

Determinación de los parámetros ambientales de mayor incidencia en las aguas residuales de la comunidad universitaria Antonio Maceo

Determination of environmental parameters of highest incidence in the wastewater of the university community Antonio Maceo

*MSc. Sandra Crombet-Grillet^I, Dr. C. Arelis Abalos-Rodríguez^{II},
Dr. C. Larisa Zamora-Matamoros^{III}*

scrombet@uo.edu.cu, aabalos@uo.edu.cu, larisa@uo.edu.cu

^I Departamento de Química, Facultad de Ciencias Naturales y Exactas, Universidad de Oriente, Cuba; ^{II} Vicerrectoría de Institucionalización, Universidad de Oriente, Cuba; ^{III} Departamento de Matemática, Facultad de Ciencias Naturales y Exactas, Universidad de Oriente, Cuba.

Recibido: 23 de octubre del 2018

Aprobado: 27 de octubre del 2018

Resumen

En este trabajo se determinan los parámetros ambientales de mayor incidencia en las aguas residuales domésticas de la comunidad universitaria, empleando la técnica de Componentes Principales y los programas estadísticos Statsoft Statistica v7.0.61.0 y IBS SPSS Statistics 22. Las aguas residuales de la comunidad universitaria Antonio Maceo se caracterizan por una gran variabilidad en los valores de sus parámetros físico-químicos. Fueron concluyentes dos componentes principales y a partir de ellos se obtuvo que los parámetros (DQO, DBO₅, turbidez, HS⁻, NH₄⁺, AyG, ST y STF) logran explicar el 87,31 % de la variabilidad total contenida en las aguas de la residencia.

Palabras clave: aguas residuales domésticas, componentes principales, parámetros medioambientales.

Abstract

In this work the environmental parameters of bigger incidence are determined in domestic wastewater of the university community, using the technique of Principals Components and the statistics programs StatSoft Statistica v7.0.61.0 and IBS SPSS Statistics 22. The wastewater from the university community Antonio Maceo is characterized by a great variability in the values of his physic-chemistry parameters. Two principal's components were conclusive and as from them it was obtained than parameters COD, BOD₅, pH, conductivity, turbidity, HS⁻, NH₄⁺, ST and STF are able to explain the 87,31 % of the total variability contained in residence's waters.

Keyword: domestic wastewater, principal's components, environmental parameters.

Introducción

El Análisis de Componentes Principales (ACP) es una técnica que se emplea, en los análisis medioambientales, para reducir el número de variables a utilizar para caracterizar un ecosistema. Pertenece a un grupo de técnicas estadísticas multivalentes [1]. Es una técnica eminentemente descriptiva, de síntesis de la información o reducción de la dimensión (número de variables) [2]. Para el análisis se parte de una matriz de datos que contiene los valores de las variables investigadas en n individuos. Para la determinación del número de componentes deben tenerse en cuenta tres criterios según Gonzáles [3].

Porcentaje explicado: consiste en fijar un porcentaje de variabilidad explicado, por ejemplo el 90 %, y considerar las sucesivas componentes principales hasta superar el porcentaje prefijado.

Criterio de Kaiser: se excluyen aquellas componentes cuyos autovalores sean menores que $\lambda = \text{tr}(\mathbf{S})/p$, siendo \mathbf{S} el estimador de la matriz de varianzas y covarianzas de las p variables investigadas, o bien menores que 1 si se han calculado las componentes a partir de \mathbf{R} , matriz de correlaciones entre las p variables, λ representa la varianza total que explica la componente y r representa el coeficiente de correlación múltiple, es equivalente al criterio basado en minimizar la traza de la matriz de varianzas-covarianzas de las variables descartadas.

Modificación de Jolliffe: se ha comprobado que cuando $p \leq 20$ el criterio de Kaiser tiende a incluir pocos componentes. La modificación de Jolliffe excluye aquellas componentes cuyos autovalores sean menores que $0,7 \lambda = 0,7 \text{tr}(\mathbf{S})/p$, es decir menores que $0,7$, si se han calculado las componentes a partir de \mathbf{R} [3].

Los pasos que se suelen seguir en la técnica de Componentes Principales son:

Calcular la matriz de correlaciones entre todas las variables (\mathbf{R}).

Extracción de los componentes necesarios para representar los datos.

Rotación de los componentes, con el objetivo de facilitar su interpretación.

Representación gráfica [4].

En términos matemáticos los componentes principales son autovectores o vectores propios de la matriz de correlación y la técnica para encontrar estos autovectores se

llama análisis propio. A cada componente le corresponde un autovalor que proporciona la cantidad de varianza total explicada por la componente. Cuando se trabaja con datos estandarizados, cada variable estandarizada tendrá una varianza igual a 1 y por lo tanto la varianza total será igual al número de variables p . La suma de los autovalores de la matriz de varianzas y covarianzas de las variables estandarizadas o de la matriz R de las variables originales también será igual a p .

En el ACP, las relaciones entre las variables se estudian a partir de correlaciones, busca hallar combinaciones lineales de las variables originales que expliquen la mayor parte de la variación total; en este, el primer factor o componente sería aquel que explica la mayor parte de la varianza total, el segundo factor sería aquel que explica la mayor parte de la varianza restante, es decir, de la que no explica el primero y así sucesivamente [5, 6].

El objetivo de este trabajo es determinar los parámetros ambientales de mayor incidencia en las aguas residuales de la comunidad universitaria Antonio Maceo, disminuyendo la cantidad de parámetros a utilizar en la evaluación de las aguas generadas y reducir los costos de la investigación por operaciones de muestreo y análisis.

Materiales y métodos

Para conocer cuáles parámetros ambientales poseen mayor incidencia en las aguas residuales de la comunidad universitaria se partió de la caracterización químico-física de dichas aguas, según los procedimientos y normas aplicados por Crombet [7]. A los datos obtenidos se le realizó un análisis descriptivo para conocer el comportamiento de cada una de las variables. Se obtuvo la matriz de correlaciones, se aplicó el criterio de Kaise-Meyer-Olkin (KMO) y de esfericidad de Bartlett para conocer la factibilidad de la aplicación del análisis de componentes principales. Por último se realizó la determinación de los componentes principales. Para el análisis de los resultados se emplearon los programa estadísticos Statsoft Statistica v7.0.61.0 Copyright ©Statsoft, Inc. y IBS SPSS Statistics 22.

Resultados y discusión

Caracterización químico-física de las aguas residuales de la comunidad universitaria

En la comunidad universitaria Antonio Maceo, perteneciente a la Universidad de Oriente, uno de los principales problemas ambientales detectados está relacionado con

el vertimiento al medio ambiente de sus residuos líquidos de forma no controlada y sin tratamiento.

Estas aguas están compuestas, fundamentalmente, por los residuales líquidos de los edificios de la residencia estudiantil, del asentamiento poblacional de trabajadores de la universidad y de la cocina-comedor de la residencia; los cuales se unen en un drenaje localizado en el área de mantenimiento de la universidad y finalmente son dirigidas a un canal localizado en el asentamiento poblacional “Los Cocos”, el cual vierte al río Yarayó [7].

Los residuales líquidos de la comunidad universitaria se caracterizan por la variabilidad de los valores de los parámetros químico-físicos, establecidos por la Norma Cubana de vertimiento al sistema de alcantarillado [8].

Para el estudio de las aguas residuales fueron realizados 10 muestreos y analizados 13 parámetros ambientales, en el horario de mayor vertimiento de agua residual (tabla 1).

TABLA 1: MATRIZ DE DATOS. TOMADOS DE CROMBET [6]

PARÁMETROS	MUESTREOS									
	1	2	3	4	5	6	7	8	9	10
T (°c)	28	27	27	27	27	27	27	27	26	26
pH(U)	7,5	7,4	7,5	7,4	7,3	7,2	7,7	6,9	7,7	6,8
Turbidez (UTN)	65	78	62	75	102	63	41	85	60	58
λ (mS/cm)	674	623	609	606	754	772	743	654	626	427
DQO dic. (mg/L)	288	240	320	261	360	236	210	190	226	206
DBO ₅ (mg/L)	88	78	101	82	112	73	65	59	68	80
HS ⁻ (mg/L)	0,0	0,4	0,0	0,0	23,2	1,0	0,3	1,1	0,0	1,5
N-NH ₄ ⁺ (mg/L)	1,8	1,8	2,3	2,0	3,7	1,6	1,3	1,3	1,5	1,3
P-PO ₄ ³⁻ (mg/L)	18,7	6,8	7,1	6,2	8,3	4,4	3,0	2,7	2,0	2,2
A y G (mg/L)	41,2	46,2	46,5	46,0	67,9	61,3	34,8	42,2	39,0	56,0
ST (mg/L)	538	671	476	569	1197	673	636	699	541	733
STF (mg/L)	276	282	282	309	850	315	365	343	298	470
STV (mg/L)	262	389	194	260	347	358	271	356	243	263

Las aguas residuales domésticas de la comunidad universitaria Antonio Maceo de la Universidad de Oriente no cumplen con los requisitos establecidos para ser descargadas directamente en el sistema de alcantarillado sin tratamiento previo, encontrándose los parámetros DQO, DBO₅, turbidez, HS⁻, NH₄⁺, A y G, ST y STF por encima de los límites establecidos por la NC 27: 2012 [8]. Se demostró que los parámetros ambientales: T, pH, conductividad eléctrica, fosfato, amonio y sulfuro no son contaminantes, por encontrarse los valores determinados por debajo del límite máximo permisible que regula la norma cubana. Cuando existe variabilidad en la composición química-física del residual a tratar es conveniente conocer en qué grado incide cada parámetro medido y cuáles son los que más aportan a esta; pudiendo seleccionarse para el monitoreo periódico de dicha agua residual sin necesidad de emplear todos los parámetros establecidos.

Análisis descriptivo de cada variable

Para realizar el estudio descriptivo se emplearon 13 variables y 10 casos, determinándose la media, los valores máximos y mínimos, varianza (s^2) y desviación estándar (s) (Tabla 2). La variabilidad de los datos fue medida por el valor máximo y mínimo de cada variable, existiendo una gran variación entre los valores obtenidos en los muestreos para cada variable. Los parámetros ST y STV fueron los de mayor desviación, corroborándose con los valores obtenidos de s^2 y s ; mientras que en las variables T y pH es donde se obtiene la menor variabilidad en las determinaciones.

TABLA 2: ESTADÍSTICOS DESCRIPTIVOS DE LAS VARIABLES ESTUDIADAS

	Media	Valor Mínimo	Valor Máximo	Varianza	Desviación estándar
T	26,9	26,0	28,0	0,32	0,567 6
pH	7,34	6,8	7,7	0,09	0,302 6
Turbidez	69	41	102	283,21	16,828 9
λ	649	427	772	9 953,07	99,765 1
DQO	253,7	190	360	2 937,34	54,197 3
DBO ₅	80,60	59	112	265,82	16,304 1
HS ⁻	2,75	0,00	23,2	51,93	7,205 9
N-NH ₄ ⁺	1,86	1,30	3,7	0,53	0,726 0
P-PO ₄ ³⁻	6,14	2,00	18,7	24,57	4,957 2
A y G	48,11	34,80	67,9	109,04	10,442 4
ST	673	476,00	1 197	40 610,90	201,521 5
STF	379	276,00	850	30 733,11	175,308 6
STV	294	194,00	389	4 004,90	63,2843

Se puede confirmar, a través del gráfico de caja y bigotes (figura 1) que para el 95 % de confianza y N=10, las variables estudiadas incluyen el valor de la media muestral dentro del intervalo de confianza. En el gráfico se evidencia, con mayor claridad, que los parámetros λ , DQO, ST, STV y STF son los que presentan mayor variabilidad de los valores obtenidos durante el período de estudio, evidenciándose a través de la amplia longitud de la caja y del bigote.

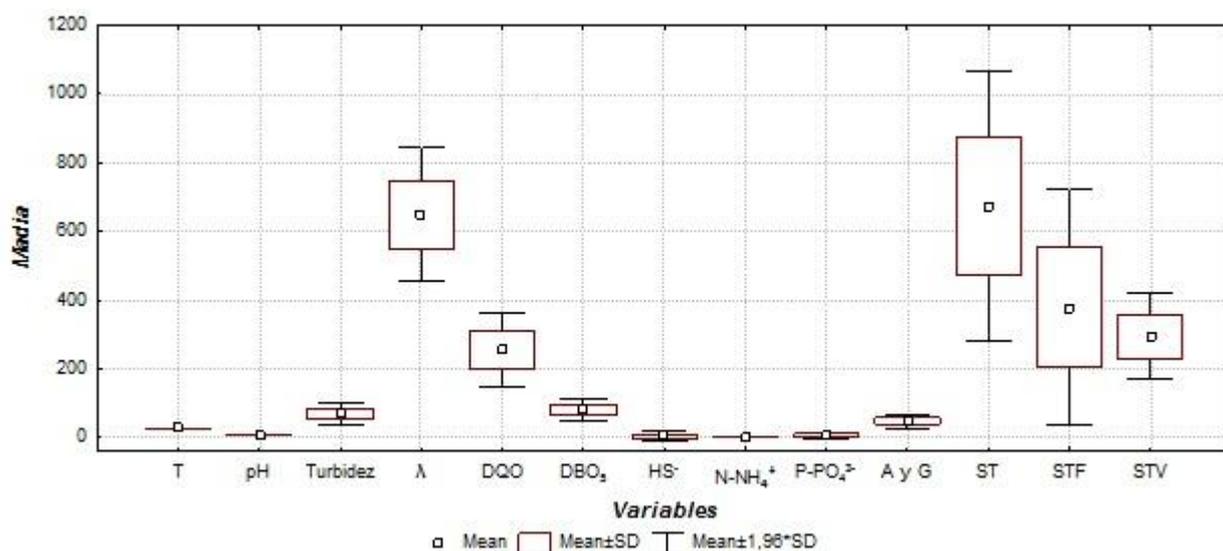


Fig. 1. Gráfico de cajas y bigotes para cada variable estudiada

Análisis de componentes principales

Matriz de correlación e índices empleados en la adecuación de la muestra al modelo

Se calculó la matriz de correlaciones para determinar éstas, entre las variables estudiadas. En la tabla 3 se establece la matriz para las variables estudiadas, señalándose el coeficiente de correlación de Pearson (r) y el p-valor. Se estableció un p-valor igual a 0,05, la significación indica la probabilidad de que se cumpla la hipótesis nula de no existencia de correlación entre las variables introducidas o lo que es equivalente de independencia entre ellas, esto conlleva a la no conveniencia de realizar el ACP.

TABLA 3. MATRIZ DE CORRELACIÓN DE LOS PARÁMETROS ANALIZADOS^A

	T	pH	Turbidez	λ	DQO	DBO ₅	HS ⁻	N-NH ₄ ⁺	P-PO ₄ ³⁻	A y G	ST	STF	STV
T	1,0000 p=---												
pH	0,2199 p=0,541	1,0000 p=---											
Turbidez	0,1849 p=0,609	-0,3439 p=0,331	1,0000 p=---										
λ	0,5294 p=0,116	0,4243 p=0,222	0,1662 p=0,646	1,0000 p=---									
DQO	0,3962 p=0,257	0,2847 p=0,425	0,4765 p=0,164	0,2884 p=0,419	1,0000 p=---								
DBO ₅	0,2473 p=0,491	0,0869 p=0,811	0,4477 p=0,194	0,0401 p=0,912	0,9506 p= 0,000	1,0000 p=---							
HS ⁻	0,0340 p=0,926	-0,1136 p=0,755	0,6951 p= 0,026	0,3485 p=0,324	0,6510 p= 0,041	0,6530 p= 0,041	1,0000 p=---						
N-NH ₄ ⁺	0,2319 p=0,519	0,1244 p=0,732	0,6899 p= 0,027	0,3295 p=0,352	0,9205 p= 0,000	0,8855 p= 0,001	0,8661 p= 0,001	1,0000 p=---					
P-PO ₄ ³⁻	0,8150 p= 0,004	0,2507 p=0,485	0,1925 p=0,594	0,2081 p=0,564	0,5812 p=0,078	0,5083 p=0,134	0,1182 p=0,745	0,3534 p=0,317	1,0000 p=---				
A y G	-0,1067 p=0,769	-0,4914 p=0,149	0,5414 p=0,106	0,0970 p=0,790	0,4656 p=0,175	0,5773 p=0,081	0,6983 p= 0,025	0,6175 p=0,057	-0,0025 p=0,994	1,0000 p=---			
ST	-0,0609 p=0,867	-0,3373 p=0,340	0,6733 p= 0,033	0,2786 p=0,436	0,3869 p=0,269	0,4375 p=0,206	0,9377 p= 0,000	0,6768 p= 0,032	-0,0471 p=0,897	0,7317 p= 0,016	1,0000 p=---		
STF	-0,1262 p=0,728	-0,2501 p=0,486	0,5751 p=0,082	0,1816 p=0,616	0,5012 p=0,140	0,5695 p=0,086	0,9602 p= 0,000	0,7412 p= 0,014	-0,0232 p=0,949	0,7000 p= 0,024	0,9530 p= 0,000	1,0000 p=---	
STV	0,1556 p=0,668	-0,3813 p=0,277	0,5509 p=0,099	0,3842 p=0,273	-0,1564 p=0,666	-0,1844 p=0,610	0,3260 p=0,358	0,1021 p=0,779	-0,0857 p=0,814	0,3909 p=0,264	0,5443 p=0,104	0,2647 p=0,460	1,0000 p=---

Las correlaciones marcadas son significativas en $p < 0,05$ y $N = 10$.

a)

Determinante = 0,000

En la tabla 3 se evidencia que para un p-valor pequeño ($p < 0,05$) se rechaza la hipótesis nula indicando que existen correlaciones entre las variables. Sin embargo, se observa también que no todas las variables están correlacionadas.

El determinante se emplea como índice del tamaño de las correlaciones, cuando su valor es bajo significa que hay correlaciones altas en la matriz y el ACP es factible, si el valor es igual a cero indica que algunas de las variables utilizadas son linealmente dependientes [9] y no se pueden realizar algunos cálculos en el ACP. Se destaca que las variables pH, λ , STV no se correlacionan con ninguna otra y en este caso el ACP no es una técnica acertada para analizarlas. Estas variables, junto con la T y P-PO₄³⁻ que solo correlacionan entre ellas, están dentro del rango establecido por la NC 27:2012 para ser vertidas al sistema de alcantarillado, por lo que no contribuyen en la contaminación del agua residual, evaluándose de forma independiente el comportamiento de las mismas [7].

Cuando $r > 0$ indica que existen correlaciones positivas (sentido directo), cuanto más próximo a 1 esté más fuerte es la correlación [9]. En la tabla 3 se evidencia que en este estudio todas las variables que se correlacionan lo hacen de forma positiva y algunas presentan un coeficiente mayor de 0,90; existiendo un vínculo energético entre ellas. Se

observa que la DQO (Demanda Química de Oxígeno) se correlaciona fuertemente con DBO₅ (Demanda Bioquímica de Oxígeno), HS⁻ (ion hidrogeno sulfuro) y N-NH₄⁺ (nitrógeno amoniacal). Esta relación es debido a que la DQO incluye todo el material orgánico o inorgánico susceptible a oxidarse. Este parámetro, a su vez, contiene a la DBO₅ que se usa como una medida de la cantidad de oxígeno requerida para la oxidación de la materia orgánica biodegradable presente en la muestra de agua, como resultado de la acción de oxidación bioquímica aerobia. Por otra parte, la presencia de HS⁻ indica descomposición de la materia orgánica y el N-NH₄⁺ se asocia a la contaminación albañal reciente.

Otra correlación detectada es del HS⁻ con los A y G (Aceites y Grasas), N-NH₄⁺, Turbidez, ST (Sólidos Totales) y STF (Sólidos Totales Fijos) evidenciando que la actual descomposición de la materia orgánica viene dada también por la presencia de material graso y por algún material inorgánico que influye en el deterioro de las aguas. La mayor correlación de los sólidos totales, los cuales incluyen los sólidos de origen orgánico e inorgánico, es con STF, contaminación de origen inorgánico en las aguas de la residencia.

El Determinante de la matriz de correlación obtenida es muy próximo a cero, lo que imposibilita el cálculo de la prueba de KMO (Kaiser-Meyer-Olkin) y el test de esfericidad de Bartlett, los cuales son índices que permiten conocer la conveniencia de emplear el ACP [10, 11]. Cuando esto sucede, es posible localizar las variables responsables de dichos valores y podrían ser eliminadas del análisis siempre y cuando su importancia teórica no lo desaconsejara.

Teniendo en cuenta los aspectos anteriormente analizados solo se dejaron las variables (DQO, DBO₅, turbidez, HS⁻, NH₄⁺, A y G, ST y STF) que están fuera del límite establecido por la NC 27:2012 para realizar la determinación de los componentes principales y así evaluar cuáles de ellas son necesarias para caracterizar las aguas residuales de la residencia. El valor del determinante obtenido para la nueva matriz de datos fue de $2,101 \times 10^{-8}$, es un valor muy bajo, lo que significa que existen variables con intercorrelaciones muy altas por lo que es factible continuar con el ACP.

En la tabla 4 se presentan los estadísticos correspondientes al estudio de la adecuación de la muestra al modelo, realizándose la prueba de KMO y test de esfericidad de Bartlett.

TABLA 4. PRUEBA DE KMO Y TEST DE ESFERICIDAD DE BARTLETT

Medida Kaiser-Meyer-Olkin de adecuación de muestreo		0,643
Prueba de esfericidad de Bartlett	Aprox. Chi-cuadrado	97,231
	gl	28
	Sig.	0,000

La prueba de esfericidad de Bartlett contrasta la hipótesis nula de que la matriz de correlaciones es una matriz identidad, en cuyo caso no existirían correlaciones significativas entre las variables y el ACP no sería pertinente. Consiste en una estimación de Chi-cuadrado a partir de una transformación del determinante de la matriz de correlaciones, si las variables no están intercorrelacionadas, entonces debe presentar un valor (significancia) superior al límite de 0,05 [10]. En este estudio (tabla 4) presentó un valor de 0,000 lo cual nos indica que la matriz de datos es válida para continuar con el proceso de ACP.

La medida de adecuación muestral KMO contrasta si las correlaciones parciales entre las variables son suficientemente pequeñas. Permite comparar la magnitud de los coeficientes de correlación observados con la magnitud de los coeficientes de correlación parcial. EL estadístico KMO varía entre 0 y 1, Kaiser, Meyer y Olkin aconsejan que si $KMO \geq 0,75$ la idea de realizar un ACP es buena; si $0,75 > KMO \geq 0,5$ la idea es aceptable y si $KMO < 0,5$ es inaceptable [10]. En el caso analizado el valor del índice de KMO obtenido es de 0,643, es un indicativo de la validez de la aplicación al caso del ACP.

Un indicador complementario al anterior es la matriz anti-imagen que contiene el complemento de las correlaciones parciales y de la matriz de covarianzas. En la matriz de correlación anti-imagen (tabla 5) se observan pocos valores elevados en términos absolutos y no existen coeficientes ceros, por lo que se recomienda llevar a cabo el análisis de componentes principales.

TABLA 5. MATRIZ DE COVARIANZA Y CORRELACIÓN ANTI-IMAGEN

		Turbidez	DQO	DBO ₅	HS ⁻	N-NH ₄ ⁺	AyG	ST	STF
Covarianza anti-imagen	Turbidez	0,195	0,018	-0,021	-0,009	-0,025	0,022	0,002	0,021
	DQO	0,018	0,006	-0,007	-0,003	-0,003	0,006	0,006	0,005
	DBO ₅	-0,021	-0,007	0,011	0,005	-0,001	-0,017	-0,007	-0,008
	HS ⁻	-0,009	-0,003	0,005	0,002	-0,001	-0,001	-0,004	-0,003
	N-NH ₄ ⁺	-0,025	-0,003	-0,001	-0,001	0,019	-6,090E-6	0,001	0,001
	AyG	0,022	0,006	-0,017	-0,001	-6,090E-6	0,282	-0,021	0,008
	ST	0,002	0,006	-0,007	-0,004	0,001	-0,021	0,015	0,004
	STF	0,021	0,005	-0,008	-0,003	0,001	0,008	0,004	0,006
Correlación anti-imagen	Turbidez	0,674 ^a	0,547	-0,444	-0,413	-0,409	0,093	0,044	0,589
	DQO	0,547	0,499 ^a	-0,916	-0,845	-0,251	0,145	0,626	0,842
	DBO ₅	-0,444	-0,916	0,516 ^a	0,877	-0,077	-0,292	-0,512	-0,898
	HS ⁻	-0,413	-0,845	0,877	0,588 ^a	-0,228	-0,035	-0,730	-0,906
	N-NH ₄ ⁺	-0,409	-0,251	-0,077	-0,228	0,934 ^a	-8,414E-5	0,078	0,089
	AyG	0,093	0,145	-0,292	-0,035	-8,414E-5	0,914 ^a	-0,322	0,190
	ST	0,044	0,626	-0,512	-0,730	0,078	-0,322	0,709 ^a	0,410
	STF	0,589	0,842	-0,898	-0,906	0,089	0,190	0,410	0,566 ^a

a. Medidas de adecuación de muestreo (MSA)

De la diagonal de la matriz de correlación anti-imagen se toman como valores mínimos y máximos respectivamente el 0 y el 1, siendo tanto mejor cuanto mayor sea el valor del MSA (superiores a 0,5), lo que corrobora que se debe continuar con el ACP [11]. En este caso el valor del MSA para cada variable es mayor que 0,5, por lo que indica que el modelo de ACP puede ser utilizado para explicar los datos.

Las pruebas realizadas muestran la validez de la matriz de datos con las variables seleccionadas, lo que permite realizar la extracción de los distintos componentes a través de la agrupación de las ocho variables originales seleccionadas en unas nuevas variables denominadas “componentes”, las cuales son combinaciones lineales de las variables originales.

Extracción de los componentes iniciales

La extracción de los componentes principales se puede ver inicialmente a partir de la tabla 6 que contiene las Comunalidades asignadas a cada variable y a las reproducidas para la solución de los componentes principales. La comunidad de una variable es la proporción de su varianza que puede ser explicada por el modelo de componentes principales obtenido, por lo que se puede explicar cuáles de las variables es peor explicada por el modelo [12].

TABLA 6. COMUNALIDADES INICIALES Y REPRODUCIDAS

Variabes	Inicial	Extracción
Turbidez	1,000	0,585
DQO	1,000	0,988
DBO ₅	1,000	0,951
HS ⁻	1,000	0,955
N-NH ₄ ⁺	1,000	0,969
AyG	1,000	0,654
ST	1,000	0,976
STF	1,000	0,905

Método de extracción: análisis de componentes principales

Todas las variables contribuyen a la creación de los componentes, las comunalidades son mayores que 0,5. En el estudio las variables Turbidez y AyG son las menos explicadas, el modelo solo es capaz de reproducir el 58 y 65 %, respectivamente, de sus variabilidades originales.

La selección del número adecuado de componentes a conservar se realizó teniendo en cuenta las reglas existentes como son: porcentaje de varianza total explicada, criterio de Kaiser y modificación de Jollife.

En la tabla 7 se muestran los porcentajes de varianza explicada, únicamente los dos primeros componentes tienen valores propios mayores que 1 y explican el 87,31 % de la varianza, produciéndose la pérdida de tan solo el 12,69 % de la información original representada por las ocho variables iniciales.

TABLA 7. PORCENTAJES DE 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	5,786	72,327	72,327	5,786	72,327	72,327	4,003	50,037	50,037
2	1,199	14,983	87,310	1,199	14,983	87,310	2,982	37,274	87,310
3	0,517	6,458	93,768						
4	0,409	5,107	98,876						
5	0,060	0,752	99,628						
6	0,017	0,215	99,843						
7	0,011	0,142	99,985						
8	0,001	0,015	100,000						

Método de extracción: análisis de componentes principales

El gráfico de sedimentación de las componentes, suele ser utilizado también como contraste gráfico para conocer el número de componentes a retener. Según este criterio se retienen todas las componentes que están situadas previamente a la zona de sedimentación, entendiendo por esta la parte del gráfico en la que los componentes empiezan a no presentar pendientes fuertes [10], que como ya hemos mencionado para los datos obtenidos, podemos comprobar que puede ser a partir de la componente tres.

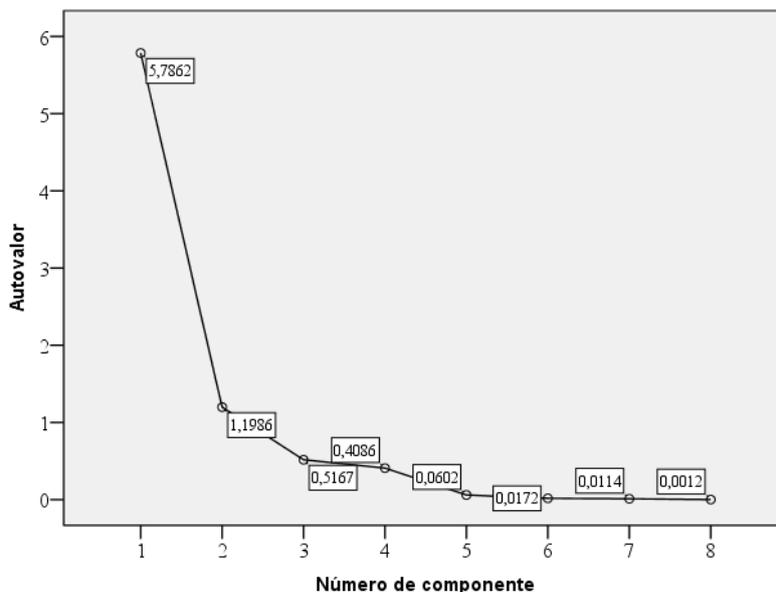


Fig. 2. Gráfico de sedimentación de los componentes

Empleando el Gráfico de Sedimentación (figura 2) se puede confirmar la cantidad de componentes que se obtienen. Se escogen las componentes cuyos valores propios sean mayores que 1. En la figura se indica que se deben extraer dos componentes principales que son los que cumplen con el requisito señalado.

TABLA 8. MATRIZ DE COMPONENTES SIN ROTAR

Variables	Componente	
	1	2
Turbidez	0,746	-0,170
DQO	0,786	0,609
DBO ₅	0,810	0,543
HS ⁻	0,958	-0,193
N-NH ₄ ⁺	0,943	0,283
AyG	0,782	-0,207
ST	0,861	-0,485
STF	0,892	-0,330

Método de extracción: análisis de componentes principales. a. 2 componentes extraídos.

En la tabla 8 se muestra la matriz de componentes sin rotar, la selección de las variables por componente se realizó teniendo en cuenta la modificación de Jollife. Sin embargo, se evidencia que existe ambigüedad en la selección de algunas variables por componente por ser sus autovalores mayor a 0,5. Estas dudas se pueden dilucidar una vez realizada la Rotación Varimax.

En un ACP idóneo la mayoría de las correlaciones reproducidas se parecerán a las correlaciones observadas y los residuos serían muy pequeños [13]. En la tabla 9 se muestran las comunalidades y las correlaciones reproducidas, así como los residuos entre las correlaciones observadas y las reproducidas. Los valores de los residuos evidencian que no existe un porcentaje elevado de diferencias superiores a 0,05 indicando que el modelo es adecuado para los datos, por tanto, las correlaciones observadas están bien estimadas, el coeficiente de correlación de Pearson es apropiado por la escala utilizada para medir las variables y se han extraído un número suficiente de componentes.

TABLA 9. MATRIZ DE CORRELACIONES REPRODUCIDAS Y RESIDUO

		Turbidez	DQO	DBO ₅	HS ⁻	N-NH ₄ ⁺	AyG	ST	STF
Correlación reproducida	Turbidez	0,585 ^a	0,483	0,512	0,747	0,655	0,618	0,725	0,722
	DQO	0,483	0,988 ^a	0,967	0,636	0,913	0,489	0,381	0,500
	DBO ₅	0,512	0,967	0,951 ^a	0,672	0,918	0,521	0,434	0,544
	HS ⁻	0,747	0,636	0,672	0,955 ^a	0,849	0,789	0,918	0,919
	N-NH ₄ ⁺	0,655	0,913	0,918	0,849	0,969 ^a	0,679	0,674	0,748
	AyG	0,618	0,489	0,521	0,789	0,679	0,654 ^a	0,773	0,766
	ST	0,725	0,381	0,434	0,918	0,674	0,773	0,976 ^a	0,928
	STF	0,722	0,500	0,544	0,919	0,748	0,766	0,928	0,905 ^a
Residuo ^b	Turbidez	-	0,006	0,064	0,052	0,035	0,077	0,051	0,147
	DQO	-0,006	-	0,017	0,015	0,007	0,023	0,006	0,001
	DBO ₅	-0,064	0,017	-	0,019	-0,032	0,056	0,003	0,026
	HS ⁻	-0,052	0,015	0,019	-	0,017	0,091	0,019	0,042
	N-NH ₄ ⁺	0,035	0,007	0,032	0,017	-	0,061	0,003	0,007
	AyG	-0,077	0,023	0,056	0,091	-0,061	-	0,042	0,066
	ST	-0,051	0,006	0,003	0,019	0,003	0,042	-	0,025
	STF	-0,147	0,001	0,026	0,042	-0,007	0,066	0,025	-

Método de extracción: análisis de componentes principales. a. Comunalidades reproducidas.

b. Los residuos se calculan entre las correlaciones observadas y reproducidas.

Existen 9 (32,0 %) residuos no redundantes con valores absolutos mayores que 0,05.

Rotación de los componentes iniciales

La extracción inicial de los componentes no permite identificar con toda claridad la relación o el modelo subyacente que se establece entre los componentes y las variables [14]. Con el objetivo de evidenciar esta relación se procede a la rotación de los componentes, que consiste en una transformación de la matriz original en otra más simple que adecúa mejor los ejes al aproximarlos a las variables correlacionadas, facilitando la interpretación de la estructura de los datos, no se altera la bondad de ajuste de la solución factorial, las comunalidades y los porcentajes de varianza explicada se mantienen inalterados, simplemente se redistribuye la varianza explicada entre los componentes [10].

El método de rotación utilizado es Varimax que busca redistribuir la varianza a lo largo de todos los componentes en la matriz de carga. Con esto se simplifica el modelo y se obtienen resultados más claros para identificar las variables que más aportan a cada componente, pues este método aproxima las cargas altas a 1 o -1 y las cargas bajas de la matriz no rotada a 0, eliminando de esta forma, las ambigüedades existentes en la matriz no rotada [11]. Podemos observar en esta matriz de componentes rotados (tabla 10) una redistribución de las variables que definen los componentes.

TABLA 10. MATRIZ DE COMPONENTES ROTADOS

Variables	Componente	
	1	2
Turbidez	0,689	0,332
DQO	0,235	0,966
DBO ₅	0,295	0,930
HS ⁻	0,869	0,447
N-NH ₄ ⁺	0,561	0,809
AyG	0,740	0,326
ST	0,975	0,157
STF	0,904	0,298

Método de extracción: análisis de componentes principales. Método de rotación: Varimax con normalización Kaiser.^a

a. La rotación ha convergido en 3 iteraciones.

El aporte más importante del ACP es la cuantificación de la significancia de las variables que explican los patrones en las propiedades inherentes de objetos individuales. Toda la información sobre los parámetros que se encuentra en el conjunto de datos originales se transforma a un espacio reducido y se puede explicar a partir de un conjunto de componentes calculados llamados componentes principales [15, 16]. En

este estudio se obtuvieron finalmente dos componentes principales, el primero logra explicar el 72,33 % de la variabilidad de las aguas residuales y el segundo el 14,98 %.

Las variables que aportan a la primera componente principal son: Turbidez, HS^- , AyG, ST y STF. Este agrupamiento refleja la significativa contaminación con material graso y materia inorgánica en descomposición, evidenciando la fuerte presencia de compuestos inorgánicos en las aguas. Todas estas variables, presentan un desvío del valor permisible, según la norma cubana, para ser vertidas al sistema de alcantarillado [8]. La segunda componente está determinada por las variables DQO, DBO_5 , $N-NH_4^+$, las mismas reflejan la alta contaminación con materia orgánica reciente, parámetros que no cumplen con los requisitos de la norma cubana.

Conclusiones

El estudio realizado demuestra que no es posible dejar fuera del análisis de las aguas residuales a ninguna de las variables evaluadas: DQO, DBO_5 , turbidez, AyG, HS^- , $N-NH_4^+$, ST y STF. Estas variables presentan una gran incidencia en las características de las aguas de la comunidad universitaria, permitiendo explicar el 87,31 % de la variabilidad total y no es posible una mayor reducción de las mismas para investigaciones posteriores.

Referencias bibliográficas

1. QUERO, M.; INCIARTE, K. “Clasificación de las Técnicas Estadísticas Multivariantes”. *Telos*. 2012, **14** (2), 275-286. ISSN 1317-0570.
2. CORONADOS, Y.; VILTRES, V. M.; SISTACH, V. “Aplicación de técnicas estadísticas multivariantes en el análisis de datos”. *INFOMED*. Revista Cubana de Medicina Física y Rehabilitación. 2017, **9** (2), 1-12. ISSN 2078-7162.
3. GONZÁLEZ, N.; TABORDA, A. *Análisis de componentes principales Sparse: formulación, algoritmos e implicaciones en el análisis de datos*. Tesis de Maestría en análisis avanzado de datos multivariantes. Universidad de Salamanca, 2015. DOI: 10.13140/RG.2.1.1448.1360.
4. POZA, C. “Técnicas estadísticas multivariantes para la generación de variables latentes”. *Revista Escuela de Administración de Negocios*. 2008, **64** (3), 89-99. ISSN: 0120-8160.

5. URRUTIA, J. A.; PALOMINO, R. “Componentes principales en la determinación de estaciones con Patrones homogéneos de temperatura en el Chocó”. *Scientia et Technica*. 2010, **16** (45), 257-262. ISSN 0122-1701.
6. MILLER, N. J.; MILLER, J. C. *Estadística y Quimiometría para Química Analítica*. 4^{ta} edición. Madrid: Pearson Educación, 2002. ISBN 9788420535142.
7. CROMBET, S.; PÉREZ, N.; ÁBALOS, A.; RODRÍGUEZ, S. “Caracterización de las aguas residuales de la comunidad "Antonio Maceo" de la Universidad de Oriente”. *Revista Cubana de Química*. 2013, **25** (2), 134-42. ISSN 2224-5421.
8. NC 27: 2012. Vertimiento de aguas residuales a las aguas terrestres y alcantarillados. Especificaciones. La Habana, 2012, p.14.
9. FRANQUET, J. M.; MARÍA, J. . *El estudio operativo de la psicología. Una aproximación matemática*. España. 1^{ra} edición. Editorial UNED-Tortosa. 2008. ISBN-13: 978-84-691-5599-8.
10. MOMTOLLA, O. “Aplicación del análisis factorial a la investigación de mercados. Caso de estudio”. *Scientia et Technica*. 2007, **13** (35), 281-286. ISSN 0122-1701.
11. LOZARES, C.; LÓPEZ, P. “El análisis de componentes principales: Aplicación del análisis de datos secundarios”. *Departamento de Sociología Universidad Autónoma de Barcelona. Papers*, 1991, 37, 31-63, DOI: <http://dx.doi.org/10.5565/rev/papers/v37n0.1595>.
12. GARCÍA, E.; GIL, J.; RODRÍGUEZ, G. *Análisis Factorial*. Cuadernos de Estadística no.7. Editorial La Muralla, 2002. ISBN9788471336965.
13. YENGLÉ, C. “Aplicación del análisis de componentes principales como técnica para obtener índices sintéticos de calidad ambiental”. *UCV – SCIENTIA*. 2012, **4** (2), 145-153, ISSN 2077-172X.
14. VISAUTA, B. *Análisis estadístico con SPSS para Windows. Volumen II Estadística Multivariante*. McGraw Hill, 2003. ISBN 9788448139933.
15. HERRERA, J.; RODRÍGUEZ, S.; SOLÍS, L. D.; CASTRO, F. “Aplicación de técnicas quimiométricas para clasificar la calidad de agua superficial de la microcuenca del río Bermúdez en Heredia, Costa Rica”. *Tecnología en Marcha*. 2009, **22** (4), 75-85. ISSN 0379-3962.

16. RESTREPO, L. F.; POSADA, S. L.; NOGUERA, R. R. “Application of the principal–component analysis in the evaluation of three grass varieties”. *Revista Colombiana de Ciencias Pecuarias*. 2012, **25** (2), 258-256. ISSN 0120-0690.