

DETERMINACIÓN DE NIVELES DE CONTAMINACIÓN DEL AGUA A PARTIR DE TÉCNICAS DE ANÁLISIS MULTIVARIABLE

DETERMINATION OF WATER POLLUTION LEVELS FROM MULTIVARIATE ANALYSIS TECHNIQUES



Deiver Saavedra Cotrina

Administrador Ambiental, Universidad Distrital Francisco José de Caldas, Bogotá D.C, Colombia

Recibido: 26/06/2018 • Aprobado: 25/07/2018

RESUMEN

El presente estudio tiene como objeto, analizar datos fisicoquímicos y microbiológicos de las aguas provenientes de las quebradas La Colosa y La Arenosa ubicadas en el Municipio de Cajamarca, Departamento del Tolima (Colombia), como herramienta para determinar, de manera multivariable, las dimensiones de la contaminación de dichos cuerpos de agua. La investigación se inicia de tipo descriptiva; sin embargo se termina como descriptiva – correlativa, debido a que se describen los resultados de información secundaria, se calculan los índices de contaminación del agua y se aplican técnicas de análisis multivariable, posteriormente se correlacionan los índices de contaminación con los resultados obtenidos con las técnicas de conglomerados y análisis factorial; permitiendo comprobar que las técnicas estadísticas seleccionadas son aplicables como herramienta para determinar los niveles la contaminación de cuerpos de agua. Lo anterior sustentado en la similitud obtenida en los resultados del cálculo de índices de contaminación y las técnicas de análisis multivariable empleadas.

Palabras clave: análisis estadístico, minería de oro, monitoreo ambiental, recurso hídrico.

ABSTRACT

The purpose of this study is to analyze physicochemical and microbiological data of the waters coming from the La Colosa and La Arenosa streams located in the Municipality of Cajamarca, Department of Tolima (Colombia), as a tool to determine, in a multivariable way, the dimensions of the contamination of said bodies of water. The investigation begins of descriptive type; however, it is finished as descriptive - correlative, because the results of secondary information are described, the water pollution rates are calculated and multivariable analysis techniques are applied, then the pollution rates are correlated with the results obtained with the techniques of clusters and factor analysis; allowing to verify that the selected statistical techniques are applicable as a tool to determine the levels of contamination of water bodies. The above is based on the similarity obtained in the results of the calculation of pollution indices and the multivariable analysis techniques used.

Keywords: environmental monitoring, gold mining, statistical analysis, water resources.

¹ deiversc@gmail.com, orcid.org/0000-0001-6720-8103

1. INTRODUCCIÓN

Las cuencas hidrográficas ofrecen numerosos servicios a la sociedad. El suministro mundial de agua dulce para usos doméstico, agrícola e industrial depende mucho de los caudales que se producen y regulan en las cuencas (FAO, 2009). Además, alrededor del mundo las cuencas hidrográficas han tenido múltiples usos, entre ellos se pueden nombrar el riego de cultivos, como espacios recreativos, generadoras de energía eléctrica, se usan como rutas fluviales de transporte, pesca, áreas de desagüe, entre otros (Tápanes, 1974).

La mayoría de los usos del agua requieren un estado de calidad óptima, que se debe evaluar de manera cuantitativa y cualitativa mediante la utilización de diferentes análisis (Romero, 2009).

La contaminación de nuestras cuencas hidrográficas degrada al medio ambiente, daña el hábitat de la flora y la fauna silvestre, afecta a la economía y a los empleos, causa impuestos y cuotas más altas, y finalmente también afecta a la salud de los seres humanos. (Manrique et al., 2007)

Según la clasificación de las cuencas que hace la corporación autónoma regional del Tolima – Cortolima, el municipio de Cajamarca cuenta con la cuenca mayor del río Coello (Cortolima, 2009), la cuenca del río Bermellón, la subcuenca de la quebrada La Guala, a la cual tributan las microcuencas de las quebradas La Arenosa y La Colosa. Dentro del área se destacan usos del agua para preservación la de la flora y la fauna, agropecuario y consumo humano mediante acueductos artesanales (Vergara, González & Gonzáles, 2013).

La calidad del agua y el estado actual de estos cuerpos hídricos juegan un papel importante para propender al equilibrio natural de la región. Tener una buena caracterización fisicoquímica y microbiológica del agua y utilizar herramientas como los índices de contaminación (ICOs) propuestos por Ramírez y Viña en su libro *Limnología colombiana: aportes a su conocimiento y estadísticas de análisis* (Ramírez & Viña, 1998), para llevarlos a una interpretación cuantitativa y cuantitativa del estado del recurso.

El análisis multivariable está formado por un conjunto de métodos estadísticos que sirven para realizar el tratamiento conjunto de datos relativos a diversas variables. Algunos de estos métodos son puramente descriptivos, se limitan a realizar un estudio de los datos recogidos a través de muestras sobre poblaciones finitas o infinitas. Otros métodos de análisis multivariable realizan inferencia estadística sobre los parámetros poblacionales, incorporando modelos con hipótesis estadísticas (Ibar & García, 2002).

La presente investigación, abarca desde la recopilación de información relacionada con resultados de análisis fisicoquímicos y microbiológicos, obtenidos a partir de monitoreos puntuales con periodicidad mensual durante 18 meses consecutivos (de enero de 2013 a junio de 2014), en las quebradas La Colosa y La Arenosa ubicadas en el municipio de Cajamarca, departamento del Tolima (Antek, 2014); pasando por el cálculo de índices de contaminación del agua y su posterior confrontación con los resultados obtenidos de la implementación de técnicas de análisis multivariable, las cuales se pretende puedan ser utilizadas como una herramienta viable para determinar las dimensiones de la contaminación de dichos cuerpos de agua, así como la comparación con metodologías existentes empleadas para la clasificación cualitativa y cuantitativa de la calidad del agua propuestas por autores como (Ramírez y Viña, 1998).

El objetivo final de este estudio, es realizar un análisis comparativo entre los resultados que arrojan los índices de contaminación del agua propuestos por Ramírez y Viña, frente a los resultados que se obtienen mediante la aplicación de diferentes técnicas de análisis multivariable; de esta manera, poder dar cuenta de la situación actual del recurso hídrico en los puntos de muestreo seleccionados de las quebradas estudiadas.

2. MATERIALES Y MÉTODOS

Tomando como base los criterios de Hernández Sampieri et al. en su libro *Metodología de la Investigación*

(Hernández, Fernández & Baptista, 1997), el presente estudio se inicia como descriptivo; sin embargo, se termina como descriptivo-correlativo dado que se busca describir los resultados de la información secundaria, calcular los índices de contaminación del agua y aplicar técnicas de análisis multivariable; posteriormente se correlacionan los índices de contaminación con los resultados de aplicar la técnica multivariable.

La metodología general de este proyecto se divide en cuatro fases: en una primera fase, se utiliza como insumo información secundaria que consiste en resultados de laboratorio de análisis fisicoquímicos y microbiológicos extraídos de monitoreos ambientales puntuales, realizados durante dieciocho meses consecutivos entre los años 2013 a 2014 en cuatro puntos distribuidos en las quebradas la Colosa y la Arenosa (Antek, 2014), estas corrientes de agua se encuentran dentro del área de influencia directa del Proyecto de exploración minera La Colosa de la compañía Anglo-gold Ashanti Colombia S.A, por lo que podrían verse afectadas de haberse aprobado a esta compañía la licencia para explotación minera (Arango, 2014).

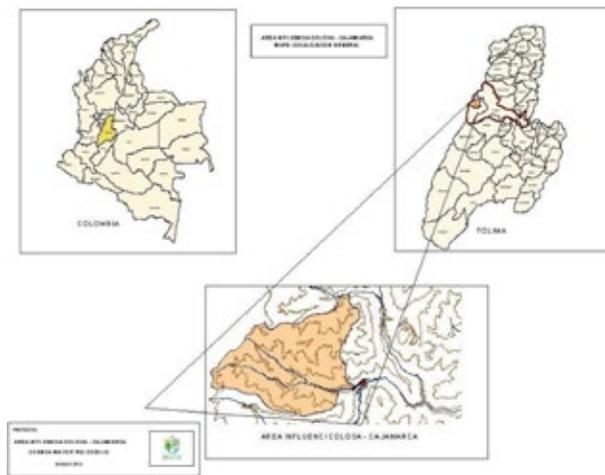


Fig. 1 Localización del área de influencia del proyecto La Colosa

El área de estudio hace parte de la cuenca del río Bermellón, subcuenca quebrada La Guala, microcuencas quebradas la Colosa y la Arenosa. La parte alta de la microcuenca de la quebrada la Colosa tiene dos tipos de drenaje: dendrítico, disectada por pequeños cursos de agua; rectangular, formado por las corrientes

menores que confluyen al cauce principal de la quebrada La Colosa, en ángulo casi perpendicular como se observa en la Figura 2.

Los puntos de muestreo se nombraron de la siguiente manera:

- Quebrada La Colosa Aguas Arriba (CAAr), ubicación en coordenadas geográficas $04^{\circ}28'02,5''$ N - $75^{\circ}29'17,9''$ W.
- Quebrada La Colosa Aguas Abajo (CAAb), ubicación en coordenadas geográficas $04^{\circ}27'00,4''$ N - $75^{\circ}28'50,5''$ W.
- Quebrada La Arenosa Aguas Arriba (AAAr), ubicación en coordenadas geográficas $04^{\circ}28'08,3''$ N - $75^{\circ}29'54,9''$ W.
- Quebrada La Arenosa Aguas Abajo (AAAb), ubicación en coordenadas geográficas $04^{\circ}27'31,0''$ N - $75^{\circ}29'53,0''$ W.

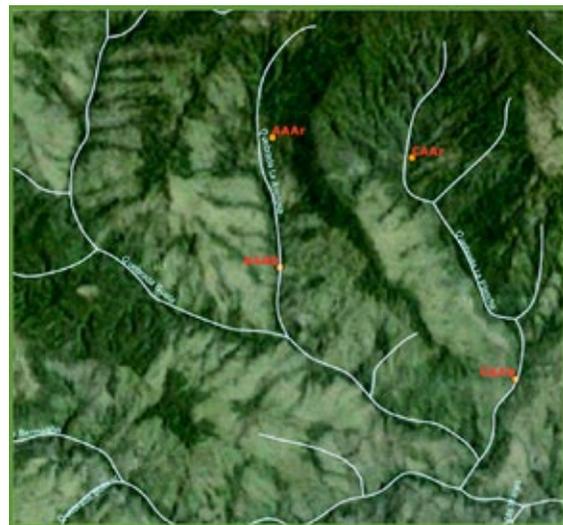


Fig. 2 Ubicación puntos de muestreo.

Ambas quebradas están ubicadas en zona rural de alta montaña en el municipio de Cajamarca, departamento del Tolima. Se utilizaron únicamente los parámetros necesarios para realizar el cálculo de Índices de Contaminación (ICOS) propuestos por Ramírez y Viña en su libro *Limnología colombiana* y que fueron concebidos para ser aplicados en Colombia (1998). Los índices permiten cuantificar el grado de contaminación de las aguas respecto a su condición general y no

a contaminantes específicos; conjugan las propiedades fundamentales de las aguas, y por esto son variables que regularmente se determinan en cualquier estudio limnológico o ambiental, muy a pesar de que la mayoría de ellas no están siquiera contempladas en la legislación nacional, razón por la cual cobran especial interés (Restrepo y Cardenosa, 1999).

Los parámetros necesarios para este estudio son:

- pH (pH)
- Conductividad (COND)
- Temperatura de Muestra (TEM)
- Oxígeno Disuelto (OD)
- Alcalinidad Total (ALKT)
- Dureza Total (DURT)
- Fósforo Total (P_TOT)
- Sólidos Suspendidos Totales (SST)
- DBO₅ (DBO)
- Coliformes Totales (COLT)

Los ICOS que a su vez se analizan en una segunda fase de resultados para poder determinar el estado de contaminación de los cuerpos de agua, son:

- Índice de contaminación por mineralización (ICOMI), evaluado en términos de la COND, DURT y ALKT.
- Índice de contaminación por sólidos suspendidos (ICOSUS) que se basa en el parámetro SST.
- Índice de contaminación por pH (ICOPH).
- Índice de contaminación por materia orgánica (ICOMO) calculado a partir de la DBO, los COLT y el porcentaje de saturación de oxígeno.
- Índice de contaminación trófico (ICOTRO) calculado a partir de la concentración del P_TOT.

En una tercera fase de resultados se seleccionan técnicas de análisis estadístico multivariable, específicamente análisis de conglomerados, también conocido como análisis clúster o de conjunto, y el análisis factorial con extracción por componentes principales las cuales se aplican a los datos fisicoquímicos y microbiológicos mediante el paquete estadístico SPSS en su versión 22 (IBM, 2013), con el objetivo de poder obtener

una clasificación de los cuatro puntos de monitoreo según las concentraciones de los parámetros analizados, así como determinar el estado de contaminación de dichos cuerpos de agua.

El análisis clúster se realizó para detectar similitudes entre los diferentes puntos de estudio que fueron seleccionados. El análisis se realizó a partir de los promedios de cada parámetro tomando entonces un solo dato por cada parámetro y punto estudiado. Además, se aplicó una estandarización de las unidades de medida por medio de la metodología de los puntajes Z.

A partir de los datos estandarizados se aplicó la técnica de análisis de clúster con el método de Ward y la distancia euclidiana al cuadrado como medida de similitud. La representación gráfica se realizó mediante un dendograma. Para el análisis factorial con extracción por componentes principales, inicialmente se estandarizaron los datos por lo que se les antepone la letra Z y se calculó una matriz de correlaciones de Pearson (Rodríguez et al., 2000).

Para finalizar, en la cuarta fase se hace una relación entre los resultados obtenidos con los índices de contaminación y los resultados obtenidos por la aplicación de las técnicas estadísticas multivariable, con el fin de comprobar la aplicabilidad y viabilidad de dichas técnicas estadísticas como herramienta para determinar las dimensiones de la contaminación de dichos cuerpos de agua.

3. RESULTADOS Y DISCUSIÓN

3.1 Recopilación de información secundaria

Luego de la extracción de la información de los reportes de resultados se puede destacar que a excepción del parámetro COLT, todos los resultados de laboratorio tienen un comportamiento uniforme durante los 18 meses estudiados. Para el caso particular de los COLT, es un parámetro que puede variar fácilmente de un mes a otro pues depende de la presencia o no de animales en la zona aguas arriba del punto de monitoreo (González, 2012).

Es importante señalar que los puntos de muestreo aguas abajo de un cuerpo de agua tienden a presentar un tanto más de afectación que los puntos aguas arriba, pues aguas abajo se tiene la carga de los puntos aguas arriba más la carga que se pueda adicionar hasta el punto aguas abajo (Roldán & Ramírez, 2008).

3.2 Cálculo de índices de contaminación

En términos generales los puntos aguas abajo de ambos cuerpos de agua presentan una mayor contaminación en relación a los puntos aguas arriba, comportamiento esperable para cuerpos de agua superficial de alta montaña ya que al estar en continuo movimiento el agua arrastra minerales y diferentes compuestos naturales a su paso, lo cual aporta al incremento de los índices de contaminación (Roldán & Ramírez, 2008).

Según los resultados de la Tabla 1, obtenidos para el cálculo de índices de contaminación la clasificación de los puntos de muestreo de menor grado contaminado a mayor grado de contaminación es la siguiente:

- Quebrada La Colosa aguas arriba
- Quebrada La Arenosa aguas arriba
- Quebrada La Arenosa aguas abajo
- Quebrada La Colosa aguas abajo.

En total se calcularon noventa (90) ICOS por punto de muestreo, de los cuales la mayoría arrojaron un grado de contaminación catalogado como “ninguno” según (Ramírez Viña, 1998), sin embargo, una proporción entre el 11% y el 15% señala un “Alto” grado de contaminación.

TABLA 1
Distribución de índices de contaminación según su grado de contaminación

Quebrada La Colosa aguas arriba			
Rango	Grado de contaminación	%	Distribución ICOS
0 – 0,2	Ninguno	74	67
>0,2 - 0,4	Bajo	7	6
>0,4 – 0,6	Medio	7	6
>0,6 – 0,8	Alto	12	11
>0,8 – 1,0	Muy Alto	0	0
Total ICOS			90
Quebrada La Colosa aguas abajo			
Rango	Grado de contaminación	%	Distribución ICOS
0 – 0,2	Ninguno	60	54
>0,2 - 0,4	Bajo	19	17
>0,4 – 0,6	Medio	5	5
>0,6 – 0,8	Alto	16	14
>0,8 – 1,0	Muy alto	0	0
Total ICOS			90
Quebrada La Arenosa aguas arriba			
Rango	Grado de contaminación	%	Distribución ICOS
0 – 0,2	Ninguno	71	64
>0,2 - 0,4	Bajo	10	9
>0,4 – 0,6	Medio	3	3
>0,6 – 0,8	Alto	16	14
>0,8 – 1,0	Muy alto	0	0
Total ICOS			90
Quebrada La Arenosa aguas abajo			
Rango	Grado de contaminación	%	Distribución ICOS
0 – 0,2	Ninguno	64	58
>0,2 - 0,4	Bajo	16	14
>0,4 – 0,6	Medio	3	3
>0,6 – 0,8	Alto	17	15
>0,8 – 1,0	Muy alto	0	0
Total ICOS			90

Fuente. autor 2018

3.3 Selección y aplicación de las técnicas de análisis multivariable

Teniendo en cuenta el tipo de variables o parámetros con que se cuenta al momento de seleccionar la información secundaria para el desarrollo del estudio, se han seleccionado las técnicas de análisis multivariable más adecuadas; todo ello a partir de los criterios establecidos por Hair, Black, Babin y Anderson (Hair et al., 2014).

Luego de analizar conceptualmente el objeto del estudio, el primer criterio para seleccionar el tipo de técnica pasa por determinar si el propósito es establecer o no relaciones de dependencia o independencia entre las distintas variables (Mira, 2006). La alternativa a este estudio en particular está en las técnicas de interdependencia, donde no existe un grupo de variables que sean definidas como dependientes o independientes, sino que las técnicas que se aplican analizan todas las variables simultáneamente, en este caso las técnicas seleccionadas son el análisis clúster y análisis factorial.

3.3.1 Análisis estadístico básico

Con el fin de tener una mejor noción de los datos obtenidos en los cuatro puntos de muestreo de los cuales se extrajo la información secundaria, se hizo un análisis de las estadísticas básicas como el rango, el mínimo, el máximo, la media, la varianza y la desviación estándar (Mira, 2006). Todos los análisis estadísticos fueron realizados con el paquete estadístico SPSS en su versión 22.

En términos generales los análisis estadísticos básicos nos dan una idea del comportamiento de los parámetros seleccionados en cada uno de los puntos sobre las quebradas La Colosa y La Arenosa. Estos cuerpos de agua evidencian una distribución uniforme para la mayoría de los parámetros analizados como se muestra en la Tabla 2.

La mayoría de los parámetros presentan sus desviaciones estándar con tendencia a cero (0); indicando de esta manera que los datos analizados no se alejan de manera significativa de su promedio. En cuanto al

parámetro COLT, se tiene mayores concentraciones en los puntos aguas abajo con relación a las reportadas en los puntos aguas arriba, siendo el que menos uniformidad presenta en su dispersión.

TABLA 2
Análisis estadístico básico

Parámetro	Rango	Mínimo	Máximo	Media	Desviación estándar	Varianza
CAAr						
TEM	3,6	9,9	13,5	11,403	1,010	1,021
pH	2,00	6,35	8,35	7,208	,558	,311
COND	57,5	32,1	89,6	58,179	14,054	197,513
OD	2,82	5,40	8,22	6,9817	,930	,865
ALKT	11,7	6,1	17,8	9,283	3,409	11,623
DURT	17,4	4,9	22,3	15,808	4,025	16,207
P_TOT	,130	,010	,140	,04206	,0366	,001
SST	9	3	12	4,61	2,615	6,840
DBO	7	2	9	3,11	1,844	3,399
COLT	13992	8	14000	1639,72	3627,06	13155626
CAAb						
TEM	3,4	12,2	15,6	14,156	,8827	,779
pH	1,53	6,68	8,21	7,425	,37259	,139
COND	83,6	109,7	193,3	147,47	24,9066	620
OD	3,09	5,00	8,09	6,8078	1,0057	1,011
ALKT	31,0	32,0	63,0	47,017	10,752	115,621
DURT	33,1	40,6	73,7	57,339	11,214	125,759
P_TOT	,199	,010	,209	,06050	,05059	,003
SST	13	4	17	9,39	3,852	14,840
DBO	5	2	7	3,72	1,526	2,330
COLT	68999	1	69000	6489	16019	256615405

Parámetro	Rango	Mínimo	Máximo	Media	Desviación estándar	Varianza
AAAr						
TEM	3,4	9,9	13,3	11,806	,9502	,903
pH	1,15	7,09	8,24	7,5544	,30157	,091
COND	75,6	31,7	107,3	61,851	19,360	374,821
OD	3,32	4,66	7,98	6,7056	,92639	,858
ALKT	47,4	1,6	49,0	23,233	11,014	121,328
DURT	41,0	9,6	50,5	24,053	12,018	144,450
P_TOT	,109	,010	,119	,05611	,03326	,001
SST	8	3	11	5,11	2,784	7,752
DBO	3	2	5	2,72	1,127	1,271
COLT	5999	1	6000	1142,6	1599	2556951
AAAb						
TEM	5,5	11,0	16,5	13,467	1,2034	1,448
pH	1,73	6,77	8,50	7,7228	,40318	,163
COND	77,1	103,8	180,9	134,58	22,822	520,854
OD	3,08	5,00	8,08	6,7661	1,0139	1,028
ALKT	73,5	1,1	74,6	52,350	18,718	350,387
DURT	54,8	19,1	73,9	55,433	12,889	166,144
P_TOT	,153	,010	,163	,06767	,04099	,002
SST	18	3	21	7,06	4,881	23,820
DBO	4	2	6	3,39	1,290	1,663
COLT	12999	1	13000	2083,3	3464	11999887

Fuente. autor 2018

3.3.2 Análisis de conglomerados (Clúster)

Ahora bien, para detectar la similitud entre los puntos analizados se generó la matriz relacionada en la Tabla 3.

Para una buena interpretación de dicha matriz, es importante precisar que entre menor sea su distancia euclídea (o euclidiana) mayor similitud, en otras palabras, entre más cercano esté el valor a 0 (cero) entre dos puntos analizados mayor es su similitud (Peña, 2002); es decir, que los parámetros analizados tienen un comportamiento similar en relación con otro punto analizado.

TABLA 3

Matriz de proximidades mediante distancia euclídea al cuadrado

Caso	Distancia euclídea al cuadrado			
	1:CAAr	2:CAAb	3:AAAr	4:AAAb
1:CAAr	,000	32,104	11,319	29,357
2:CAAb	32,104	,000	25,732	7,845
3:AAAr	11,319	25,732	,000	13,643
4:AAAb	29,357	7,845	13,643	,000

Esto es una matriz de disimilaridad.

Fuente. autor 2018

En el caso de la Tabla 3, se evidencia que los puntos con mayor similitud son precisamente los puntos aguas abajo entre sí y aguas arriba entre sí. La distancia euclídea al cuadrado entre los puntos CAAb (caso 2) y AAAb (caso 4) es de 7,845 mientras que los puntos CAAr (caso 1) y AAAr (caso 3) es de 11,319. Lo anterior posiblemente se relaciona con el nivel de contaminación de estos puntos, pues es probable que los puntos aguas abajo reporten mayores concentraciones de algunos parámetros, dado que el agua a su paso ha recogido descargas o materiales que afecten el comportamiento de los parámetros analizados (Roldán & Ramírez, 2008).

Para poder jerarquizar los puntos de muestreo según su nivel de contaminación se optó por hacer una comparación de las distancias euclídeas en cada una de las relaciones existentes de la siguiente manera (Peña 2002):

- Para identificar el punto con mayor y menor grado de contaminación sólo se deben buscar los puntos con mayor distancia euclídea, es decir

los que menos similitud tienen entre sí. La mayor distancia euclídea se presenta entre los puntos CAAr y CAAb con un valor de 32,104, lo que indica que el punto CAAr es el que menos grado de contaminación posee; por ende, el punto CAAb debe ser el de mayor grado de contaminación, pues es el de menos similitud respecto al menos contaminado.

- La segunda mayor distancia euclídea se presentó entre los puntos denominados CAAr y AAAb con un valor de 29,357. Como se constató anteriormente el punto CAAr es el de menor grado de contaminación, por lo tanto, el punto AAAb debe ser el segundo más contaminado por tener la segunda mayor distancia euclídea respecto al menos contaminado.
- Aunque luego de tener la distribución de los tres puntos anteriores según su grado de contaminación, es correcto afirmar que el punto restante AAAr es el segundo menos contaminado; sin embargo, esta afirmación se puede sustentar al observar que la distancia euclídea entre el punto menos contaminado (CAAr) y el punto AAAr es la menor entre las distancias posibles que se relacionen con el punto menos contaminado.

En este orden de ideas, la clasificación de los puntos de muestreo del menos contaminado al más contaminado sería la siguiente:

1. Quebrada la Colosa aguas arriba (CAAr)
2. Quebrada la Arenosa aguas arriba (AAAr)
3. Quebrada la Arenosa aguas abajo (AAAb)
4. Quebrada la Colosa aguas abajo (CAAb)

3.3.3 Análisis factorial (extracción por componentes principales)

En la Tabla 4 de varianza total explicada, se observan los autovalores de la matriz de varianzas y del porcentaje de varianza que representa cada uno de ellos. Los autovalores muestran la varianza total que está explicada por cada componente. El programa extrae autovalores iniciales mayores a 1.

En la tabla referenciada hay 2 autovalores mayores que 1 (7,205 y 2,371), por lo que el procesamiento de la información extrae 2 componentes que consiguen explicar un 95,765 % de la varianza de los datos originales. El resultado obtenido de esta tabla es importante para tomar una decisión sobre el número de componentes que deben extraerse (Pérez, 2004).

Una manera gráfica de importancia para tomar la decisión de cuantos componentes se debe extraer, es mediante el análisis de un gráfico de sedimentación que relaciona el número de autovalores con la cantidad de componentes que se pueden extraer mediante distancia euclídea al cuadrado (Pérez, 2004).

En la Figura. 3 se muestra la relación que se explicó en la Tabla 3.

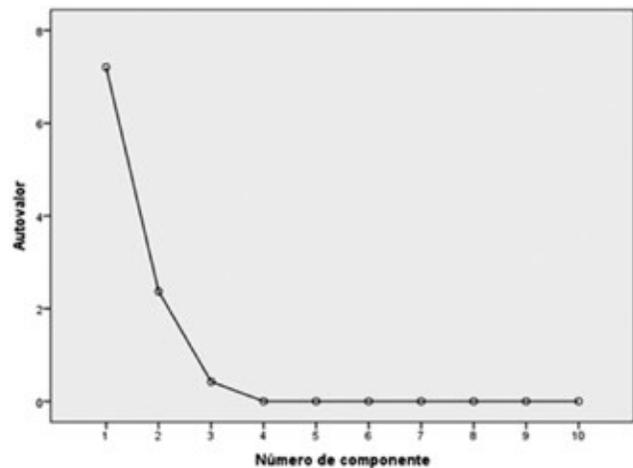


Fig. 3 Gráfico de sedimentación

En el primer componente se extrae el autovalor 7,205 y el segundo el autovalor 2,371. La gráfica muestra que, al extraer el segundo componente se evidencia un cambio brusco en su pendiente lo que da indicio que debemos escoger dos componentes que expliquen la varianza.

La Tabla 5 muestra la solución de la técnica de análisis factorial (extracción por componentes principales), en donde en una nota al pie de la tabla indica que fueron extraídos 2 componentes principales. En el primer componente se muestra que las saturaciones

de cada variable hacen pensar que dicho componente está constituido por los parámetros ZTEM, ZCOND, ZALKT, ZDURT, ZFOST, ZSST, ZDBO y ZCOLT; mientras que el componente dos agrupa los parámetros ZOD y ZpH.

Sin embargo, para una mejor interpretación de los componentes se recomienda hacer una rotación de la matriz como se puede ver en la Tabla 6. La matriz rotada es mucho más clara en las relaciones de los parámetros del componente 2, evidenciando una fuerte

relación positiva con el ZpH y el ZFOST y una fuerte relación negativa con el ZOD que puede tener relación con procesos de demanda de oxígeno disuelto. Para el caso de los componentes extraídos, se puede observar que los parámetros relacionados son en la mayoría los mismos utilizados para la elaboración de índices de contaminación, lo cual nos da pie para hacer esta comparación y, de esta manera, establecer las dimensiones de la contaminación de los cuerpos de agua analizados.

TABLA 4
Varianza total explicada

Componente	Autovalores iniciales			Sumas de extracción de cargas al cuadrado			Sumas de rotación de cargas al cuadrado		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	7,205	72,054	72,054	7,205	72,054	72,054	5,873	58,734	58,734
2	2,371	23,712	95,765	2,371	23,712	95,765	3,703	37,032	95,765
3	,423	4,235	100,000						
4	7,864E-16	7,864E-15	100,000						
5	3,934E-16	3,934E-15	100,000						
6	5,360E-17	5,360E-16	100,000						
7	-4,907E-17	-4,907E-16	100,000						
8	-1,752E-16	-1,752E-15	100,000						
9	-2,985E-16	-2,985E-15	100,000						
10	-9,867E-16	-9,867E-15	100,000						

Método de extracción: análisis de componentes principales.

Fuente. autor 2018

TABLA 5
Componentes principales extraídos

Matriz de componente ^a		
	Componente	
	1	2
ZTEM	,991	,136
ZpH	,575	-,805
ZCOND	,982	,144
ZOD	-,460	,790
ZALKT	,973	-,190
ZDURT	,995	,000
ZFOST	,863	-,500
ZSST	,937	,296
ZDBO	,797	,576
ZCOLT	,723	,596
Método de extracción: análisis de componentes principales.		
a. 2 componentes extraídos.		

Fuente. autor 2018

TABLA 6
Componentes principales rotados

Matriz de componente rotado ^a		
	Componente	
	1	2
ZTEM	,915	,404
ZpH	,067	,987
ZCOND	,912	,392
ZOD	,023	-,914
ZALKT	,729	,673
ZDURT	,847	,523
ZFOST	,472	,879
ZSST	,953	,240
ZDBO	,981	-,072
ZCOLT	,928	-,128
Método de extracción: análisis de componentes principales. Método de rotación: Varimax con normalización Kaiser ^a .		
a. La rotación ha convergido en 3 iteraciones.		

Fuente. autor 2018

3.4 Confrontación de resultados y determinación de aplicabilidad y viabilidad de las técnicas de análisis multivariable seleccionadas

La confrontación de resultados se hizo mediante matrices de relación que pretenden comparar, la clasificación del grado de contaminación de cada punto de muestreo respecto a su ICO, con relación a la distribución obtenida con las técnicas estadísticas multivariable, en este caso particular mediante el análisis de conglomerados (clúster).

Para la correcta interpretación de la Tabla 7, es importante señalar que la clasificación de los puntos de estudio según el grado de contaminación señala el número 1, como el punto menos contaminado y 4 como el más contaminado, tanto para índices de contaminación como para análisis de conglomerados.

Los números de la parte izquierda de cada celda hacen referencia a la clasificación perteneciente a ICOS, mientras que el de la derecha a la clasificación obtenida mediante el análisis de conglomerados.

TABLA 7
Matriz de relación grados de contaminación

		Análisis de conglomerados (clústers)			
Índices de contaminación	Punto de muestreo	CAAr	CAAb	AAAr	AAAb
		CAAr	1		
1					
CAAb			4		
			4		
AAAr				2	
				2	
AAAb					3
					3

Fuente. autor 2018

En la tabla anterior, se puede observar que el análisis de conglomerados es una técnica estadística aplicable y viable para determinar una clasificación coherente de cuerpos de agua según su grado de contaminación. Lo anterior, debido a que como se hace evidente, la clasificación de los puntos de muestreo según su grado de contaminación, es coincidente con la clasificación obtenida mediante el cálculo de índices de contaminación.

Para poder comparar el comportamiento de los parámetros fisicoquímicos y microbiológicos independientemente del punto de muestreo, en la Tabla 8 se relacionan los parámetros empleados por cada índice de contaminación, confrontándolos con los componentes extraídos mediante la técnica multivariable de análisis factorial por componentes principales (AF/CP).

TABLA 8

Matriz de relación agrupación de parámetros

Índice de contaminación relacionado	Parámetros empleados	Componente extraído (AF/CP)	Parámetros extraídos
ICOMI	Alcalinidad total Dureza total Conductividad	Componente 1	Alcalinidad total Dureza total Conductividad
ICOSUS	Sólidos suspendidos totales		Temperatura de muestra Sólidos suspendidos totales
ICOMO	DBO ₅ Coliformes totales % Saturación de oxígeno		DBO ₅ Coliformes totales
ICOpH	pH	Componente 2	pH Oxígeno disuelto Fósforo total
ICOTRO	Fósforo total		

Fuente. autor 2018

De la tabla anterior, se destaca que la técnica de análisis factorial extrae en un primer componente los mismos parámetros que se emplean para la generación del ICOMI, los sólidos suspendidos totales empleados para el cálculo del ICOSUS y los parámetros utilizados para el cálculo del ICOMO (DBO y COLT), así como el parámetro temperatura de muestra que se emplea para el cálculo del porcentaje de saturación de oxígeno y junto con el parámetro oxígeno disuelto y la altitud del punto de muestreo sirven para calcular el ICOMO. En un segundo componente se extrae el parámetro pH utilizado para el cálculo del ICOpH, el fósforo total utilizado para calcular el ICOTRO, así como el parámetro oxígeno disuelto.

Es importante precisar que la técnica de análisis multivariable por componentes principales relacionó el parámetro temperatura de muestra junto con los parámetros que miden la mineralización del agua, pues estos tienen una fuerte relación, dado que los sólidos y sales disueltos por efecto de la radiación solar hacen que la temperatura del agua aumente.

Para el caso particular del parámetro oxígeno disuelto, el análisis factorial por componentes principales no lo extrajo junto con los demás parámetros utilizados para el cálculo del ICOMO, esto se puede explicar porque dicho índice no utiliza el parámetro como tal, sino el porcentaje de saturación de oxígeno, en donde también son necesarias la temperatura de la muestra y la altura sobre el nivel del mar del punto en que se toma la muestra.

El resultado final de este análisis comparativo muestra que tras el cálculo los índices de contaminación propuestos por Ramírez y Viña, aplicables para el territorio colombiano, sugieren el estado real de los cuerpos de agua en estudio, siendo acertados, en la medida en que al confrontarlos con los resultados obtenidos mediante las técnicas estadísticas de análisis multivariable seleccionadas se asemejan en gran medida entre sí.

4. CONCLUSIONES

El comportamiento individual de cada variable analizada, señaló que en términos generales los puntos aguas abajo presentaron mayores concentraciones en los parámetros seleccionados que los puntos aguas arriba, especialmente aquellos relacionados con sales disueltas y sólidos; esto tiene relación con que el agua arrastra a su paso materiales ricos en estos parámetros haciendo que aguas abajo se tenga una mayor carga de los mismos.

Los índices de contaminación permitieron generar una clasificación de los puntos de muestreo de menor a mayor grado de contaminación, siendo el punto CAAR el menos contaminado, seguido por el punto AAAr, el punto AAAb y por último el punto CAAb.

El análisis clúster evidenció que los puntos que presentan mayor similitud son precisamente los puntos aguas abajo entre sí y aguas arriba entre sí. Con esta técnica multivariable también se logró jerarquizar los puntos de muestreo según su nivel de contaminación a partir de la comparación entre las distancias euclídeas al cuadrado. Lo anterior permitió clasificar los puntos de muestreo del menos contaminado al más contaminado, siendo el punto CAAR el menos contaminado, en segundo lugar, el punto AAAr, seguido por el punto AAAb y por último el punto CAAb.

La segunda técnica de análisis multivariable aplicada fue el análisis factorial, la cual se empleó para ver la manera en que los parámetros analizados se relacionan entre sí, arrojando como resultado que el comportamiento de los parámetros es similar entre aquellos relacionados con el nivel de mineralización del agua. Del mismo modo se encontró relación en el comportamiento entre la DBO_5 y los coliformes totales que son parámetros relacionados con la presencia de materia orgánica en el medio.

Se pudo determinar que la clasificación de los puntos de muestreo según su grado de contaminación, por medio de cluster es coincidente con la clasificación obtenida mediante el cálculo de índices de contaminación.

La técnica de análisis factorial con extracción por componentes principales, fue ideal para determinar la relación en el comportamiento que existe entre los parámetros seleccionados, dado que en un primer componente se pudo extraer los mismos parámetros que se emplean para la generación del ICOMI, los sólidos suspendidos totales empleados para el cálculo del ICOSUS y los parámetros utilizados para el cálculo de ICOMO. También, relacionó el parámetro temperatura de muestra junto con los parámetros que miden la mineralización del agua, pues estos tienen una fuerte relación dado que los sólidos y sales disueltas por efecto de la radiación solar hacen que la temperatura del agua ascienda. En un segundo componente se extrae el parámetro pH utilizado para el cálculo del ICOPH, el fósforo total utilizado para calcular el ICOTRO, así como el parámetro oxígeno disuelto.

Se concluye que, este análisis comparativo muestra el estado real de contaminación de los cuerpos de agua La Colosa y La Arenosa, dado que, tras el cálculo de los índices de contaminación propuestos por Ramírez y Viña, se encontraron grandes semejanzas al confrontarlos con los resultados obtenidos mediante las técnicas estadísticas de análisis multivariable seleccionadas.

REFERENCIAS

- Antek S.A. (2014). Reportes de resultados de laboratorio. Bogotá.
- Arango Mendoza, J. (2014). Proyecto de minería de oro La Colosa, identificación ambiental de la zona de explotación y sus impactos. Tesis Maestría en Gestión Ambiental. Bogotá: Pontificia Universidad Javeriana.
- Cortolima (2009). *Plan de acción trienal 2007-2009*. Recuperado de: <https://www.cortolima.gov.co/plan-acci%C3%B3n-trienal-2007-2009>
- FAO (2009). ¿Por qué invertir en ordenación de las cuencas hidrográficas? Organización de las Naciones Unidas para la Agricultura y la Alimentación, Roma. Recuperado en: <http://www.fao.org/3/a1295s/a1295s00.htm>
- González Leal, G. R. (2012). *Microbiología del agua. Conceptos y aplicaciones*. Bogotá: Escuela Colombiana de Ingeniería.
- Hair, J. F., Black, W. C., Babin, B. J. & Anderson, R. E. (2014). *Multivariate Data Analysis*, Séptima Ed. New York: Pearson New International Edition.

- Hernández Sampieri, R., Fernández Collao, C. & Baptista Lucio, P. (1997). *Metodología de la investigación*, México: Mc Graw-Hill.
- Ibar Alonso R. & García Centeno, M. del C. (2002). El análisis multivariante en la enseñanza universitaria. Departamento de Métodos Cuantitativos para la Economía. Universidad de San Pablo, CEU. Recuperado de: https://www.researchgate.net/publication/26428288_El_Analisis_Multivariante_en_la_ensenanza_universitaria
- IBM (2013). IBM SPSS Statistics 22 Core System. Recuperado de: file:///C:/Users/USUARIO/Downloads/IBM_SPSS_Statistics_Core_System_User_Guide.pdf
- Manrique, F., Manrique, D., Manrique, R. & Tejedor, M. (2007) Contaminación de la cuenca alta del río Chicamocha y algunas aproximaciones sobre la salud humana. *Rev Salud Hist y sanidad*, 2 (1), 03-13. Recuperado de: <https://repositorio.uptc.edu.co/bitstream/001/1352/1/RED-124.pdf>
- Mira Olarte, L. C. (2006). *Análisis estadístico multivariado de los parámetros de calidad del agua en vertimientos urbanos en la ciudad de Bogotá*. Bogotá: Universidad de los Andes.
- Peña, D. (2002). *Análisis de datos multivariantes*. Madrid: McGraw Hill.
- Pérez, J. R. (2004). Que es el análisis multivariante. *Psicología Experimental*, Recuperado de: <https://psicologiaexperimental.files.wordpress.com/2010/03/analisis-factorial.pdf>
- Ramírez, R. & Cardeñoso, M. (1999). Índices de contaminación para caracterización de aguas continentales y vertimientos. Formulación, *CT&F - Ciencia, Tecnología y Futuro*, 1(5), 89-99.
- Ramírez A. & Viña, G. (1998). *Limnología colombiana: aportes a su conocimiento y estadísticas de análisis*. Bogotá: Universidad Jorge Tadeo Lozano.
- Rodríguez Rodríguez, M., Benavente Herrera, J., Cruz Sanjulián, J. J. & Torres Ruiz, F. (2000). Análisis multivariable aplicado a datos hidroquímicos de las zonas húmedas de la provincia de Málaga. *Geogaceta*, 28, 125-128.
- Roldán Pérez G. & Ramírez Restrepo, J. J. (2008). *Fundamentos de limnología neotropical*, vol. 2. Medellín: Universidad de Antioquia.
- Romero, J. A. (2009). *Calidad del agua*. Bogotá: Escuela Colombiana de Ingeniería.
- Tápanes, J. J. (1974). Hidrología de la desembocadura del Caucho y regiones adyacentes. *Voluntad Hidráulica*, 32, 10-12.
- Vergara, C., González, A. & González, C. (2013). Evaluación de impacto ambiental y estudios previos a una valoración contingente. Caso La Colosa, Cajamarca, Tolima, Colombia. *Ensayos de Economía*, 42, 191-222, 2013.

