). Foram coletadas 74 amostras de águas subterrâneas, classificadas por tipo de água e razão molar $\text{Cl}^- - \text{HCO}_3^-$. Do total, 36 amostras do tipo $\text{Ca}^{2+} - \text{Cl}^-$ e 32 do tipo $\text{Na}^+ - \text{Cl}^-$ (representando 91,9%), foram identificadas como influenciadas pela água do mar. Ao classificar as amostras com base na razão molar $\text{Cl}^- / \text{HCO}_3^-$, constatou-se que 40 das 74 amostras apresentavam uma razão molar de 2,8 ou superior, indicando uma forte influência da água marinha nas águas subterrâneas.

Através da Análise de Componentes Principais, observou-se que a influência da água oceânica ocupou a primeira componente, representando 54,1%. Esses resultados evidenciam que as amostras com uma significativa proporção de mistura de água do mar provêm de regiões onde a influência marinha é mais pronunciada. Tal constatação confirma que as águas subterrâneas na área de estudo foram impactadas pela intrusão salina.

3.5 Estudos Complexos - Simultâneas Heterogeneidades Hidrogeoquímicas

No estudo conduzido por Kazakis (2017), foram aplicadas técnicas de estatística multivariável a dados hidroquímicos de três regiões distintas do norte da Grécia, caracterizadas por diferentes condições hidrogeológicas, climatológicas e usos do solo (Figura 8). O trabalho utilizou métodos como o coeficiente de correlação de Pearson, a Análise de Fator (AF) no modo R, e a Análise de Agrupamento Cluster (HCA), visando avaliar a qualidade das águas subterrâneas e os processos hidroquímicos nessas regiões.

De maneira inovadora, os dados das três regiões foram analisados de forma conjunta, sem presumir a similaridade dos processos hidroquímicos. A análise fatorial, após a rotação Varimax, revelou dois fatores que explicaram 70,6% da variância total.

Os resultados indicaram que, na primeira região, a contaminação estava relacionada a atividades industriais e agrícolas; na segunda região, a intrusão de água do mar e fluidos geotérmicos eram os principais fatores contaminantes; enquanto na terceira região, a contaminação era atribuída a impactos antrópicos, notadamente o intenso uso de agrotóxicos.

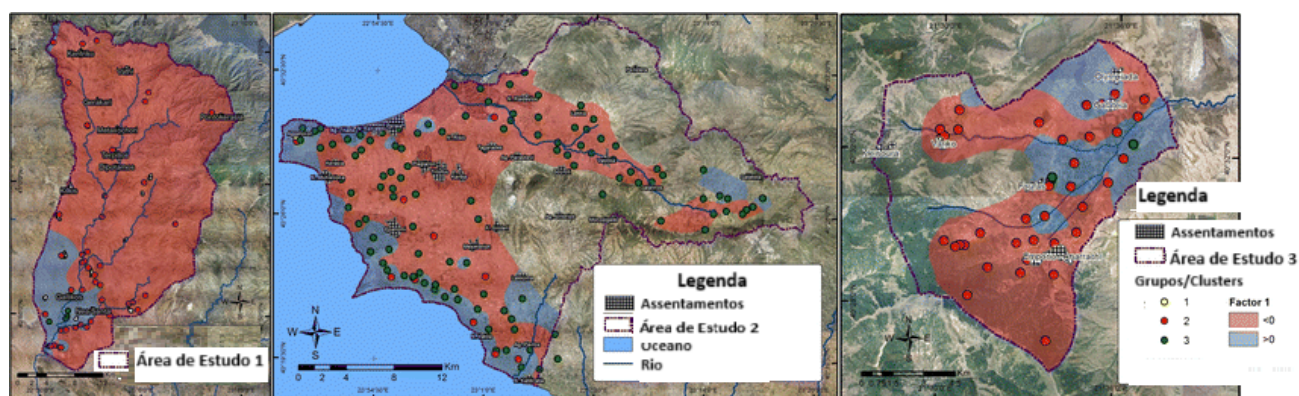


Figura 8 – Resultado da aplicação da Análise de Cluster em 3 regiões da Grécia, com diferentes usos e ocupação de solo e sobre diferentes aquíferos.

Fonte: Mapa modificado de Kazakis (2017).

4. Considerações finais

A análise estatística multivariada (AEM) emerge como uma ferramenta essencial para lidar com a vasta quantidade de dados, proporcionando integração, interpretação e representação eficientes dos resultados. Esta metodologia possibilita uma abordagem quantitativa que viabiliza a classificação de amostras aquíferas, a análise de correlações entre variáveis (parâmetros físico-químicos) e a avaliação das heterogeneidades nos sítios de amostragem dessas águas, entre outras aplicações relevantes.

O sucesso da AEM em estudos hidrogeoquímicos destaca-se ao identificar fatores da água, compreender variações temporais e espaciais em suas características, e delimitar sítios com base em sua origem ou nos processos aos quais estão submetidos. Essas informações são fundamentais para uma gestão eficaz da qualidade da água.

É crucial ressaltar que a utilização adequada da AEM exige um sólido conhecimento das técnicas e de suas limitações. Este método não estabelece relações de causa e efeito, mas oferece insights valiosos sobre as relações que podem ser inferidas. Com base nas pesquisas analisadas, concluímos que a abordagem estatística multivariada é universalmente aplicável, independente do tipo de litologia. As principais ferramentas empregadas são a Análise dos Componentes Principais (PCA) e a Análise de Agrupamento Hierárquico (HCA), muitas vezes aplicadas em conjunto.

Referências

- AHMED, N.; BODRUD-DOZA, MD.; ISLAM, S. M. D.; CHOUDHRY, M. A.; MUHIB, MD. I.; ZAHID, A.; HOSSAIN, S.; MONIRUZZAMAN, MD.; DEB, N.; BHUIYAN, MD. A. Q. 2019. *Hydrogeochemical evaluation and statistical analysis of groundwater of Sylhet, north-eastern Bangladesh*. Acta Geochim 38, 440-455. <https://doi.org/10.1007/s11631-018-0303-6>
- AHMED, A. H., RAYALEH, W. E., ZGHIBI, A., & OUDDANE, B. 2017. *Assessment of chemical quality of groundwater in coastal volcano-sedimentary aquifer of Djibouti, Horn of Africa*. Journal of African Earth Sciences, 131, 284–300. doi:10.1016/j.jafrearsci.2017.04.010
- ANDRADE, A. I. A. S. S. & STIGTER, T. Y.. 2013. *The distribution of arsenic in shallow alluvial groundwater under agricultural land in central Portugal: insights from multivariate geostatistical modeling*. Sci Total Environ. DOI: 10.1016/j.scitotenv.2013.01.033
- BHUIYAN M.A.H, BODRUD-DOZA M, ISLAM A.T, RAKIB M.A, RAHMAN M.S, RAMANATHAN A.L.. 2016. *Assessment of groundwater quality of Lakshimpur district of Bangladesh using water quality indices, geostatistical methods, and multivariate analysis*. Environ Earth Sci 75:1020. <https://doi.org/10.1007/s12665-016-5823-y>
- CHE, Q.; SU, X.; WANG, S.; ZHENG, S.; LI, Y. 2021. *Hydrochemical Characteristics and Evolution of Groundwater in the Alluvial Plain (Anqing Section) of the Lower Yangtze River Basin: Multivariate Statistical and Inversion Model Analyses*. Water, 13, 2403. <https://doi.org/10.3390/w13172403>

- CHAI, Y.; XIAO, C.; LI, M., LIANG, X. 2020. *Hydrogeochemical Characteristics and Groundwater Quality Evaluation Based on Multivariate Statistical Analysis*. Revista Water, 12(10):2792, DOI:10.3390/w12102792
- CLOUTIER, V.; LEFEBVRE, R.; THERRIEN, R.; SAVARD, M. M. 2008. *Multivariate statistical analysis of geochemical data as indicative of the hydrogeochemical evolution of groundwater in a sedimentary rock aquifer system*. Journal of Hydrology 353, 294– 313.
- DAVIS, J.C., 1986. *Statistics and Data Analysis in Geology*. John Wiley & Sons Inc., New York.
- DAWDY, D.R. AND FETH, J.H., 1967. *Application of factor analysis in the study of chemistry of groundwater quality, Mojave River Valley, California*. Water Resour. Res., 3(2): 505-510.
- FARNHAM, I.M., JOHANNESSON, K.H., SINGH, A.K., HODGE, V.F., STETZENBACH, K.J., 2003. *Factor analytical approaches for evaluating groundwater trace element chemistry data*. Analytical Chimica Acta 490, 123–138.
- FARNHAM, I.M., STETZENBACH, K.J., SINGH, A.K., JOHANNESSON, K.H., 2000. *Deciphering groundwater flow systems in Oasis Valley, Nevada, using trace element chemistry, multivariate statistics, and geographical information system*. Mathematical Geology 32, 943–968.
- GÜLER, C., THYNE, G.D., MCCRAY, J.E., TURNER, K.A., 2002. *Evaluation of graphical and multivariate statistical methods for classification of water chemistry data*. Hydrogeol. J.10:455–474. <http://dx.doi.org/10.1007/s10040-002-0196-6>.
- GÜLER, C., KURT, M.A., ALPASLAN, M., AKBULUT, C., 2012. *Assessment of the impact of anthropogenic activities on the groundwater hydrology and chemistry in Tarsus coastal plain (Mersin, SE Turkey) using fuzzy clustering, multivariate statistics and GIS techniques*. J. Hydrol. 414, 435–451.
- HELENA, B., PARDO, B., VEGA, M., BARRADO, E., FERNANDEZ, J.M., FERNANDEZ, L., 2000. *Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis*. Water Res. 34 (3), 807–816.
- JIANG, Y.; GUOA, H.; JIAA, Y.; CAO, Y.; HUB, C. 2020. *Principal component analysis and hierarchical cluster analyses of arsenic groundwater geochemistry in the Hetao basin, Inner Mongolia*. Chemie der Erde 75 (2015) 197–205.
- KAISER H., F. 1958. *The varimax criteria for analytical rotation in factor analysis*. Psychometrika, 23(3):187 –200.
- KAZAKIS, N., MATTAS, C., PAVLOU, A., PATRIKAKI, O., VOUDOURIS, K. 2017. *Multivariate statistical analysis for the assessment of groundwater quality under different hydrogeological regimes*. Environ Earth Sci, 76:349 DOI 10.1007/s12665-017-6665-y.
- KEITA, S., & ZHONGHUA, T. 2017. *The assessment of processes controlling the spatial distribution of hydrogeochemical groundwater types in Mali using multivariate statistics*. Journal of African Earth Sciences, 134, 573–589. doi:10.1016/j.jafrearsci.2017.07.023
- KUMAR DAS, T.; AHMED, S.; HOSSEN, A.; RAHAMAN, M.D.H.; RAHMAN, M., M. 2022. *Multivariate statistics and hydrogeochemistry of deep groundwater at southwestern part of Bangladesh*, Heliyon, Volume 8, Issue 10, e11206, ISSN 2405-8440, <https://doi.org/10.1016/j.heliyon.2022.e11206>.
- LIU CW, LIN KH, KUO YM. 2003. *Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan*. Sci. Total Environ. 2003;313(1):77–89. [http://dx.doi.org/10.1016/S0048-9697\(02\)00683-6](http://dx.doi.org/10.1016/S0048-9697(02)00683-6).
- MAO, H.; WANG, G.; RAO, Z.; LIAO, F.; SHI, Z.; HUANG, X.; CHEN, X.; YANG, Y. 2021. *Deciphering spatial pattern of groundwater chemistry and nitrogen pollution in Poyang Lake Basin (eastern China) using self-organizing map and multivariate statistics*, Journal of Cleaner Production, Volume 329, 129697, ISSN 0959-6526, <https://doi.org/10.1016/j.jclepro.2021.129697>.

- MELLOUL, A., COLLIN, M., 1992. *The 'principal components' statistical method as a complementary approach to geochemical methods in water quality factor identification; application to the Coastal Plain aquifer of Israel*. Journal of Hydrology 140, 49–73.
- MOHAMED,N. A.; WACHEMO, A. C.; KARUPPANNAN, S. & DURAISAMY,K. 2022. *Spatio-temporal variation of groundwater hydrochemistry and suitability for drinking and irrigation in Arba Minch Town, Ethiopia: An integrated approach using water quality index, multivariate statistics, and GIS*, Urban Climate, Volume 46,2022,101338, ISSN 2212-0955, <https://doi.org/10.1016/j.uclim.2022.101338>.
- MOYA, C.E., RAIBER, M., TAULIS, M., COX, M.E., 2015. *Hydrochemical evolution and ground-water flow processes in the Galilee and Eromanga basins, Great Artesian Basin, Australia: a multivariate statistical approach*. Sci. Total Environ. 508:411–426. <http://dx.doi.org/10.1016/j.scitotenv.2014.11.099>.
- NARANY, S. T., RAMLI, M. F., ARIS, A. Z., SULAIMAN, W. N. A., & FAKHARIAN, K. 2014. *Spatiotemporal variation of groundwater quality using integrated multivariate statistical and geostatistical approaches in Amol–Babol Plain, Iran*. Environmental Monitoring and Assessment, 186(9), 5797–5815. doi:10.1007/s10661-014-3820-8
- ODAT, S. 2015. Cluster and Factor Analysis of Groundwater in Mafrqa Area, Jordan, Current World Environment Vol. 10(2), 422-431.
- ORAKWE, L. C., CHUKWUMA, E.C. 2015. *Multivariate Analysis Of Ground Water Characteristics Of Geological Formations Of Enugu State Of Nigeria*. International Journal Of Scientific & Technology Research /Volume 4, Issue 08, August.
- PINGHENG, Y., DAOXIAN, Y., WENHAO, Y., YINGLUN, K., PENG, Y., E QIUFANG H. 2010. *Formations of groundwater hydrogeochemistry in a karst system during storm events as revealed by PCA*. Chinese Sci Bull May 2010 Vol.55 No.14: 1412–1422 doi: 10.1007/s11434-010-0083-9
- RAKOTONDRABE, F., NGOUPAYOU, J. R. N., MFONKA, Z., RASOLOMANANA, E. H., ABOLO, A. J. N. E AKO, A. A. 2018. *Water quality assessment in the Bétaré-Oya gold mining area (East-Cameroon): Multivariate Statistical Analysis approach*. Science of the Total Environment 610–611 (2018) 831–844
- REVENGA, C. 2000. *Will the be enough water? In: Mock G. Ed. "Pilot Analysis of Global Ecosystems: Freshwater Systems*. EarthTrends World Resources Institute. p. 1-6.
- RAVIKUMAR, P.; SOMASHEKAR, R.K; PRAKASH, K.L. 2015. *A comparative study on usage of Durov and Piper diagrams to interpret hydrochemical processes in groundwater from SRLIS river basin, Karnataka, India*. www.elixirpublishers.com. Elixir Earth Sci. 80 (2015) 31073-31077.
- ROMANOWISKI, JOANA PAULIN & ENS, ROMILDA TEODORA (2006). *As pesquisas denominadas do tipo “estado da arte” em educação*. In Diálogo Educacional. Curitiba, v.6, n. 19, p. 37-50, set./dez., 2006.
- SAS - CARY, N. C., 1982. SAS Institute Inc. SAS User's Guide.
- SAE-JU, J., CHOTPANTARAT, S., & THITIMAKORN, T. (2019). *Hydrochemical, geophysical and multivariate statistical investigation of the seawater intrusion in the coastal aquifer at Prachuap-Khiri-Khan Province, Thailand*. Journal of Asian Earth Sciences, 104165. doi:10.1016/j.jseaes.2019.104165
- SCHOT, P.P., VAN DER WAL, J., 1992. *Human impact on regional groundwater composition through intervention in natural flow patterns and changes in land use*. Journal of Hydrology 134, 297–313
- SEYHAN, E. AND KEET, B., 1981. *Multivariate statistical analysis (part 1), Application to hydromorpho-metrica! data (case-study; Ahr River basin, Bolzano, Italy)*. Aqua-VU series A, No. 8. Rodopi, Amsterdam.
- SEYHAN, E., VAN DE GRIEND, A.A. AND ENGCLIN, G.B., 1985. *Multivariate analysis and interpretation of the hydrochemistry of a dolomite reef aquifer, Northern Italy*. Water Resour. Res., 21: 1010-1024.

-
- SHIN, K.; KOH,D-C.; JUNG, H.;AND LEE, J. 2020. *The Hydrogeochemical Characteristics of Groundwater Subjected to Seawater Intrusion in the Archipelago*, Korea. *Water*, 12, 1542; doi:10.3390/w12061542.
- SLAMA, F., & BOUHLILA, R. 2017. *Multivariate statistical analysis and hydrogeochemical modelling of seawater-freshwater mixing along selected flow paths: Case of Korba coastal aquifer Tunisia*. *Estuarine, Coastal and Shelf Science*, 198, 636–647. doi:10.1016/j.ecss.2016.10.005
- STEINHORST, R.K., WILLIAMS, R.E., 1985. *Discrimination of groundwater sources using cluster analysis, MANOVA, canonical analysis and discriminant analysis*. *Water Resources Research* 21, 1149– 1156.
- TESSEMA, S. G., MIRGORODSKY ,D., MERTEN , D., HILDEBRANDT , A., ATTINGER S., BU´CHEL, G. 2012. Hydrochemical characterization of groundwater system of biodiversity experimental field: Jena, Germany. *Environ Earth Sci*. DOI: 10.1007/s12665-012-2171-4
- USUNOFF, E.J., GUZMA´N-GUZMA´N, A., 1989. *Multivariate analysis in hydrochemistry: an example of the use of factor and correspondence analyses*. *Ground Water* 27, 27–34.
- WANDA, E.; MONJEREZI, M.; MWATSETEZA, J. F.; KAZEMBE, L. N. 2011. *Hydrogeochemical appraisal of groundwater quality from weathered basement aquifers in Northern Malawi*. *Physics and Chemistry of the Earth*, Volume 36, Issue 14-15, p. 1197-1207.
- WALLIS, J. R. 1965. *Multivariate statistical methods in hydrology--a comparison using data of known functional relationship*. *Water Resources Res.*, 1(4), 447-462.
- YANG, Q.; ZHANG, J.; WANG, Y.; FANG, Y; & MARTÍN, J. D. 2015. *Multivariate Statistical Analysis of Hydrochemical Data for Shallow Ground Water Quality Factor Identification in a Coastal Aquifer*. *Pol. J. Environ. Stud.* Vol. 24, No. 2 (2015), 769-77.
- ZHANG, Q.; GUO, J.; YU, L. E ZHAO, R. 2018. *Hydrochemical Characteristics and Multivariate Statistical Analysis of Natural Water System: A Case Study in Kangding County, Southwestern China*; *Water*, 10, 80; doi:10.3390/w10010080
- ZGHIBI, A., MERZOUGUI, A., ZOUHRI, L., & TARHOUNI, J. 2014. *Understanding groundwater chemistry using multivariate statistics techniques to the study of contamination in the Korba unconfined aquifer system of Cap-Bon (North-east of Tunisia)*. *Journal of African Earth Sciences*, 89, 1–15. doi:10.1016/j.jafrearsci.2013.09.