Application of multivariate statistical methods to water quality assessment in Arroyo Plata, Colombian Caribbean

Aplicación de métodos estadísticos multivariables para La evaluación de la calidad del agua en Arroyo Plata, Caribe colombiano

Recibido: 15 Ago 2017 – **Revisado:** 30 Sep 2017
Aceptado: 30 Oct 2017 – **Publicado:** 30 Dic 2017

Ildefonso Baldiris Navarro
Researcher, Ciptec Research Group, Fundación Universitaria Tecnológico Comfenalco, Sede A Barrio España Cra 44 D Nº 30ª -91, Cartagena, Colombia.
ibaldiris@tecnocomfenalco.edu.co

Jorge Hernán Sánchez Aponte
Researcher, Sena Cinaflup Research Group, SENA Centro internacional Náutico, fluvial y portuario, Mamonal km 5, Cartagena, Colombia.
jhsancheza@sena.edu.co

Abstract: Arroyo Plata is a ramification of dique canal that flows into the Caribbean Sea near Cartagena, discharging pollutants that may influence the quality of water. Environmental data was analyzed using hypothesis testing and confidence intervals and also using multivariate statistical methods including Pearson's correlation and Principal components analysis (PCA). Biochemical oxygen demand (BOD_5), Temperature, Phosphorus and salinity are out of the range recommended for the good health of marine flora and fauna. Four factors were identified as responsible for the data structure, explaining 65,66% of the total variance. The first factor is the physical and non-organic chemical parameters explaining 20,22% of the total variance. The second and third factors are, respectively, the microbiological (18%) and physico-chemical (16.2%) and the fourth factor, the agro industrial explains 11,54 %. This study shows that multivariate statistical methods can help the water managers to understand the factors affecting the water quality.

Keywords: Arroyo Plata; dique canal; pollution; hypothesis testing; principal component analysis.

Resumen: Arroyo Plata es una ramificación del dique canal del Río Magdalena, que desemboca en el Mar Caribe, cerca de Cartagena, y que descarga contaminantes que influyen en la calidad del agua. Los datos ambientales se analizaron mediante pruebas de hipótesis e intervalos de confianza y también utilizando métodos estadísticos multivariados, incluida la correlación de Pearson y análisis de componentes principales (PCA). La demanda bioquímica de oxigeno (DBO_5), Temperatura, Fósforo y salinidad están fuera del rango recomendado para la buena salud de la flora y fauna marina. Se identificaron cuatro factores como responsables de la estructura de datos, lo que explica el 65,66% de la varianza total. El primer factor son los parámetros químicos físicos y no orgánicos que explican el 20,22% de la varianza total. El segundo y el tercer factor son, respectivamente, el microbiológico (18%) y fisicoquímico (16.2%) y el cuarto factor, el agroindustrial explica el 11,54%. Este estudio muestra que los métodos estadísticos multivariados pueden ayudar a los administradores del recurso a comprender los factores que afectan la calidad del agua.

Palabras clave: Arroyo Plata; dique canal; contaminación; prueba de hipótesis; análisis de componentes principales.

1. Introduction

River mouths and their ecosystems are being threatened by the rise of nutrients, solids and bacteria proceeding from communities near rivers basins (Barakat et al. 2009). Magdalena river is the longest river in Colombia and at the end of its journey has several canals that transport part of the river flow into the Caribbean Sea, the most important is Dique canal.

Dique canal flows into the Caribbean Sea through arroyo Plata and others canals, providing an annual volume of sediments of approximately $397 m^3 \cdot s^{-1}$. At sea, most of the sediments are transported to the Rosario and San Bernardo Islands coral reefs. These islands are the habitat of extensive coral reefs and therefore of a great variety of fauna and flora, this environment may be affected by the sediment discharges, which causes the bleaching and necrosis of certain parts of the living tissue and subsequent death of the coral colonies (Restrepo et al. 2006).

Basic statistics is a not enough robust tool to evaluate large environmental data (Wunderlin et al. 2001), Multivariable statistic technics have been extensively used to evaluate water quality, identify the latent sources that influence surface water, and offer a valuable tool for reliable management of water resources as well as effective solutions to pollution in the last years. Taoufik et al. (2017) studied the variation of water quality in the river Wadi El Bey during two years using PCA y cluster analysis. Hajigholizadeh & Melesse (2017) used cluster analysis and discriminant analysis to study southern Florida water quality, analyzing a 15-year database and over 35,000 observations to assess the state of water pollution and its time-space variation. Yoon et al. (2016) studied the temporal and spatial variation in Chilika lagoon using PCA and data collected from 1999 to 2009. Jung et al. (2016) analyzed the behavior of water quality in the Nakdong River basin by principal component analysis and cluster analysis.

In this paper, variability of marine water quality was evaluated during 16 years, 2001-2016, according to Colombian law in Decree 1594 de 1984, using parameters such as: pH, salinity, dissolved oxygen (DO), total suspended solids (TSS), nitrates, total coliforms (TC), total phosphorus (TP), ammonium (NH_4^+), biochemical oxygen demand (BOD_5) and temperature(T), we also assessed the possible threat for coral reefs in the zone.

1.1 Study Area

Arroyo plata mouth is located at coordinates 848706-1604346, to the south of Cartagena bay, in front of the Rosario and San Bernardo Islands Natural Park, as shown in Fig. 1. Dry period extends from December to April, characterized by strong winds from the North - Northeast sector and weak and scarce rains. Rainy season extends from May to November; it is characterized by weak winds, of variable orientation and by a regime of copious rains.

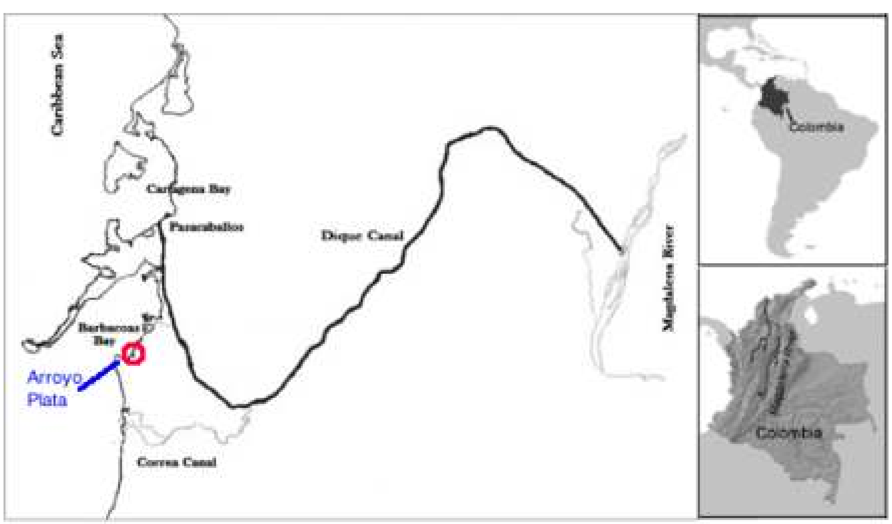

Fig 1. Arroyo Plata sampling station.

Marine water samples collection, receptacles, stabilization, and transportation to the laboratory, as well as samples storage, were done in accordance with APHA protocols. Samples were taken at 50 cm depth, pH, temperature, dissolved oxygen and salinity were measured in situ. refrigerated water samples, preserved with sulfuric acid and EDTA were also taken for laboratory tests.

2. ANALYTICAL METHODS

The analytical methods used for the determination of the concentrations are deposited in Table 1.

Table 1. Standard Methods

Parameter	Method	Standard code	Units
Temperature	Termometer	SM-2550-B	ºC
pH	Potenciométric	SM-4500-H+B	NA
Dissolved oxygen	Oximeter	SM-4500-0 G	$mg \cdot L^{-1}$
Salinity	Saltmeter	SM-2520-A	%o
DBO5	Winkler	SM-4500-OA	mg·L-1
Total suspended solids	Gravimetry	SM-2520-B	mg·L-1
Total Coliforms	Multiple tubes	SM-9222B	MPV
Ammonium	Distillation-volumetry	SM-4500-NH_3 B,C	mg·L-1
Nitrate	Cadmium reduction	SM-4500-NO3	mg·L-1
Total phosphorus	Ascorbic acid	SM-4500-P B, E	mg·L-1

2.1 Statistical Techniques

In order to compare the behavior of data with the law threshold values, hypothesis testing was used. Multivariate analysis of the water quality data-set was performed through PCA. Since methods of classification used here are non-parametric, no assumptions about underlying statistical distribution of the data was made, therefore no evaluation of normal distribution is necessary (Razmkhah et al. 2010). Mathematical and statistical computations were made using Microsoft Office Excel and minitab 16.

2.2 Hypothesis Testing and Confidence Interval

A statistical hypothesis is a statement about the parameters of one or more populations. Hypothesis testing is an effective tool to compare the mean of a population to a specified value. A confidence interval gives an estimated range of values which is likely to include an unknown population parameter, the estimated range being calculated from a given set of sample data (Montgomery 2017).

2.3 Correlation Matrix

Dataset analysis was performed using the Pearson correlation analysis in order to evaluate the relationship between water quality variables. This minimizes the effect of between-stations correlations and between-sampling campaigns relationships. A correlation coefficient near −1 or 1 means a strongest or negative or positive relationship between two variables and its value closet to 0 means no linear relationship between them at a significant level of $p < 0.05$. However, it should be noted that with larger samples, a low strength of correlation, for example $r < 0.50$, might be highly statistically significant at $p < 0.01$ (Montgomery et al. 2009).

2.4 Principal Component Analysis

Principal component analysis is a multivariate analysis technique that can be used to find new variables represented by a linear combination of variables having correlations via the variance-covariance matrix of several multivariate variables; it explains most of the total variations with some important principal components. The new axes lie along the directions of maximum variance. PCA provides an objective way of finding indices of this type so that the variation in the data might be accounted for as concisely as possible (Azhar 2015).

3. RESULTS AND DISCUSSION

Descriptive statistical analysis for Arroyo Plata is summarized in table 2. In this table may be observed the values recommended by the Colombian law for the protection of flora and fauna in marine ecosystems (guide levels).

Dissolved oxygen defines the biodiversity and survival of the biotic community. Almost all fishes tolerate a low concentration of dissolved oxygen for a certain period of time. However, reductions below the percentage of saturation generate negative effects on biodiversity, growth, reproduction and activity (Apha 2015).

Table 2. Descriptive Statistics for arroyo Plata

	Min	Max	Mean	Std. Dev.	Guide levels
BOD (mg/L)	1,20	8,96	4,76	1,89	<3
P (mg/L)	0,01	1,05	0,19	0,21	0,03
NO_3^- (mg/L)	0,01	0,92	0,07	0,18	5
OD (mg/L)	1,23	10,40	5,66	2,12	>4
pH	5,92	8,41	7,66	0,56	6,5–8,5
T(°C)	29,00	35,80	30,98	1,59	27-30
TSS (mg/L)	8,00	184	55,36	47,27	90
TC(NMP/100mL)	2,00	7900	494,5	1580,63	<5000
NH_4^+ (mg/L)	0,07	0,93	0,31	0,29	<1

In arroyo Plata OD presented a minimum value of 1.23 mg/L and a maximum of 10.40 mg/L, and an average value of 5.66 mg/L with a deviation of 2.12 mg/L. During the analysis period, depicted in fig. 2, the variable presented values below the established by the regulations[8]. Therefore, a hypothesis and confidence interval test were carried out. The test concluded that there is insufficient evidence to decide that the mean is less than 4 mg/L at a significance level of 0.05. In addition, the true mean is between 4.12 mg/L and 5.39 mg/L with a 90% confidence (see fig 3).

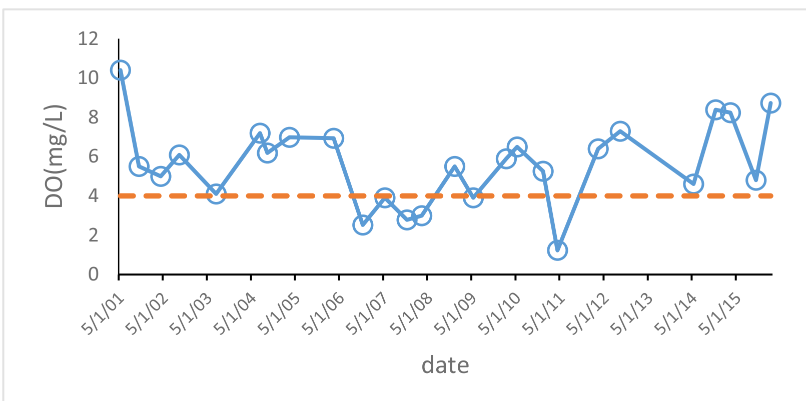

Fig 2. DO Variation in Arroyo Plata

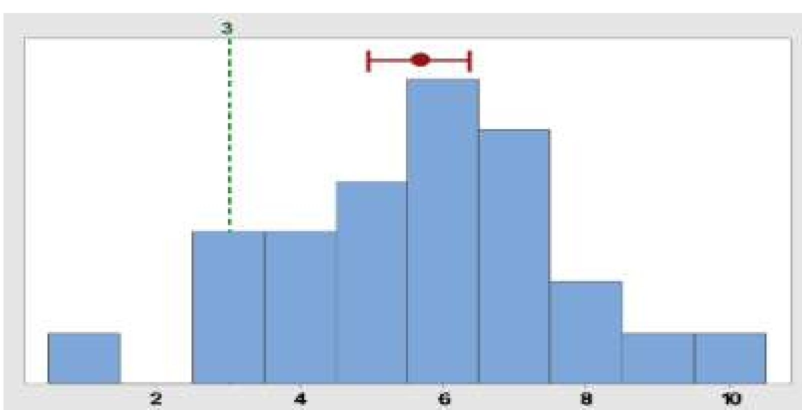

Fig 3. Hypothesis testing for OD in arroyo Plata

Biochemical oxygen demand is a test that measures the oxygen used during a specific incubation period for the biochemical degradation of organic matter -demand carbonaceous- and oxygen used to oxidize inorganic matter -sulfides and ferrous iron-. It can also measure the oxygen used to oxidize reduced forms of nitrogen -nitrogen demand- unless this oxidation is prevented by an inhibitor (Baldiris & Acosta 2015).

Hypothesis testing for BOD_5 mean showed that there is not enough evidence to conclude that the mean is less than 4 mg/L at a level of significance of 0.05, which indicates that it is outside the value stipulated by the law for this parameter. The confidence interval for BOD_5 was [4.13-5.39] mg/L and it was less than 5.39 mg/L 95% of the time it was measured as shown in fig.4-5.

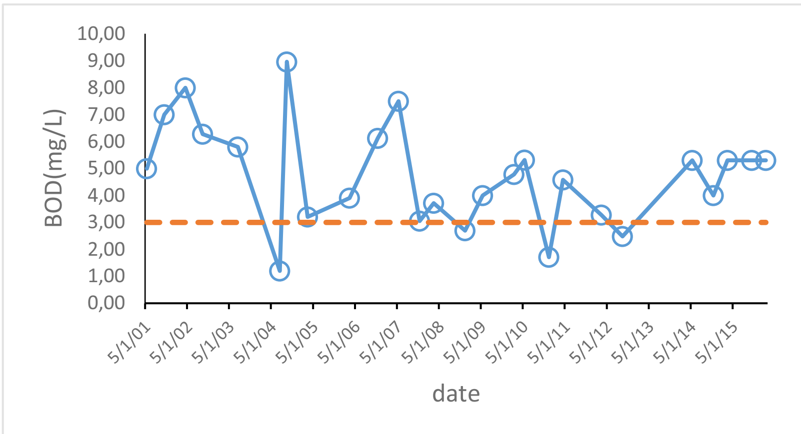

Fig 4. BOD_5 variation in Arroyo Plata

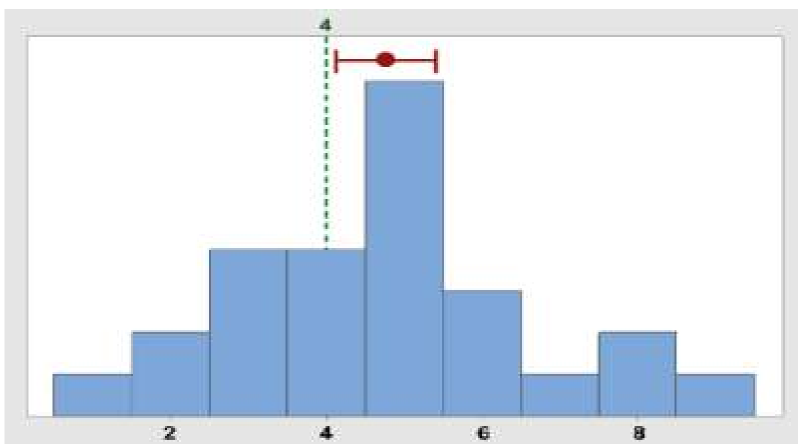

Fig 5. Hypothesis testing for BOD_5 in Arroyo Plata

Temperature affects physical processes -solubility of salts and gases such as oxygen, etc.-, chemical processes -pH, ionization equilibrium or ammonia concentration, reaction rates- and biological processes -metabolic rate, decomposition of organic matter- and, therefore, has an impact on the effect of pollutants (Pérez & Rodríguez 2008).

Temperature reached a maximum of 35.80°C and a minimum of 29°C as seen in fig.6, with an average of 30.98°C and standard deviation of 1.59°C. The hypothesis testing showed that the mean exceeds the maximum recommended value with a 95% confidence. The true mean is in the interval [30,45°C -31,53°C] and there is 95% confidence that it is greater than 30,45°C. pH is a common variable among the water quality indexes, due to its potential as an indicator of the degree of affection by pollutants and the extension of a pollution trail produced by discharges of an effluent. In addition, pH affects the toxicity of some compounds, such as ammonia, by controlling their ionization, as well as the biological availability of certain contaminants, such as heavy metals (Kamble & Vijay 2011).

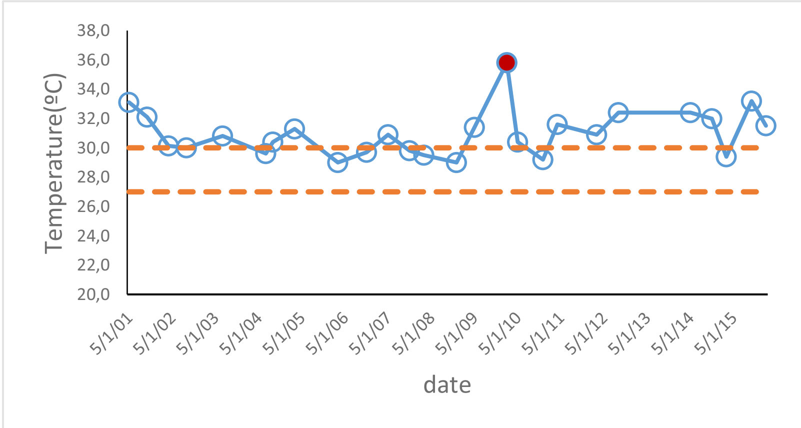

Fig 6. Temperature variation in Arroyo Plata

For pH, the confidence interval was [7.47-7.84] as may be seen in fig. 7, also the mean was greater than 7.47 and less than 8.6 with 95% confidence. This confirms that this parameter complies with the established for the preservation of the flora and fauna of marine ecosystems.

Ammonium is a water-soluble gas that exists at low levels in natural waters. NH_4^+comes from the nitrogen-containing organic material and gas exchange between water and atmosphere. It also comes from the biodegradation of waste and inputs from domestic, agricultural and industrial, and that is why it is a good indicator of contamination of water ways[4].

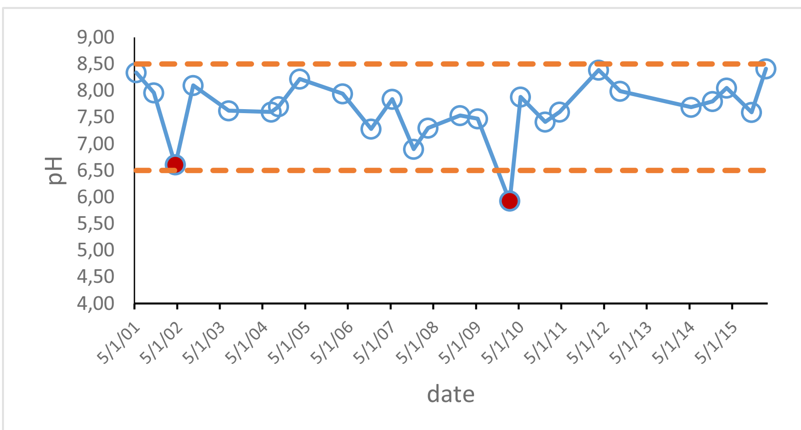

Fig 7. pH Variation in Arroyo Plata

Ammonium showed a range from 0.07 mg / L to 0.93 mg/L, with a mean of 0.31 mg / L and a deviation of 0.29 mg/L (fig. 8). At the level of significance of 0.05 it may be concluded that the mean is less than 1 mg / L. On the other hand, the 95% confidence interval for the mean is [0.216-0.409] mg/L, complying in this way with the current regulations for ammonium in Colombia. Nitrates concentration was included to make visible the content of fertilizers and their capacity to favor the processes of anthropogenic eutrophication. Nitrate sources may be wastewater and runoff from highly fertilized soils (Gavio et al. 2010).

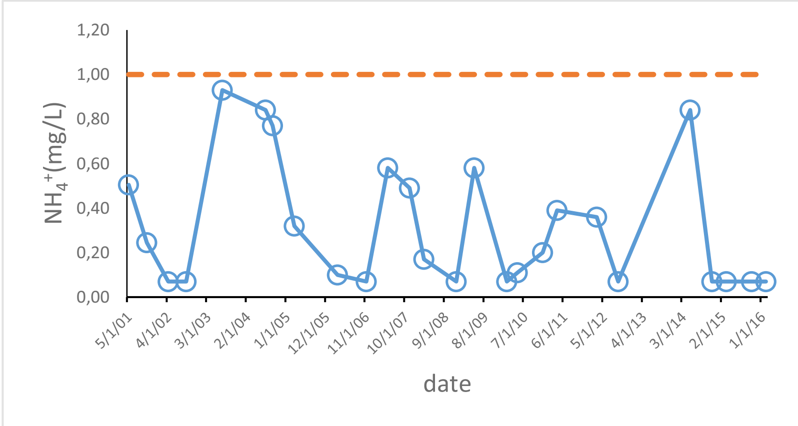

Fig 8. Ammonium variation in arroyo Plata

Nitrates achieved a maximum value of 0.92 mg/L and a minimum value of 0.01 mg/L. The mean was 0.07 mg/L with a deviation of 0.18 mg/L -see Fig 9-. It can be concluded that the mean is less than 0.5 mg/L at the significance level of 0.05 and is between 0.008 and 0.127 mg/L. These values are according to the Colombian law. Total phosphorus is an essential component of the biological cycle in water bodies and is generally the growth limiting agent of algae and aquatic plants, so its concentration serves as a standard to recognize the eutrophic state of a body of water. In agricultural areas, the entrainment of phosphorus linked to eroded soil -introduced by the application of fertilizers containing orthophosphates- is the main source of contamination.

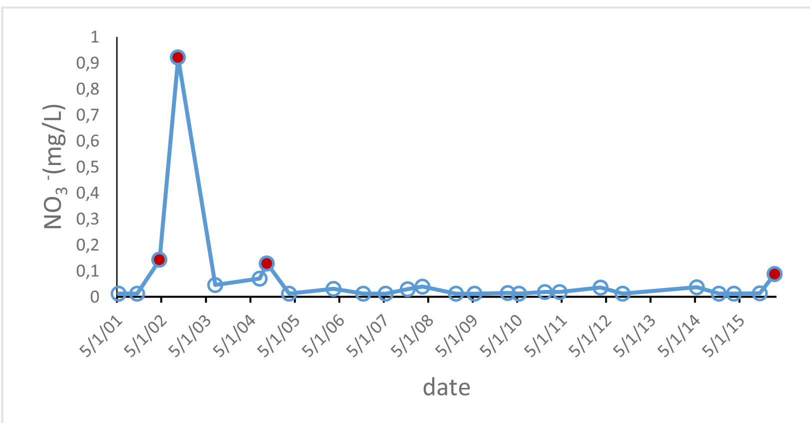

Fig 9. Nitrates variation of in arroyo Plata

In this study, phosphorus ranged from 0.01 to 1.05 mg/L, a mean of 0.19 mg/L and a deviation of 0.21 mg/L -see Fig 10-. To test compliance with the standard, a hypothesis test was conducted. it showed that there is not enough information to conclude that the mean is less than 0.03 mg/L. 90% confidence interval was [0.122-0.262] mg/L and there is a 95% confidence that it is less than 0.262 mg/L, which means that it does not comply with what is stated in the law for this variable.

Nevertheless, to prevent the enrichment of nutrients in coastal waters, the biologically available nitrogen (nitrate and ammonium) must be below 0.014 ppm and the biologically available phosphorus must be less than 0.003 ppm (Fabricius 2005), in arroyo Plata these variables are over this threshold values.

Total suspended solids are used to evaluate the effect on water, of the erosion caused by agricultural practices and the transport of material during the runoff of rainwater or irrigation. The solids generate problems of clogging and sedimentation may form deltas upstream of the reservoir or even destroy habitats for aquatic organisms by decreasing the water column (Severiche et al. 2017).

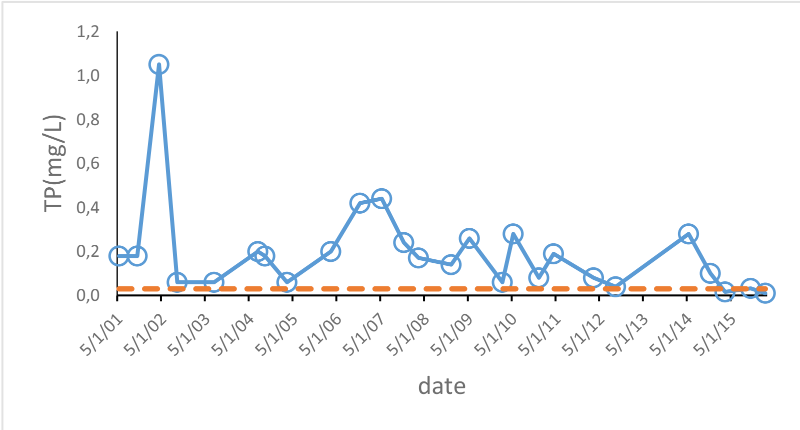

Fig 10. Phosphorus variation in arroyo Plata

TSS showed in fig.11 a minimum of 8 mg/L and a maximum value of 184 mg/L. The mean was 55.36 mg/L, with a standard deviation of 47.27 mg/L. The variable showed several outliers above the law. The hypothesis testing concluded with a 90% confidence that the SST are below 90 mg/L. For this variable, the true mean is between 39.52 mg/L and 71.19 mg/L and with it is less than 71.19 mg/L with a 90% of confidence, which complies with the Colombian law for this parameter.

To analyze salinity, the value was taken as a base between 33% and 36% as the optimum value for the health of the coral reefs in marine ecosystems. Abrupt changes or below this range might cause diseases in these organisms[19].

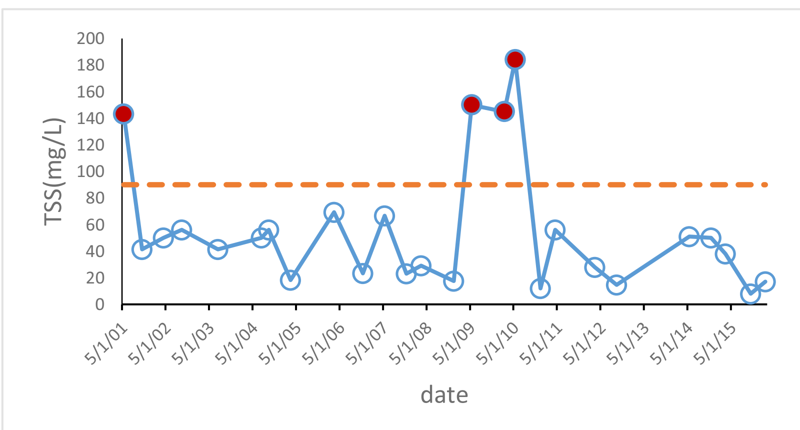

Fig 11. Total suspended solids variation in arroyo Plata

In arroyo Plata salinity reached a maximum value of 39.10‰ and a minimum value of 0.1‰ -see Fig 12-, with an average value of 22.63‰ and standard deviation of 11.59‰. 90% confidence interval was established as [18.68‰ - 26.42‰] and with 95% confidence is greater than 18.68‰. According to the performed hypothesis testing there is not enough evidence to conclude that the mean is greater than 32‰ at the level of significance of 0.05, this fact may represent a threat to coral reefs present in the area.

Polluted waterbodies may contribute to waterborne illness, especially in children and people with compromised immune systems. Annually, an estimated millions gastrointestinal illnesses are experienced globally due to contact with polluted coastal waters, resulting in billions public health costs. It is believed that the majority of the illnesses are caused by exposure to pathogenic bacteria and viruses of human fecal origin. World Health Organization (WHO) has identified at least 20 pathogens found in recreational water that can cause severe health effects, including adenoviruses, hepatitis viruses, pathogenic Escherichia coli, Campylobacter spp., and Salmonella spp. (USEPA 2010).

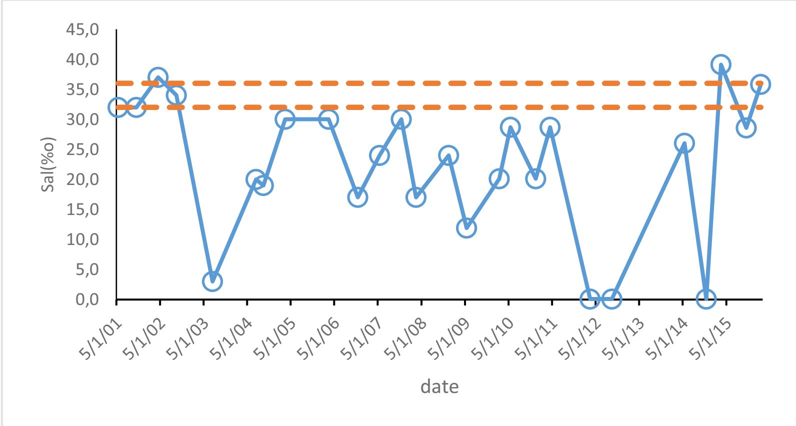

Fig 12. Salinity variation in arroyo Plata

Total coliforms varied in a range from 2 NMP to 7900 NMP (see fig. 13), with a mean value of 494.50 NMP and a standard deviation of 1580.63 NMP. The 90% confidence interval was [0 - 1024 MPN] and there is a 95% confidence that it is less than 1024 MPN. The variable is behaving according to the law, the hypothesis testing established that the mean is less than 5000 MPN at the significance level of 0.05.

In order to determine the existence of relationships between environmental variables, a Pearson correlation matrix was made -see table 3-

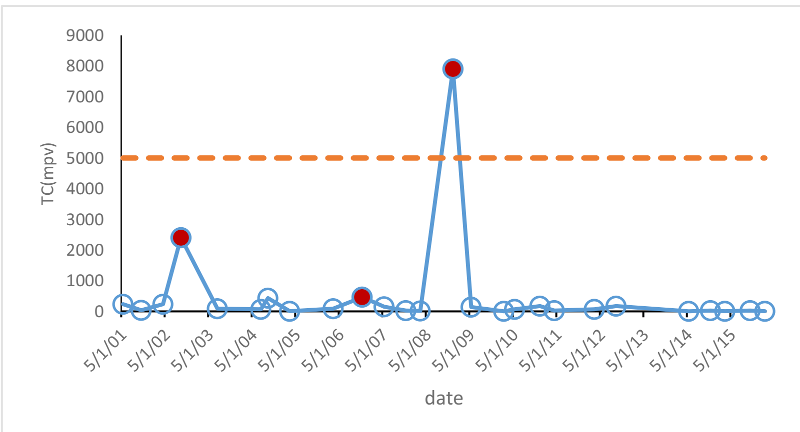

Fig 13. Variation of total coliforms in arroyo Plata

Tabla 3. Correlation matrix

	BOD	P	NO_3	OD	pH	T	TSS	Sal	TC	NH_4
BOD	1,000									
P	**0,431**	1,000								
NO_3	**0,243**	-0,031	1,000							
OD	-0,108	**-0,309**	0,051	1,000						
pH	-0,043	**-0,407**	0,127	**0,491**	1,000					
T	0,135	**-0,212**	-0,169	0,174	-0,174	1,000				
TSS	0,174	0,144	-0,026	0,165	-0,156	**0,344**	1,000			
Sal	**0,312**	**0,236**	**0,226**	0,078	0,010	-0,127	0,073	1,000		
TC	-0,137	-0,055	**0,222**	-0,009	-0,005	**-0,300**	-0,154	0,070	1,000	
NH_4	0,068	0,028	-0,122	-0,181	0,070	0,004	0,109	**-0,215**	**-0,207**	1,000

It may be seen in the correlation matrix, that there is a significant positive relationship between BOD_5 and phosphorus (r=0.431), BOD_5 and nitrates (r=0.243), BOD_5 and salinity (r=0.312), temperature and solids (r=0.344), dissolved oxygen and pH (r=0.491), phosphorus and salinity (r=0.236), nitrates and salinity (r=0.226) and nitrates and total coliforms (r=0.222) and a significant negative relationship between the variables phosphorus and OD (r=-0.309), phosphorus and pH (r=-0.407), phosphorus and temperature (r =-0.212), temperature and total coliforms (r=-0.3), salinity and NH_4^+ (r=-0.215) and total coliforms and NH_4^+ (r=-0.207). All these results confirm that is correct the use of PCA to find relations between water variables.

3.1 Principal Components Analysis

The scree plot (Fig. 14) was used to identify the number of PCs arroyo Plata. Four components were retained, which have eigenvalues greater than unity and explain 65,66% of the information contained in the original data set.

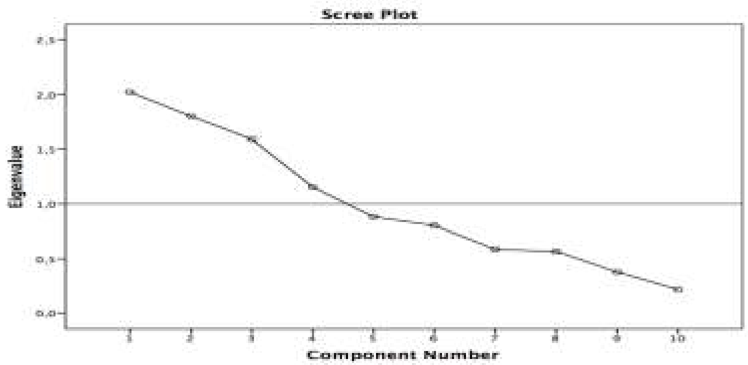

Fig 14. Principal components scree plot

Loading of four retained PCs are presented in Table 4. The first factor (PC1) is the bio-chemical and non-organic chemical parameters accounting 20,22% of the variance, showed high positive loadings of P, moderate and positive with BOD_5, moderate negative loading of OD and pH. This factor is dominated by phosphorus concentration, which might be caused by excess of fertilizer during agricultural activities in the zone and also by sewage coming from communities near the mouth of the canal and also by organic waste coming from the communities near the canal.

Table 4. Loadings of 10 experimental variables on five significant principal components

	Components			
Parameter	1	2	3	4
BOD	**0,614**	0,118	0,483	0,303
TP	**0,809**	0,184	-0,048	0,062
NO_3^-	0,010	**0,571**	0,361	0,139
DO	**-0,566**	-0,066	**0,634**	-0,073
pH	**-0,660**	0,149	0,346	**0,514**
T	0,045	**-0,684**	0,382	-0,330
TSS	0,304	-0,422	0,495	-0,164
Sal	0,305	0,485	0,492	-0,060
TC	-0,177	**0,606**	-0,181	-0,310
NH_4^+	0,121	-0,391	-0,197	**0,731**

The second factor (PC2) explains 18% of the variance and is highly positive contributed to nitrates and total coliforms, also is negatively contributed to by temperature. This factor might be attributed to anthropogenic pollution sources like wastewater discharges. The third factor (PC3) explains 15,91% of variance and it is positively contributed to by dissolved oxygen and represents physico-chemical source of the variability and influences from natural and anthropogenic inputs. The fourth factor (PC4) explains 11,54 % of the information and it is positively contributed to by pH and ammonium. It is likely to represent the sewage and manure discharges to water bodies, influences from agricultural runoff and byproducts from agro industrial processes.

4. Conclusions

Hypothesis testing and confidence intervals applied to the information from arroyo Plata showed that BOD_5, temperature, phosphorus, salinity and solids were over the threshold of the Colombian regulations for the preservation of the flora and fauna of the marine ecosystems. Solids, temperature and salinity behavior might be caused by temporal natural phenomena, while anthropogenic effects rule DOB_5, phosphorus and nitrogen levels, such as the discharges of domestic and agro-industrial waters and the excessive use of fertilizers.

Correlation and principal components analysis helped to detect the factors that may cause the degradation of water quality. The cause of water quality variations is mainly related to contamination load from diffuse (non-point) sources due to natural factors and agroindustry.

Nutrients concentration are in the law allowed range, but these conditions might be a threat for coral reefs. Studies had shown that nutrient enrichment, abrupt changes of salinity and excessive solids has a negative effect on coral reproduction, and its calcification, decreasing the growth rate of the reef and bleaching. High levels of BOD and nutrients might become a threat not only to animals and plants but to humans in the zone due to trophic chain. It is necessary to train communities on good agricultural practices and wastewater management in the zone in order to improve the levels of these variables.

References

Acosta, J. C., Baldiris, I., & Pacheco, H. P. (2015). ANÁLISIS DE LA VARIACIóN EN LA CALIDAD DEL AGUA EN LA BAHÍA DE BARBACOAS-CARTAGENA DURANTE EL PERIODO 2001-2014. *Ingeniería e Innovación*, *3*(1).

Alberto, W. D., del Pilar, D. M., Valeria, A. M., Fabiana, P. S., Cecilia, H. A., & de los Ángeles, B. M. (2001). Pattern Recognition Techniques for the Evaluation of Spatial and Temporal Variations in Water Quality. A Case Study: Suquía River Basin (Córdoba–Argentina). *Water research*, *35*(12), 2881-2894.

Azhar, S. C., Aris, A. Z., Yusoff, M. K., Ramli, M. F., & Juahir, H. (2015). Classification of river water quality using multivariate analysis. *Procedia Environmental Sciences*, *30*, 79-84.

Barakat, A., El Baghdadi, M., Rais, J., Aghezzaf, B., & Slassi, M. (2016). Assessment of spatial and seasonal water quality variation of Oum Er Rbia River (Morocco) using multivariate statistical techniques. *International Soil and Water Conservation Research*, *4*(4), 284-292.

Fabricius, K. E. (2005). Effects of terrestrial runoff on the ecology of corals and coral reefs: review and synthesis. *Marine pollution bulletin*, *50*(2), 125-146

Gavio, B., Palmer-Cantillo, S., & Mancera, J. E. (2010). Historical analysis (2000–2005) of the coastal water quality in San Andrés Island, SeaFlower Biosphere Reserve, Caribbean Colombia. *Marine Pollution Bulletin*, *60*(7), 1018-1030.

Hajigholizadeh, M., & Melesse, A. M. (2017). Assortment and spatiotemporal analysis of surface water quality using cluster and discriminant analyses. *Catena*, *151*, 247-258.

Jung, K. Y., Lee, K. L., Im, T. H., Lee, I. J., Kim, S., Han, K. Y., & Ahn, J. M. (2016). Evaluation of water quality for the Nakdong River watershed using multivariate analysis. *Environmental Technology & Innovation*, *5*, 67-82.

Kamble, S. R., & Vijay, R. (2011). Assessment of water quality using cluster analysis in coastal region of Mumbai, India. *Environmental monitoring and assessment*, *178*(1-4), 321-332.

Kim, J. Y., Bhatta, K., Rastogi, G., Muduli, P. R., Do, Y., Kim, D. K., ... & Joo, G. J. (2016). Application of multivariate analysis to determine spatial and temporal changes in water quality after new channel construction in the Chilika Lagoon. *Ecological Engineering*, *90*, 314-319.

Montgomery, D. C. (2017). *Design and analysis of experiments*. John wiley & sons.

Montgomery, D. C., Runger, G. C., & Hubele, N. F. (2009). *Engineering statistics*. John Wiley & Sons.

Pérez-Castillo, A. G., & Rodríguez, A. (2008). Índice fisicoquímico de la calidad de agua para el manejo de lagunas tropicales de inundación. *Revista de Biología tropical*, *56*(4), 1905-1918.

Razmkhah, H., Abrishamchi, A., & Torkian, A. (2010). Evaluation of spatial and temporal variation in water quality by pattern recognition techniques: a case study on Jajrood River (Tehran, Iran). *Journal of Environmental Management*, *91*(4), 852-860.

Restrepo, J. D., Zapata, P., Díaz, J. M., Garzón-Ferreira, J., & García, C. B. (2006). Fluvial fluxes into the Caribbean Sea and their impact on coastal ecosystems: The Magdalena River, Colombia. *Global and Planetary Change*, *50*(1-2), 33-49.

Severiche, C., Baldiris, I., Acosta, J., Bedoya, E., Castro, I., & Pacheco, H. (2017). Multivariate Analysis of Water Quality in Rosario Islands National Park (Colombia). *American Journal of Engineering Research*, *6*(6), 136-144.

Taoufik, G., Khouni, I., & Ghrabi, A. (2017). Assessment of physico-chemical and microbiological surface water quality using multivariate statistical techniques: a case study of the Wadi El-Bey River, Tunisia. *Arabian Journal of Geosciences*, *10*(7), 181.

USEPA, “Assessment of the Extra-Enteric Behavior of Fecal Indicator Organisms in Ambient Waters,” Washington, DC, 2010.

W. Apha, “AWWA, 1998,” *Stand. Methods Exam. Water Wastewater, 20th ed. Am. Public Heal. Assoc. Washington, DC*, 2008