iN 20

DOI: 10.12461/20.02

Modelo de calidad del agua subterránea mediante el uso combinado del análisis de componentes principales (ACP) y regresiones lineales múltiples (RLM). Caso de estudio: Acuíferos de Maturín, Monagas, Venezuela.

Groundwater quality model through the combined use of principal component analysis (PCA) and multiple linear regression (MLR). Case study: Maturín's Aquifers, Monagas, Venezuela.

Modelo de qualidade das águas subterrâneas através do uso combinado da análise de componentes principais (ACP) e regressões lineares múltiplas (RLM). Estudo de caso: Aquíferos Maturín, Monagas, Venezuela.

RECIBIDO: 10/3/2020 → **APROBADO:** 21/5/2020 jalexgil2005@hotmail.com

GIL MARÍN, JOSÉ ALEXANDER

DEPARTAMENTO DE INGENIERÍA AGRÍCOLA, ESCUELA DE INGENIERÍA AGRONÓMICA (EIA), UNIVERSIDAD DE ORIENTE (UDO), MATURÍN, ESTADO DE MONAGAS, VENEZUELA.

RESUMEN

Se estudió un conjunto de datos recolectados durante el año 2016 en ochos pozos de la ciudad de Maturín, Estado de Monagas, Venezuela, en un intento por evaluar y determinar las aportaciones de la fuente que afectan la calidad del agua. Una técnica precisa de regresiones lineales múltiples (RLM) se utilizó como herramienta avanzada para el modelado y pronóstico de la calidad de las aguas subterráneas. Igualmente, se utilizó el análisis de componentes principales (ACP) para simplificar y comprender la compleja relación entre el agua y los parámetros de calidad. Se encontraron seis componentes principales responsables del 86.57% de la variación total. Se comprobó que la mayor fuente de contaminación de los acuíferos de Maturín la constituyen las descargas residuales con altos valores de coliformes fecales. El resultado del estudio mostró que el uso del ACP como entrada mejoró

DOI: 10.12461/20.02

la predicción del modelo RLM al reducir su complejidad y eliminar la colinealidad de datos, donde el valor de R^2 en este estudio fue de 0.99, indicando que el 99% de la variabilidad de los índices de calidad del agua (ICA) es explicada por las once variables independientes utilizadas en el modelo.
PALABRAS CLAVE: antropogénicos, análisis estadístico multivariado, monitoreo ambiental, contaminación.

ABSTRACT

A set of data collected during 2016 was studied in eight wells in the city of Maturín, Monagas, Venezuela, in order to evaluate and determine the contributions of the source that affect water quality. An accurate multiple linear regression (MRL) technique was used as an advanced tool for modeling and forecasting groundwater quality. Likewise, principal component analysis (PCA) was used to simplify and understand the complex relationship between water and quality parameters. Six main components responsible for 86.57% of the total variation were found. It was proved that the main source of contamination of Maturín aquifers was the residual discharges with high values of fecal coliforms. The result of the research showed that the use of PCA as inputs improved the prediction of the MRL model by reducing its complexity and eliminating the collinearity of data, where the value of R^2 in this study was 0.99, indicating that 99% of the variability of the water quality index (WQI) is explained by the eleven independent variables used in the model.
KEYWORDS: antropogenic, multivariate statistical analysis, environmental monitoring, pollution.

RESUMO

Um conjunto de dados coletados durante 2016 em oito poços na cidade de Maturín, Monagas, Venezuela, foi estudado na tentativa de avaliar e determinar as contribuições da fonte que afetam a qualidade da água. Uma técnica precisa de regressão linear múltipla (RLM) foi usada como uma ferramenta avançada para modelar e prever a qualidade das águas subterrâneas. Da mesma forma, a Análise de Componentes Principais (ACP) foi usada para simplificar e entender a complexa relação entre água e parâmetros de qualidade. Seis componentes principais foram encontrados, responsáveis por 86,57% da variação total. Verificou-se que a principal fonte de contaminação dos aqüíferos de Maturín foram as descargas residuais com altos valores de coliformes fecais. O resultado do estudo mostrou que o uso de ACP como insumos melhorou a previsão do modelo RLM, reduzindo sua complexidade e eliminando a colinearidade de dados, onde o valor de R^2 neste estudo foi de 0,99, indicando que 99% da variabilidade dos índices de qualidade da água (ACI) é explicada pelas onze variáveis independentes usadas no modelo.
PALAVRAS-CHAVE: antropogénica, análise estatística multivariada, monitoramento ambiental, contaminação.

DOI: 10.12461/20.02

INTRODUCCIÓN

El agua es vital para la naturaleza y es un recurso limitante para los seres humanos y otros seres vivos. Se requiere agua de calidad y en cantidades adecuadas para satisfacer las crecientes necesidades domésticas, industriales y agrícolas (Azaza, et al., 2011; Pazand, et al., 2012). Dependiendo de su uso y consumo, puede ser un recurso renovable o no renovable. Por lo tanto, los problemas de calidad del agua y sus opciones de gestión deben recibir mayor atención en los países. El agua subterránea es la principal fuente de agua para fines domésticos, agrícolas e industriales en el mundo. Se estima que aproximadamente un tercio de la población mundial usa agua subterránea para beber (Simge y Aysen, 2014). Entre las diversas razones, las más importantes son la falta de disponibilidad de agua superficial potable y la creencia general de que el agua subterránea es más pura y segura que el agua superficial debido a las cualidades protectoras de la cubierta del suelo (Mishra, et al., 2005; Arumugam y Elangovan, 2009). Sin embargo, en los últimos años, debido a la llegada del crecimiento industrial, la aplicación a gran escala de fertilizantes sintéticos para la producción agrícola y el uso de pesticidas e insecticidas para la producción han causado una gran preocupación respecto a la susceptibilidad de contaminación del agua subterránea. Los cambios naturales en la calidad del agua subterránea se deben a la interacción roca-agua y a las reacciones de óxido-reducción durante la filtración del agua a través de los acuíferos. Además de estos procesos, los agentes patógenos transmitidos por el agua, los contaminantes tóxicos y no tóxicos son los principales parámetros de calidad del agua que se transportan desde el área de recarga al área de descarga a través de los acuíferos por el movimiento del agua subterránea (Simge y Aysen, 2014). La calidad del agua se identifica en términos de sus parámetros físicos, químicos y biológicos. Las aguas subterráneas contaminadas no pueden lograr un ecosistema equilibrado. Un ecosistema equilibrado es aquel en el que los seres vivos y el medio ambiente interactúan de forma beneficiosa entre sí. La calidad del agua obviamente juega un papel crítico en esta relación (Ntengwe, 2006), ya que es clave para el mantenimiento de un entorno bien equilibrado. Un problema particular en el caso del monitoreo de la calidad del agua es la complejidad asociada al análisis de la gran cantidad de variables medidas. Los conjuntos de datos contienen abundante información sobre el comportamiento de los recursos hídricos. La clasificación, el modelado y la interpretación de los datos monitoreados son los pasos más importantes en la evaluación de la calidad del agua (Boyacioglu, 2006). La aplicación de diferentes técnicas estadísticas multivariadas, como el análisis de conglomerados (AC), el análisis de componentes principales (ACP), el análisis factorial (AF) y el análisis de regresión múltiple (ARM) ayudan a identificar componentes o factores importantes, representando la mayoría de las variaciones de un sistema (Ouyang, et al., 2006; Shrestha y Kazama, 2007). Están diseñados para reducir el número de variables a un pequeño número de índices al intentar preservar las relaciones presentes en los datos originales. El objetivo de este estudio consistió en utilizar los ACP y ARM en los estudios de calidad del agua para extraer los parámetros más relevantes que permitan evaluar las variaciones en la calidad de las aguas subterráneas de la ciudad de Maturín.

DOI: 10.12461/20.02

MATERIALES Y MÉTODOS

Sitio de estudio

El trabajo se llevó a cabo en la ciudad de Maturín, municipio de Maturín, estado de Monagas, situado geográficamente a 9°46'46" de latitud norte y 63°11'05" de longitud oeste, y a una altura de 63.5 m.s.n.m (Figura 1). Holdridge, citado por Gil y otros (2012), señalan que el clima de la zona es del tipo bosque seco tropical, caracterizado por presentar una estación lluviosa de mayo a diciembre y una estación seca de enero a abril, con una precipitación media anual de 1219.6 mm, una temperatura media anual de 25.9 °C, una evapotranspiración potencial de 1372 mm y una evaporación de 1573 mm. Tiene una superficie de 13.352 km^2 y es la sexta urbe más grande del país con 542.259 habitantes (Instituto Nacional de Estadística, 2011).

Figura 1. Ubicación geográfica de la Ciudad de Maturín. Estado Monagas. Venezuela. Fuente: (Gil, et al., 2012).

Según el Ministerio del Poder Popular del Ambiente (2009), los acuíferos de Maturín se encuentran ubicados en la región hidrológica de la provincia del Orinoco, en el sistema acuífero Mesa-Las Piedras, y se caracterizan por presentar una profundidad media de 40 m, un nivel medio de 10 m y un caudal de 10.5 l/s. Constituyen uno de los reservorios subterráneos con mayor potencial del país con unidades litológicas de sedimentos poco o nada consolidados (gravas, conglomerados, areniscas, con intercalaciones de arcillas y lutitas), que cubren un área de 352.000 km^2 (Duran, 2011).

Lugares de muestreo

Se realizó una encuesta detallada dentro de un radio de 10 km en Maturín para identificar los pozos de perforación activos que se utilizan con fines agrícolas y

DOI: 10.12461/20.02

consumo humano. En consecuencia, se seleccionaron 8 pozos de perforación en diferentes lugares (Figura 2) para el muestreo del agua, de los cuales 5 lugares (P1, P2, P3, P5, y P7) se ubicaron en zonas urbanas, y 3 en zonas rurales y agrícolas (P4, P6 y P8).

Figura 2. Ubicación de los pozos de muestreo.

Conjunto de datos

En este estudio, veinte parámetros de calidad del agua fueron observados a lo largo de los pozos de monitoreo del acuífero. Un total de 960 observaciones se utilizaron para la distribución de fuentes y técnicas de modelado. El índice de calidad del agua (ICA) se determinó para evaluar el estado de la calidad del agua y la clasificación de la clase del acuífero en base a los datos de calidad del agua, utilizando la metodología empleada por Gil y otros (2018). Los veinte parámetros de calidad del agua consistieron en: fenoles, hierro (Fe), calcio (Ca), magnesio (Mg), manganeso (Mn), sodio (Na), potasio (K), nitrito (NO_2), nitrato (NO_3), sulfato (SO_4), bicarbonatos (HCO_3), carbonatos (CO_3), coliformes totales (CT), coliformes fecales (CF), temperatura, conductividad eléctrica (CE), pH, total de sólidos disueltos (TDS), dureza (mg/l) y cloro (Cl). El ICA proporciona una manera útil de predecir cambios y tendencias en la calidad del agua considerando múltiples parámetros. Por lo tanto, en este estudio se formó con catorce variables seleccionadas: coliformes fecales (CF), temperatura, conductividad eléctrica (CE), pH, sodio (Na), manganeso (Mn), hierro (Fe), fenoles, potasio (K), nitrito (NO_2), nitrato (NO_3), sulfato (SO_4), dureza (mg/l) y cloro (Cl).

Tratamiento previo del conjunto de datos

Los datos se dispusieron inicialmente según las estaciones y el mes de muestreo. A todos los datos se les aplicó la prueba de normalidad de Wilk-Shapiro (Azkue

DOI: 10.12461/20.02

y Soto, 2004), con una probabilidad del 95%. La hipótesis nula de que los datos son normales es rechazada al nivel de significancia α ($\alpha = 0.05$), si la probabilidad de la prueba es inferior al nivel de significancia ($p \leq 0.05$). Esta se llevó a cabo con el paquete estadístico SAS, versión 8.0, y el procedimiento "univariate". Luego, el cálculo estadístico de ACP y RLM se realizó utilizando el software SPSS, versión 16.0.

Análisis de componentes principales (ACP)

El análisis de componentes principales presenta los parámetros más significativos y describe la reducción de datos de una representación de conjunto de datos completos, con la pérdida mínima de información original (Helena, et al., 2000 y Wunderlin, et al., 2001). Los ACP son sensibles a los valores atípicos, a los datos que faltan y a la mala correlación lineal entre las variables debido a la insuficiencia de las variables asignadas (Sarbu y Pop, 2005). Por lo tanto, debe realizarse un tratamiento previo detallado al conjunto de datos para obtener una imagen más clara en datos complejos. Es una técnica prominente para el reconocimiento de patrones en los intentos de explicar la varianza de un gran conjunto de variables correlacionadas y la transformación en un conjunto más pequeño de variables independientes (no correlacionadas) (componentes principales). El componente principal (CP) se expresa como:

$$Y_{ij} = Z_{i1}X_{1j} + Z_{i2}X_{2j} + Z_{i3}X_{3j} + ... + Z_{im}X_{mj} .....(1)$$

Donde Z es la carga del componente, Y es la puntuación del componente, X es el valor medido de una variable, i es el número del componente, j es el número de la muestra y m es el número total de variables. Mientras tanto, el análisis factorial (AF) intenta extraer una estructura lineal de dimensión inferior del conjunto de datos. Reduce aún más la contribución de las variables menos significativas obtenidas de ACP y el nuevo grupo de variables conocidas como varifactores (VF), que se extrae mediante la rotación del eje definido por ACP. En el AF, el concepto básico se expresa en la ecuación 2.

$$Y_{ji} = Z_{p1}P_{1i} + Z_{p2}P_{2i} + Z_{p3}P_{3i} + ... + Z_{pr}P_{ri} + C_{pj} .....(2)$$

Donde Y es el valor medido de una variable, i se refiere a la carga del factor, P es la puntuación del factor, C es el término residual que representa los errores u otras fuentes de variación, i es el número de muestra, j es el número de la variable y r es el número total de factores. Principalmente, los métodos ACP y AF se expresan en ecuaciones similares; la diferencia está en CP, donde se expresa como una combinación lineal de variables medidas. Mientras que en AF, la variable medida se expresa como una combinación de factores y la ecuación contiene el término residual y, por lo tanto, un VF puede incluir variables latentes no observables, hipotéticas (Helena, et al., 2000, Wunderlin, et al., 2001 y Liu, et al., 2003). El análisis de componentes principales o el análisis factorial se realizó en la matriz de correlación de los datos reorganizados (todas las observaciones para cada grupo de sitios), por lo tanto explica la estructura del conjunto de datos subyacentes. La matriz de coeficientes de correlación mide qué tan bien puede explicarse la varianza de cada constituyente en función de la relación

DOI: 10.12461/20.02

entre sí. El ACP de las variables normalizadas (conjunto de datos de calidad del agua) se realizó para extraer CP significativos y reducir aún más la contribución de las variables con menor importancia; estos CP se sometieron a rotación varimax (en bruto) generando VF.

Los CP resultantes de ACP a veces no se interpretan fácilmente y la rotación de varimax debe realizarse para reducir la dimensionalidad de los datos e identificar las nuevas variables más significativas. El coeficiente del factor varimax (VF) con una correlación de > 0.75 se explica como una carga de factor significativo fuerte (Liu, et al., 2003). Mientras que los rangos de correlación de 0.75-0.50 y 0.50-0.30 se consideran carga de factores moderada y débil respectivamente. En pocas palabras, el análisis de componentes principales tiene como objetivo descubrir un conjunto más subyacente de factores que explican el patrón principal de todas las variables originales (Saima, et al., 2009).

Análisis de regresión múltiple

El análisis de regresión lineal múltiple (RLM) nos permite establecer la relación que se produce entre una variable dependiente Y y un conjunto de variables independientes (X1, X2,... XK). El análisis de regresión lineal múltiple, a diferencia del simple, se aproxima más a situaciones de análisis reales; puesto que los fenómenos, hechos y procesos naturales son complejos por definición y, en consecuencia, deben ser explicados en la medida de lo posible por la serie de variables que, directa e indirectamente, participan en su concreción (Karim y Abdolrahman, 2011).

La anotación matemática del modelo o ecuación de regresión lineal múltiple es la siguiente:

$$Y_u=\beta_0+\beta_1 X_{u1}+\beta_2+X_{u2}+\ldots\ \ldots.+\beta_k X_{uk}+\boldsymbol{\mathcal{E}}_u\ldots\ldots(3)$$

Donde Y_u denota la respuesta observada en el intento u o variable dependiente, X_{u1} representa el nivel del factor i en el intento u o variables independientes, lasβ son parámetros desconocidos y $\boldsymbol{\mathcal{E}}_u$ representa el error aleatorio en Y_u.

En el análisis de regresión lineal múltiple, la construcción de su correspondiente ecuación se realizó con el método *"stepwise"* o *paso a paso*, seleccionando las variables una a una. La finalidad fue buscar entre todas las posibles variables explicativas (parámetros físicos, químicos y biológicos) aquellas que más y mejor expliquen la variable dependiente (ICA), sin que ninguna de ellas sea una combinación lineal de las restantes. Este procedimiento implica tres pasos: en el primero, solo se introduce aquella variable que cumple con los criterios de entrada; en segundo lugar, se valora si la misma cumple con los criterios de salida; y por último último, se calculan los factores de bondad de ajuste de los datos al modelo de regresión lineal y los parámetros del modelo verificado. El proceso se inicia sin ninguna variable independiente en la ecuación de regresión y el proceso concluye cuando no queda ninguna variable fuera de la ecuación que satisfaga el criterio de selección (garantiza que las variables seleccionadas son significativas) y/o el criterio de eliminación (garantiza que una variable seleccionada no es redundante). El análisis comenzó con 20 variables explicativas y terminó recomendando solo 11 variables.

DOI: 10.12461/20.02

RESULTADOS Y DISCUSIONES

Las estadísticas descriptivas relacionadas con la calidad del agua se presentan en la Tabla 1. Los valores de calidad del agua para el manganeso, nitrito, nitrato, fenol y coliformes fecales sobrepasan los límites permisibles según la Organización Mundial de la Salud (Organización Mundial de la Salud, 2006) y la normativa nacional Decreto N° 883 (Venezuela, 1995). Para el pH, el valor promedio se encuentra un poco por debajo del valor umbral. Los valores de dureza, sodio (Na), conductividad eléctrica (CE) y cloruro (Cl) parecen ser de desviación estándar muy alta. Estas variables cambian considerablemente en los pozos con respecto a sus valores promedios, pero no influyen en la calidad de sus aguas porque no sobrepasan los valores límites establecidos en las normas planteadas.

Tabla 1. Estadísticos descriptivos básicos de los parámetros físicos, químicos y biológicos estudiados en 8 pozos del Municipio Maturín. Período marzo-mayo, 2016.

		RANGO (Nº DE POZOS)			LMP		Desviación Estándar
Parámetro	Unidad	Mínimo	Máximo	Media	Decreto 883*	OMS	
Temperatura	ºC	28.20(P-4)	31.20(P-3)	29.51(P-3,5,8)			0.66
pH		4.12(P-6)	7.56(P-1)	5.57	6.0-8.5		1.01
Dureza	MG/L	15.15(P-4)	217.17(P-1)	27.73	500.0		63.08
CE	MS/CM	30.00(P-4)	600.00(P-1)	152.19		3000	177.20
Fe	MG/L	0.00(P-6)	0.54(P-1)	0.09(P-3)	0.1	0.3	0.14
Mn	MG/L	0.00(P-1,2,3,4,5,6,7)	0.50(P-8)	0.09	0.1	0.5	0.15
Na	MG/L	2.53(P-4)	90.62(P-)	30.35	200		22.38
K	MG/L	5.00(P-2)	40.00(P-1,4,7)	25.00		50	8.34
NO_2	MG/L	0.00(P-2,3,4,5,6,7,8)	22.00(P-8)	4.04	5	3	6.07
NO_3	MG/L	4.50(P-3)	59.50(P-2)	25.13	5	50	12.80
SO_4	MG/L	0.00 (P-1,2,3,4,5,6,7,8)	45.00(P-1)	5.42	400	250	12.24
Cl	MG/L	30.00 (P-2)	150.00(P-6)	67.92	600	250	27.97
Fenoles	MG/L	0.0026(P-2)	0.0224(P-3)	0.0075	0.002		0.063
CF	NM-P/100ML	0 (P-1,2,3,4,5,6,7,8)	4300(P-1)	260.75	0		913.1636

LMP=Límite máximo permitido por la normativa venezolana, Decreto 883 (Venezuela, 1995) y la Organización Mundial de la Salud, 2006.

DOI: 10.12461/20.02

Análisis de Componentes Principales (ACP)

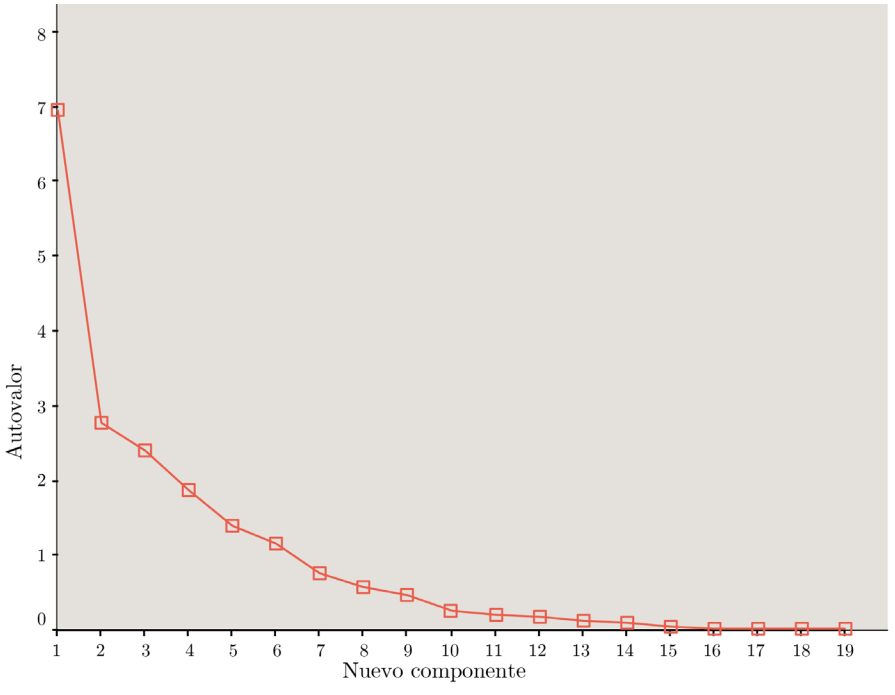
Antes de ejecutar los análisis de componentes principales (ACP), se aplicó la prueba de Kolmogorov Smirnov (K-S) a todo el conjunto de datos, para comprobar si se ajustaron a una distribución normal. La prueba K-S indicó que todas las variables siguen una distribución normal de registro. Por otra parte, a fin de confirmar la idoneidad de los datos para el análisis de componentes principales se aplicaron las pruebas de Kaiser-Mayer-Olkin (KMO) y de esfericidad Barlett. La prueba de KMO mide adecuación del muestreo y señala la fracción de variables que tienen varianza común. Los valores de KMO superiores a 0.5 son considerados como satisfactorios para ACP. En el presente estudio, el valor de 0.575 (Tabla 2) indica que el conjunto de datos estuvo en forma para el análisis de componentes principales. En tanto, la prueba de esfericidad de Barlett indica si las variables están significativamente relacionadas y si la matriz de correlación es una matriz de identidad. El nivel de significancia de 0.00 para este estudio indicó claramente que la matriz de correlación no es la matriz de identidad y que las variables están significativamente relacionadas. El análisis de componentes principales se aplicó a los datos obtenidos de los ocho sitios de muestreo, utilizando el software SPSS 16.0. Los resultados del ACP en la matriz de correlación y factores fueron extraídos por el método del centroide rotado a través de la rotación Varimax (Ahmed, 2005). Existen reglas para determinar el adecuado número de factores a conservar. Entre ellas se encuentra el que se conoce como criterio de Kaiser (Kaiser, 1960), que indica que hay que conservar los componentes principales cuyos valores propios sean mayores que la unidad o que el porcentaje de varianza total explicada por cada componente o factor (varianza acumulada) sea lo suficientemente alto (Holland y Witthuser, 2009; Belkhiri, et al., 2010; Kanade y Gaikwad, 2011). Bajo estos criterios, solo los factores con valores propios mayores o iguales a 1, fueron aceptados como posibles fuentes de varianza en los datos, con la más alta prioridad atribuida al factor que tiene la suma más grande del eigenvector (Belkhiri, et al., 2010). Los autovalores indican la importancia del factor, valores mayores a 1 son considerados significativos. Los resultados del ACP señalaron que los primeros seis valores de los autovalores son mayores a 1 y por lo tanto fueron considerados como significativos. En la figura 3 se puede observar que la mayor parte de la variación en los datos originales se explica por los primeros seis factores.

Tabla 2. Prueba de KMO y Bartlett.

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		0.575
Prueba de esfericidad de Bartlett		
Chi-cuadrado aproximado	571.638	
Gl	171	
Sig.	0.00	

DOI: 10.12461/20.02

Figura 3. Gráfico de sedimentación.



Los valores propios para diferentes factores, varianza porcentual, variación porcentual acumulada y componentes cargados (sin rotar y Varimax rotadas), son presentados en la Tabla 3. Este análisis condujo a la explicación del 86.57% de las varianzas en los datos. Los parámetros de carga para los seis componentes del ACP del conjunto de datos se dan en la Tabla 4. Liu y otros (2003) clasificaron a las cargas factoriales como "fuertes", "moderadas" y "débiles", correspondientes a valores de carga absolutos de > 0.75, 0.75–0.50 y 0.50–0.30, respectivamente. A modo de determinar el factor relevante para cada factor solo se tomaron en cuenta valores superiores a 0.65.

DOI: 10.12461/20.02

Tabla 3. Análisis de componentes principales por método de extracción.

Varianza total explicada									
Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Sumas de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total	% de la varian-za	% acumulado
1	6.930	36.475	36.475	6.930	36.475	36.475	5.966	31.400	31.400
2	2.766	14.558	51.033	2.766	14.558	51.033	2.934	15.443	46.843
3	2.392	12.587	63.620	2.392	12.719	63.620	2.098	11.043	57.887
4	1.843	9.719	73.338	1.847	9.719	73.338	2.095	11.024	68.911
5	1.375	7.238	80.577	1.375	7.238	80.577	1.976	10.400	79.311
6	1.139	5.996	86.573	1.139	5.996	86.573	1.380	7.262	86.573
7	.732	3.854	90.427						
8	.732	2.943	93.370						
9	.443	2.330	95.700						
10	.250	1.313	97.014						
11	.182	.958	97.972						
12	.161	.846	98.818						
13	.118	.621	99.439						
14	.069	.365	99.804						
15	.018	.095	99.900						
16	.012	.063	99.964						
17	.007	.034	99.998						
18	.000	.001	99.999						
19	.000	.001	100.000						

DOI: 10.12461/20.02

Tabla 4. Matriz de componentes rotados.

Matríz de componentes rotados.

	Componente					
	1	2	3	4	5	6
Temperatura	.114	-.003	.060	.079	.186	.922
pH	.794	.013	.394	.123	-.075	.304
CE	.946	.258	.046	-.017	.111	-.014
TDS	.944	.251	.062	-.033	.111	-.018
Dureza	.973	.195	.052	.041	.013	-.025
Fe	.569	.563	.013	-.045	.023	.194
Ca	.971	.213	.060	.019	.010	.004
Mg	.015	.056	.637	.072	.563	.057
Mn	.055	-.207	.233	.195	.878	-.056
Na	-.178	-.095	-.923	.096	-.167	.044
K	.006	.430	.132	.536	.119	-.467
NO_2	.105	.121	-.012	-.235	.868	.242
NO_3	.185	-.058	.057	-.812	.030	-.032
SO_4	.580	.512	.047	.194	-.034	-.117
CI	.359	.098	-.482	.671	.059	-.059
HCO_3	.830	-.168	-.261	-.234	.118	.117
Fenol	.025	-.200	.537	.685	-.068	.259
CT	.172	.956	.058	.026	-.050	-.112
CF	.233	.944	-.023	.031	-.013	.028

Método de extracción: Análisis de componentes principales.
Método de rotación: Normalización Varimax con Kaiser.
[a.] La rotación ha convergido en 8 iteraciones.

El primer componente (CP1), denominado componente geológico, explica el 36.48% de la varianza total, presentando una fuerte correlación positiva con las variables pH, CE, Ca, dureza, TDS, y HCO_3. Esta combinación de variables destaca el grado natural de mineralización del agua y podría estar relacionada con la geología del área estudiada, derivada de los procesos de evaporación y disolución. Por otro lado, esto indica que las variables CE, Ca, dureza, TDS y HCO_3 en el área de estudio están muy influenciadas por los valores del pH. Los valores de pH presentaron un valor mínimo de 4.12 (P6) y un máximo de 7.56 (P1), con un promedio de 5.57, que les otorga a las aguas subterráneas de la ciudad de Maturín una naturaleza ligeramente ácida. A pH bajo, una parte importante del carbono inorgánico disuelto está como ácido carbónico (H_2CO_3) o CO_3 disuelto. Además, las aguas subterráneas suelen contener cantidades apreciables de ácido silícico no disociado, que se suele expresar como sílice disuelta. En general, un agua con un pH menor a 6.5 podría ser ácida y corrosiva. La dureza depende de la presencia de Ca^2+ y Mg^2+ en el agua. Un patrón similar en la variación estacional de dureza y HCO_3 se debe posiblemente a la disolución de minerales como la calcita y la dolomita a través del proceso de meteorización HCO_3. Este proceso está representado por la siguiente reacción simplificada (Cai, et al., 2007).

DOI: 10.12461/20.02

$$CaCO_3 + CO_2 + H_2O <-> Ca^2+ + 2HCO_3-$$

La reacción del $CaCO_3$ mineral con dióxido de carbono disuelto (ácido carbónico) libera HCO_3- y Ca^{2+} en el agua del acuífero. Existe una relación inversa entre la temperatura y el CO_2 disuelto (Zare, et al., 2011), por lo tanto, la intensidad de la intemperie de carbonatos y la concentración de HCO_3 y dureza es mayor en las estaciones frías, es decir, en invierno, y menor en las estaciones cálidas o de verano.

El segundo componente (CP2) representa el 14.56% de la varianza total, y fue correlacionado de manera fuertemente positiva con las variables CT y CF, lo que revela la intensiva descarga de aguas residuales en el acuífero. Este componente principal representa un factor microbiológico de la calidad del agua. En esta investigación, se observó un valor promedio general de coliformes fecales de 260.75 ± 913.16; con un valor máximo de 4,300 y un valor mínimo de 0 NMP/100ml, que sobrepasa en más de un 4,000% al máximo permitido por la normativa vigente. Estos altos valores de coliformes fecales tienen como causa principal la filtración de aguas contaminadas hacia los acuíferos debido a una mala disposición de sus desechos fecales (pozos sépticos, napas freáticas y otros) y a las actividades pecuarias de la zona. Las bacterias coliformes se excretan en grandes cantidades en las heces de los humanos y otros animales de sangre caliente. En consecuencia, el agua contaminada por materia fecal se identifica como potencialmente peligrosa debido a los organismos indicadores que coexisten con Escherichia coli (E. coli), que causa el cólera (Cansu, et al., 2008). Algunos ejemplos de enfermedades causadas por beber o nadar en agua contaminada con heces son diarrea, cólera, disentería e infecciones de piel, ojos, oídos, nariz y garganta (Organización Mundial de la Salud, 2006).

El CP3, denominado factor mineral, representa el 12.59% de la varianza total del conjunto de datos del agua del acuífero y muestra una fuerte correlación negativa con el sodio (Na) y una moderada correlación positiva con el magnesio (Mg). Este factor contiene variables hidrogeoquímicas (Mg y Na), que se originan en la mineralización de los componentes geológicos de los acuíferos. La contribución de Mg a este factor puede considerarse como resultado de los procesos de intercambio de cationes en la interfaz suelo-agua (Guo y Wang 2004) y la disolución de minerales con calcio que se encuentran en la región (Memet y Bulent, 2009). Según Beita-Sandí y Barahona (2010), el Na en aguas naturales varía de manera considerable dependiendo de las condiciones geológicas locales y las descargas de aguas residuales. La correlación negativa del Na indica que su introducción en el agua subterránea no se debe a la disolución natural del suelo o la roca, y la fuente de este parámetro puede estar relacionada con una contaminación del tipo antropogénica.

El CP4 representa el 9.72% de la varianza total y muestra una fuerte correlación negativa con los nitratos (NO_3^-) y una moderada correlación positiva con Fenoles y Cl^-. Este componente principal representa un factor de contaminación orgánica de la calidad del agua, que se produce por desechos domésticos y nutrientes. Mientras se produzca nitrificación en el agua, un aumento en NO_3^- y una reducción en NH_4^+ es un evento normal. En cuanto al nitrato (NO_3^-), se observó su presencia en todos los puntos de muestreo, resultando con mayor concentración el P2, con valores de 59.50 mg/l, y un valor mínimo para el P3 con 4.50 mg/l. A pesar de que las aguas subterrá-

DOI: 10.12461/20.02

neas presentaron nitrato, solo en P2 se observaron valores por encima de la normativa (Organización Mundial de la Salud, 2006). El NO_3- presente en las aguas subterráneas tiene un origen orgánico. Los altos valores de esta especie se asocian con la actividad antrópica, excreto de animales, uso de fertilizantes (Cardona, et al., 2004 y Jin, et al., 2004). La concentración de esta especie en aguas subterráneas no contaminadas varía ampliamente, aunque no suele sobrepasar los 10 mg/l (Porras, et al., 1985). Para el fenol, (C_6H_5OH) se observaron diferencias significativas en los meses de muestreo y en los pozos estudiados, dando como resultados valores mayores a 0.002 mg/l en todos los muestreos, valor umbral indicado por la normativa Decreto N° 883 (Venezuela, 1995). Las concentraciones de este compuesto en las aguas subterráneas de la ciudad de Maturín son alarmantes ya que las estudiadas durante el periodo de lluvia incrementaron los valores del compuesto químico, sobrepasando los niveles permitidos. Esto es sumamente inquietante ya que estudios científicos señalan que un exceso de este compuesto en el agua para consumo puede generar toxicidad, debido a su acción cancerígena y mutagénica en los seres vivos (Nieves, et al., 2005), además de alteraciones en el intestino y en muchos casos la muerte, según la concentración de ingesta. Su presencia en las aguas subterráneas de Maturín es una característica que se debe analizar con más detalle ya que su incorporación a los acuíferos puede estar asociado a procesos de contaminación de la fuente por desechos industriales, aguas servidas o residuales y/o agentes químicos que presenten este compuesto fenólico (Díaz, et al., 2005). Por otra parte, el cloruro en el agua potable se origina de fuentes naturales, aguas residuales, efluentes industriales y escorrentía urbana. Su concentración excesiva aumenta la velocidad de corrosión de los metales en el sistema de distribución de agua y depende de la alcalinidad del agua; y esto puede llevar a un aumento de las concentraciones de metales en el suministro (Organización Mundial de la Salud, 2006).

El quinto componente (CP5) explica el 7.24% de la varianza total, presentando una fuerte correlación positiva con las variables NO_2- y Mn; y se puede asumir como un componente antropogénico tóxico. En el proceso de nitrificación, NO_2- es un intermediario inestable que tiende a transformarse en NO_3. El manganeso (Mn) no existe como un metal libre en la naturaleza, pero forma parte de otros minerales, algunos de los cuales incluyen diferentes óxidos, silicatos, boratos, sulfuros, carbonatos y fosfatos (Nádaská, et al., 2010). El Mn se introduce en ambientes acuáticos a través de fuentes antropogénicas, como descargas de aguas residuales municipales, lodos de depuración, emisiones de las industrias de aleaciones, hierro y acero, y procesamiento de minerales y minería. Además de ser un elemento esencial para el ser humano y otros animales, tanto la carencia como la sobreexposición pueden causar efectos adversos, como lo indican diversos estudios que han notificado efectos neurológicos adversos tras la exposición prolongada a concentraciones muy altas en el agua de consumo (Calla y Cabrera, 2010). La presencia de Mn puede derivarse de la generación de procesos de reducción (bajo potencial redox) en las corrientes contaminadas con aguas residuales; estos procesos solubilizan metales como el Mn (Chávez-Alcántar, et al., 2011). Las aguas residuales son también fuente de metales pesados (Chávez-Alcántar, et al., 2011). A pesar de la ausencia del elemento en algunos pozos, se evidenció su presencia en otros, y en concentraciones que sobrepasan la normativa sanitaria (0.1 mg/l). Los pozos P2, P4, P5 y P6 no evidenciaron la presencia de este

DOI: 10.12461/20.02

elemento, por lo que el promedio general de manganeso estuvo en 0,09 ± 0,15 mg/l, con un valor máximo de 0,50 mg/l para el P8 y un mínimo de 0,00 mg/l.

El último componente (CP6), denominado componente físico, explica el 5.99% de la varianza total, presentando una fuerte correlación positiva con la variable temperatura. La temperatura es una de las variables fundamentales a ser medidas en ambientes acuáticos, debido a su influencia en todos los procesos hidrológicos, químicos y biológicos. Tiene gran efecto en la solubilidad de las sales y las reacciones biológicas de los seres vivos, y también acelera la putrefacción y la demanda de oxígeno (Benítez, 2004). Además influye en el pH, afecta los procesos físicos y químicos (pH, equilibrio iónico, velocidades de reacción) y los procesos biológicos (tasa metabólica, descomposición de materia orgánica); y por lo tanto, repercute en el efecto de los agentes contaminantes (Villa, 2011). La temperatura es un parámetro muy importante para un acuífero. En este sentido, Gil (2003) señala que, por lo general, las temperaturas de las aguas subterráneas permanecen constantes a menos de que existan factores externos que intervengan.

Análisis de la regresión lineal múltiple (RLM)

Básicamente, la RLM se basa en un proceso lineal de ajuste de mínimos cuadrados y requiere un elemento traza o propiedad a ser determinada para cada fuente o categoría de fuente (Mohd, et al., 2011). La RLM se utilizó para identificar las posibles fuentes de contaminación en los pozos muestreados. Dos tipos de modelos básicos de receptores que generalmente se aplican a la fuente son el balance químico de masa (BQM) y las técnicas multivariadas (Gordon, 1988). Usando el paquete estadístico SPSS, versión 16.0, se realizó el análisis de regresión múltiple para determinar los factores que contribuyen con el índice de calidad del agua. En este análisis se fijó el índice de calidad del agua (ICA) como variable dependiente, y pH, CE, Fe, Mn, Na, NO_2, NO_3, SO_4, Cl, Fenol y CF como variables independientes (o predictoras).

El procedimiento *Stepwise* de SAS, con la opción de Backward, que comenzó con las 20 variables tomadas en consideración para el análisis de los ICA, finalmente nos redujo el número de variables a 11 sin perder la precisión en el análisis de datos, con un R^2 igual a 0.9999. Según Grivas y Chaloulakou (2006), el procedimiento de regresión múltiple por pasos se usa comúnmente para producir un modelo parsimonioso que maximiza la precisión, con un número opcionalmente reducido de variables predictoras. El coeficiente β representa el coeficiente en el modelo para predecir la calidad del agua. ET representa el error estándar o típico del coeficiente β, mientras que el β estandarizado es el peso de cada variable independiente. Esto es más importante, ya que el objetivo principal del estudio es evaluar la influencia de los predictores en la calidad del agua en lugar de predecir la calidad del agua.

Los resultados del análisis estadístico se presentan en la Tabla 5.

DOI: 10.12461/20.02

Tabla 5. Análisis de regresión múltiple para determinar los factores que contribuyen con Índice de calidad del Agua (ICA).

Parámetros	Coeficiente β	Error Típico (ET)	β Estandarizado	Posición	t student´s	Valor p
(Constante)	-39.477	0.487			-81.102	<0.001a
pH	6.373	0.084	0.053	IV	76.024	<0.001a
CE	0.005	0.001	0.007	VIII	5.905	<0.001a
Fe	9.417	0.436	0.011	VII	21.614	<0.001a
Mn	-31.618	0.434	0.04	V	72.895	<0.001a
Na	0.017	0.003	0.003	X	5.107	<0.001a
NO_2	2.104	0.01	0.105	III	217.646	<0.001a
NO_3	0.136	0.004	0.014	VI	32.413	<0.001a
SO_4	0.02	0.005	0.002	XI	3.731	<0.003a
CI	0.025	0.003	0.006	IX	7.526	<0.001a
Fenol	6318.542	10.973	0.332	II	575.81	<0.001a
CF	0.126	0	0.95	I	1919.101	<0.001a

a Altamente significativo

Los once parámetros considerados para la evaluación del ICA están contribuyendo significativamente con el ICA. Los CF, el fenol y los nitritos presentan los valores más altos de los pesos de los β estandarizados y tienen una marcada influencia en el índice de calidad del agua en el área de Maturín. Para el resto de los parámetros de calidad del agua, aunque estadísticamente significativo ($p < .001$), su contribución al ICA es insignificante en comparación con CF, Fenol y nitritos, basado en el coeficiente β estandarizado obtenido del análisis de regresión múltiple realizado para definir el ICA. La mala calidad del agua de Maturín se puede determinar en orden por CF, Fenol y nitritos. Este estudio brinda una visión general de la calidad del agua potable de aguas subterráneas en esta zona. La identificación de los indicadores de calidad del agua ayudaría a concentrarse en pocos elementos (atributos) durante un estudio (el chequeo) regular de la calidad del agua; por lo tanto, solo se determinarían los parámetros de CF, Fenol y nitritos. Esto ahorra tiempo, energía y gasto en gran medida sin comprometer la calidad de los resultados.

La Tabla 6 muestra el resumen del modelo del ICA. El valor R^2 ajustado es alto y el cuadrado medio residual es bajo para el modelo. Los coeficientes de regresión derivados no son cero ni menos que el error estándar. Para un modelo, el R^2 ajustado aumenta si la adición de la variable reduce el cuadrado medio residual. Además de esto, no es bueno retener variables insignificantes, es decir, variables con

DOI: 10.12461/20.02

coeficientes cero o coeficientes menores que sus correspondientes errores estándar (Montgomery, et al., 2003).

Tabla 6. Análisis de varianza del modelo de regresión múltiple.

Fuente	Suma de cuadrados	GL	Cuadrado medio	Cociente-F	P-Valor
Modelo	338160.39	11	30741.854	1157469	0.000
Residuo	0.319	12	0.027		
Total (Corr.)	338,160.71	23			
R-cuadrado=99.9997 porcentaje					
R-cuadrado(ajustado para g.l)=99.9995 porcentaje					
Error estándar de est.=0.278912					
Error absoluto medio=0.16287					
Estadístico de Durbin-Watson=2.455 (P=0.0294)					
Autocorrelación residual en Lag1=-0.231793					

El estadístico R-cuadrado indica que el modelo explica un 99.9997% de la variabilidad en el ICA. El estadístico R-cuadrado ajustado, que es más conveniente para comparar modelos con diferentes números de variables independientes, es 99.9995%. El error estándar de la estimación muestra la desviación típica de los residuos que es 0.278912. Este valor puede usarse para construir los límites de predicción para las nuevas observaciones. El error absoluto medio (MAE) de 0.16287 es el valor medio de los residuos. El estadístico Durbin-Watson (DW) examina los residuos para determinar si hay alguna correlación significativa basada en el orden en que se introdujeron los datos en el fichero. Dado que el p-valor es inferior a 0.05, hay indicio de una posible correlación serial. Para decidir la simplificación del modelo, se tomó en cuenta que el p-valor más alto en las variables independientes fue 0.003, perteneciendo a SO_4. Puesto que el p-valor es inferior a 0.01, ese término es estadísticamente significativo para un nivel de confianza del 99%. Por lo tanto, no vale la pena quitar ninguna variable del modelo.

Por último, los coeficientes de regresión del modelo para la ecuación ICA se muestran a continuación:

ICA = -39.477 + 6.373*pH + 0.005*CE (µS/cm) + 9.417*Fe (mg/l) + 31.618*Mn (mg/l) +0.017*Na (mg/l) +2.104*NO2 (mg/l) + 0.136*NO3(mg/l) + 0.02*SO4 (mg/l) + 0.025*Cl (mg/l) +6318.54*Fenol (mg/l) + 0.126*CF (NMP/100ml)..... (3).

El análisis de varianza (ANOVA) para el modelo indicó que el valor observado F fue 1,1157,469 para ICA, mientras que el valor crítico de tabla F (0.01, 11, 12) fue 4.23, revelando que el valor observado es muchas veces mayor que el valor crítico. Para que un modelo de regresión sea útil como predictor, la relación F observada

DOI: 10.12461/20.02

debe ser al menos cuatro o cinco veces mayor que el valor crítico de F, como lo señalan Montgomery y otros (2003).

Gupta y Roy (2012), para aguas subterráneas en una zona minera de la India, reportan una ecuación múltiple con 9 parámetros fisicoquímicos (arsénico, magnesio, dureza, COD, turbidez, dureza total, BOD, DO, nitrato y sulfato). Mientras Pius y otros (2012), también para aguas subterráneas en una zona industrial al sur de la India, definen una ecuación con pH, alcalinidad, TDS, Ca, Mg, cloruro, sulfato y nitrato como variables independientes (o predictores).

CONCLUSIONES

La mayor fuente de contaminación de los acuíferos de Maturín la constituyen las descargas residuales con altos valores de coliformes fecales infiltradas, producto de una mala disposición de sus desechos fecales (pozos sépticos, napas freáticas y otros) y de las actividades pecuarias de esta zona. En segundo lugar, se ubican las actividades de la industria petrolera y la agricultura, en las cuales se identifican a los parámetros de fenoles y NO3- como las mayores cargas de contaminación. Los ACP dieron como resultado seis factores que explicaron el 86.57% de la varianza acumulada en el conjunto de datos. El primer factor explicó el 36.48% de la varianza acumulada; el segundo factor, el 14.56%; el tercer factor, el 12.58%; el cuarto, el 9.72%; el quinto, el 7.24%; y por último, el sexto, el 5.99% de las varianzas acumuladas. Los resultados del análisis de regresión lineal múltiple (RLM) revelan que pH, CE, Fe, Mn, Na, NO_2, NO_3, SO_4, Cl, Fenol y CF son los parámetros más importantes utilizados para evaluar los cambios en la calidad del agua subterránea. Los resultados anteriores muestran que el modelo APC-ARM proporciona una buena precisión para el pronóstico del ICA, donde el valor R^2 es 0.99 y el modelo indica que el 99.99% de variabilidad del ICA se explica por las once variables independientes utilizadas en el modelo. El análisis estadístico multivariado gestiona numerosos y complejos datos para disminuir el número variable y proporcionar datos completos. En este estudio, los métodos estadísticos multivariantes se expresaron como bastante exitosos en la determinación de la calidad del agua. Numerosas variables fueron explicadas con un menor número de variables. Al mismo tiempo, estas variables proporcionaron información más útil sobre la calidad del agua subterránea del área de Maturín. Por lo tanto, esta técnica se puede aplicar con mucho éxito para la planificación y el seguimiento de la calidad en el futuro.

REFERENCIAS

Ahmed, S., Hussain, M. y Abderrahman, W., 2005. Using multivariate factor analysis to assess surface/logged water quality and source of contamination at a large irrigation project at Al-Fadhli, EasternProvince, Saudi Arabia. En: *Bulletin of Engineering Geology and the Environment,* 64, pp.232–315. DOI: 10.1007/s10064-005-0277-6.

Arumugam, K. y Elangovan, K., 2009. Hydrochemical characteristics and groundwater quality assessment in Tirupur Region Coimbatore District, Tamil Nadu, India. En: *Environ Geol,* 58, pp.1509–1520. https://doi.org/10.1007/s00254-008-1652

Azaza, F.H., Ketata, M., Bouhlila, R., Gueddari, M. y Riberio, L., 2011. Hydrogeochemical characteristics and assessment of drinking water quality in Zeuss-Koutine aquifer, southeastern Tunisia. En: *Environ Monit Assess,* 174, pp.283–298. https://doi.org/10.1007/s10661-010-1457-9

Azkue, M. y Soto, E., 2004. Régimen de precipitación en Montalbán, Estado Carabobo, Venezuela. 2004. En: *Agronomía Tropical,* 54(4), pp.359-370.

Beita-Sandí, W. y Barahona-Palomo, M., 2010. Físico-química de las aguas superficiales de la Cuenca del río Rincón, Península de Osa, Costa Rica. En: *Cuadernos de Investigación UNED*, 2(2), pp.157-179. DOI: 10.22458/urj.v2i2.156.

Belkhiri, L., Boudoukha, A., Mouni, L. y Baouz, T., 2010. Multivariate statistical characterization of groundwater quality in Ain Azel plain, Algeria. En: *African Journal of Environmental Science and Technology,* 4(8), pp.526-534. DOI: 10.4314/ajest.v4i8.71307.

Benítez, S., 2004. *Determinación de la carga orgánica impuesta al río Manzanares en el periodo febrero-junio 2004*. Cumaná: IUT. (Trabajo de Grado).

Boyacioglu, H., 2006. Surface water quality assessment using factor analysis. En: *Water S.A.*, 32(3), pp.389–393. DOI: 10.4314/wsa.v32i3.5264.

Calla, L.H. y Cabrera, C.C., 2010. Calidad del agua en la cuenca del río Rímac, sector de San Mateo, afectado por las actividades mineras. En: *Revista del Instituto de Investigaciones de la Facultad de Geología, Minas, Metalurgia y Ciencias Geográficas,* 13(25), pp.87-94. https://doi.org/10.15381/iigeo.v13i25.399

Cai, W.J., Guo, X.H., Chen, C.T.A., Dai, M.H., Zhang,L.J., Zhai, W.D., Lohrenz, S.E. y Yin, K., 2007. A comparative overview of weathering intensity and HCO3-flux in the world's major rivers with emphasis on the Changjiang, Huanghe, Zhujiang (Pearl) and Mississippi Rivers. En: *Cont. Shel. Res.*, 28(12), pp.1538-1549.DOI: 10.1016/j.csr.2007.10.014.

Cansu Filik Iscen, Özgür, Emiroglu, Semra Ilhan, Naime, Arslan, Veysel, Yilmaz y Seyhan, Ahiska, 2008. Application of multivariate statistical techniques in the assessment of surface water quality in Uluabat Lake, Turkey. En: *Environ Monit Assess*, 144(1-3), pp.269–276. DOI: 10.1007/s10661-007-9989-3.

Cardona, A., Carrillo-Rivera, J., Huizar-Alvarez, R. y Graniel-Castro, E., 2004. Salinization in coastal aquifer sofarid zones: anexamplefrom Santo Domingo, Baja California Sur, México. En: *Environmental Geology,* 45, pp.350-366. https://doi.org/10.1007/s00254-003-0874-2

Chávez-Alcántar, A.L., Velázquez-Machuca, M., Pimentel-Equihua, J.L., Venegas-González, J., Montañez-Soto, J.L. y Vázquez-Gálvez, G., 2011. Hidroquimica de las aguas superficiales de la Ciénega de Chapala e índice de calidad de agua. *En: Terra Latinoamericana*, 29(1), pp.13.

Díaz, A., Rincón, N., Marín, J., Behling, E., Chacin E. y Fernández N., 2005. Degradación de fenoles totales durante el tratamiento biológico de aguas de producción petroleras. En: *Ciencia*, 13(3), pp.281-291.

Duran, Liliana, 2011. Las políticas hídricas en Venezuela en la gestión del agua subterránea. En: *Revista Voces: Tecnología y Pensamiento,* 5(1-2), pp.93-106.

Gil, J., 2003. *Recursos hidrogeológicos* [En línea].[s.l.]: [s.n.]. [Consulta 11 de mayo de 2015]. Disponible en: http://pdfhumanidades.com/sites/default/files/apuntes/recursos.pdf

Gil Marín, José Alexander, Lozada García, Beatriz I., López Pérez, Nadiezhda, Márquez, Luis y Del Valle Salazar Gil, Marielin, 2012. Evidencias de la variabilidad y cambios climáticos en Maturín, estado Monagas, Venezuela. En: *Revista Científica UDO Agrícola,* 12(2), pp.389-399.

Gil Marín, José Alexander, Vizcaíno, G.C. y Montaño-Mata, Nelson. P.D., 2018. Evaluación de la calidad del agua superficial utilizando el índice de calidad del agua (ICA). Caso de estudio: cuenca del Río Guarapiche, Monagas, Venezuela. En: *Anales Científicos,* 79(1), pp.111-119.

Gordon, G.E., 1988. Receptor models. En: *Environ. Sci. Technol.*, 22, pp.1132-1142.

Grivas, G. y Chaloulakou, A., 2006, Artificial neural net-work models for prediction of PM10 hourly concentrations, in the Greater area of Athens, Greece. En: *Atmos. Environ.*, 40(7), pp.1216-1229. https://doi.org/10.1016/j.atmosenv.2005.10.036

Guo, H. y Wang, Y., 2004. Hydrogeochemical processesing shallow quaternary aquifers from the northern part of the Datong Basin, China. En: *AppliedGeochemistry,* 19, pp.19–27. https://doi.org/10.1016/S0883-2927(03)00128-8

Gupta, P. y Roy, S., 2012. Evaluation of spatial and seasonal variations in groundwater quality at Kolar Gold Fields, India. En: *American Journal of Environmental Engineering,* 2(2), pp.19-30. DOI: 10.5923/j.ajee.20120202.04.

Helena, B., Pardo, R., Vega, M., Barrado, E., Fernández, J.M. y Fernández, J., 2000. Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principle component analysis. En: *Wat. Res.*, 34(3), pp.807-816. https://doi.org/10.1016/S0043-1354(99)00225-0

Holland, M. y Witthuser, K.T., 2009. Geochemical characterization of karst groundwater in the Cradle of Humankind World Heritage Site, South Africa. En: *EnvironmentalGeology,* 57, pp.513-524. https://doi.org/10.1007/s00254-008-1320-2

Instituto Nacional de Estadística, 2011. XIV *Censo nacional de población y vivienda resultados por entidad federal y municipio del Estado Monagas.* Caracas: INE. 97 pp.

Jin, Z., Chen, Y., Wang, F. y Ogura, N., 2004. Detection of nitrate sources in urban groundwater by isotopic and chemical indicators, Hangzhou City, China. En: *Environmental Geology,* 45(7), pp.1017-1024. DOI: 10.1007/s00254-004-0962-y.

Kaiser, H.F., 1960. The application of electronic computers to factor analysis. En: *Educational and Psychological Measurement,* 20(1), pp.141-151. https://doi.org/10.1177/001316446002000116

Kanade, S.B. y Gaikwad, V.B., 2011. A multivariate statistical analysis of bore well chemistry data - Nashik and Niphad Taluka of Maharashtra, India. En: *Universal Journal of Environmental Research and Technology,* 1(2), pp.193-202.

Karim, Zare y Abdolrahman, Rasekh, 2011. Diagnostic measures for linear mixed measurement error models. En: *Statistics and Operations Research Transactions,* 35(2), pp.125-144.

Krishna, A.K., Satyanarayanan, M. y Govil, P.K., 2009. Assessment of heavy metal pollution in water using multivariate statistical techniques in an industrial area: A case study from Patancheru, Medak District andhraPradesh,India. En: *J. Hazardous Materials,* 167(1-3), pp.366-373. https://doi.org/10.1016/j.jhazmat.2008.12.131

Liu, C.W., K.H. Lin y Kuo, Y.M., 2003. Application of factor analysis in the assessment of groundwater quality in a black foot disease area in Taiwan. En: *The Science of the Total Environ.*, 313(1-3), pp.77-89. https://doi.org/10.1016/S0048-9697(02)00683-6

Memet, Varol y Bülent, Sen, 2009. Assessment of surface water quality using multivariate statistical techniques: a case study of Behrimaz Stream, Turkey. En: *EnvironMonitAssess,* 159, pp.543–553. https://doi.org/10.1007/s10661-008-0650-6

Ministerio del Poder Popular del Ambiente, 2009. *Plan nacional de gestión integral de las aguas.* Caracas: MPPA. 200p.

Mishra, P.C., Behera, P.C. y Patel, R.K., 2005. Contamination of water due to major industries and open refuse dumping in the steel city of Orissa—a case study. En: *J Environ Sci Eng*, 47(2), pp.141–154.

Mohd, Fahmi, Mohd, Nasir, Mohd, Saiful, Samsudin, Isahak, Mohamad, Mohammad, Roshide, Amir, Awaluddin, Muhd, Ariffin, Mansor, Hafizan, Juahir y Norlafifah, Ramli, 2011. River water quality modeling using combined principle component analysis (PCA) and multiple linear regressions (MLR): a case study at Klang River, Malaysia. En: *World Applied Sciences Journal,* 14, (Exploring Pathways to Sustainable Living in Malaysia: Solving the Current Environmental Issues), pp.73-82.

Montgomery, D.C., Peck, E.A. y Vining, G.G., 2003. *Introduction to linear regression analysis.* New York: John Wiley & Sons, Inc.

Nádaská, G., Lesný, J. y Michalík, I., 2010. Environmental aspect of manganese chemistry. En: *Hungarian Journal of Sciences,* ENV-100702-A, pp.1-16.

Nieves, K., Riera, A., Vega, C. y Manganiello, L., 2005. Desarrollo de sistemas de manipulación de muestras para la determinación de fenol y pentaclorofenol en agua potable. En: *Revista Ingeniería UC*, 22(2), pp.88.

Ntengwe, F.W., 2006. Pollutant loads and water quality in streams of heavily populated and industrialised towns. En: *Physics and Chemistry of the Earth,* 31, pp.832–839. https://doi.org/10.1016/j.pce.2006.08.025

Organización Mundial de la Salud, 2006. *Guías para la calidad del agua potable. Primer apéndice a la tercera edición. Volumen 1. Recomendaciones.* Ginebra: OMS. 393p.

Ouyang, Y., Nkedi-Kizza, P., Wu, Q.T., Shinde, D. y Huang, C.H., 2006. Assessment of seasonal variations in surface water quality. En: *Water Research,* 40, pp.3800–3810. https://doi.org/10.1016/j.watres.2006.08.030

Pazand, K., Hezarkhani, A., Ghanbari, Y. y Aghavali, N., 2012. Geochemical and quality assessment of groundwater of Marand Basin, East Azarbaijan Province, Northwestern Iran. En: *EnvironEarthSci*, 67, pp.1131–1143.DOI: 10.1007/s12665-012-1557-7.

Pius, A., Jerome, C. y Nagaraja, S., 2012. Evaluation of groundwater quality in and around Peenya industrial area of Bangalore, South India using GIS techniques. En: *EnvironMonitAssess*, 184(7), pp.4067–4077. doi: 10.1007/s10661-011-2244-y.

Porras, J., Nieto López-Guerrero, Pedro, Álvarez-Fernández, Ceferino, Fernández Uría, Antonio y Gimeno, María Victoria, 1985. *Calidad y contaminación de las aguas subterráneas en España.* Madrid: Instituto Geológico y Minero de España.

Saima, S., Osman, R., Spian, D.R.S.A, Jaafar, M.Z., Juahir, H., Abdullah, M.P. y Ghani, F.A., 2009. Chemometric approach to validating faecal sterols as

source tracer for faecal contamination in water. En: Water Res., 43, pp.5023-5030. DOI: 10.1016/j.watres.2009.08.052.

Sarbu, C. y Pop, H.F., 2005. Principal component analysis versus fuzzy principal component analysis a case study: the quality of danube water (1985–1996). En: *Talanta*, 65, pp.1215-1220. doi: 10.1016/j.talanta.2004.08.047

Shrestha, S. y Kazama, F., 2007. Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin; Japan. En: *Environmental Modelling & Software*, 22, pp.464–475.https://doi.org/10.1016/j.envsoft.2006.02.001

Simge, Varol y Aysen, Davraz, 2014. Evaluation of the groundwater quality with WQI (Water Quality Index) and multivariate analysis: a case study of the Tefenniplain(Burdur/Turkey). En: *Environmental Earth Sciences*, 73(4), pp.1725–1744. https://doi.org/10.1007/s12665-014-3531-z

Zare Garizi, A., Sheikh, V. y Sadoddin, A., 2011. Assessment of seasonal variations of chemical characteristics in surface water using multivariate statistical methods. En: *Int. J. Environ. Sci. Tech.,* 8(3), pp.581-592.https://doi.org/10.1007/BF03326244

Venezuela. Decreto 883.1995, de 11 de octubre de 1995. *Gaceta Oficial Extraordinaria*, 18 de diciembre de 1995, No 5.021.

Villa, A.M.A., 2011. *Evaluación de la calidad del agua en la subcuenca del río Yacuambi. Propuestas de tratamiento y control de la contaminación.* Cádiz: Departamento de Química Analítica. (Tesis de Máster).

Wunderlin, D.A., Diaz, M.P., Ame, M.V., Pesce, S.F., Huedy, A.C. y Bistoni, M.A., 2001. Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquia river basin (Cordoba, Argentina). En: *Water Res.*, 35, pp.2881-2894. https://doi.org/10.1016/S0043-1354(00)00592-3