



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

River Water Quality Modeling Using Combined Principle Component Analysis (PCA) and Multiple Linear Regressions (MLR). Case study: Guarapiche River Basin, Monagas, Venezuela

José Alexander Gil-Marín^{1*}; Aliser González C.¹

¹ Departamento de Ingeniería Agrícola, Escuela de Ingeniería Agronómica (EIA), Universidad de Oriente (UDO). Avenida Universidad, *Campus* Los Guaritos, Maturín, 6201, estado Monagas, Venezuela. E-mail: jalexgil2005@hotmail.com.

Recepción: 21/04/2019; Aceptación: 15/05/2020

Resumen

Se estudió un conjunto de datos colectados durante el año 2011 en el río Guarapiche, Maturín, Monagas, Venezuela, en un intento para 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 una herramienta avanzada para el modelado y pronóstico de la calidad de las aguas superficiales. 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 siete componentes principales responsables de la estructura de datos provisionalmente nombrados como componente de disolución, erosión del suelo, variabilidad fisicoquímica, microbiológico o residuos fecales, nutriente o residuos industriales o municipales, químico y antropogénico tóxico, que explican el 83% de la variación total para todos los conjuntos de datos. Mientras tanto, la contaminación doméstica urbana se considera el mayor contribuyente de contaminación al río Guarapiche. Por lo tanto, el avance de modelo receptor se aplicó con el fin de identificar las principales fuentes de contaminantes en el río Guarapiche. El resultado mostró que el uso de ACP como entradas mejoró 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 ICA

Forma de citar el artículo: Gil-Marín, J.; González, A. 2020. Modelo de calidad del agua de un río mediante el uso combinado de análisis de componentes principales (ACP) y regresiones lineales múltiples (RLM). Caso de estudio: Cuenca del río Guarapiche, Maturín, Monagas, Venezuela. *Anales Científicos* 81 (1):152- 172 (2020). <http://dx.doi.org/10.21704/ac.v81i1.1586>

DOI: <http://dx.doi.org/10.21704/ac.v81i1.1586>

Autor de correspondencia (*): José Alexander Gil-Marín. Email: jalexgil2005@hotmail.com.

© Universidad Nacional Agraria La Molina, Lima, Perú.

es explicada por las seis variables independientes utilizadas en el modelo. Esta evaluación presenta la importancia y las ventajas que plantea el análisis estadístico multivariante, de grandes y complejas bases de datos, con el fin de obtener información mejorada sobre la calidad del agua y luego ayudar a reducir el tiempo de muestreo y el costo de los reactivos utilizados antes de los análisis.

Palabras clave: análisis de componentes principales; calidad del agua; regresión lineal múltiple.

Abstract

A set of data collected during 2011 in the Guarapiche River, Maturín, Monagas, Venezuela, was studied in an attempt to assess and determine the contributions of sources affecting the water quality. An accurate linear multiple regression (RML) technique was used as an advanced tool for modeling and forecasting the quality of surface water. Likewise, the principal components analysis (PCA) was used to simplify and understand the complex relationship between water and quality parameters. Seven main components were found responsible for the structure of data provisionally named as a component of dissolution, soil erosion, physicochemical, microbiological or fecal variability, nutrient or industrial or municipal waste, chemical and anthropogenic toxic, which explain 83% of the variation total for all data sets. Meanwhile, urban domestic pollution accounted as the highest pollution contributor to the Guarapiche River. Thus, the advancement of receptor model was applied in order to identify the major sources of pollutant at Guarapiche River. Result showed that the use of PCA as inputs improved the MLR model prediction by reducing their complexity and eliminating data colinearity where R^2 value in this study is 0,99 and the model indicates that 99% variability of WQI explained by the five independent variables used in the model. This assessment presents the importance and advantages poses by multivariate statistical analysis of large and complex databases in order to get improved information about the water quality and then helps to reduce the sampling time and cost for reagent used prior to analyses.

Keywords: principle component analysis; water quality; multiple linear regressions.

1. Introducción

La extensa industrialización, la urbanización descontrolada y el rápido desarrollo económico alrededor de muchas ciudades, especialmente Maturín, han incrementado los niveles de contaminación al medio ambiente. Particularmente el río Guarapiche sufre mucho, ya que fluye por el estado de Monagas y Maturín. La contaminación proviene principalmente de infraestructuras municipales que plantean numerosas preocupaciones ambientales a lo largo del río Guarapiche, mientras tanto, las actividades humanas intensivas han resultado en sustanciales deformaciones

hidrológicas. Según [Tong y Chen \(2002\)](#), los tipos de uso de la tierra se correlacionan con la mayoría de las características de calidad del agua. La escorrentía de la superficie terrestre transporta los residuos de la tierra al sistema fluvial, lo que se conoce como contaminación de fuentes no puntuales. La expansión de las áreas urbanas en cualquier cuenca hidrográfica afecta el medio ambiente por el aumento de la carga contaminante en el sistema fluvial y cambios en la calidad del agua superficial. Sin embargo, una modificación sustancial en la escorrentía y calidad del agua se encuentra que es aportada por el desarrollo urbano ([Mohd et al., 2011](#)).

Costa *et al.* (2003), encontraron que la conversión de la vegetación interrumpirá el ciclo hidrológico de una cuenca de drenaje al alterar el equilibrio entre la precipitación y la evaporación del área. Los ríos urbanos también están contaminados con la descarga de aguas residuales de las plantas de tratamiento y el desbordando de las aguas residuales causadas por la lluvia (Nix y Merry, 1990), lo que origina una contaminación fecal, que es una preocupación importante en el río cerca del área de la ciudad donde los residentes locales utilizan el agua superficial. No obstante, los desechos industriales y domésticos que se descargan directamente o a través de fugas en los sistemas de alcantarillado fluirán hacia las fuentes de agua, causando una contaminación excesiva de las aguas superficiales y subterráneas (Mohd *et al.*, 2011). Los programas de monitoreo con muestreos frecuentes de agua y la determinación de parámetros fisicoquímicos pueden proporcionar representativamente el estado de la calidad del agua superficial. Desde 2010, el Departamento de Ingeniería Agrícola, de la Universidad de Oriente (UDO) de Venezuela, ha realizado acciones de monitoreo que resultan en grandes matrices de datos y requiere herramientas estadísticas avanzadas, como inteligencia multivariada y artificial para una ilustración excepcional de los datos. Inicialmente, el programa cubrió toda la cuenca fluvial del río Guarapiche, que involucraba principalmente el muestreo manual y las mediciones in situ y en sus laboratorios de las muestras de agua del río. En la actualidad el programa se enfoca a otras cuencas y a las aguas subterráneas de la zona. A pesar de que el programa de monitoreo regular proporciona conjuntos de datos ambientales complejos, sin embargo, todavía hay una falta de aplicación de los métodos estadísticos multivariantes para intentar extraer toda la información posible de los conjuntos de datos de calidad del agua

obtenidos del río Guarapiche, por lo que no se puede determinar la fuente principal de influencia en la clase de río. Por lo tanto, las técnicas de estadística multivariante y el análisis de datos exploratorios son las herramientas adecuadas para una reducción de datos sobresaliente y la interpretación de las mediciones físicas, químicas y biológicas de múltiples componentes. En general, la calidad del agua se refiere a las características del agua, ya sea física, química o biológica. Sobre la base de estos datos, se desarrolló en Venezuela, un índice de calidad del agua (ICA) para evaluar el estado de calidad del agua y la clasificación de sus ríos. El ICA proporciona una predicción de cambios y tendencias en la calidad del agua considerando múltiples parámetros. Aunque, a menudo no son convincentes para la interpretación grandes conjuntos de datos con muchas variables; un enfoque ambiental es necesario para comprender las variaciones de los datos. Además, las herramientas matemáticas avanzadas se pueden utilizar como un medio para advertir a las personas, así como a las agencias ambientales relacionadas con la protección y mantenimiento, el estado actual de contaminación de una fuente. Este estudio intenta predecir los valores ICA utilizando el modelo RLM a partir de los factores varimax generados por ACP, implementando un entendimiento absoluto sobre qué tan bueno es el modelo, que delinee o reduzca el mejor modelo para la predicción del ICA en el río Guarapiche.

2. Materiales y métodos

Área de estudio: el río Guarapiche es uno de los ríos más importantes del estado Monagas, Venezuela, donde fluye directamente a través de las poblaciones agrícolas de Miraflores, Tristé, San Félix de Caicara, Merecure, Jusepín, Candelaria, San Vicente y Maturín (Capital del Estado); que son representativas del nivel socio

productivo del estado (MARNR, 1980); alto crecimiento demográfico y por la intervención de la actividad petrolera. El río Guarapiche presenta una longitud de aproximadamente 238,04 km y forma parte de la Región Hidrográfica Oriental de Venezuela (MINAMB, 2006). A lo largo de su cauce principal se ubicaron las siete estaciones de muestreo (Figura 1). En el *Tabla 1* se observan las coordenadas geográficas y UTM de las estaciones de muestreo.

El río Guarapiche enfrenta enormes amenazas de varias fuentes a través de los últimos años debido a diversos tipos de actividades industriales, urbanas, agrícolas y petroleras. De hecho, un río que fluye a través de un área densamente poblada se asocia comúnmente con fuentes puntuales y no puntuales, por lo que es difícil rastrear la carga de contaminantes en el río. Por lo tanto, las convocatorias de investigación fortalecidas con herramientas matemáticas avanzadas están altamente consolidadas.



Figura 1. Ubicación geográfica de la Cuenca del Río Guarapiche

Tabla 1. Coordenadas Geográficas de las Estaciones de Muestreo en el Río Guarapiche, Estado Monagas, Venezuela. Período Febrero –Julio 2011

Estación	Localidad	Altitud (m)	Coordenadas Geográficas	
1	Miraflores	524	10°10'46"	63°42'0"
2	Tristé	471	10°09'22"	63°42'27"
3	San Félix de Caicara	235	09°57'35"	63°39'31"
4	Merecure	110	09°46'18"	63°33'54"
5	Jusepín	80	09°10'47"	63°28'33"
6	Bajo Guarapiche	33	09°45'34"	63°11'0"
7	Palmonagas	12	09°52'08"	63°03'57"

Datos extraídos del dispositivo GPS modelo Magellan, 2011. Datum Regven.

Conjunto de datos: En este estudio, veinte parámetros de calidad del agua fueron observados a lo largo de las estaciones de monitoreo del río Guarapiche. Las estaciones de muestreo se ubicaron en puntos cercanos a las comunidades que se abastecen del río. Un total de 1118 observaciones se utilizaron para la distribución de fuentes y técnicas de modelado. El ICA se desarrolló para evaluar el estado de la calidad del agua y la clasificación de la clase del río en base a los datos de calidad del agua. Los veinte parámetros de calidad del agua consistieron en oxígeno disuelto (OD), 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 seis variables de calidad del agua seleccionadas, como coliformes fecales (CF), temperatura, conductividad eléctrica (CE), pH, sodio (Na) 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 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 11.5.

Análisis de componentes principales (ACP):

El análisis de componentes principales presenta los parámetros más significativos, que describen 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.*, 2003). ACP son sensibles a los valores atípicos, los datos que faltan y la mala correlación lineal entre las variables debido a la insuficiencia de las variables asignadas (Sarbu y Pop, 2005). Por lo tanto, un tratamiento previo detallado al conjunto de datos debe realizarse 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:

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

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 (FA) 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 FA, el concepto básico se expresa en la ec. (2)

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

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, n es el número de muestra, j es el número de la variable y r es el número total de factores. Principalmente, los dos métodos que son ACP y FA se expresan en ecuaciones similares, sin embargo, la diferencia está en CP donde se expresa como una combinación lineal de variables medidas. Mientras que FA, la variable medida se expresa como una combinación de factores y la ecuación contiene el término residual y , por lo tanto, una FV puede incluir variables latentes no observables, hipotéticas (Helena *et al.*, 2000, Wunderlin *et al.*, 2003 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 entre sí. 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 (Saime *et al.*, 2009). Por lo tanto, el componente principal a menudo presenta información sobre los parámetros confiables más significativos, que definen todo el conjunto de datos que ofrece una reducción de datos con una pérdida mínima de información original (Krishna *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 (X_1, X_2, \dots, X_k). El análisis de regresión lineal múltiple, a diferencia del simple, se aproxima más a situaciones de análisis real puesto que los fenómenos, hechos y procesos naturales, por definición, son complejos 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 que sigue:

$$Y_u = \beta_0 + \beta_1 X_{u1} + \beta_2 X_{u2} + \dots + \beta_k X_{uk} + \varepsilon_u \quad (3)$$

Donde Y_u denota la respuesta observada en el intento u o variable dependiente;

X_{ui} representa el nivel del factor i en el intento u o variables independientes;

las betas son parámetros desconocidos y

ε_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, “paso a paso”. La finalidad perseguida es buscar de entre todas las posibles variables explicativas (parámetros físicos, químicos y biológicos) aquellas que más y mejor expliquen a la variable dependiente (ICA) sin que ninguna de ellas sea combinación lineal de las restantes. Este procedimiento implica que: (1) en cada paso solo se introduce aquella variable que cumple unos criterios de entrada; (2) una vez introducida, en cada paso se valora si alguna de las variables cumplen criterios de salida; y (3), en cada paso se valora la bondad de ajuste de los datos al modelo de regresión lineal y se calculan los parámetros del modelo verificado en dicho paso. 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 (garantizar que una variable seleccionada no es redundante). El análisis comenzó con 20 variables explicativas y termino recomendado solo 6 variables.

3. Resultados y discusiones

En la [Tabla 2](#), se muestran las estadísticas descriptivas básicas, relacionadas con la calidad del agua, donde los valores de calidad del agua de CF, Dureza, Cl⁻ y CE, parecen ser de distribución estándar muy alta. Estas variables cambian considerablemente en el río Guarapiche con respecto a su valor promedio, pero la dureza, Cl⁻ y CE no influyen en la calidad de sus aguas porque no sobrepasan los valores límites de las normas venezolanas ni internacionales.

Tabla 2. Estadísticos descriptivos básicos de los parámetros físicos, químicos y biológicos estudiados en el río Guarapiche. Período febrero - julio, 2011

Variable	Unidad	Rango (N° Estación)		Media	LMP		Desviación Estándar
		Mínimo	Máximo		Decreto 883	OMS	
Temperatura	°C	21,1(1)	29,6(7)	26,31			2,76
pH		7,5 (7)	8,41(1)	8,02	6.0-8,5		0,27
Dureza	mg/l	270,29(6)	445,08(2,4)	344,14	500		47,34
CE	S/cm	218 (7)	380(2,5)	312,12		3000	36,59
Fe	mg/l	0(1)	255(7)	0,38	1,0	0,3	0,51
Mn	mg/l	0(1,2)	4,6(2,3)	0,32	0,1	0,5	1,07
Na	mg/l	0(1,2,3,4)	25,1(7)	5,58	200		8,60
K	mg/l	20(1,2)	65	23,84(4)		50	7,89
NO ₂	mg/l	0(1,2)	14(2,4)	0,83	5	3	2,40
NO ₃	mg/l	0(1,2)	13,2(4)	0,87	5	50	2,48
SO ₄	mg/l	0(1)	40(4,5)	23,51	400	250	8,71
Cl	mg/l	0(1)	77(6,7)	5,27	600	250	12,33
OD	mg/l	2,45(7)	9,9(1)	4,01	>3		3,09
CF	NMP/ 100 ml	0(1)	11000(6)	496,21			1720,58

LMP=límite máximo permitido por la norma venezolana, Decreto 883(1995)* y Organización Mundial de la Salud (OMS). ()Entre paréntesis Estación de Muestreo.

Análisis de Componentes Principales (ACP)

Antes de ejecutar los Análisis de Componentes Principales (ACP), se aplicó a todo el conjunto de datos, la prueba de Kolmogorov Smirnov (K-S), para probar si los datos se ajustaron a una distribución normal. La prueba K-S nos indicó que todas las variables siguen una distribución normal de registro. Por otra parte, para probar la idoneidad de los datos para el análisis de componentes principales se aplicaron las pruebas de Kaiser-Mayer-Olkin (KMO) y la prueba de esfericidad de Bartlett. La prueba de KMO mide la adecuación del muestreo y señala la fracción de variables que tienen varianza común. Los valores de KMO superiores a 0,5 son considerados como satisfactorio para PCA. En el presente estudio, este valor de 0,588 (Tabla 3), indica que el conjunto de datos está en forma para el análisis de componentes principales. La prueba de esfericidad de Bartlett 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 indica 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 siete 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 *et al.*, 2005). Existen reglas para determinar el número de factores adecuado a conservar, por ejemplo, el que se conoce como criterio de Kaiser (Kaiser, 1960), el cual indica que hay que conservar los componentes principales cuyos valores propios son 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; Belkhir *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 (Belkhir *et al.*, 2010). Los autovalores indican la importancia del factor, valores mayores de uno son considerados significativos. Los resultados del ACP señalan que los primeros siete valores de los autovalores son mayores de uno y por lo tanto son considerados como significativos. En la Figura 2 se puede observar que mayor parte de la variación en los datos originales se explica por los primeros siete Factores.

Tabla 3. Prueba de KMO y Bartlett.

Medida de adecuación muestral de Kaiser-Meyer-Olkin.	
0,588	
Prueba de esfericidad de Bartlett	
Chi-cuadrado aproximado	408,929
Gl	105
Sig.	0,00

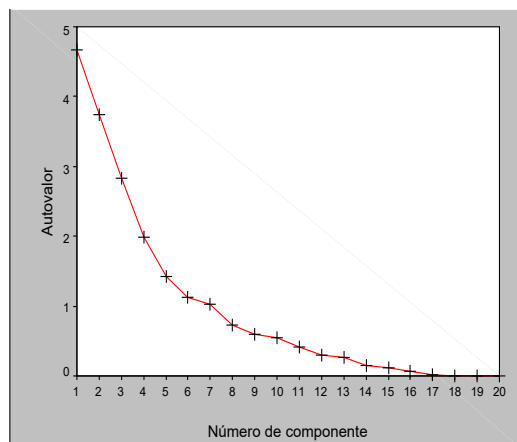


Figura 2. Gráfico de Sedimentación Río Guarapiche

Los valores propios para diferentes factores, varianza porcentual, variación porcentual acumulada y componentes cargados (sin rotar y Varimax rotadas), son presentados en la [Tabla 4](#). Este análisis condujo a la explicación del 83,99% de las varianzas en los datos. Las variables de carga para los siete componentes del ACP del conjunto de datos se dan en el [Tabla 5](#). Liu *et al.* (2003), clasifica a las cargas factoriales como “fuerte”, “moderado” y “débil” correspondientes a valores de carga absolutos de $> 0,75$; $0,75-0,50$ y $0,50-0,30$, respectivamente.

El primer componente (CPI), denominado componente de disolución, explica el 23,36 % de la varianza total, presentando una fuerte correlación positiva con las variables sodio (Na^+) y temperatura; y negativa para pH. Por otra parte, hay una moderada correlación positiva con Hierro (Fe^+) y negativa con el bicarbonato (HCO_3^-). Este componente representa la capacidad de disolución de los elementos químicos Fe^+ , Na^+ y HCO_3^- , ya que un incremento de la temperatura y disminución del pH favorece la disolución de Fe^+ y Na^+ , pero no la del HCO_3^- .

Por otro lado, el hierro y su comportamiento químico en las aguas

naturales es originada principalmente por la reacción oxidación-reducción, que se da por medio de la interacción con compuestos orgánicos e inorgánicos del medio; así también es gracias a este tipo de reacciones que se reduce el contenido de este elemento en las aguas naturales (Jenkins, 2009). El ión Fe^{+2} se encuentra mayoritariamente dentro de los causes, sin embargo una acumulación excesiva de este metal puede llegar a ocasionar cambios en la coloración y sabor del agua, haciendo que no sea gusto del consumidor (Postawa, 2013).

En cuanto al sodio, las concentraciones de este ion en aguas naturales superficiales varían de manera considerable dependiendo de las condiciones geológicas locales y descargas de aguas residuales. El sodio es un factor importante en el agua, tanto para consumo como para riego, y de acuerdo a sus concentraciones puede llegar a afectar funciones renales, cardíacas y circulatorias (Beita-Sandí y Barahona-Palomo, 2010). La presencia de bicarbonato en el agua de origen profundo, y su elevada concentración, se debe a la disolución de minerales de carbonato de calcio, calcita, principalmente (Appelo y Postma, 1994), o de la reducción de sulfato (Hounslow, 1995).

Tabla 4. Análisis de Componentes principales por Método de extracción.

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 varianza	% acumulado
1	4,672	23,36	23,36	4,672	23,36	23,36	3,832	19,161	19,161
2	3,734	18,671	42,031	3,734	18,671	42,031	2,879	14,393	33,554
3	2,831	14,154	56,184	2,831	14,154	56,184	2,657	13,285	46,84
4	1,983	9,916	66,1	1,983	9,916	66,1	2,648	13,239	60,078
5	1,421	7,104	73,204	1,421	7,104	73,204	1,81	9,048	69,127
6	1,127	5,634	78,838	1,127	5,634	78,838	1,779	8,895	78,021
7	1,032	5,159	83,997	1,032	5,159	83,997	1,195	5,975	83,997

Tabla 5. Matriz de componentes rotados

Variables	Componente						
	1	2	3	4	5	6	7
Fe	0,636	0,208	-0,069	0,255	-0,054	-0,425	-0,022
Ca	-0,155	0,887	0,092	-0,336	0,013	-0,039	0,125
Mg	0,206	0,235	-0,227	0,759	0,248	0,291	0,058
Mn	-0,363	-0,086	-0,343	0,141	-0,062	0,192	0,587
Na	0,801	-0,136	-0,12	0,142	-0,092	-0,203	0,227
K	0,074	0,267	0,043	0,07	0,774	0,209	0,071
NO2	-0,183	-0,122	0,02	-0,08	0,779	-0,153	0,062
NO3	0,06	0,341	0,051	0,081	0,663	0,48	-0,175
SO4	0,261	0,029	0,203	-0,053	0,117	0,856	0,03
HCO3	-0,742	-0,133	0,004	-0,161	0,012	-0,245	0,15
CO3	0,433	0,811	-0,094	0,128	0,178	0,022	-0,141
OD	-0,087	0,174	-0,806	0,12	-0,041	0,24	0,146
CF	0,132	-0,166	-0,089	0,931	-0,061	-0,068	0,086
pH	-0,927	0,025	0,001	-0,11	0,043	-0,115	-0,102
Tempe	0,84	-0,041	0,086	0,049	0,055	0,336	-0,027
CE	-0,071	0,083	0,934	-0,074	0,016	0,239	-0,03
SDT	-0,081	0,086	0,927	-0,12	0,031	0,239	-0,021
Dureza(mg/l)	-0,071	0,963	0	-0,028	0,111	0,078	0,145
Cl	0,315	0,247	-0,027	0,098	0,115	-0,101	0,795
CT	0,183	-0,16	-0,085	0,93	-0,056	-0,156	0,059

En cuanto al pH, este es controlado por el balance entre los iones bicarbonato y carbonato, y el dióxido de carbono. En la mayoría de las aguas naturales, el pH es controlado principalmente por el sistema carbonato – bicarbonato (Stumm y Morgan 1996); y estas, tienen cierta capacidad de amortiguar (o resistir) los cambios de pH. En los ríos esta capacidad se debe sobre todo, a la presencia de iones bicarbonato (HCO_3^-), aportados al sistema por la disolución de rocas y suelos (Drever, 1988; Stumm y Morgan, 1996). En aguas de ríos no contaminados los valores de pH varían y su determinación es importante, ya que tiene una gran influencia en muchos sistemas biológicos. Valores superiores o inferiores a este ámbito

producen limitaciones en el desarrollo y fisiología de los organismos acuáticos, en general y en la biota de los humedales en particular. Dichos cambios del pH en el agua pueden ocurrir por diversas razones, una de ellas la constituyen las prácticas agrícolas que conducen a la lixiviación de aniones hacia el subsuelo, más allá de la zona de las raíces. Típicamente los nitratos, aunque también bicarbonato y aniones orgánicos, son lixiviados en asociación con cationes (calcio, magnesio y potasio), dejando un exceso de iones H^+ en la superficie del suelo, que luego por escurrimiento son transportados al flujo de agua (CCME, 1999; Beita-Sandí y Barahona-Palomo, 2010).

Por último se tiene que 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, ya que ésta tiene gran efecto en la solubilidad de las sales y las reacciones biológicas de los seres vivos, también acelera la putrefacción y la demanda de oxígeno (Benítez, 2004). Esta 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).

El segundo componente (CP2), representa el 18,67 % de la varianza total, y se correlaciona de manera fuertemente positiva con las variables Ca, CO₃ y Dureza. Este componente principal representa una variable Mineral de la calidad del agua. Donde la dureza total revela la influencia de la lixiviación del suelo y erosión. Se puede apreciar que la disolución de los suelos de piedra caliza y yeso en las cuencas fluviales es una de las principales fuentes de minerales en el agua (Ismail *et al.*, 2014). De esta manera, la dureza del agua varía considerablemente de un sitio a otro, y es causada por la presencia de cationes metálicos multivalentes. Por otra parte, la misma es evaluada como una fuente de variabilidad fisicoquímica. El clima y las características geológicas de la región juegan un papel muy importante en la configuración de la dureza. Ésta es un reflejo de la naturaleza de las formaciones geológicas con las cuales ha estado en contacto (Arcos, 2010). La dureza, dependiendo del contenido de Ca, se conoce como una función del agua salada. Se sabe que la alcalinidad total y la dureza total son procesos naturales de disolución de los componentes del suelo, en particular, la formación de carbonato de calcio (Varol y

Şen, 2009). Las fuentes minerales principales de la dureza provienen del suelo y de acuerdo a la composición de éste, el agua será más o menos dura. Las aguas duras, aguas con elevado contenido de calcio y magnesio, se asocian con cuencas de captación de rocas sedimentarias, de las cuales las más comunes son las de piedra caliza y creta. Las aguas blandas, aguas con bajo contenido de calcio y magnesio, suelen haber estado en contacto con rocas impermeables como el granito. Las aguas superficiales en general, suelen ser más blandas que las aguas subterráneas (Neira, 2006).

En cuanto al calcio, se sabe que es uno de los cationes más abundantes en aguas subterráneas y superficiales, y su presencia en aguas naturales varía de acuerdo con la proximidad a la fuente natural. Típicamente se encuentra en mayores concentraciones en aguas cercanas a rocas carbonatadas. Por otro lado, las aguas residuales contribuyen de manera importante a las concentraciones de calcio en las aguas superficiales; y la lluvia ácida puede incrementar el lixiviado de los iones calcio del suelo (Beita-Sandí y Barahona-Palomo, 2010).

Haciendo referencia al ion carbonato, su origen predominante son las rocas calizas, las cuales están constituidas predominantemente de carbonato de calcio (CaCO₃) que es casi insoluble en el agua. Las aguas naturales que están expuestas a la caliza se dominan aguas calcáreas. El ion carbonato actúa como una base, produciendo iones bicarbonato e hidróxido en el agua (Teves, 2016).

El CP3 representa el 14,15% de la varianza total del conjunto de datos del agua del río, muestra una fuerte correlación positiva entre la conductividad eléctrica (CE) y los sólidos disueltos totales (SDT) y una negativa en el oxígeno disuelto (OD). Este componente principal representa una fuente de variabilidad fisicoquímica de la

calidad el agua, en donde altos contenidos de SDT en el agua afectan adversamente el OD y representa las contaminaciones orgánicas y de nutrientes de las aguas domésticas, aguas residuales de plantas de tratamiento, aguas residuales industriales, así como de las actividades agrícolas. Por otro lado, la conductividad eléctrica en el agua se convierte en una medida de SDT (Siamak y Srikantaswamy, 2009), y niveles elevados corresponden una mezcla de aguas de alcantarillados en el agua del río. En este sentido, los SDT presentes en el agua de consumo proceden de fuentes naturales, aguas residuales, escorrentía urbana y aguas residuales industriales (Argandoña y Macías, 2013). También se tiene que el estándar de calidad para esta variable se define por sólidos provenientes de fuentes de aguas usadas (Sole, 2005); y debido a las solubilidades de diferentes minerales, las concentraciones de SDT en el agua varían considerablemente de unas zonas geológicas a otras (Argandoña y Macías, 2013).

De igual forma el oxígeno disuelto es importante en la calidad del agua. La cantidad de oxígeno disuelto en el agua depende de muchos factores ambientales, y su presencia en las aguas es esencial para evaluar los posibles efectos potenciales de algunos agentes contaminantes, así como para sustentar organismos de vida acuática (Marañón-Reyes *et al.*, 2014). Cuando el Oxígeno Disuelto se encuentra en bajas concentraciones puede ser un indicador de que el agua tiene una alta carga orgánica, provocada por aguas residuales. El conjunto de residuos orgánicos, producidos por los humanos, animales y demás seres vivos, como heces y otros materiales orgánicos, que se vierten en los cuerpos de agua, son descompuestos por bacterias aeróbicas, que consumen oxígeno. Es por ello que cuando estos desechos se encuentran en exceso, ocasionan mayor proliferación de

bacterias y agotan el Oxígeno, haciendo escasa la vida de muchas especies acuáticas (Sotil y Flores, 2016). Es importante destacar que este no es calificado como un contaminante, sin embargo su exceso o escasez dan como resultado condiciones no favorables para el cuerpo de agua (Pérez, 2003). Y como se sabe, el oxígeno disuelto en el agua está estrechamente relacionado con la conductividad, y la solubilidad del mismo en el agua disminuye al aumentar la conductividad (Wetzel, 2001; Manahan, 2011). La correlación negativa del CP3, con las concentraciones de oxígeno disuelto muestra que a bajos valores de OD facilita la liberación de sales desde los depósitos de fondo en el cuerpo de agua.

El CP4, denominado microbiológico, representa el 9,92% de la varianza total, mostró una fuerte correlación positiva con Coliformes Fecales (CF), Coliformes Totales (CT) y Magnesio (Mg); lo que revela la intensiva descarga de aguas residuales en el río Guarapiche. Los niveles de coliformes fecales alcanzaron su punto máximo en la estación 6 (1963,33 NMP/100mL) y el mes más crítico con un promedio de 2006,86 NMP/100mL es el mes de mayo, que coincide con el comienzo del período lluvioso, que favorece la proliferación de estos microorganismos. Este aumento puede ser causado por el incremento del caudal y escorrentía producido por las lluvias. Otra causa es la ausencia de plantas de tratamiento de aguas servidas en los asentamientos adyacentes al río donde se observan el uso de pozos sépticos o hay un vertido directo de desechos al río. Se puede inferir que al estar formado este grupo microbiano exclusivamente por bacterias presentes en el tracto digestivo de animales superiores, la fuente principal de contaminación del río Guarapiche proviene de las aguas servidas vertidas a su cauce sin el conveniente tratamiento previo.

Las elevadas concentraciones de CT pueden deberse al aporte de bacterias procedentes de los residuales domésticos de las comunidades locales y a las precipitaciones, que arrastran bacterias y materia orgánica; estos factores además de incrementar la población bacteriana en el agua favorecen su supervivencia (Ferguson *et al.*, 1996; Bitton y Harvey, 1992; Fuentes, 2008). Las coliformes constituyen un grupo de bacterias heterogéneas con hábitat primordialmente intestinal y es utilizado como indicador de contaminación fecal en agua (Castro, 2002). La presencia de coliformes fecales (CF) en el agua posiblemente indica contaminación reciente de la misma por efluentes o estiércol animal (Hunter *et al.*, 2000). Una de las fuentes de contaminación por CF más importantes son los sitios donde se acumula estiércol, como por ejemplo las pasturas bajo pastoreo intensivo y zonas de bebedero animal (Smith *et al.*, 1996).

En cuanto al magnesio, presente en las aguas como ion Mg^{2+} , se encuentra también en la clorofila, en diferentes compuestos organometálicos, en materia orgánica y es un elemento también esencial para los seres vivos (Bolaños *et al.*, 2015).

El CP5 representa el 7,10% de la varianza total, mostró una fuerte correlación con el potasio (K^+) y nitrito (NO_2^-), y una correlación moderada con el nitrato (NO_3^-). Este componente se puede asumir como un factor “nutriente” del agua de río. Si bien es cierto, estos compuestos forman parte del ciclo natural del nitrógeno, las actividades humanas incrementan sus niveles, principalmente en el suelo, y es debido a su solubilidad en agua, por lo que llega a alcanzar concentraciones importantes en ríos (Bolaños-Alfaro *et al.*, 2017). Las principales rutas de ingreso de nitrógeno a las masas de agua son a través de aguas residuales industriales o municipales, por

tanques sépticos o descargas de corrales ganaderos, residuos animales (incluyendo aves y peces) y también por las descargas de la emisión de gases de vehículos (Calderón, 2004).

Los incrementos de nitritos y nitratos, podrían estar representado por descargas continuas de compuestos nitrogenados en desechos animales, escurrimiento agrícola, óxidos de nitrógeno emitidos durante la quema de huesos, cuernos y pieles, etc. Los nitratos son compuestos con alto contenido de óxido de compuestos nitrogenados y generalmente están presentes en las aguas superficiales. Es el producto final de la descomposición aerobia de nitrógeno orgánico materia presente en los desechos animales. Además, el nitrógeno atmosférico es un nivel de nitrato y un contribuyente a las floraciones de algas que privan a los peces y otros organismos acuáticos de oxígeno (Schiffman, 1995). La acumulación de exceso de nutrientes durante un período de tiempo también podría conducir a la eutroficación y la pérdida de microorganismos acuáticos y de otro tipo en el río.

Las altas concentraciones de nitrato pueden ser atribuidas al crecimiento de la población y a la expansión de las ciudades, los cuales traen como consecuencia un incremento de las actividades que contribuyen con la degradación de los ecosistemas; especialmente, la agricultura, ya que los fertilizantes en general contienen gran cantidad de nitrógeno, que puede llegar a la masa de agua por procesos de absorción, drenaje y escorrentías (Yu *et al.*, 2003). La concentración de nitratos en las aguas superficiales y subterráneas es normalmente baja, pero puede alcanzar un alto nivel como resultado de la lixiviación o la escorrentía de tierras agrícolas o la contaminación por desechos humanos o animales (WHO, 2003). Por otro lado, en aguas naturales las concentraciones de potasio (K) son bajas,

aunque pueden aumentar por deposición atmosférica, escorrentía de terrenos agrícolas e influencia de drenajes de irrigación (Beita-Sandí y Barahona-Palomo, 2010).

El CP6 representa el 5,63% de la varianza total, mostrando una correlación fuerte del sulfato, este componente principal representa una variable química de la calidad del agua. Es importante destacar que los sulfatos (SO_4^{2-}) son muy abundantes en la naturaleza, y su presencia en el agua es muy variable. Los niveles de sulfato en agua de lluvia y agua superficial se correlacionan con las emisiones de dióxido de azufre ocasionadas por la actividad humana (Keller y Pitblade, 1986), principalmente provenientes de motores que utilizan al diésel como combustible. Los sulfatos en concentraciones superiores favorecen la corrosión de los metales y cambian el sabor al agua (en menor medida que los cloruros y carbonatos), lo que también incrementa la cantidad de plomo disuelto, proveniente de las tuberías de plomo (Gallego *et al.*, 2013). Su afectación en el ser humano tiene que ver principalmente con el efecto laxante que surge al ingerir una alta concentración, y como efecto secundario provoca deshidratación, producto de un *Tabla* diarreico, el cual es más crítico en niños y adultos mayores, siendo muy común después de una ingestión de más de cinco gramos al día (Bolaños-Alfaro *et al.*, 2017).

El último componente (CP7) representa el 5,16% de la varianza total, muestra una fuerte correlación del Cloro, y una correlación débil del Manganeseo y se puede asumir como un componente antropogénico tóxico. El Manganeseo (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 principalmente 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, 2010).

Por otra parte, el cloruro en el agua potable se origina de fuentes naturales, aguas residuales, efluentes industriales y escorrentía urbana. La concentración excesiva del mismo, aumenta la velocidad de corrosión de los metales en el sistema de distribución de agua y depende de la alcalinidad del agua; esto puede llevar a un aumento de las concentraciones de metales en el suministro (WHO, 2003a).

Análisis de Regresión Múltiple (ARM)

Básicamente el ARM 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). Los ACP y ARM se combinaron para identificar las posibles fuentes de contaminación en el río Guarapiche. 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). El procedimiento Stepwise de SAS con la opción de Backward, que comenzó con las 14 variables tomadas en consideración para el análisis de componentes principales, finalmente nos reduce el número de variables a 6 sin perder la precisión en el análisis de datos, con un R^2 igual a 0,9914.

Tabla 6. Análisis de regresión Múltiple

Variable dependiente: ICA				
Parámetro	Estimación	Error estándar	Estadístico T	P-Valor
CONSTANTE	-1191,37	287,4620	-4,14446	0,0002
CF	0,125982	0,0021162	59,5323	0,0000
CE	0,399656	0,0993281	4,02359	0,0003
CI	0,529844	0,294182	1,80107	0,0803
Temperatura	5,17665	2,186290	2,36778	0,0236
pH	120,428	29,69550	4,05542	0,0003
Na	2,53128	0,680873	3,71770	0,0007

Tabla 7. Análisis de varianza (Anova).

Fuente	Suma de cuadrados	GL	Cuadrado medio	Cociente-F	P-Valor
Modelo	1914960,00	6	319160	679,37	0,000
Residuo	16442,5	35	469,786		
Total (Corr.)	1931400,00	41			

R-cuadrado = 99,1487 porcentaje

R-cuadrado (ajustado para g.l.) = 99,0027 porcentaje

Error estándar de est. = 21,6745

Error absoluto medio = 14,1486

Estadístico de Durbin-Watson = 2,09401 (P=0,2426)

Autocorrelación residual en Lag 1 = -0,0502132

La salida muestra los resultados del ajuste a un modelo de regresión lineal múltiple para describir la relación entre ICA y 6 variables independientes (Tablas 6 y 7).

La ecuación del modelo ajustado es:

$$ICA = -1191,37 + 0,125982*CF + 0,399656*CE + 0,529844*CI + 5,17665*Temp + 120,428*pH + 5,53128*Na \dots\dots\dots 4$$

El modelo lineal está dominado por las variables pH, Temperatura y Na (Tabla 6). Dado que el p-valor en el Tabla 7

ANOVA es inferior a 0,01, existe relación estadísticamente altamente significativa entre las variables para un nivel de confianza del 99%. El criterio más comúnmente usado para evaluar el desempeño del modelo es el coeficiente de determinación (R²) (Pearson, 1986); sin embargo R² no es buena medida de comparación de diferentes modelos, ya que los valores de R² solo proporcionan lo bueno que se ajusta el modelo a los datos utilizado para construir los modelos y no lo bien que funciona en datos externos (Aertsen *et al.*, 2010). El estadístico R-cuadrado indica que el modelo explica un 99,15%

de la variabilidad en ICA. El estadístico R-cuadrado ajustado, que es más conveniente para comparar modelos con diferentes números de variables independientes, es 99,0027%. El error estándar de la estimación, muestra la desviación típica de los residuos que es 21,6745. Este valor puede usarse para construir los límites de predicción para nuevas observaciones. El error absoluto medio (EAM) de 14,1486 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 el que se han introducido los datos en el fichero. Dado que el p-valor es superior a 0,05, no hay indicio de autocorrelación serial en los residuos. Para decidir la simplificación del modelo, se tuvo en cuenta que el p-valor más alto en las variables independientes es 0,0803, perteneciendo a Cl. Puesto que el p-valor es inferior a 0,01, ese término es estadísticamente significativo para un nivel de confianza del 99%. Dependiendo del nivel de confianza con el que quiera trabajar, se puede decidir quitar o no a la variable Cl del modelo.

4. Conclusiones

- El Análisis de componentes principales (ACP) en los estudios de calidad del agua es un medio auxiliar en la identificación de sus recursos efectivos. La calidad del agua requiere ser constantemente monitoreada por cambios debidos especialmente a las degradaciones y contaminaciones. La principal causa de degradación del Río Guarapiche es la descarga de la agricultura y desechos del hogar. Todas las estaciones son afectadas por factores agrícolas intensivas como fertilización, agua de riego, agua de drenaje, y residuos de ganado. También son seriamente afectados negativamente por los residuos de las zonas residenciales cercanas a la cuenca del río. Los ACP dieron como resultado siete factores explicando el 83,99%

de La varianza acumulada en el conjunto de datos. El primer factor explicó el 23,36% de la varianza acumulada, el segundo factor 18,67%, el tercer factor 14,15%, el cuarto 9,92%, el quinto 5,71%, el sexto 5,63% y el séptimo 5,16 % de las varianzas acumuladas, respectivamente. Los resultados del Análisis de regresión múltiple (ARM) revelan que el pH, Temperatura, Na, conductividad eléctrica, coliformes fecales y cloro son los parámetros más importantes utilizados para evaluar los cambios en la calidad del agua del río. La calidad del agua es considerada el factor principal para la Salud del hombre y de los animales. La calidad del agua es en gran parte determinada por los procesos naturales y por los aportes antropogénicos. 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 proporcionan información más útil sobre la calidad del agua del río Guarapiche. Esta técnica se ha aplicado con éxito para la planificación y el seguimiento de la calidad en el futuro.

5. Literatura citada

- Aertsen, W.; Kinta, V.; Orshovena, J.; Özkan, K.; Muysa, B. 2010. Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests *EcologicalModelling*, 221: 1119-1130.
- Ahmed, S.; Hussain, M.; Abderrahman, W. 2005. "Using multivariate factoranalysis to assess surface/ logged water quality and source of contamination at a large

- irrigation project at Al-Fadhli, Eastern Province, Saudi Arabia,” *Bulletin of Engineering Geology and the Environment* 64: 232–315.
- Akcaý, H.; Oguz, A.; Karapire, C. 2003. Study of heavy metal pollution and speciation in Buyak Menderez and Gediz river sediments. *Water Research* 37: 813-822.
- Appelo, C.A.J.; Postma, D. 1994. *Geochemistry, groundwater and pollution*. Rotterdam, Netherlands, Balkema. 536 p.
- Argandoña, L.E.; Macías, R.G. 2013. “Determinación De Sólidos Totales, Suspendidos, Sedimentados Y Volátiles, En El Efluente De Las Lagunas De Oxidación Situadas En La Parroquia Colón, Cantón Portoviejo, Provincia De Manabí, Durante El Período De Marzo A Septiembre 2013”. Tesis De Grado Previo A La Obtención Del Título De Ingeniero Químico. Universidad Técnica De Manabí. Facultad De Ciencias Matemáticas, Físicas Y Químicas. Portoviejo, Manabí. Ecuador.
- Arcos, A.O. 2010. Consideraciones básicas de ingeniería para el diseño de una planta potabilizadora para cerro Colorado con un proceso de nanofiltración. Tesis Para Optar Por El Grado De: Maestro En Ingeniería Ingeniería Ambiental – Agua. Programa De Maestría Y Doctorado En Ingeniería. Universidad Nacional Autónoma De México.
- Azkue, M.; Soto, E. 2004. Régimen de precipitación en Montalbán, Estado Carabobo, Venezuela. 2004. *Agronomía Tropical* 54(4): 359-370.
- Bitton, G.; Harvey, R. 1992. Transport of pathogens through soil and aquifers. In: *Environmental Microbiology*. Wiley-Liss (eds.). New York. pp 103-124.
- Beita-Sandí, W.; Barahona-Palomo, M. 2010. Físico-química de las aguas superficiales de la Cuenca del río Rincón, Península de Osa, Costa Rica. *Cuadernos de Investigación UNED* 2(2):157-179.
- Belkhir, L.; Boudoukha, A.; Mouni, L.; Baouz, T. 2010. Multivariate statistical characterization of groundwater quality in Ain Azel plain, Algeria: *African Journal of Environmental Science and Technology* 4(8):526-534.
- Benítez, S. 2004. Determinación de la carga orgánica impuesta al río Manzanares en el período febrero-junio 2004. Trabajo de Grado (T.S.U. Química Aplicada). IUT-Cumaná, Venezuela.
- Bitton, G.; Harvey, R. 1992. Transport of pathogens through soils and aquifers. In *Environmental Microbiology*. 19th ed. by Ralph Mitchell. New York, USA, 103-123.
- Bolaños, J.; Montero, N.; Rodríguez, N.; Sánchez, A. 2015. Calidad de aguas superficiales: estudio de la quebrada Estero, ubicada en el cantón de San Ramón, Costa Rica.
- Bolaños-Alfaro, J.; Cordero-Castro, G.; Segura-Araya, G. 2017. Determinación de nitritos, nitratos, sulfatos y fosfatos en agua potable como indicadores de contaminación ocasionada por el hombre, en dos cantones de Alajuela (Costa Rica). *Tecnología en Marcha* 30-4. Pág 15-27.
- Calderón, R. 2004. “Evaluación de la comunidad de macro invertebrados bentónicos y la calidad físicoquímica del agua en la parte alta de la quebrada El Carracá, del municipio de Los Santos”, tesis de licenciatura,

- Universidad Industrial de Santander, Colombia.
- Calla, L.; Cabrera, C. 2010. Calidad del agua en la cuenca del río Rímac, sector de San Mateo, afectado por las actividades mineras. Revista del Instituto de Investigaciones de la Facultad de Geología, Minas, Metalurgia y Ciencias Geográficas. 13(25).
- Castro, M. 2002. Técnicas para el Análisis Microbiológico. 2ª edición. Facultad de Química, UNAM, México, pp. 29.
- CCME (Canadian Council of Ministers of the Environment). 1999. Canadian Water Quality Guidelines for the Protection of Aquatic Life. Winnipeg, Canada.
- Costa, M.H.; Botta, A.; Cardille, J.A. 2003. Effects of large-scale changes in land cover on the discharge of the Tocantins River, South eastern Amazonia. *J. Hydrol.*, 283: 206-217.
- Decreto N° 883. Normas para la clasificación y el control de la calidad de los cuerpos de agua y vertidos o efluentes líquidos. Gaceta Oficial de la República de Venezuela, N° 5.021 Extraordinarios del 18 de diciembre de 1995.
- Drever, J.L. 1988. *The Geochemistry of Natural Waters: Surface and Groundwater Environments*. Prentice-Hall, Englewood Cliffs, Nueva Jersey, EEUU.
- Ferguson, C.; Coote, B.; Ashbolt, N.; Stevenson, I. 1996. Relationships between indicators, pathogens and water quality in an estuarine system. *Water Res.* 30 (9): 2045-2054.
- Fuentes, Y. 2008. Calidad Bacteriológica del agua de las Cuencas Alta y Media del Río Manzanares, Estado Sucre, Venezuela. Trabajo De Grado Presentado Como Requisito Parcial Para Optar Al Título De Licenciada En Biología. Universidad de Oriente Núcleo de Sucre. Escuela de Ciencias. Sucre-Cumaná. Venezuela.
- Gallego, A.; Garcinuño, Morcillo, M. 2013. Experimentación en química analítica. Universidad Nacional de Educación a Distancia. Edición electrónica. 362-6530-9
- Gil, M.J.A.; Belloso de Herrera, G.; Vizcaino, G.C.; Maza, I.J.; Sánchez, C.M.C.; Bolívar, C.E.; Martínez, P.D. 2013. Evaluación de la calidad microbiológica y niveles de nitratos y nitritos en las aguas del río Guarapiche, estado Monagas, Venezuela. *Revista Científica UDO* 154 Agrícola 13(1):154-163.
- Gil, M. J. A.; Vizcaíno, G.C.; Montañomata, N. 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. *Anales Científicos* 79(1): 111 - 119.
- Gordon, G.E. 1988. Receptor models. *Environ. Sci. Technol.* 22: 1132-1142.
- Helena, B.; Pardo, R.M.; Vega, R.M.; Barrado, R.; Fernandez, J.; Fernandez, J. 2000. Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principle component analysis *Wat. Res.* 34(3): 807-816.
- Holland, M.; Witthuser, K.T. 2009. Geochemical characterization of karst groundwater in the Cradle of Humankind World Heritage Site, South Africa: *Environmental Geology* 57: 513-524.
- Hounslow, A.W. 1995. *Water Quality Data*. New York, USA, CRC Lewis Publishers. 49-59.

- Hunter, C.; Perkins, J.; Tranter, J.; Hardwick, P. 2000. Fecal bacteria in the water of and upland área in Derbyshire England: The influence of agricultural land use. *J. Environ. Qual.* 29: 1253-1261.
- Ismail, A.H.; Basim S.A.; Shahla, A.Q. 2014. "Application of Multivariate Statistical Techniques in the surface water quality Assessment of Tigris River at Baghdad stretch," *Iraq Journal of Babylon University/ Engineering Sciences* 22(2).
- Jenkins, D. 2009. *Química del Agua: Reacciones de Oxidación-Reducción*. México D.F. México. Editorial LIMUSA, S.A. 351-466 p.
- Kaiser, H.F. 1960. The application of electronic computers to factor analysis: *Educational and Psychological Measurement* 20:141-151.
- Kanade, S.B.; Gaikwad, V.B. 2011. A Multivariate Statistical Analysis of Bore Well Chemistry Data-Nashik and Niphad Taluka of Maharashtra, India: *Universal Journal of Environmental Research and Technology* 1(2): 193-202.
- Karim, Z.; Abdollahman, R. 2011. Diagnostic measures for linear mixed measurement error models. *Statistics and Operations Research Transactions* 35(2):125-144.
- Keller, W.; Pitblade, J.R. 1986. "Water quality changes in Sudbury area lakes: a comparison of synoptic surveys in 1974-1976 and in 1981-1983", *Water, Air and Soil Pollution* 29: 285.
- Krishna, A.K.; Satyanarayanan, M.; 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 andhra Pradesh, India. *J. Hazardous Materials* 167: 366-373.
- Liu, C.W.; Lin, K.H.; Kuo, Y.M. 2003. Application of factor analysis in the assessment of groundwater quality in a black foot disease area in Taiwan. *The Science of the Total Environ.* 313: 77-89.
- Marañón-Reyes, A.M.; Perez-Pompa, N.E.; Dip-Gandarilla, A.M.; González-Marañón, A.; Pérez-Silva, R.M.; Ruiz-Estrella, A. 2014. Evaluación temporal de la calidad de las aguas del río. Los Guaos de Santiago de Cuba.
- Ministerio del Ambiente y de los Recursos No Renovables [MARNR]. 1980. *Uso actual y perspectivas de aprovechamiento del río Guarapiche*. Maturín: Serie informes técnicos Zona 12/IT. Maturín, Estado Monagas.
- Ministerio del Ambiente. 2006. *Recursos hídricos de Venezuela*. Caracas. Fundambiente.
- Miyabara, Y.; Suzuki, J.; Suzuki, S. 1994. Classification of urban rivers on the basis of water pollution indicators in river sediment. *Bulletine Environmental Contamination Toxicol.* 52: 1-8.
- Mohd Fahmi Mohd Nasir, Mohd, Saiful Samsudin, IIsahak Mohamad, IMohammad Roshide Amir Awaluddin, 1Muhd Ariffin Mansor, 1Hafizan Juahir, 2Norlafifah Ramli. 2011. River Water Quality Modeling Using Combined Principle Component Analysis (PCA) and Multiple Linear Regressions (MLR): A Case Study at Klang River, Malaysia. *World Applied Sciences Journal* 14 (Exploring Pathways to Sustainable Living in Malaysia:

- Solving the Current Environmental Issues): 73-82.
- Nádaská, G.; Lesný, J.; y Michalík, I. 2010. Environmental aspect of manganese chemistry, *Hungarian Journal of Sciences, ENV-100702-A*, pp 1-16.
- Neira, G.M.A. 2006. Dureza En Aguas De Consumo Humano Y Uso Industrial, Impactos Y Medidas De Mitigación. Estudio De Caso: Chile. Memoria para optar al título de Ingeniero Civil. Universidad de Chile. Santiago de Chile.
- Nix, P.G.; Merry, C.H. 1990. Use of sediment bags as a monitor of fecal pollution in streams. *Bulletine Environmental Contamination Toxicol.* 45:864-869.
- Organización Mundial de la Salud [OMS]. 2006. Guías para la calidad del agua potable. Primer apéndice a la tercera edición. Volumen 1. Recomendaciones. Organización Mundial de la Salud, Genève 27, Suiza ; 393 pp.
- Pearson, K. 1986. Regression, heredity, panmixia. In *Mathematical contributions to the theory of evolution.* *Philos. Trans. R. Soc. Lond.* 187: 253-318.
- Peréz, G.A. 2003. Bioindicación de la calidad del agua en Colombia. Medellín, Colombia: Universidad de Antioquia.
- Postawa, A. 2013. Best Practice Guide on the Control of Iron and Manganese in Water Supply. London. United Kingdom. British Library. 30 p.
- Ramakrishnaiah, C. R.; Sadashivaiah, C.; Ranganna, G. 2009. Assessment of Water Quality Index for the Ground Water in Tumkur Taluk. *E-Journal of Chemistry* 6(2):523-530.
- Saima, S.; Osman, R.; Spian, D.; Jaafar, Juhair, H.; Abdullah, M.P.; Ghani, F.A. 2009. Chemometric approach to validating faecal sterols as source tracer for faecal contamination in water. *Water Res.* 43: 5023-5030.
- Sarbu, C.; Pop, H. 2005. Principal component analysis versus fuzzy principal component analysis a case study: the quality of danube water (1985–1996) *Talanta* 65: 1215-1220.
- Schiffman, S. 1995. "Effects of Livestock Industry on the Environment," *Environmental Health Perspectives* 102: 1096-1099.
- Siamak Gholami and S. Srikantaswamy. 2009. Statistical Multivariate Analysis in the Assessment of River Water Quality in the Vicinity of KRS Dam, Karnataka, India. *Natural Resources Research* 18(3):235-247.
- Singh, K.P.; Malik, A.; Sinha, S. 2005. Water quality assessment and apportionment of pollution sources of Gomti River (India) using multivariate statistical techniques: a case study. *Analytica Chimica Acta* 35: 3581-3592.
- Smith, J.U.; Bradbury, N.J.; Adiscott, T. M. 1996. Sundial: APC-based system for simulating nitrogen dynamics land. *Agron. J.* 88: 38-43.
- Sole, A. C. 2005. Instrumentación Industrial. (Vol. 7ma Edición). España: MARCOMBO S.A.
- Sotil, R.; Flores, V, H.I. 2016. Determinación de parámetros físicos, químicos y bacteriológicos del contenido de las aguas del río Mazán-Loreto, 2016. Tesis para optar el título profesional de Ingeniero Químico. Universidad Nacional De La Amazonía Peruana facultad De Ingeniería Química. Perú.
- Stanley, E. Manahan. 2011. Introducción A La Química Ambiental. Primera

- Edición. Editorial Reverte. España. 725 pag.
- Stumm, W.; y Morgan, J.J. 1996. *Aquatic Chemistry*. Wiley, Nueva York, Nueva York, EEUU.
- Teves, A.B.M. 2016. *Estudio Físicoquímico De La Calidad Del Agua Del Río Caca, Región Lima. Tesis Para Optar El Grado De Magister En Químico Pontificia Universidad Católica Del Perú Escuela De Posgrado.*
- Tong, S.T.Y.; Chen, W 2002. Modeling the relationship between land use and surface water quality. *J. Environ. Manage* 66: 377-393.
- Tong, S.T.Y. 1990. The hydrologic effects of urban land use: a case study of the Little Miami River Basin. *Landscape Urban Plan*, 19: 99-105.
- Varol, M.; Şen, B. 2009. Assessment of surface water quality using multivariate statistical techniques: a case study of Behrimaz Stream, Turkey. *Environment Monitoring Assessment* 159: 543–553.
- 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. Tesis de Máster. Departamento de Química Analítica. Cádiz-España.*
- Wetzel, R.G.G.E. 2001. *Likens. Limnological Analysis*. 3aed. Ed. Springer Science. New York (USA). Pp.289-304.
- WHO. 2003. "Nitrate and Nitrite in Drinking Water. Background Document for Preparation of WHO Guidelines for Drinking Water Quality," World Health Organization, Geneva, 2003.
- WHO. 2003a. "Guidelines for Drinking Water Quality," World Health Organization, Geneva, 2003.
- Wunderlin, D.A.; Diaz, M.P.; Ame, M.V.; Pesce, M. Hued, A.C. and 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). *Water Res.* 35: 2881-2894.
- Yu, P.; Li, Z.; Zhang, H.; Cao, Y.; Li, D.; Al, E. 2003. Sampling strategy of primary core collection of common wild rice (*Oryza rufipogon* Griff.) in China. *J. China Agr.* 8, 37-41.