



APLICACIÓN DEL ANÁLISIS FACTORIAL (AF), EN EL ESTUDIO DE LA CALIDAD DE AGUA

¹ Gonzalo I. Carrasco O. y ^{2,3} Marisela del C. Castillo G.

¹Universidad de Panamá, Facultad de Ciencias Naturales Exactas y Tecnología, Departamento de Estadística. Centro de Investigación y Consultoría Estadística.
E-mail: gcarrasco27@hotmail.com

²Universidad de Panamá, Facultad de Ciencias Naturales Exactas y Tecnología. Programa de maestría en Estadística Aplicada. E-mail: Microcystissp@hotmail.com

³Universidad de Panamá, Facultad de Ciencias Naturales Exactas y Tecnología, Programa de Maestría en Microbiología Ambiental.
E-mail: Microcystissp@hotmail.com

RESUMEN

La calidad del agua está definida por una serie de características físicas, químicas y biológicas. En los conjuntos de datos de calidad de agua, las variables o grupos de variables se relacionan, y desde magnitudes diferentes presentan similitudes. Esto podría ser porque más de una se rige por el mismo principio, permitiendo evaluar el funcionamiento en un sistema acuático. En el presente estudio se simplificó el conjunto de variables: temperatura, pH, transparencia, turbiedad, nitratos, ortofosfatos, fósforo y nitrógeno total, clorofila a, radiación solar, oxígeno disuelto y microcistinas; que determinan la calidad de agua en los sitios de muestreo, Gamboa (643343E; 1007468N) y Paraíso (651144E; 997879N), ubicados en el embalse Gatún, con el objetivo de encontrar grupos homogéneos de variables que expliquen la calidad de agua desde una dimensión reducida. Esto se realizó mediante la aplicación del análisis factorial como método de análisis multivariante. Se realizaron mediciones en campo y se colectaron muestras para análisis de laboratorio. Los muestreos se realizaron durante febrero a mayo y julio a diciembre de 2009. Un total de doce (12) variables de calidad de agua fueron medidas. Los resultados sugieren que la aplicación del análisis factorial extrae cinco (5), factores que consiguen explicar un 73.06% de la varianza de los datos originales, con lo cual se logra reducir y resumir doce variables a cinco factores que representan las nuevas variables de calidad de agua.

PALABRAS CLAVES

Calidad de agua, análisis factorial, homogeneidad de variables.

ABSTRACT

Water quality is defined by a series of physical, chemical and biological characteristics. In a group of data with several water quality variables, the variables or the set of variables often change or move together. This may occur because more than one variable is governed by the same principle and they may be measuring the behavior of a system. This study simplified a set of variables that determine water quality in two sampling points located in Gatun Lake, to find homogeneous sets of variables that better explain water quality at the sampling points. This was conducted by applying factor analysis as a multivariate analysis method. Field analyses were performed and samples were collected for laboratory analysis. Sampling was conducted during the periods from February to May and July to December 2009. A total of twelve (12) water quality variables were measured. Results show that the multivariate analysis method extracted five (5) factors that explain 73,06% of the variance from the original data; a reduction from twelve to five variables was achieved where the new five variables represent the new water quality variables.

KEYWORDS

Water quality, factorial analyses, homogeneity of variables.

INTRODUCCIÓN

El agua dulce es un recurso limitado y su calidad está bajo presión constante. Calidad de agua es el término utilizado para describir las condiciones físicas, químicas y biológicas que tiene un cuerpo de agua en un punto y tiempo dado (ANAM, 2006). Concentraciones de diferentes elementos, tales como, metales pesados, hidrocarburos, nutrientes, detergentes, temperatura, pH, turbiedad, oxígeno disuelto, demanda bioquímica de oxígeno, sólidos en suspensión, entre otros, son ejemplos de variables que determinan la calidad del agua.

Ventajosamente, los conjuntos de variables de calidad de agua a menudo “se mueven juntas”; es decir, comparten información desde dimensiones diferentes. Por ejemplo, la turbiedad, la transparencia del agua y los sólidos en suspensión son variables disímiles con unidades

diferentes, que sugieren acerca de la “claridad del agua”. El color y la turbiedad son dos conceptos que difieren entre sí, se tratan conjuntamente, ya que interfieren en la trasmisión de la luz en las aguas naturales y, por consiguiente, regulan los procesos biológicos que en ella se realizan (Roldan, 2008).

La disponibilidad de equipos de instrumentación sofisticados para medición y análisis de calidad de agua, permite medir decenas de variables, en ocasiones de forma simultánea. Cuando esto sucede, se puede tomar ventaja de esta redundancia de información para simplificarla mediante la aplicación de análisis estadísticos multivariantes que sustituyan un grupo de variables por una nueva y sola variable. La aplicación de análisis multivariantes resultan ser métodos estadísticos requeridos para determinar la contribución de varios factores (visualizar los datos que poseen muchas variables), en un simple evento o resultado (Pérez, 2004). Algunas de las técnicas de reducción de dimensiones que ofrecen los paquetes estadísticos son: el análisis de componentes principales, escalamiento óptimo, análisis de correspondencias y el análisis factorial.

En el presente estudio, se utilizó el análisis factorial porque los datos utilizados cumplían con los supuestos requeridos del modelo, además este, permite hacer rotaciones que facilitan la interpretación de las variables en caso de que la solución inicial no es del todo clara y porque la mayoría de los paquetes estadísticos disponibles cuentan con este análisis. El análisis factorial (AF), se suele utilizar en la reducción de los datos para identificar un pequeño número de factores que explique la mayoría de la varianza observada en un número mayor de variables manifiestas (Pérez, 2008).

El objetivo de esta investigación era encontrar grupos homogéneos en variables de calidad de agua de dos sitios del embalse Gatún (Gamboa y Paraíso), partiendo de doce mediciones realizadas durante el 2009. Se busca seguir el principio de parsimonia, el cual establece que la solución más simple suele ser la mejor.

MATERIALES Y MÉTODOS

Los registros utilizados para el análisis, corresponden a las mediciones de los parámetros físico-químicos y biológicos que determinan la calidad de agua en los sitios de muestreo, Gamboa (643343; 1007468) y Paraíso (651144; 997879), ubicados en el embalse Gatún.

Se colectaron tres (3), muestras en cada sitio a un metro de la superficie. El muestreo se realizó de manera sistemática cada siete días hasta completar los catorce primeros días del mes; durante febrero a mayo y julio a diciembre de 2009. Se estableció un horario de muestreo entre las 11:00 am a 2:30 pm. En campo, se realizaron mediciones *in situ* con la sonda multiparamétrica YSI, modelo 556 MPS la cual fue previamente verificada y/o calibrada en laboratorio. Se emplearon envases apropiados para cada muestra (vidrio, plástico de polipropileno, etc.). Las muestras colectadas fueron puestas en neveras con hielo hasta su transporte al laboratorio. El acceso a los sitios se realizó vía terrestre.

Se colectaron muestras para realizar determinaciones analíticas en el laboratorio de la Unidad de Calidad de Agua, División de Agua del Canal de Panamá, además, muestras para análisis de nitrógeno y fósforo total fueron colectadas y preservadas (H_2SO_4 al 20%), enviadas y analizadas en el laboratorio de la Universidad Tecnológica de Panamá, contratado por el canal de Panamá.

Las metodologías de análisis corresponden a las descritas en el Standard Method for the Examination of Water and Wastewater (APHA, AWWA, WEF, 21ª Edición). (Cuadro 2). Un total de doce (12), parámetros de calidad de agua, los cuales incluyen: temperatura ($^{\circ}C$), pH, transparencia (m), turbidez (NTU), nitratos (mg/l), ortofosfato (mg/l), fósforo total (mg/l), nitrógeno total (mg/l), clorofila ($\mu g/l$), radiación solar (Langley), toxina (microcistina, $\mu g/l$), y oxígeno disuelto (mg/l), fueron medidos en campo y laboratorio.

Cuadro 2. Métodos analíticos empleados según parámetros de calidad de agua.

Parámetro	Método	Unidades de medida	Límite detección
¹ Oxígeno disuelto	SM 4500-O C	mg/l	0,1 mg/l
N-Nitratos	SM 4500-NO ₃ - E	mg/l	0,01 mg/l
P-Fosfatos	SM 4500-P E	mg/l	0,02 mg/l
Fósforo total	SM 4500-P D	mg/l	0,03 mg/l
Nitrógeno total	SM 4500-Norg-B	mg/l	0,03 mg/l
¹ pH	SM 4500-H+B	pH units	0,1 pH units
¹ Temperatura	SM 2550-A	°C	0,1°C
Turbiedad	SM 2130-B	NTU	0,05 NTU
¹ Transparencia	Profundidad de Secchi	m	0,1m
Clorofila a	SM 10200-H, modificación USEPA 445.0	µg/l	0,05 µg/l

¹ Mediciones *in situ*.

Las muestras para análisis de microcistinas fueron colectadas en envases de vidrio ámbar, y las mediciones se realizaron utilizando el método Quanti tube kit ET039 de Envirologix. Los registros de radiación solar fueron suministrados por la Sección de Recursos Hídricos de la División de Agua del Canal de Panamá, estos corresponden a la estación SRFAA, que se encuentra ubicada en Balboa, encima de la caseta próxima al mercado de abastos.

Se utilizó el análisis factorial como método de análisis multivariante; para reducir los datos a partir de las correlaciones y encontrar un número reducido de factores que lo expliquen. Las etapas comprendían la medición y captura de variables, posteriormente se computaron las correlaciones y se arreglaron en una matriz para extraer los factores.

Para una mejor interpretación de los resultados se efectuó la rotación ortogonal, varimax.

Se realizó una estandarización debido a que las variables de calidad de agua tienen distintas unidades. Todas las variables analizadas son cuantitativas continuas.

RESULTADOS Y DISCUSIÓN

Estadística descriptiva

Al evaluar la dispersión de los datos originales mediante diagramas de caja para las variables de calidad de agua, se puede indicar que la radiación solar y el oxígeno disuelto son las características que presentan mayor variabilidad en los registros (Fig. 1 y 2). Nótese que las variables fósforo y nitrógeno total, además de la turbiedad, clorofila y microcistinas presentan valores atípicos y extremos (Fig. 1 y 2).

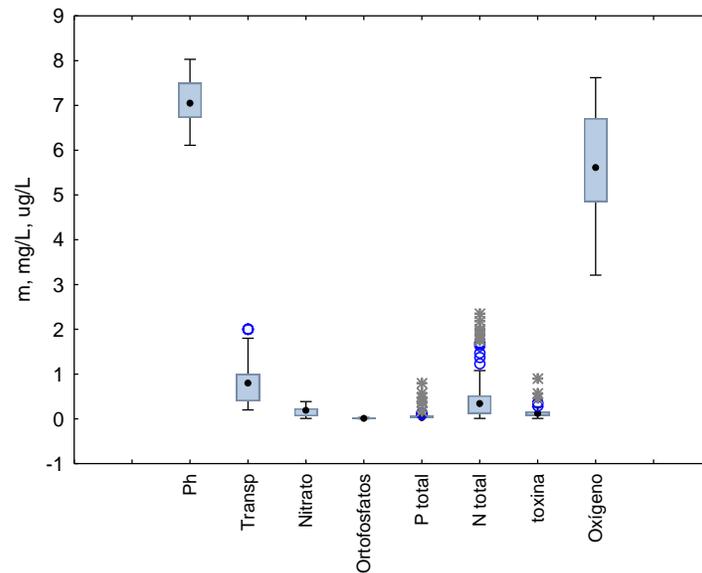


Fig. 1. Diagrama de caja de las variables de calidad de agua. Febrero-mayo y julio-diciembre, 2009.

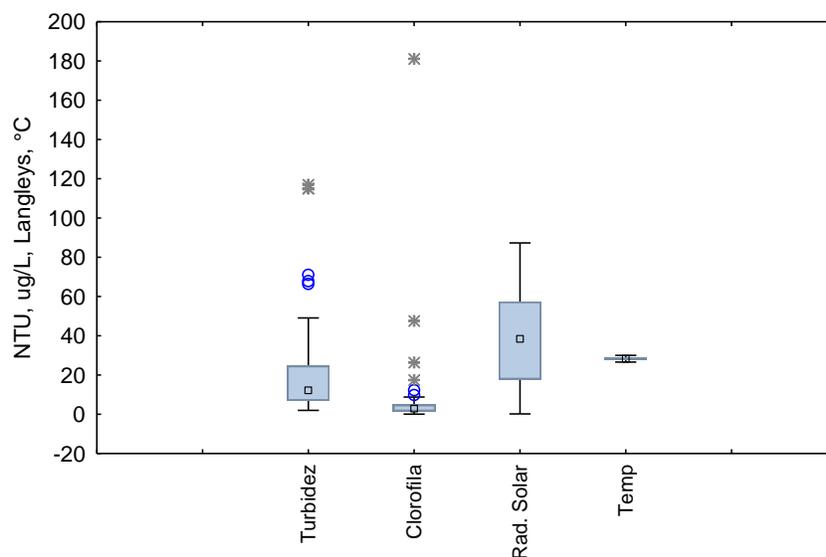


Fig. 2. Diagrama de caja de las variables de calidad de agua. Febrero-mayo y julio-diciembre, 2009.

El análisis factorial es conveniente cuando la matriz contiene grupos de variables que están fuertemente correlacionadas entre sí y donde los niveles de significancia entre las variables son significativos.

Para tener una idea preliminar de que el análisis factorial es aplicable; se evaluó la matriz de correlaciones. Esta nos indica que el diseño factorial es aplicable, ya que el determinante de la matriz está próximo a cero (0.008), lo que sugiere correlación entre las variables.

La prueba de esfericidad Bartlett contrasta la hipótesis nula de que la matriz de correlaciones observada es en realidad la matriz identidad (Cuadro 1). Como el valor p, es menor que el alfa (establecido), 0,05, se rechaza la hipótesis nula de que la matriz de correlaciones observada sea la matriz identidad. Por otro lado, la medida de bondad de ajuste KMO (Kaiser, Meyer, Olkin), se encuentra entre 0.5 y 0.75 por lo que la idea del análisis factorial es aceptable para este estudio.

Cuadro 1. Estadístico KMO y prueba de Bartlett para las variables de calidad de agua. Febrero-mayo y julio-diciembre, 2009.

KMO y prueba de Bartlett		
Medida de adecuación muestral de Kaiser-Meyer-Olkin.		0.608
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	550.813
	gl	66
	Sig.	Tienda a cero

Cuadro 2. Comunalidades de la solución factorial para las variables de calidad de agua. Febrero-mayo y julio-diciembre, 2009.

Variables	Extracción
Temp.	0.794
pH	0.731
Transparencia	0.762
Turbidez	0.827
Nitrato	0.879
Ortofosfato	0.743
P total	0.688
N total	0.572
Clorofila	0.729
Rad. Solar	0.667
toxina	0.645
Oxígeno	0.73

Para llegar a la solución factorial, se ha utilizado el método de extracción denominado componentes principales. En el cuadro 2 se muestran las comunalidades asignadas inicialmente a las variables; esta comunalidad es la proporción de su varianza que puede ser

Gonzalo I., C. O. y M. del C. Castillo G.

explicada por el modelo factorial obtenido. Nótese, que las variables: P total (fósforo total), N total (nitrógeno total), Rad Solar (radiación solar), y la toxina son las menos explicadas; ya que el modelo sólo es capaz de reproducir el 68.8%, 57.2%, 66.7% y 64.5% respectivamente de su variabilidad original.

En el cuadro 3 de varianza total explicada se observa que hay 5 autovalores mayores que uno por lo que el procedimiento extrae 5 factores que consiguen explicar un 73.06% de la varianza de los datos originales. Con esta información podríamos pensar sobre el número idóneo de factores que deben extraerse y cumplir con el objetivo de reducir el número de dimensiones necesarias para explicar los datos.

Cuadro 3. Varianza total explicada de la solución factorial para las variables de calidad de agua. Febrero-mayo y julio-diciembre, 2009.

Componente	Varianza total explicada					
	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	3.667	30.557	30.557	3.667	30.557	30.557
2	1.556	12.963	43.520	1.556	12.963	43.520
3	1.326	11.053	54.573	1.326	11.053	54.573
4	1.210	10.086	64.659	1.210	10.086	64.659
5	1.008	8.400	73.059	1.008	8.400	73.059
6	.907	7.562	80.622			
7	.725	6.038	86.659			
8	.589	4.910	91.569			
9	.474	3.950	95.519			
10	.249	2.075	97.594			
11	.176	1.467	99.060			
12	.113	.940	100.000			

Método de extracción: Análisis de Componentes principales.

Comparando las saturaciones relativas de cada variable en cada uno de los 5 factores podemos indicar que el primer factor esta constituido por las variables pH, transparencia, turbidez, nitrato, ortofosfato y oxígeno, todas ellas saturan en un único factor porque constituyen un grupo

diferenciado de variables dentro de la matriz de correlaciones. El segundo factor recoge al grupo de variables N total (nitrógeno total), P total (fósforo total), y radiación solar. En el tercer, cuarto y quinto factor tenemos a las variables: temperatura, toxina y clorofila respectivamente conformadas por una sola variable y que se supone son independientes entre sí (Cuadro 4).

Cuadro 4. Matriz de componentes para las variables de calidad de agua. Febrero-mayo y julio-diciembre, 2009.

Matriz de componentes					
Variables	Componente				
	1	2	3	4	5
Temperatura	-.032	-.040	.827	.141	-.297
pH	-.649	.233	.281	-.381	.177
Transparencia	-.732	.073	-.349	.234	-.213
Turbidez	.712	.354	.264	-.290	.203
Nitrato	.883	-.202	.136	-.189	-.068
Ortofosfato	.820	-.066	-.221	-.044	.122
P total	.341	.491	.290	.396	-.298
N total	.087	-.733	.003	-.163	-.017
Clorofila	-.021	-.180	.245	.308	.736
Rad. Solar	.256	.668	-.289	.110	.244
toxina	.093	-.239	.071	.742	.157
Oxígeno	-.743	.098	.270	-.169	.260

Método de extracción: Análisis de componentes principales.

a. 5 componentes extraídos

En ocasiones, los factores obtenidos en los diferentes tipos de análisis factorial no se traducen en una agrupación fácilmente interpretable de variables con significado; en este caso, la rotación mejora el significado e interpretación de los factores obtenidos originalmente (Martínez, Sánchez & Faulin, 2009). Se procedió a realizar la rotación

varimax, rotación ortogonal que asume que los factores generados son independientes unos de otros (no están correlacionados).

En la figura 2 se observa que lo ideal sería escoger dos factores, ya que el cambio o punto de inflexión se acentúa en el factor dos, sin embargo, se considera que la cantidad de varianza explicada por estos dos factores es muy baja (43.52%), por lo tanto, se escoge a los 5 factores que agrupan a las variables, explicando así el 73.06% de la varianza.

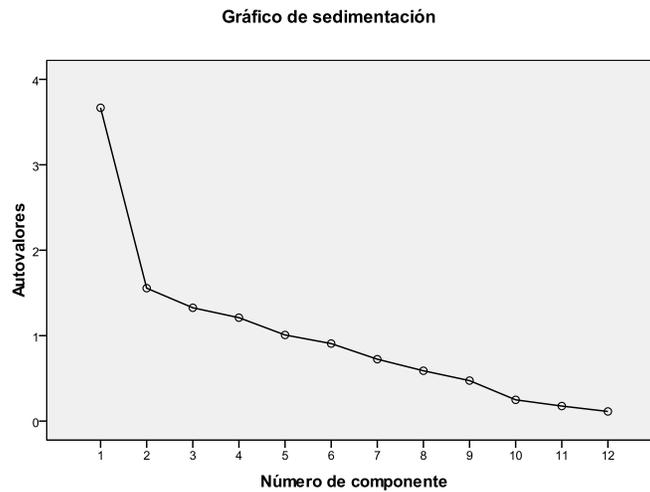


Fig. 2. Gráfico de sedimentación de la solución factorial. Febrero-mayo y julio-diciembre, 2009.

Rotación

Comparando ambas soluciones factoriales: sin rotación y con rotación, podemos señalar que esta última ha mejorado en la saturación de las variables agrupadas. Por ejemplo, originalmente la variable pH estaba contenida en el primer factor con un puntaje de -0.649, después de la rotación esta ha sido colocada en el segundo factor con una puntuación de 0.836, es decir, ha mejorado.

Después de la rotación, se excluyó a las variables pH y oxígeno del primer factor; la agrupación resultante contempla a las variables:

turbidez, transparencia, nitratos y ortofosfato. Esta agrupación puede ser etiquetada como “aparición física del agua y su componente inorgánico” tratándose la aparición física del agua como una condición visual apreciable (transparente, clara, turbia, muy turbia), y su componente inorgánico, como la carga de iones (en este caso: nitratos y ortofosfatos), asociada a las partículas en suspensión. Las mayores puntuaciones en este factor corresponden a las variables turbidez y transparencia (0.846 y -0.840 respectivamente), éstas son variables físicas medidas en el agua. El efecto es causado por una gran variedad de materiales en suspensión, que varían en tamaño desde dispersiones coloidales hasta partículas gruesas, entre otros, arcillas, limo, materia orgánica e inorgánica finamente dividida, organismos planctónicos, microorganismos, etc. Por otra parte, los nutrientes, tales como nitratos y ortofosfato; en un cuerpo de agua, se presentan y aumentan considerablemente debido al arrastre de sedimentos por las lluvias en los suelos erosionados, por el vertimiento de contaminantes domésticos e industriales y por las actividades agrícolas que se den alrededor del embalse o de la cuenca.

El factor 2, el cual está compuesto por las variables pH y oxígeno, las cuales se encontraban en el factor 1 (sin rotación), son indicadores inmediatas (*in situ*), de las “condiciones de vida acuática” Estas variables (oxígeno y pH), son importantes para el soporte adecuado, protección y conservación de la vida acuática (Romero, 1999).

El factor 3, contiene a las variables radiación solar, nitrógeno total y fósforo total; esta agrupación nos revela variables que explican una “condición de eutrofización”. El concepto de eutrofización puede definirse como el proceso de sobreproducción de algas y macrófitos (que utilizan la energía solar para su fotosíntesis), en cuerpos de agua

enriquecidos con nutrientes; principalmente, nitrógeno y fósforo (Carlson, 1977). La eutrofización puede producir dificultades en los cuerpos de agua destinados a diversos usos: suministro de agua potable por alteración de sus propiedades organolépticas (olor, sabor), corrosión del equipo hidroeléctrico y distintos trastornos en los procesos de tratamiento potabilizador por disminución del oxígeno,

obstrucción en las bombas de succión de agua por exceso de vegetación acuática, etc. Aunque es un proceso que, de forma lenta, puede tener un origen natural, hoy en día es, fundamentalmente, de carácter cultural, acelerado por el aporte continuo de nutrientes (nitrógeno y fósforo total), de origen antropogénico.

El factor 4, esta conformado por la única variable: temperatura la cual es una variable que tiene implicaciones importantes en los procesos fisiológicos en los organismos, tales como metabolismo, respiración, fotosíntesis, y patrones de actividad y comportamiento. En este sentido, organismos de vida libre en las aguas naturales de ambientes relativamente estables aún considerando fluctuaciones en la temperatura, experimentan un amplio rango de temperatura, que varía con los periodos diurnos y de estacionalidad. El quinto factor, esta compuesto por las variables clorofila a y toxina (microcistina), se considera a esta agrupación como “potencial de riesgo” sobre la base de un crecimiento excesivo de microorganismos llamados algas y cianobacterias, organismos que poseen clorofila a; que pueden estar predominando en el cuerpo de agua. Algunos géneros de cianobacterias, tienen la capacidad de producir potentes toxinas, en este caso la microcistina, cuya presencia en el agua puede tener implicaciones negativas en cuanto a la salud del mismo ecosistema como al uso que se le de al agua.

Cuadro 5. Matriz de componentes rotados para las variables de calidad de agua. Febrero-mayo y julio-diciembre, 2009.

Variables	Matriz de componentes rotados ^a				
	Componente				
	1	2	3	4	5
Turbidez	.846				
Transp.	-.840				
Nitrato	.766				
Ortofosfato	.597				
pH		.836			
Oxígeno		.787			
Rad. Solar			.743		
N total			-.729		
P total			.551		
Temp.				.861	
Clorofila					.819
toxina					.633

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización Varimax con Kaiser.

a. La rotación ha convergido en 10 iteraciones.

CONCLUSIONES

Los métodos estadísticos multivariados, entre ellos análisis de factor y componentes principales se pueden utilizar para comprender la naturaleza compleja en cuestiones de calidad del agua y determinar variables que establezcan prioridades para mejorar la calidad del agua.

De un total de 12 variables originales se pudo reducir la base de datos a 5 factores que explican todas las variables de manera agrupada. Con esto se logra el objetivo de reducir la dimensión.

El análisis factorial es una herramienta muy poderosa que permitió analizar y agrupar a las variables de acuerdo a una explicación en

común. Las nuevas variables (factores), marcan escenarios naturales que ocurren en el ambiente, por ejemplo: los factores 1, 3 y 5, (aparición física del agua y sus componente inorgánico, condición de eutrofización y potencial de riesgo), agrupan variables que explican como se acarrea partículas (en este caso: iones: nitratos y ortofosfato), a través de la erosión y escorrentía a un cuerpo de agua (factor 1). Por otro lado, como la entrada de nutrientes, favorecida por la luz solar influye en el crecimiento excesivo de vegetación acuática lo que induce al proceso de eutrofización (factor 3), y como una concentración elevada de clorofila nos indica una sobrepoblación del fitoplancton que puede poner en riesgo la calidad del agua y por ende la salud del ecosistema (factor 5). Mientras que los factores 2 y 4, son más relacionados a variables que indican condiciones para el sostenimiento de la vida acuática (biota), desarrollada en el ecosistema.

REFERENCIAS

Autoridad Nacional del Ambiente (ANAM). 2006. Manual de curso de capacitación, Elaboración de las normas de calidad ambiental para aguas naturales. Calidad de aguas naturales características principales, su gestión y control. Contrato BID N° PAN61 – 2005, CSI Ingenieros. Panamá. Pág. 15.

Carlson, R. E., 1977. A trophic state index for lakes. Limnological Research, University of Minnesota, Minneapolis.

Eaton, D., S. Clesceri, A. Lenore & B. Grenn. 1999. Standard Methods for the Examination of Water and Wastewater. 21st Edition. Publication office: American Health Association 1015 Fiftieth Street, NY Washington, DC 20005.

Martínez, M. A., A. Sánchez & J. Faulin. 2009. Bioestadística Amigable. 2^{da} edición Ediciones Díaz de Santos. España, 919 págs.

Pérez, C. 2008. Técnicas de Análisis Multivariantes de Datos, Aplicaciones con SPSS. Pearson Education, S. A., Madrid, 2004.646 págs.

Pérez., C. 2004. Técnicas de Análisis Multivariantes de Datos., Aplicaciones con SPSS., Pearson, Prentice Hall. España. 646 págs.

Roldan, G. 2008. Fundamentos de limnología neotropical., Segunda edición. Editorial Universidad de Antioquia. Colombia.

Romero, J. 1999. Calidad del Agua. Segunda Edición. Editorial Escuela Colombiana de Ingeniería. 276 págs.

Recibido febrero de 2011, aceptado octubre de 2012.