

UNIVERSIDAD NACIONAL DE TUMBES
ESCUELA DE POSGRADO

DOCTORADO EN ESTADÍSTICA MATEMÁTICA APLICADA



**Modelo estadístico de calidad del agua del Estero El
Macho, Machala-El Oro**

TESIS

**Para optar al grado académico de Doctor en Estadística
Matemática Aplicada**

Autor, MSc. Rafael Gilberto Zhindón Almeida

Tumbes, 2024

UNIVERSIDAD NACIONAL DE TUMBES
ESCUELA DE POSGRADO
DOCTORADO EN ESTADÍSTICA MATEMÁTICA APLICADA



Modelo estadístico de calidad del agua del Estero El
Macho, Machala-El Oro

Tesis aprobada en forma y estilo por:

Mg/Dr NAPOLEON PUÑO LECARNAQUE (presidente)
ORCID: 0000000250088085

Mg/Dr LUIS JHONY CAUCHA MORALES (secretario)
ORCID: 0000-0002-4786-9008

Mg/Dr JUAN SANTIAGO BLAS PEREZ (vocal)
ORCID: 0000-0002-97413164

Mg/Dr HUMBERTO LORENZO NARVA RONCAL (vocal)
ORCID: 0009-0006-5668-4191

Tumbes, 2024

UNIVERSIDAD NACIONAL DE TUMBES

ESCUELA DE POSGRADO

DOCTORADO EN ESTADÍSTICA MATEMÁTICA APLICADA



Modelo estadístico de calidad del agua del Estero El
Macho, Machala-El Oro

**Los suscritos declaramos que el Informe de tesis es original en su
contenido y forma:**

MSc. Rafael Gilberto Zhindón Almeida (**Autor**)

Dr. Walter Javier Castañeda Guzmán (**Asesor**)
ORCID: 0000-0002-9483-0210

Dr. Raúl Alfredo Sánchez Ancajima (**Co-Asesor**)
ORCID: 0000-0003-3341-7382

Tumbes, 2024

ACTA DE SUSTENTACIÓN DEL INFORME DE TESIS



UNIVERSIDAD NACIONAL DE TUMBES
Licenciada
Resolución del Consejo Directivo N° 155-2019-SUNEDU/CD
ESCUELA DE POSGRADO
Tumbes - Perú

UNIVERSIDAD NACIONAL DE TUMBES ESCUELA DE POSGRADO	
Expediente N°:	001193
Fecha:	15 OCT 2024
Recibido:	

"Año del Bicentenario, de la consolidación de nuestra Independencia, y de la Comemoración de las Heroicas Batallas de Junín y Ayacucho"

ACTA DE SUSTENTACIÓN DE TESIS

En Tumbes, a los treinta días de setiembre del dos mil veinticuatro, siendo las veinte horas, se reunieron en el aula N° 02, los miembros del jurado calificador constituido con la Resolución N° 139-2024/UNTUMBES-EPG-D, presidido por el Dr. Napoleón Puño Lecarnaque (Presidente), Dr. Luis Jhony Caucha Morales (Secretario), Dr. Juan Santiago Blas Perez (vocal), Dr. Humberto Lorenzo Narva Roncal (vocal), Dr. Walter Javier Castañeda Guzmán (asesor) y Dr. José Modesto Carillo Sarango (asesor).

Instalado el jurado, se procedió a la evaluación, deliberación y calificación del acto de la sustentación de la tesis titulada: "**Modelo estadístico de calidad del agua del estero el macho, machala – el Oro**" presentada por egresado **Rafael Gilberto Zhindon Almeida** para optar el grado académico de Doctor en Estadística Matemática Aplicada.

Concluida la sustentación y absueltas las preguntas, por parte la sustentante y después de la correspondiente, deliberación el jurado, conforme a lo normado en el artículo n° 111 del reglamento de tesis de la universidad nacional de tumbes, declara al sustentante **APROBADO**, con el calificativo de **SOBRESALIENTE**.

Por lo anterior, el sustentante está expedito para iniciar los trámites correspondientes y conducentes a la obtención del grado académico de Doctor en Estadística Matemática Aplicada, en conformidad con lo normado en la Ley Universitaria N° 30220, el Texto Único Ordenado del Estatuto, El Reglamento General, el Reglamento General de Grados Títulos y el Reglamento de Tesis de la Universidad Nacional de Tumbes.

Siendo las veinte y uno horas, del mismo día, se dió por concluido la ceremonia académica, procediendo a firmar el acta en presencia de público asistente.

Dr. Napoleón Puño Lecarnaque
(Presidente),
DNI: 00225904
Código ORCID: 0000000250088085

Dr. Luis Jhony Caucha Morales
(Secretario)
DNI: 41378388
Código ORCID: 0000-0002-4786-9008

Dr. Juan Santiago Blas Perez
(vocal)
DNI: 32845810
Código ORCID: 0000-0002-97413164

Dr. Humberto Lorenzo Narva Roncal
(vocal)
DNI: 18850169
Código ORCID: 0009-0006-5668-4191

Dr. Walter Javier Castañeda Guzmán
(asesor)
DNI: 17804496
Código ORCID: 0000-0002-9483-0210

C.c.
Jurado de Tesis (03).
Asesor
Interesado
Coasesor (es)
Unidad de Investigación.
Archivo (Director EPG).

Informe de originalidad Turnitin

Modelo estadístico de calidad del agua del Estero El Macho, Machala-El Oro

por Rafael Gilberto Zhindon Almeida



Fecha de entrega: 15-oct-2024 01:35p.m. (UTC-0500)

Identificador de la entrega: 2486272999

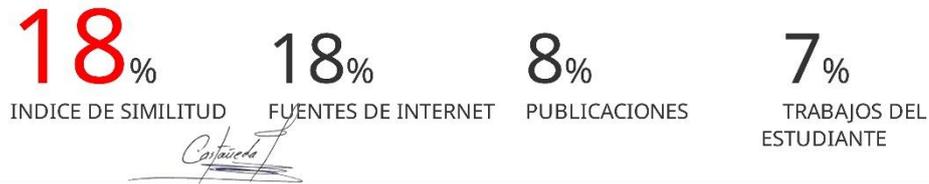
Nombre del archivo: Tesis_Aprobada_Rafael_Zhindon_3.docx (4.52M)

Total de palabras: 20029

Total de caracteres: 109360

Modelo estadístico de calidad del agua del Estero El Macho, Machala-El Oro

INFORME DE ORIGINALIDAD



FUENTES PRIMARIAS

1	www.produccioncientificaluz.org Fuente de Internet	3%
2	www.coursehero.com Fuente de Internet	1%
3	dokumen.pub Fuente de Internet	1%
4	hdl.handle.net Fuente de Internet	1%
5	repositorio.untumbes.edu.pe Fuente de Internet	1%
6	revistes.ub.edu Fuente de Internet	<1%
7	1library.co Fuente de Internet	<1%
8	docplayer.es Fuente de Internet	<1%
9	www.repositorio.usac.edu.gt Fuente de Internet	

		<1 %
10	JORGE BLUHM GUTIÉRREZ. "Modelación unidimensional de la calidad del agua en embalses. Análisis comparativo de modelos y multivariantes.", Universitat Politecnica de Valencia, 2008 Publicación	<1 %
11	Submitted to Universidad Andrés Bello Trabajo del estudiante	<1 %
12	es.slideshare.net Fuente de Internet	<1 %
13	qdoc.tips Fuente de Internet	<1 %
14	repositorio.ug.edu.ec Fuente de Internet	<1 %
15	repositorio.unasam.edu.pe Fuente de Internet	<1 %
16	www.slideshare.net Fuente de Internet	<1 %
17	Submitted to uncedu Trabajo del estudiante	<1 %
18	Submitted to Fakultas Ekonomi Universitas Indonesia Trabajo del estudiante	<1 %

19	dspace.uces.edu.ar:8180 Fuente de Internet	<1 %
20	repositorio.ucv.edu.pe Fuente de Internet	<1 %
21	rstudio-pubs-static.s3.amazonaws.com Fuente de Internet	<1 %
22	doczz.es Fuente de Internet	<1 %
23	Submitted to Universidad Internacional de la Rioja Trabajo del estudiante	<1 %
24	Submitted to Universidad Nacional de Tumbes Trabajo del estudiante	<1 %
25	ojs.latu.org.uy Fuente de Internet	<1 %
26	repositorio.unh.edu.pe Fuente de Internet	<1 %
27	idus.us.es Fuente de Internet	<1 %
28	estadistica-dma.ulpgc.es Fuente de Internet	<1 %
29	bibdigital.epn.edu.ec Fuente de Internet	<1 %

30	docplayer.fr Fuente de Internet	<1 %
31	www.crautonomia.gov.co Fuente de Internet	<1 %
32	dspace.esPOCH.edu.ec Fuente de Internet	<1 %
33	dspace.unl.edu.ec Fuente de Internet	<1 %
34	text-id.123dok.com Fuente de Internet	<1 %
35	libros.umariana.edu.co Fuente de Internet	<1 %
36	repositorio.utn.edu.ec Fuente de Internet	<1 %
37	vdocuments.com.br Fuente de Internet	<1 %
38	www.cepis.org.pe Fuente de Internet	<1 %
39	repositorio.utc.edu.ec Fuente de Internet	<1 %
40	snap3.uas.mx Fuente de Internet	<1 %
41	Submitted to Pontificia Universidad Catolica del Ecuador - PUCE	<1 %

Trabajo del estudiante

42	Submitted to Universitas Diponegoro Trabajo del estudiante	<1 %
43	dspace.ucuenca.edu.ec Fuente de Internet	<1 %
44	www.controlsanitario.gob.ec Fuente de Internet	<1 %
45	Submitted to UCL Trabajo del estudiante	<1 %
46	grad.uprm.edu Fuente de Internet	<1 %
47	ri.ues.edu.sv Fuente de Internet	<1 %
48	Submitted to Fundaci3n Universitaria Cat3lica del Norte Trabajo del estudiante	<1 %
49	dspace.uah.es Fuente de Internet	<1 %
50	tel.archives-ouvertes.fr Fuente de Internet	<1 %
51	pdffox.com Fuente de Internet	<1 %
52	repository.untag-sby.ac.id Fuente de Internet	<1 %

53	worldwidescience.org Fuente de Internet	<1 %
54	gabrielaeman.wordpress.com Fuente de Internet	<1 %
55	repositorio.untrm.edu.pe Fuente de Internet	<1 %
56	repositorio.unu.edu.pe Fuente de Internet	<1 %
57	revistas.unal.edu.co Fuente de Internet	<1 %
58	1library.net Fuente de Internet	<1 %
59	Margaret Ferreira dos Santos. "Subsídios para o planejamento conservacionista da bacia hidrográfica do Ribeirão Feijão (São Carlos, Itirapina, Analândia, SP)", Universidade de São Paulo. Agência de Bibliotecas e Coleções Digitais, 1993 Publicación	<1 %
60	Submitted to Universidad Andina del Cusco Trabajo del estudiante	<1 %
61	repositorio.uap.edu.pe Fuente de Internet	<1 %
62	bindani.izt.uam.mx Fuente de Internet	<1 %

63	dspace.ups.edu.ec Fuente de Internet	<1 %
64	repositorio.uancv.edu.pe Fuente de Internet	<1 %
65	repositorio.uladech.edu.pe Fuente de Internet	<1 %
66	repositorio.unemi.edu.ec Fuente de Internet	<1 %
67	repositorio.upagu.edu.pe Fuente de Internet	<1 %
68	riull.ull.es Fuente de Internet	<1 %
69	digilib.poliupg.ac.id Fuente de Internet	<1 %
70	repository.unipasby.ac.id Fuente de Internet	<1 %
71	ria.utn.edu.ar Fuente de Internet	<1 %
72	scholar.uprm.edu Fuente de Internet	<1 %
73	www.periodicotribuna.com.ar Fuente de Internet	<1 %
74	Submitted to Universidad Cesar Vallejo Trabajo del estudiante	<1 %

75	boletin.invemar.org.co Fuente de Internet	<1 %
76	doaj.org Fuente de Internet	<1 %
77	tesis.ipn.mx Fuente de Internet	<1 %
78	ojs.unud.ac.id Fuente de Internet	<1 %

Excluir citas Activo
Excluir bibliografía Activo

Excluir coincidencias < 15 words

DEDICATORIA

Dedico este trabajo de investigación a mi familia, quienes han sido mi mayor apoyo y motivación a lo largo de este arduo camino hacia el doctorado. Su amor incondicional, comprensión y sacrificio han sido la fuerza impulsora detrás de cada paso que he dado. A ustedes les debo mi éxito y les agradezco profundamente por estar siempre a mi lado, brindándome aliento en los momentos difíciles y celebrando conmigo cada logro alcanzado. Esta tesis es un tributo a su amor y apoyo inquebrantables. Gracias por ser mi roca en este viaje hacia la excelencia académica.

AGRADECIMIENTO

Agradezco profundamente a los distinguidos docentes de la Universidad de Tumbes, en especial a mi asesor Dr. Walter Javier Castañeda Guzman y Co-asesor Dr. Raul Alfredo Sanchez Ancajima, por su invaluable orientación, dedicación y apoyo durante mi trayectoria académica. Su compromiso con la excelencia educativa y su pasión por el conocimiento han sido una fuente constante de inspiración para mí. Gracias a su guía experta, he crecido tanto personal como intelectualmente. Su influencia perdurará en mi vida, y estoy profundamente agradecido por el impacto positivo que han tenido en mi desarrollo como investigador y como persona.

INDICE GENERAL

DEDICATORIA.....	xiv
AGRADECIMIENTO.....	xv
INDICE DE TABLAS	xix
INDICE DE FIGURAS	xxi
RESUMEN.....	xxii
ABSTRACT	xxiii
RESUMO.....	xxiv
CAPITULO I.....	25
INTRODUCCIÓN	25
CAPÍTULO II.....	28
PLANTEAMIENTO DEL PROBLEMA.....	28
2.1. Situación problemática	28
2.2. Formulación del problema.....	28
2.3. Justificación	29
2.4. Objetivos.....	29
2.4.1. Objetivo general	29
2.4.2. Objetivos específicos	30
CAPÍTULO III.....	31
ESTADO DEL ARTE	31
3.1. Antecedentes	31
3.2. Bases teórico-científicas	39
3.2.1. Control estadístico de calidad.....	39
3.2.2. Modelo estadístico	40
3.2.2.1. Modelos multivariantes.....	40
3.2.2.2. Análisis de Componentes Principales (Principal Component Analysis, PCA) 41	

3.2.3.	Validación del modelo estadístico	42
3.2.3.1.	Linealidad	42
3.2.3.2.	Normalidad	43
3.2.3.3.	Homocedasticidad	43
3.2.3.4.	Independencia de los errores	43
3.2.4.	Contaminación del agua	43
3.2.4.1.	Principales contaminantes del agua.....	44
3.2.4.2.	Calidad de agua.....	44
3.3.	Definición de términos básicos	45
CAPÍTULO IV		47
METODOLOGÍA.....		47
4.1.	Tipo de estudio y diseño de contrastación de hipótesis	47
4.1.1.	Población, muestra y muestre	48
4.2.	Formulación de la hipótesis	48
4.2.1.	Hipótesis general	48
4.2.2.	Hipótesis específicas	48
4.3.	Definición y operacionalización de Variables.....	48
4.3.1.	Operacionalización de Variables	49
4.4.	Métodos, técnicas e instrumentos de recolección de datos	50
4.5.	Plan de procesamiento y análisis de datos.....	50
CAPÍTULO V		52
RESULTADOS		52
5.1.	Parámetros iniciales	52
5.2.	Parámetros estudiados (data 2019).....	61
a)	Parámetros de biológicos	65
b)	Parámetros físicos	65
c)	Parámetros microbiológicos	67

d) Parámetros químicos.....	68
5.3. Análisis de Componentes Principales (ACP).....	70
5.4. Modelo de regresión lineal múltiple.....	76
5.5. Validación del modelo	87
CAPITULO VI	98
CAPITULO VII	101
CONCLUSIONES.....	101
CAPITULO VIII	102
RECOMENDACIONES	102
CAPITULO IX.....	103
REFERENCIAS BIBLIOGRÁFICAS.....	103
I. ANEXOS.....	112
1.) Matriz de consistencia.....	112
2.) Mapa de la ubicación del Estero El Macho respecto de la ciudad de Machala. Ubicación de sitios de toma de muestras.	113
3.) Normativa de calidad de agua	114
4.) Cuadro de coeficientes de correlación.....	116

INDICE DE TABLAS

Tabla 1 Operacionalización de las variables	49
Tabla 2 Resultados iniciales de monitoreo de agua.	53
Tabla 3 Matriz de correlaciones iniciales.	59
Tabla 4 Valores de los parámetros fisicoquímicos y biológicos en el sector 1	61
Tabla 5 Valores de los parámetros fisicoquímicos y biológicos en el sector 2.....	62
Tabla 6 Valores de los parámetros fisicoquímicos y biológicos en el sector 3.....	63
Tabla 7 Valores de los parámetros fisicoquímicos y biológicos en el sector 4.....	64
Tabla 8 Matrix de correlacion ^a	70
Tabla 9 Prueba KMO y Bartlett.....	71
Tabla 10. Comunalidades.....	72
Tabla 11. Varianza total explicada.....	72
Tabla 12. Matrix de componente ^a	74
Tabla 13 Comunalidades.....	74
Tabla 14 Varianza total explicada.....	75
Tabla 15. Matrix de componentes ^a	76
Tabla 16 Resumen de casos	77
Tabla 17 Prueba de normalidad	78
Tabla 18 Correlaciones	79
Tabla 19 Variables ingresadas/eliminadas ^a	84
Tabla 20 Model Summary	84
Tabla 21. ANOVA ^a	84
Tabla 22 Coeficientes ^a	85
Tabla 23. Variables ingresadas/eliminadas ^a	85
Tabla 24 Resumen del modelo.....	85
Tabla 25 ANOVA ^a	86
Tabla 26. Coeficientes ^a	86
Tabla 27 Resumen del modelo.....	87
Tabla 28 ANOVA ^a	88
Tabla 29 Coeficientes ^a	88
Tabla 30 Residuales ^a	90
Tabla 31 Resumen del modelo. Software IBM SPSS ^b	91
Tabla 32 Resumen del modelo. Software R DW	92
Tabla 33 Prueba Breusch-Godfrey	92
Tabla 34 Prueba de L jung-Box.....	92
Tabla 33 Resumen del modelo. Software R	94

Tabla 34. Prueba de Kolmogorov-Smirnov	96
Tabla 35 Prueba de normalidad Lilliefors (Kolmogorov-Smirnov). R	96
Tabla 36 Coeficientes ^a	97
Tabla 37 Coeficientes.....	97
Tabla 34. Criterios de calidad admisibles para la preservación de la vida acuática y silvestre en aguas dulces, marinas y de estuarinos.	114

INDICE DE FIGURAS

Figura 1: Demanda Bioquímica Oxígeno.....	54
Figura 2: Potencial de hidrógeno pH (situación actual)	54
Figura 3: Demanda química de oxígeno.....	55
Figura 4: Oxígeno disuelto	55
Figura 5: Nitritos	56
Figura 6: Nitratos (situación actual)	56
Figura 7 : Arsénico	56
Figura 8: Cadmio.....	57
Figura 9: Cobre	57
Figura 10: Hierro	57
Figura 11: Coliformes fecales y coliformes totales (situación actual).....	58
Figura 12: Demanda Bioquímica Oxígeno.....	65
Figura 13: Solidos disueltos totales	65
Figura 14: Temperatura del agua.....	66
Figura 15: Turbidez	66
Figura 16: Aceites Grasa.....	67
Figura 17: Coliformes fecales y coliformes totales	67
Figura 18: Potencial de hidrógeno pH	68
Figura 19: Nitritos	68
Figura 20: Nitratos.....	69
Figura 21: Hierro	69
Figura 22: Sulfatos	69
Figura 23: Figura De Sedimentación	73
Figura 24: ACP2 Sedimentación.....	75
Figura 25: Poca relación positiva entre las dos variables. Software IBM SPSS ...	90
Figura 26: Poca relación positiva entre las dos variables. Software R.....	91
Figura 27: Poca relación positiva entre las dos variables.....	93
Figura 28: Poca relación positiva entre las dos variables.....	93
Figura 29: Agrupamientos débiles. Software IBM SPSS.....	94
Figura 30: Verificación de la homocedasticidad del modelo.....	94
Figura 31: Variables cumplen principio de normalidad.....	95
Figura 29. Ubicación del canal y punto de toma de muestras.....	113

RESUMEN

Este estudio empleó técnicas estadísticas multivariadas para construir un modelo que prediga la calidad del agua en el Estero El Macho, ubicado en Machala, El Oro. Mediante un enfoque correlacional, se aplicaron análisis de regresión lineal múltiple y de componentes principales. Los hallazgos revelaron que indicadores como coliformes fecales, demanda bioquímica y química de oxígeno, hierro y oxígeno disuelto superan los niveles permitidos. El modelo de regresión, centrado en el oxígeno disuelto y los nitritos, explica el 88.20% de la variabilidad en la calidad del agua, mientras que con el análisis de componentes principales alcanza una explicación del 73.815%.

Palabras clave: calidad de agua, componentes principales, modelos estadísticos, regresión lineal múltiple.

ABSTRACT

This study utilized multivariate statistical techniques to develop a model for predicting water quality in the Estero El Macho, located in Machala, El Oro. Employing a correlational approach, multiple linear regression and principal component analyses were conducted. The findings indicated that parameters such as fecal coliforms, biochemical oxygen demand, chemical oxygen demand, iron, and dissolved oxygen exceeded permissible levels. The regression model, centered on dissolved oxygen and nitrites, explained 88.2% of the variability in water quality, while principal component analysis accounted for 73.815%.

Keywords: water quality, principal components, statistical models, multiple linear regression.

RESUMO

Este estudo utilizou técnicas estatísticas multivariadas para desenvolver um modelo que preveja a qualidade da água no Estero El Macho, localizado em Machala, El Oro. Utilizando uma abordagem correlacional, foram realizadas análises de regressão linear múltipla e de componentes principais. Os resultados indicaram que parâmetros como coliformes fecais, demanda bioquímica de oxigênio, demanda química de oxigênio, ferro e oxigênio dissolvido excederam os níveis permitidos. O modelo de regressão, centrado no oxigênio dissolvido e nitritos, explicou 88,2% da variabilidade na qualidade da água, enquanto a análise de componentes principais representou 73.815%.

Palavras-chave: qualidade da água, componentes principais, modelos estatísticos, regressão linear múltipla.

CAPITULO I

INTRODUCCIÓN

El agua es un elemento esencial para la vida, el mismo que es utilizado en la alimentación, agricultura, industria, entre otras. El monitoreo orientado al control de la calidad de los recursos hídricos conlleva a un diagnóstico de su estado a través de análisis de indicadores químico-físicos de la calidad de agua, obtenidos a través de mediciones y observaciones sistemáticas de las variables de las aguas superficiales. Este recurso se encuentra en constante movimiento que, a través de un ciclo permanente, va formando en la superficie de la Tierra los lagos, lagunas, ríos, arroyos, entre otras aguas en circulación o en reposo (Banco Mundial, 2011).

Dado que el agua es fundamental para la vida, es crucial conocer su calidad en función de los diversos usos que se le asigna, como el riego, la recreación o el consumo humano. En esta investigación, nos centraremos en analizar la calidad del agua y su relación con los diferentes usos que se le pueden dar. El monitoreo continuo de la calidad del agua permite obtener información valiosa sobre sus características físicas, químicas y biológicas en un determinado cuerpo de agua, facilitando así su gestión y protección (Gil Gómez, 2014).

La evaluación de la calidad del agua se realiza tradicionalmente mediante el cálculo de índices de calidad, los cuales permiten cuantificar las desviaciones de los valores considerados normales y detectar posibles riesgos para la salud. Estos índices se basan en el análisis de diversos parámetros como el oxígeno disuelto, el pH, la temperatura, los nitratos, los coliformes fecales, los sólidos disueltos y la turbidez del agua.

Las actividades humanas, como las industriales y agrícolas, pueden contaminar las aguas superficiales con sustancias químicas o biológicas. Para evaluar adecuadamente el impacto de esta contaminación y comprender su variabilidad en el tiempo y el espacio, es indispensable complementar las mediciones de calidad del agua con análisis estadísticos. Estos análisis nos permiten identificar patrones y tendencias en los datos, lo que es fundamental para tomar decisiones informadas sobre la gestión de los recursos hídricos (Torres et al., 2009).

Los modelos estadísticos se erigen como herramientas indispensables para predecir la evolución de la calidad del agua en cuerpos superficiales. Al traducir los datos de laboratorio en ecuaciones matemáticas, estos modelos nos permiten simular escenarios futuros y anticipar los efectos de diferentes intervenciones, lo que resulta crucial para diseñar estrategias de remediación y gestión eficientes (Gómez Miranda & Peñuela Mesa, 2016).

Según Torres, Hernán y Patiño (2009), la estadística es una herramienta esencial para comprender la dinámica de la calidad del agua en ríos, arroyos y esteros. Al analizar los índices de calidad del agua, la estadística permite identificar patrones espaciales y temporales, detectar problemas específicos y establecer prioridades para una gestión más eficiente de estos recursos.

El análisis multivariante ofrece un conjunto de herramientas estadísticas que nos permiten estudiar de forma integral datos complejos, considerando múltiples variables al mismo tiempo. Estas técnicas no solo describen los datos, sino que también nos permiten realizar inferencias sobre la población de la cual provienen, lo que es fundamental para tomar decisiones basadas en evidencia (Closas et al., 2013).

La presente investigación doctoral se enfoca en abordar la problemática de la contaminación del Estero El Macho, un cuerpo de agua de vital importancia para la ciudad de Machala. A través de la aplicación de técnicas estadísticas multivariantes, se desarrollará un modelo predictivo que permita evaluar la calidad del agua y anticipar posibles escenarios futuros. Considerando que el estero recibe un significativo aporte de aguas residuales domésticas y drenajes agrícolas, provenientes de aproximadamente el 48% de la población de Machala y de 800 hectáreas de cultivos de banano, respectivamente (Sarie Palas, 2020), esta

investigación busca contribuir a una gestión más sostenible de este recurso hídrico.

A través de una representación matemática con un enfoque de estadística multivariante se puede deducir los fenómenos que influyen sobre los parámetros de calidad de este recurso natural importante para los seres vivos.

CAPÍTULO II

PLANTEAMIENTO DEL PROBLEMA

2.1. Situación problemática

La cantidad de aguas residuales que llega al estero producto de las actividades antrópicas están cambiando su calidad; el no existir un mecanismo de fácil acceso que permita conocer el estado de contaminación de este cuerpo de agua, influye para que no se tomen medidas para su control.

El Estero El Macho, en Machala, recibe un caudal significativo de aguas residuales domésticas provenientes de aproximadamente el 48% de la población de la ciudad, a través de 13 puntos de descarga que suman un total de 422.40 litros por segundo. Adicionalmente, el estero funciona como drenaje para 800 hectáreas de cultivos de banano, lo que agrava aún más la contaminación del cuerpo de agua (Aguas Machala EP, 2022); siendo 139.52 lt/s de aguas servidas producto de esta actividad. Y, además, el drenaje de 700 ha de camarón que están asentadas en su margen a partir de la mitad hasta la zona baja y que aportan 1350 l/s; esta situación ha contaminado las aguas de ese ecosistema.

Con la finalidad de realizar un control se propone la realización de un estudio basado en un modelo estadístico multivariante de regresión lineal múltiple y análisis de componentes principales. Este tipo de investigaciones resulta fundamental para comprender y predecir la dinámica de la calidad del agua en el Estero El Macho. Al analizar las relaciones entre las diversas variables, podemos anticipar cómo evolucionarán los parámetros físico-químicos y biológicos de este cuerpo de agua, lo que permitirá tomar decisiones más informadas para su gestión y protección.

2.2. Formulación del problema

El objetivo principal de este estudio es desarrollar un modelo estadístico sencillo y accesible para evaluar la calidad del agua del Estero El Macho. Mediante el análisis de indicadores fácilmente medibles, se busca brindar a las autoridades y a la

comunidad una herramienta útil para tomar decisiones informadas sobre la gestión del recurso hídrico. Esta propuesta surge ante la dificultad actual de realizar análisis complejos de calidad del agua, debido a limitaciones en cuanto a tiempo, costo y disponibilidad de laboratorios especializados (Yamashita et al., 2022). En este sentido, se ha planteado la siguiente interrogante cuestión:

¿El desarrollo de un modelo estadístico permitirá determinar en forma significativa la calidad del agua del estero El Macho de la ciudad de Machala, provincia de El Oro?

2.3. Justificación

Las descargas directas de las aguas residuales domésticas y agroindustriales, están cambiando las características hídricas del estero El Macho, poniendo en grave riesgo a este ecosistema y a todas las actividades económicas que depende de él (López, 2015). Por lo que, es importante conocer su nivel de contaminación para establecer la toma de decisiones tendientes a generar políticas de acción para evitar que se siga deteriorando este ecosistema (Abarca, 2007). En este sentido el modelo matemático planteado facilitaría este trabajo.

Al aplicar la estadística multivariante en este estudio, podemos desentrañar las complejas relaciones entre las diversas variables que influyen en la calidad del agua del Estero El Macho. Analizando simultáneamente múltiples datos, obtenemos una visión más completa y precisa del fenómeno, lo que nos permite comprender mejor las interacciones entre los diferentes factores que afectan la calidad del agua (Closas et al., 2013).

Dada la importancia de comprender la calidad del agua del Estero El Macho, este estudio busca identificar el modelo estadístico más adecuado para analizar sus parámetros físico-químicos y biológicos. Esto permitirá obtener una evaluación precisa de su estado y tomar decisiones informadas para su gestión.

2.4. Objetivos

2.4.1. Objetivo general

Desarrollar un modelo estadístico matemático para determinar la calidad del agua del estero El Macho de la ciudad de Machala, provincia de El Oro.

2.4.2. Objetivos específicos

- a) Analizar de forma estadística los parámetros físicos, químicos y biológicos que inciden en la calidad del agua del estero El Macho de la ciudad de Machala.
- b) Realizar un Análisis de Componentes Principales (ACP) para establecer el agrupamiento que mejor defina al modelo estadístico matemático para conocer la calidad de las aguas del Estero El Macho.
- c) Encontrar un modelo estadístico matemático utilizando regresión lineal múltiple que defina la calidad de agua del Estero EL Macho.
- d) Validar el modelo estadístico.

CAPÍTULO III

ESTADO DEL ARTE

3.1. Antecedentes

En el marco de las investigaciones sobre la calidad del agua, a continuación se exponen algunos referentes teóricos:

Pérez, Nardini y Galindo (2018) en su investigación cuyo propósito fue comparar los diferentes índices de calidad del agua (ICA) para el mejor desempeño tanto de los sectores económicos, así como para el consumo. Esta investigación es importante en relación a que es necesario diseñar o seleccionar adecuados parámetros e índices para la aplicación en un modelo estadístico, a fin de evitar/reducir subjetividades en los análisis de causa-efecto.

Méndez, Arcos y Cazorla (2020) señalan que un índice de calidad del agua comprende un valor numérico que sintetiza la información de determinados parámetros, ya sean estos físicos, químicos o microbiológicos. El estudio de los ICA, ayudan a establecer el origen y cuantificación de los factores contaminantes de los recursos hídricos y por ende si se está cumpliendo con las normativas sanitarias de una localidad.

Moreno, Toro y Carvajal (2008) realizan una revisión respecto a las metodologías para el monitoreo de la calidad del agua, considerando en primer lugar definir los objetivos del estudio, ya sean estos para: a) evaluar tendencias; b) valorar estándares específicos de calidad; c) determinar causas externas; d) examinar variaciones de parámetros en momentos específicos, entre otros. Posteriormente se debe elegir las variables, que pueden ser sobre caudales, temperatura, pH, turbiedad, grupos bióticos, minerales, etc. Asimismo, es necesario ubicar los lugares de monitoreo el cual está en relación con las técnicas de análisis estadístico.

En su estudio, Arroyave, Moreno, Toro, Gallego y Carvajal (2013) utilizaron el modelo QUAL2KW para simular el comportamiento de la calidad del agua en un río y evaluar el impacto de diferentes medidas de gestión. Los autores destacaron la relevancia de los parámetros cinéticos en este tipo de análisis.

Según Samboni, Reyes y Carvajal (2011) la evaluación de la calidad del agua a través de indicadores físico-químicos y biológicos es fundamental para garantizar la protección de los recursos hídricos en diferentes ámbitos. Además, estos autores resaltan la relevancia de los monitoreos periódicos para detectar y mitigar los riesgos de contaminación.

Caho y López (2017) señalan que el índice de calidad del agua es una herramienta que ayuda a reconocer determinadas propiedades del recurso hídrico en función a ciertos estándares ambientales, biológicos o de utilidad. En este sentido, determinaron la calidad del agua empleando metodologías UWQI para consumo humano y CWQI para cuencas de ríos. Asimismo, con los datos obtenidos, utilizaron la técnica de análisis de componentes principales para comprender el comportamiento de los factores físico, químicos y microbiológicos de las aguas analizadas.

Vigueras, Carbajal, Sánchez y Hernández (2016) establecieron un modelo computacional para evaluar la calidad de agua para la pesca en cautiverio. Los parámetros empleados fueron de tipo físico-químicos examinados bajo una escala categorizada en cuatro niveles, utilizando análisis de correlaciones simples que no integran parámetros medioambientales importantes para este tipo de cultivos.

Carrillo y Duran (2018) desarrollaron un método de análisis sensorial para determinar la calidad de agua. Este método consiste en procesos analíticos en laboratorio usando un sistema multisensorial para gases químicos. Este procedimiento puede ser utilizado como una alternativa para la toma y control de muestras de calidad del agua.

Torres, Hernán y Patiño (2009) analizan la importancia del ICA como herramienta para evaluar las características fisicoquímicas y microbiológicas de los recursos hídricos, permitiendo conocer variaciones espacio-temporal de las cualidades del agua; así como proporcionar valores que, con la aplicación de técnicas estadísticas, permiten identificar tendencias.

Pauta y Chang (2016) exponen resultados sobre la calidad de agua de un río. Para ello, emplearon el índice de calidad de agua sobre parámetros físico-químico y bacteriológicos. Este tipo de índice permite el análisis de la tendencia de la calidad, importantes para la adecuada gestión de las fuentes hídricas.

En base a estos artículos se puede comprender la importancia de la aplicación y estudio de índices de calidad de agua para la obtención de datos de las cualidades y características de las fuentes hídricas. En los siguientes artículos se presenta la importancia de las técnicas y métodos estadísticos.

Osorio y Suárez (2014) abordan la temática de la importancia de la estadística para la generación de información técnica en la toma de decisiones. Dentro de los aspectos relevantes de la estadística está la posibilidad de efectuar estudios de control de calidad, permitiendo medir, predecir y mejorar los procesos u objetos de análisis. Para lo cual es posible aplicar diferentes procedimientos como análisis exploratorios, muestreos, series de tiempo, técnicas de regresión, entre otros, fundamentales para la comprensión de fenómenos complejos o la solución de problemas.

Poza (2008) hizo una investigación considerando la necesidad de medir y simplificar los datos de un conjunto de variables para un análisis multivariado. En este sentido, explica técnica estadística del Análisis Factorial (AF), procedimiento orientado a la reducción de datos a variables no observables, llamadas factores, eliminando la redundancia informativa. Estos factores son muy importantes para interpretar, así como para dotar de sentido a una matriz de datos sobre un fenómeno o evento.

Escobar (2008) efectúa una descripción sobre la aplicabilidad del análisis multivariante en función a estimaciones de indicadores de calidad ambiental. Entre las técnicas que establecen correlaciones de matrices multivariantes se encuentran: análisis de componentes principales, análisis por factores, análisis de regresión, el análisis de correlación canónica, análisis de variables canónicas, análisis discriminante canónico, análisis por agrupación, análisis multivariado de la varianza, entre las principales. Estas técnicas están dirigidas a reducir la dimensión del conjunto de datos e identificar nuevas variables que expliquen con facilidad el fenómeno.

En relación con el desarrollo de estudios sobre modelos estadísticos para el control de calidad de aguas, a continuación, se presenta algunos de los principales trabajos

de tesis de referencia, tales como:

Guerrero (2019) en su investigación acerca de un análisis estadístico multivariado para evaluar la calidad del agua en una quebrada, se establecieron como propósitos seleccionar técnicas estadísticas para estudiar correlaciones, agrupaciones y variabilidades que permitieron diagnosticar las propiedades del recurso hídrico de forma espacial y temporal según la normativa colombiana. En este sentido, se escogieron como indicadores de calidad del agua parámetros de tipo fisicoquímicos, hidrobiológicos y microbiológicos.

La propuesta consistió en el diseño de un modelo estadístico para la interpretación de múltiples datos que ayudaron a identificar relaciones entre los parámetros elegidos, así como conocer la evolución en la escala espaciotemporal de la cuenca hídrica. La información del modelo estadístico proporcionó una base técnica de calidad del agua para la toma de decisiones respecto al nivel de protección ambiental, además de regulaciones para el ordenamiento territorial que permitan preservar este recurso en el tiempo.

Guano y Velasteguí (2014) desarrollaron un manual práctico para la recopilación y análisis de datos hidrológicos con el fin de simular la calidad del agua en ríos utilizando el software HEC-RAS. Su investigación se motivó por la creciente preocupación sobre el deterioro de la calidad del agua en el río Machángara, en Quito, debido a las descargas de contaminantes.

En este sentido, los modelos estadísticos y matemáticos, comprende una herramienta de relevancia para anticipar los impactos en las propiedades del agua, ayudando de esta forma a mejorar políticas de manejo y control de los recursos hídricos. Los resultados del trabajo investigativo contribuyeron a establecer simulaciones de parámetros de interés como demanda bioquímica de oxígeno, oxígeno disuelto, temperatura que fueron relevantes para identificar puntos contaminantes y sus causas.

Sotomayor (2016) destaca la relevancia de emplear técnicas estadísticas multivariantes para evaluar la calidad del agua en ríos, como el Paute. Según el autor, la creciente urbanización y las intensificadas actividades industriales y agrícolas ejercen una presión cada vez mayor sobre los recursos hídricos, deteriorando su

calidad. Por tanto, la implementación de programas de monitoreo de la calidad del agua genera volúmenes considerables de datos físico-químicos y microbiológicos que requieren ser analizados mediante técnicas estadísticas para modelar y evaluar los impactos de la urbanización en los cuerpos de agua.

Con el fin de asegurar una evaluación rigurosa de la calidad del agua, el estudio aplicó el análisis de componentes principales (PCA) a los datos multivariados, con el objetivo de obtener información sólida que sustente la implementación de estrategias efectivas para la gestión y recuperación de los ecosistemas ribereños.

Valencia (2007) en su estudio estadístico de la calidad del agua sobre una de las vertientes del río Ebro, en España, se elaboró una metodología estadística de clasificación de series temporales multivariantes irregulares de parámetros básicos como el pH, conductividad, sulfatos, demanda biológica de oxígeno, cloruros, coliformes totales, nitratos, entre otros factores de análisis de calidad del agua. Para las estimaciones estadísticas se realizaron 78 observaciones a lo largo de un período determinado de tiempo, las mismas que fueron analizadas a partir de tres técnicas:

- 1) Métodos de descomposición a fin de analizar las series temporales y relacionar los componentes existentes entre sí.
- 2) Método de alisamiento exponencial o medias móviles para verificar tendencias en base a las últimas observaciones.
- 3) Modelo ARIMA que es una metodología de análisis dinámico que emplea datos de series temporales, con el propósito de identificar y estimar parámetros.

Mediante este clúster de técnicas de análisis estadístico fue posible clasificar y obtener elementos importantes sobre concentraciones de sodio, cloro, así como el grado de contaminación biológica.

Gómez y Peñuela (2016) destacan que el análisis clúster, el análisis de componentes principales, el análisis factorial y el análisis discriminante son técnicas estadísticas idóneas para interpretar conjuntos de datos relacionados con la calidad del agua. La aplicación de técnicas estadísticas como el análisis de correlación canónica, modelos jerárquicos y modelos de ecuaciones estructurales permite desarrollar modelos

estadísticos sofisticados para el estudio de la calidad del agua. Estos modelos permiten analizar relaciones complejas entre variables químicas, físicas y factores ambientales, proporcionando una visión integral de los ecosistemas acuáticos.

La aplicación de estos métodos y modelos permite generar información crucial para realizar diagnósticos precisos y realizar predicciones confiables, lo que es fundamental para la toma de decisiones informadas en la gestión de la calidad del agua.

Bécher y Blarasin (2017) consideraron en su investigación sobre procesos de gestión de calidad de agua subterránea la aplicación de técnicas estadísticas y modelación numérica, el denominado análisis de agrupamiento (clúster), con el propósito de determinar vínculos entre los factores químicos y los parámetros geomorfológicos de la zona, estableciendo el grado de influencia de las propiedades hidroquímicas, así como posibilitando la anticipación de escenarios futuros sobre la calidad del agua subterránea.

Caballero (2009) señala la importancia de la aplicación de la estadística para la realización de análisis comparativos acerca de la calidad del agua y su relación con los residuos líquidos provenientes de la minería. En este sentido, las técnicas estadísticas contribuyeron al desarrollo de una modelación matemática que permitió el análisis, mediante una escala de valoración cualitativa, sobre el impacto de los efluentes mineros a lo largo un arroyo.

López y Palací (2014), en su investigación sobre el estudio multivariante de la calidad del agua, indican que es necesario efectuar un diagnóstico de tipo longitudinal y transversal. Para ello, que aplicó en primer lugar el Análisis de Componentes Principales (PCA) con la finalidad de reducir los datos de las variables de investigación. Posteriormente se utilizó el Análisis Discriminante mediante Mínimos Cuadrados Parciales (PLS-DA), teniendo por objetivo establecer una matriz de datos X – Y para de esta forma buscar relaciones entre los datos multivariantes de la matriz. Este procedimiento contribuyó con la detección de observaciones anómalas en las características de calidad del agua a lo largo del río en distintos lugares a través del tiempo.

Carrera, Guevara, Tamayo y Guallichico (2015) en su análisis multivariado de calidad de agua para uso agrícola, tuvo por propósito conocer las concentraciones de elementos inorgánicos a través de la aplicación de la técnica de componentes Principales (PCA), utilizando como parámetros de comparación la norma INEN 2169 y realizando 26 tomas de muestras, los cuales permitieron determinar las características de las aguas y su incidencia en los suelos de cultivo.

Casares y De Cabo (2018) en su estudio sobre análisis de los indicadores de calidad del agua, consideraron el análisis estadístico de parámetros inorgánicos y factores endógenos en una serie de tiempo aplicando la prueba de Mann-Kendall, la cual permitió establecer un conjunto de tendencias relevantes sobre la evolución de las características y condiciones del agua de la cuenca Matanza-Riachuelo, Argentina.

Rodríguez y Giménez (2017) en su investigación sobre los efectos de los residuos orgánicos en las aguas subterráneas, se utilizó el análisis de componentes principales (PCA) para evaluar las concentraciones de nitritos, nitratos y bacterias coliformes. El procedimiento investigativo se conformó por un análisis químico, bacteriológico y estadístico. Este último, consistió en la realización de una matriz con los datos químicos y bacteriológicos, utilizando la técnica multivariante descriptiva PCA. Los resultados del estudio permitieron determinar un alto riesgo sanitario, recomendando el establecimiento de un alto riesgo sanitario en el manejo de las aguas subterráneas.

Guzmán, Nava y Díaz (2015) orientan su estudio en relacionar la calidad del agua de consumo humano con la morbimortalidad infantil, ya que las cualidades del agua presentan una relación con la salud y aparición de enfermedades. El proceso de análisis consistió en la aplicación de métodos estadísticos descriptivos de correlación entre calidad de agua y un conjunto de enfermedades, así como un análisis de riesgos entre la variable calidad de agua y eventos en salud, utilizando la técnica de análisis de componentes principales, la cual permitió reducir el tamaño de los datos y poder identificar de esta forma los riesgos de mortalidad infantil.

En conclusión, se pudo demostrar el impacto de la calidad del agua en la salud de las personas, siendo necesario implementar políticas que mejoren el suministro de agua en la población.

Pérez, López, Hernández y González (2009) desarrollaron una investigación acerca de un modelo matemático para determinar la calidad del agua. Este modelo consistió en una evaluación del comportamiento de indicadores físico-químicos y microbiológicos, mediante el análisis de regresión.

Lorbes, García, Ohep y Milla (2014) realizaron una investigación utilizando una técnica multivariante, tal como el método del análisis discriminante canónico, para evaluar la calidad del agua de riego de tres pozos, considerando parámetros físico-químicos. Los resultados permitieron establecer relaciones significativas, así como discriminar o diferenciar claramente los tratamientos entre las variables de las tres fuentes de agua.

Ledesma, Bonansea, Rodríguez y Sánchez (2013) realizaron un análisis multivariado de componentes principales (ACP) generando una matriz que permitió establecer correlaciones significativas respecto de variables contaminantes del agua y variables fisicoquímicas. Los resultados de la investigación ayudaron al mejorar los procesos de monitoreo y gestión del agua.

Baldiris, Sánchez, Cabeza y Acosta (2017) efectuaron un estudio estadístico multivariado en función a una base de datos de 11 años, con el propósito de extraer información sobre la evolución e identificación de los factores contaminantes de la calidad de agua en la ciudad de Cartagena. Para el procesamiento de la información se utilizó el análisis de componentes principales, el cual ayudó a identificar factores principales de degradación de las aguas provenientes de residuos agrícolas y domésticos, así como la progresividad de las fuentes contaminantes en el tiempo.

Gil, Vizcaino y Montaña (2018) plantean como propósito de investigación efectuar una evaluación de la calidad del agua, aplicando indicadores conformados por 14 parámetros de carácter fisicoquímicos y microbiológicos. El procedimiento estadístico consistió en método del índice aritmético ponderado, así como a través de un ANOVA (Análisis de Varianza). Los resultados ayudaron al establecimiento de medidas para el tratamiento del agua de consumo humano.

Aparicio, Espinoza, Milla y Reyes (2013) desarrollaron un modelo estadístico para predecir la calidad del agua para el consumo de las personas. Para ello, se aplicó un modelo lineal de probabilidad de correlación serial, sobre una base de datos

cuantitativos conformado por parámetros físico, químicos y microbiológicos.

Gil Marin (2020) propone un modelo innovador para evaluar la calidad del agua subterránea, combinando el análisis de componentes principales para simplificar las relaciones entre los parámetros y la regresión lineal múltiple para realizar predicciones más precisas del índice de calidad del agua (ICA). Esta combinación metodológica permite una aproximación más realista a los procesos naturales.

Rubio y otros (2014) realizaron un estudio de calidad de agua para una represa. Por tanto, seleccionó como procedimiento estadístico un análisis de varianza (ANOVA), y, para el cálculo del ICA se emplearon parámetros fisicoquímicos y ambientales. Los resultados de la evaluación demostraron que el comportamiento de los parámetros físicos depende directamente de factores ambientales.

Delgado y otros (2020) efectuaron estudio de calidad ambiental de las aguas de un estuario. Para analizar la calidad del agua se aplicó un análisis de varianza ANOVA sobre parámetros fisicoquímicos. Este procedimiento ayudó a identificar el origen los contaminantes.

Los estudios consultados coinciden en señalar que el análisis de componentes principales, el análisis clúster, el análisis factorial y la regresión lineal múltiple son las técnicas estadísticas más empleadas en la evaluación de la calidad del agua. Estos métodos permiten desarrollar modelos predictivos que facilitan la toma de decisiones informadas en la gestión de los recursos hídricos para diversos usos.

3.2. Bases teórico-científicas

3.2.1. Control estadístico de calidad

El control estadístico hace referencia a la aplicación del método científico para analizar la calidad de un producto fabricado o un recurso de la naturaleza, con la finalidad de identificar desviaciones en función de determinados parámetros para mejorarlos, remediarlos o evitar que se generen algún tipo de problema (Barreto-Villanueva, 2012).

El control estadístico está orientado fundamentalmente en predecir un proceso o fenómeno en el tiempo, siendo un mecanismo de ayuda para la toma de decisiones y

mejora permanente. Aunque esta herramienta es implementada principalmente en los procesos industriales, su utilidad ha avanzado en procesos investigativos para la salud, seguridad alimentaria, conservación de recursos naturales, entre otras áreas de interés para el bienestar humano (Otero et al., 2003).

3.2.2. Modelo estadístico

Los modelos estadísticos son herramientas, basadas en ecuaciones matemáticas, que proporcionan una aproximación sobre determinadas realidades, así como para efectuar predicciones con un grado razonable de exactitud.

El proceso de modelación estadística depende de las variables que se requieren analizar o describir. Las variables son aquellas características o propiedades que tienen la capacidad de cambiar en función de sí mismo o por otras variables, y dicho cambio es medido y observado. Las variables pueden ser independientes e dependientes (Balzarini et al., 2015).

La variable independiente es la que se emplean para evaluar la incidencia, relación o influencia sobre la o las variables dependientes. La variable dependiente corresponde a aquella cualidad que se quiere explicar, medir o predecir. De esta forma, según las variables a examinar se debe escoger un modelo estadístico que mejor se adapte a los objetivos de la investigación (Pereyra, 2022).

3.2.2.1. Modelos multivariantes

La estadística multivariante también denominada estadística multivariada es aquella que se encarga de forma simultánea de observar y analizar más de una variable, por tanto, su utilización en para estudiar aquellos fenómenos que presentan gran cantidad de variables (Benavides Alarcón, 2023).

La estadística multivariante tiene como objetivo principal comprender las interrelaciones existentes entre múltiples variables. Este tipo de metodología aborda además los procedimientos de distribuciones de probabilidad multivariadas, para determinar la representación de los datos observados, así como para establecer qué datos son de interés para desarrollar inferencias (Aldas Manzano & Uriel Jiménez, 2017).

Según varios aspectos de análisis, los modelos multivariantes pueden dividirse en:

- **Análisis factorial:** técnica de reducción de datos a fin de obtener información relevante e individual de las variables. Los estudios de tipo factorial están orientados a explicar las correlaciones de diferentes variables en función de un menor número de variables o factores, dividiéndose en análisis de componentes principales y análisis de correspondencia (López-Aguado & Gutiérrez-Provecho, 2019).
- **Análisis de regresión:** estudia la relación o incidencia de dos tipos de variables, siendo una de ellas la variable explicativa, predictora o independiente y la otra la variable respuesta o dependiente. La primera indica la situación real en función de los datos recabados, la segunda presenta los cambios producto de la modificación de la variable independiente (Baeza-Serrato & Vázquez-López, 2014).

La regresión lineal que puede ser múltiple ayuda a determinar la relación que se genera entre una variable dependiente y un conjunto de variables independientes (Giraud Herrera & Morantes Quintana, 2017). La aplicación regresión lineal múltiple tiene entre sus diversas utilidades, observar de qué forma un conjunto de variables independientes pueden predecir la variable dependiente (Closas et al., 2013), el esquema simplificado del análisis de regresión múltiple es:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_q X_{iq} + \varepsilon_i$$

Importancia del análisis multivariante: este análisis, pilar fundamental de la investigación cuantitativa, destaca por su eficacia en la exploración y comprobación de fenómenos complejos. Sin embargo, dada su intrincada naturaleza, requiere el empleo de software especializado para el procesamiento y análisis de grandes volúmenes de datos (Aldas Manzano & Uriel Jiménez, 2017).

3.2.2.2. Análisis de Componentes Principales (Principal Component Analysis, PCA)

Al presentarse un problema donde se han identificado múltiples variables, surge la necesidad de encontrar una forma estructurada y sencilla de comprender las relaciones que existen en ellas. Por tanto, se requiere en primer lugar simplificar al

máximo los datos obtenidos de las variables, y posteriormente visualizar el resultado para así observar y describir cómo se ha agrupado la información facilitando su interpretación, así como la determinación de inferencias (Corradi et al., 2020).

En el método estadístico de análisis de componentes principales (PCA) lo que se pretende es reducir la complejidad de los datos de una muestra, conservando a su vez las dimensiones de la información. Por ejemplo, en una muestra de n objetos cada uno con una cantidad q de variables ($X_1, X_2, X_3, \dots, X_q$). El análisis de componentes principales busca un número inferior de factores ($p < q$) que expliquen de forma aproximada lo mismo que la cantidad q de variables del inicio de la investigación. En consecuencia, cuando se requería describir o analizar el objeto n con q variables, ahora es posible analizarlo exactamente igual con p factores, que es un valor simplificado. Las nuevas p variables obtenidas toman el nombre de componente principal.

Por lo tanto, el método de análisis de componentes principales ayuda a condensar gran cantidad de datos obtenida por múltiples variables, transformándolas en unos pocos elementos de análisis. Así este método es muy útil previo a la utilización de otras técnicas estadísticas, sin olvidar que siempre es necesario contar con los valores de las variables originales (Amat Rodrigo, 2017).

3.2.3. Validación del modelo estadístico

La validación de modelos hace referencia al proceso de verificación que todos los datos o parámetros a fin de considerarlos como válidos. La validación permite que los procesos del modelo estén en condición para ejecutarse. Dentro de las técnicas estadísticas, la validación del modelo puede realizarse por medio pruebas de: linealidad, normalidad homocedasticidad e independencia de los errores (Poza Lara, 2008).

3.2.3.1. Linealidad

Esta prueba indica la condición lineal de la relación existente entre las variables independientes y la variable dependiente del modelo. En un esquema de regresión lineal múltiple los diagramas de dispersión de los residuos permiten verificar que se está cumpliendo esta condición (Forero Gómez & Martínez Lozano, 2020).

3.2.3.2. Normalidad

Los contrastes de normalidad, como los denomina Amat Rodrigo (2017), cuantifican la discrepancia entre la distribución empírica de los datos y una distribución normal teórica con parámetros idénticos. La simetría en torno a la media es una característica distintiva de la distribución normal y un aspecto clave en estos análisis.

3.2.3.3. Homocedasticidad

Esta prueba muestra la uniformidad en las variaciones de los residuos a lo largo de los valores pronosticados. Esto supone que los residuos son independientes a los valores pronosticados, lo cual es posible comprobar mediante la realización de los diagramas de dispersión.

3.2.3.4. Independencia de los errores

Esta prueba consiste en que los errores en las mediciones de las variables independiente no tengan relación entre sí, es decir, demuestra que los residuos son independientes respecto de las variables independientes y los pronósticos. Si no se cumple esto existe un problema de autocorrelación (Forero Gómez & Martínez Lozano, 2020).

3.2.4. Contaminación del agua

El agua elemento vital para vida, en muchas ocasiones sufre de los efectos de la contaminación por sustancias químicas o biológicas, esto como consecuencia del entorno que rodea a las aguas superficiales producto de los desechos de las industrias, de las aguas residuales, explotación de los suelos, agriculturas, entre otras actividades del ser humano (Torres et al., 2009).

Por tanto, la contaminación del se la puede definir como la pérdida de pureza en relación con unas determinadas condiciones especiales o estandarizadas necesarias para ser utilizadas, lo cual puede generar resultados adversos para la salud de los seres vivos, en especial de las personas, debido a la alteración de las propiedades físicas, químicas y biológicas del agua (Samboni et al., 2011).

Por lo general, las zonas rurales y las poblaciones de escasos recursos económicos están expuestas están a la contaminación del agua, la mismo que puede ser de tipo biológica y química.

La contaminación biológica, comprende la presencia de microorganismos, tales como: virus, bacterias y protozoos, así como coliformes fecales, las cuales son generadoras de enfermedades no solo por el consumo, sino también por el contacto con las mismas. Este tipo de contaminación es ocasionada principalmente por las descargas de aguas residuales o servidas en las distintas fuentes hídricas.

La contaminación química, hace referencia a la presencia de elementos químicos, en estado natural o producidos por el hombre, tales como presencia de nitratos, sólidos disueltos, entre otros factores que generan turbiedad o alteran el pH y la temperatura. Este tipo de contaminación es ocasionada principalmente por los desechos de las actividades industriales (Doménech, 2002).

Bajo estas circunstancias, al ser el agua superficial o dulce un recurso escaso, es importante lograr que las autoridades sanitarias realicen los respectivos controles, los cuales deben ser analizados con criterios profesionales y procedimientos estadísticos pertinentes para la conservación de los *recursos hídricos* (Sotomayor, 2016).

3.2.4.1. Principales contaminantes del agua

Según Brousett-Minaya y otros (2018), entre los principales contaminantes de agua, que requieren un constante valoración y análisis se encuentran:

- a) **Agentes Patógenos:** comprende los organismos biológicos tales como bacterias, virus y protozoos que generan múltiples enfermedades tanto a las personas, animales, como a plantas. Generalmente provienen de desechos orgánicos que son vertidos sin ningún tratamiento en ríos, embalses o esteros.
- b) **Sustancias químicas inorgánicas:** comprenden los metales como el mercurio y el plomo, así como los ácidos.
- c) **Sustancias químicas orgánicas:** hace referencia a los conformados por carbono, oxígeno, nitrógeno, fósforo tales como petróleo, aceites, alcoholes, entre otros.
- d) **Sustancias radiactivas:** son átomos que se desintegran de forma natural. Pueden emitir partículas alfa, partículas beta y radiación gamma.

3.2.4.2. Calidad de agua

La calidad del agua es un concepto vinculado al cumplimiento de determinadas propiedades o características físicas, químicas y microbiológicas que no producen

una afectación en la salud de los seres vivos, en especial para las personas. Comprende una garantía razonable de salubridad que permite el uso y consumo del recurso líquido (Gómez Miranda & Peñuela Mesa, 2016).

Para lograr niveles de calidad de agua, según los requerimientos de los organismos de salud, es necesario un monitoreo constante desde las distintas fuentes hídricas, con la finalidad de minimizar riesgos de posibles enfermedades. En resumen, la calidad del agua se comprueba cotejando las características físicas, químicas y bacteriológicas, en correspondencia a unas directrices o estándares de calidad, proporcionadas por las autoridades competentes (Suárez-Medina, 2020).

En consecuencia, es relevante la existencia de planes de salubridad vinculados con el recurso del agua, a fin de mitigar o eliminar los posibles peligros latentes en ella, así como también contar con procesos de control y seguimiento, basados en análisis estadísticos que permitan predecir escenarios de contaminación, para futuras acciones de remediación (Recabarren Santibáñez, 2016).

3.3. Definición de términos básicos

Estadística: es una rama de la matemática que tiene el propósito de recabar, analizar, describir y estudiar un conjunto de datos buscando de esta forma una explicación o predicción aproximada de determinados fenómenos de interés, facilitando la toma de decisiones y el establecimiento de soluciones.

Estadística descriptiva: es la encargada de resumir y organizar la información cuantitativa a un conjunto mínimo de valores que permiten describir las principales características de un hecho, tales como: el promedio, la mediana, varianzas, desviaciones, entre otras (García Merino et al., 2023).

Estadística inferencial: utiliza una serie de técnicas y procedimientos que permiten la realización de generalizaciones. La base de este tipo de estadística son las técnicas descriptivas. En general la estadística inferencial está orientada a: buscar conclusiones, así como determinar relaciones y predicciones entre variables (Ramírez Ríos & Polack Peña, 2020).

Estadística multivariante: comprende aquellos métodos que posibilitan el estudio simultáneo sobre el comportamiento de tres o más variables de investigación. Es

utilizada para establecer correlaciones entre múltiples elementos, así como buscar una distribución probabilística representativa de las variables que ayude a simplificar los modelos estadísticos para la comprensión de un fenómeno (Aldas Manzano & Uriel Jiménez, 2017).

Regresión lineal: técnica estadística que estudia la relación entre variables. Su función es lineal de la forma $y = ax + b$. La regresión lineal es empleada para el estudio de una gran variedad de fenómenos o problemas de distintos contextos de la ciencia.

Regresión lineal simple: mediante este modelo se trata de explicar la relación o incidencia entre la variable dependiente y una única variable independiente. Su representación gráfica es una línea recta (Forero Gómez & Martínez Lozano, 2020).

Regresión lineal múltiple: mediante este modelo se trata de explicar la relación o incidencia entre la variable dependiente y dos o más variables independientes. Su representación gráfica es un plano de dispersión (Mendoza et al., 2023).

Análisis de componentes principales: es un método de simplificación o reducción de datos de múltiples variables, examinando la magnitud o representatividad de los variables originales, obteniendo de esta forma unas nuevas que expliquen o describan más fácilmente un fenómeno (Corradi et al., 2020).

Índice de calidad del agua: es un número que representa las condiciones físicas, químicas y biológicas del agua más adecuadas para su consumo o uso, es decir, es un valor que señala los indicios de contaminación del recurso hídrico (Rubio Arias et al., 2014).

Parámetros de calidad microbiológica: hace referencia de la presencia de microorganismos patógenos de diferentes tipos, tales como: Bacterias, Virus, Protozoo, Coliformes fecales.

Parámetros de calidad físico-químicos: los parámetros físicos comprenden características como el olor, sabor y apariencia o turbidez. Los parámetros químicos constituyen elementos adicionales a la composición del agua que alteran sus propiedades físicas, tales como: Potencial de hidrogeno, Oxígeno Disuelto, Temperatura, Salinidad. (Flores Suárez & Pozo Merejildo, 2023)

CAPÍTULO IV

METODOLOGÍA

4.1. Tipo de estudio y diseño de contrastación de hipótesis

El propósito principal de este estudio es profundizar en los fundamentos teóricos de la modelación estadística aplicada a la calidad del agua, con el fin de establecer una base sólida para futuras investigaciones. Complementariamente, se llevará a cabo una aplicación práctica de los conceptos teóricos, mediante el diseño de un modelo estadístico matemático que describa y pronostique la calidad del agua del estero El Macho.

Según Baeza-Serrato y Vázquez-López (2014) el tipo de enfoque de investigación puede clasificarse como descriptiva debido a su enfoque en detallar las características del agua. Sin embargo, al emplear un diseño correlacional y modelos multivariantes, también busca identificar relaciones entre las variables estudiadas.

El modelo de regresión lineal múltiple, planteado con q variables, mantiene la siguiente estructura matemática:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_q X_{iq} + \varepsilon_i$$

Donde:

Y = variable dependiente, valor observado

X = variable independiente, variable predictora

β = magnitud de las observaciones

ε = error por factores no controlados

4.1.1. Población, muestra y muestre

Las aguas del estero El Macho en Machala serán el objeto de nuestro análisis. Para ello, se recolectarán muestras en cuatro puntos específicos del estero el Macho.

4.2. Formulación de la hipótesis

4.2.1. Hipótesis general

El modelo estadístico determina significativamente la calidad del agua del estero El Macho de la ciudad de Machala, provincia de El Oro.

4.2.2. Hipótesis específicas

- a) El modelo estadístico multivariante de regresión lineal múltiple describe en forma significativa los parámetros físico-químicos de calidad del agua del estero El Macho de la ciudad de Machala, provincia de El Oro.
- b) El modelo estadístico multivariante de regresión lineal múltiple describe en forma significativa los parámetros biológicos de calidad del agua del estero El Macho de la ciudad de Machala, provincia de El Oro.
- c) El análisis de componentes principales explica la correlación de los parámetros de calidad del agua del estero El Macho de la ciudad de Machala, provincia de El Oro.

4.3. Definición y operacionalización de Variables

- **Variable independiente Y:**
Modelo Estadístico
- **Variable dependiente:**
Calidad de agua del estero el Macho de la ciudad de Machala

4.3.1. Operacionalización de Variables

Tabla 1: Operacionalización de las variables

Variable	Dimensiones	Indicadores	Descriptorios	Instrumento	Método
Variable independiente: Modelo estadístico	Parámetros químicos Parámetro biológico	- Modelo estadístico multivariante de regresión lineal múltiple - Análisis de componentes principales	<p>“Acuerdo Ministerial Número 097-A, Anexo 1, Norma de calidad ambiental y de descarga de efluentes al recurso agua.</p> <p>✓ Tabla 2: Criterios de calidad admisibles para la preservación de la vida acuática y silvestre en aguas dulces, marinas y de estuarios.”^a</p> <p>Relación: $\frac{DBO}{DQO} = CALIDAD$</p>	<p>Softwares estadísticos: Programa estadístico SPSS</p> <p>Análisis de Laboratorio.</p>	<p>Regresión lineal múltiple</p> <p>Análisis de componentes principales</p>
Variable dependiente: Parámetros de calidad del agua	Parámetros físicos	- Sólidos Disueltos Totales - Aceites y grasa - Temperatura - Turbidez - PH - Oxígeno disuelto			
	Parámetros químicos	- Potencial Hidrógeno - Oxígeno Disuelto Fosfatos - Nitritos - Nitratos - Sulfatos - Hierro			
	Parámetros Microbiológicos y biológicos	- Coliformes Fecales - Coliformes Totales - Demanda Biológica de oxígeno			

Nota. ^aCODA, 2017, acuerdo ministerial 097-A.

4.4. Métodos, técnicas e instrumentos de recolección de datos

Para la recolección de datos se procederá de la siguiente manera:

- Observación y selección de los puntos de toma de muestras de agua.
- Para la recolección de muestras de agua del estero se hará uso de equipos portátiles de análisis de laboratorio.
- El procedimiento de toma de muestras se lo realizará en cuatro puntos y en en marea baja y durante el día, en cada punto se tomarán 15 muestras.
- Las muestras serán analizadas en un laboratorio calificado con la colaboración de Aguas Machala EP.

4.5. Plan de procesamiento y análisis de datos

Siguiendo la metodología propuesta por García Merino y otros (2023), el procesamiento y análisis de datos se realizará de la siguiente manera:

- 1) Para llevar a cabo un análisis estadístico de los parámetros que afectan la calidad del agua del estero El Macho, se seguirá el siguiente procedimiento:
 - Se seleccionarán 4 puntos de recolección de muestras.
 - Obtención de parámetros de tipo físico-químicos, biológicos y microbiológico de las muestras del agua (Aguas Machala EP entregara la data año 2018 - 2019).
 - Organización y clasificación de los datos recabados.
 - Se establecerán como parámetros de comparación los establecidos en la norma INEN 2169. (ver anexo 3).
 - Se establecerá mediante comparación la variación de los parámetros estudiados mediante comparación de las datas
- 2) Análisis de componentes principales y modelo estadístico multivariante de regresión lineal usando la data 2019 por ser la más actual y completa.
 - Se usará la relación $\frac{DBO}{DQO}$ para determinar el nivel de contaminación (biológico o químico) que tienen las aguas del estero estudiado

- Se realizará un Análisis de Componentes Principales (ACP) para determinar los agrupamientos (componentes) que expliquen el comportamiento de la variable dependiente.
- Con base ACP se generará el modelo estadístico de regresión lineal múltiple, se generaran varias inter acciones hasta encontrar el modelo más eficiente.
- Para el Análisis de Componentes Principales y el modelo estadístico multivariante de regresión lineal múltiple se usará el software IBM SPSS por su capacidad para manejar datos cuantitativos y cualitativos y por el reconocimiento que tiene en el ámbito académico.

3) Validación del modelo estadístico.

- Se generará un segundo modelo con el 80% de los datos de la data y con este segundo modelo se pronostica el 20% excluido de la data.
- Se probarán los supuestos que debe cumplir un modelo de regresión lineal múltiple esto es linealidad, independencia, homocedasticidad, normalidad y no colinealidad; los datos obtenidos con el software IBM SPSS se los comparara con los obtenidos con el software R.

4) Análisis y discusión de resultados, establecimiento de conclusiones y recomendaciones.

CAPÍTULO V

RESULTADOS

5.1. Parámetros iniciales

Como un preámbulo para desarrollar el modelo, se realizó una selección de datos relevantes del monitoreo efectuado por la Empresa Pública Aguas Machala en 2018. Esta selección se centró en los puntos de muestreo que coincidían con las zonas de interés de nuestro estudio, es decir, los sectores: A-1 Drenaje Macho El Cambio, A-2 La Primavera 1, A-3 La Primavera 2, A-4 Estero Salinas.

Tabla 2: Resultados iniciales de monitoreo de agua.

		A-1	A-2	A-3	A-4
Parámetro biológico	a "Demanda Bioquímica Oxígeno" (mg/l)	13,59	4,75	10,71	37,07
Parámetros físicos	a "Sólidos Disueltos Totales" (mg/l)	1280,00	638,00	1046,00	22784,00
	a "Temperatura"	25,00	24,80	25,00	21,10
	a "Turbidez" NTU	9,50	20,70	17,80	15,30
	a "Aceites Grasas IR" (mg/l)	0,2	0,2	0,2	0,2
Parámetros químicos	a "Potencial Hidrógeno"	7,87	7,44	7,17	7,38
	a "Demanda Química Oxígeno"	30,80	10,00	22,00	75,30
	a "Oxígeno Disuelto" (mg/l)	9,75	6,48	1,49	3,33
	a "Fosfatos" (mg/l)	5,54	1,23	1,74	1,56
	a "Nitritos" (mg/l)	0,01	0,032	0,01	0,012
	a "Nitratos" (mg/l)	1	1	1	1
	a "Sulfatos" (mg/l)	151,20	102,50	103,50	1582,00
	a "Arsénico"	0,008	0,005	0,006	0,003
	a "Cadmio"	0,0004	0,0004	0,0008	0,0004
	a "Cobre"	0,002	0,0024	0,002	0,001
	a "Hierro" (mg/l)	0,54	0,8	0,9	0,56
Parámetros microbiológicos	a "Coliformes Fecales" NPM/100 ml	13,00	54,00	68,00	24,00
	a "Coliformes Totales" NPM/100 ml	27,00	920,00	330,00	40,00

Nota. ^a(Zhindon, 2024)

A continuación, se analiza los diferentes parámetros al 2018:

Parámetro biológico

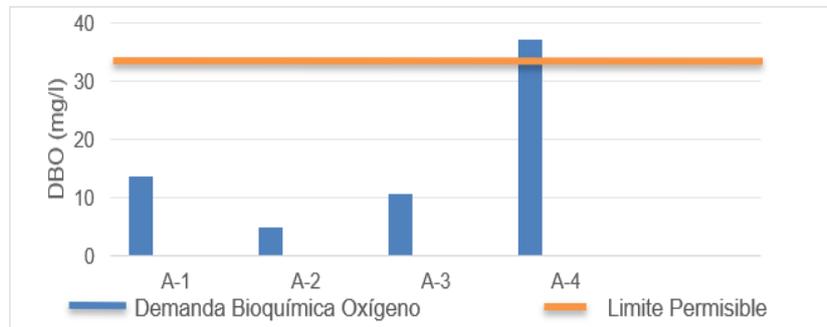


Figura 1: Demanda Bioquímica Oxígeno

Fuente: Aguas Machala EP (2019)

Dentro del parámetro biológico de demanda biológica de oxígeno, que mide la cantidad de oxígeno requerido para oxidar la materia orgánica en una muestra de agua, se evidencia que en el sector A-4, en el año 2018, ha sobrepasado los límites permisibles. Mientras que los otros puntos de análisis se observan que aún permanecen dentro de los niveles aceptables, siendo el sector A-1 el más susceptible de llegar a un punto de contaminación alto.

Parámetros químicos

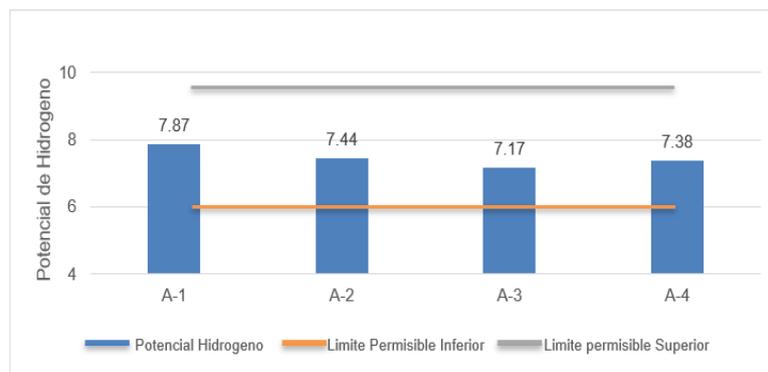


Figura 2: Potencial de hidrógeno pH (situación actual)

Fuente: Aguas Machala EP (2019).

En relación con el potencial de hidrógeno, los límites permitidos se encuentran entre 6,5 y 9,5 según la norma de calidad ambiental del ministerio de ambiente, por tanto, se considera que las aguas del estero El Macho aún se mantiene dentro de los rangos permitidos.

En la figura 3, se muestra los resultados de la demanda química de oxígeno, evidenciándose que en el sector A-4 se ha superado los límites permitidos, por tanto, existe un alto grado de contaminación en esa área

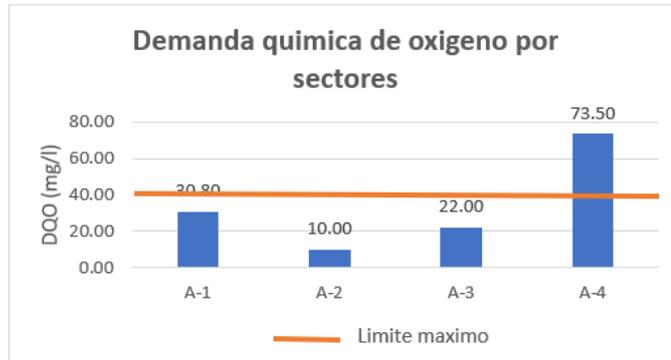


Figura 3: Demanda química de oxígeno
Fuente: Aguas Machala EP (2019)

En la figura 4, sobre el oxígeno disuelto, se muestra que en las zonas A-2, A-3 y A-4, los niveles de este parámetro están bajo el límite permitido, lo que indica que no hay suficiente oxígeno para soportar una diversidad de vida acuática.

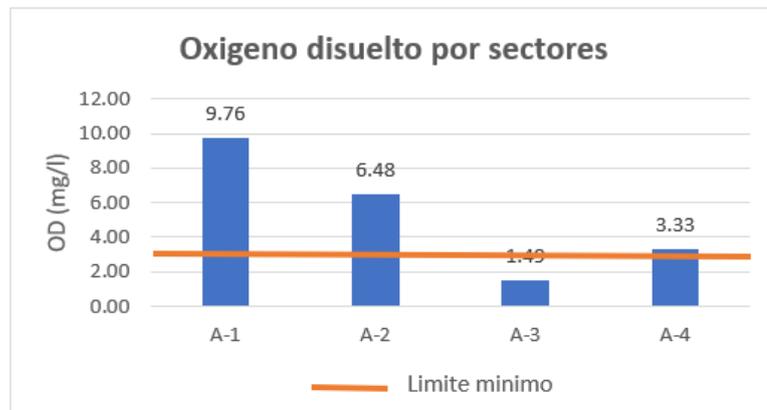


Figura 4: Oxígeno disuelto
Fuente: Aguas Machala EP (2019).

Los nitritos y nitratos se encuentran dentro de los límites permisibles para aguas marinas de estuario (ver figura 5 y 6).

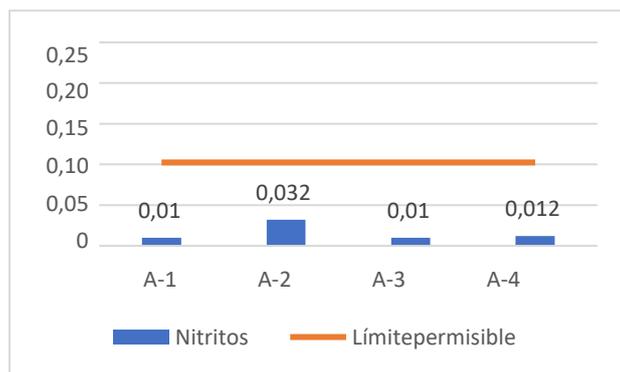


Figura 5: Nitritos

Fuente: Aguas Machala EP (2019).

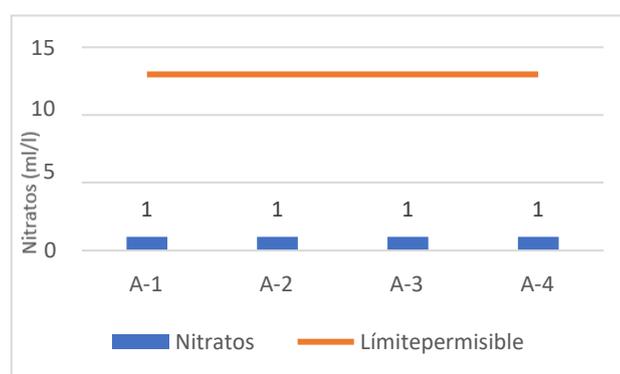


Figura 6: Nitratos (situación actual)

Fuente: Aguas Machala EP (2019).

Las concentraciones de arsénico y cadmio se encuentran dentro del rango permitido para aguas marinas de estuario (ver figura 7 y 8), siendo importante monitorear el elemento cadmio en el sector A-3 debido a que empieza a acercarse al límite permisible.

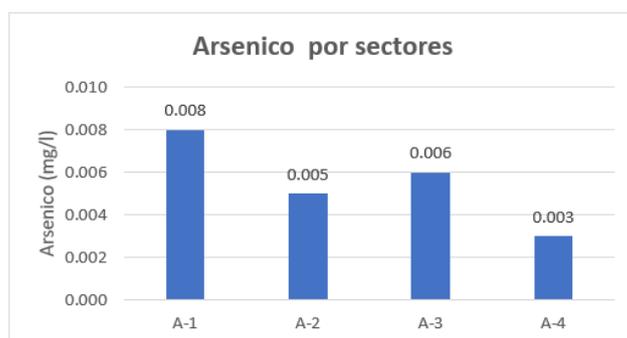


Figura 7 : Arsénico

Fuente: Aguas Machala EP (2019).

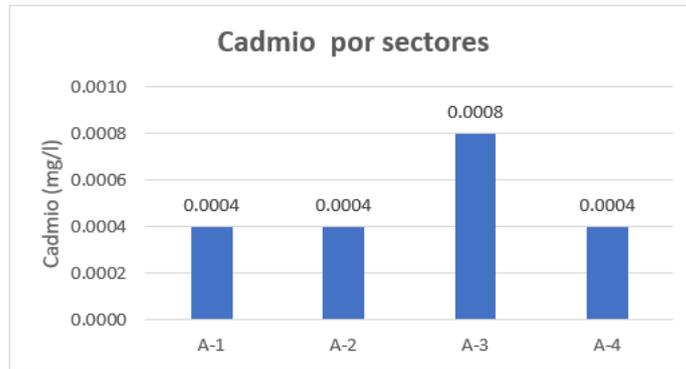


Figura 8: Cadmio
Fuente: Aguas Machala EP (2019)

En todas las zonas analizadas, el cobre se mantiene dentro del rango manejable para aguas marinas de estuario (ver figura 9).

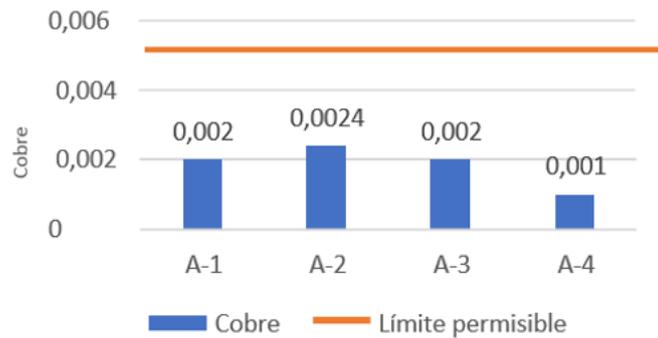


Figura 9: Cobre
Fuente: Aguas Machala EP (2019)

Las concentraciones de hierro están por encima de los límites permisibles, ocasionando cambio en las características físicas de color y sabor metálico indeseable (ver figura 10).

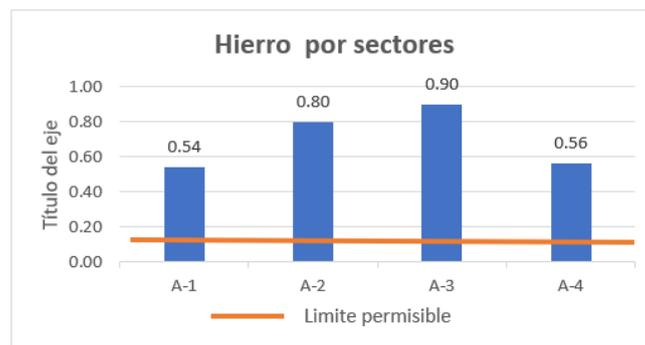


Figura 10: Hierro
Fuente: Aguas Machala EP (2019)

Parámetros microbiológicos

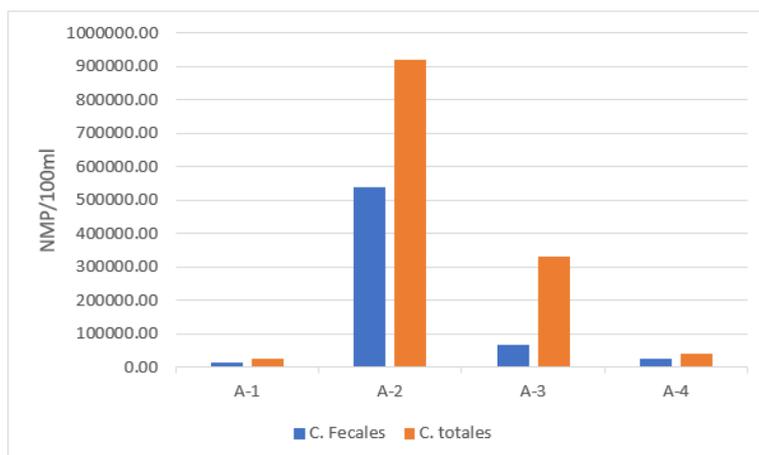


Figura 11: Coliformes fecales y coliformes totales (situación actual)
Fuente: Aguas Machala EP (2019)

Respecto a estos parámetros, se evidencia que la zona A-1 presenta niveles altos de coliformes fecales y totales, seguido por la zona A-3. Su presencia indica que existe una vía de contaminación de bacterias en las aguas superficiales, ya sea debido al sistema séptico, desechos animales, entre otros.

Análisis de componentes principales en la prueba preliminar

Tabla 3: *Matriz de correlaciones iniciales.*

	Demanda Bioquímica Oxígeno	Sólidos Disueltos Totales	Demanda Química Oxígeno	Temperatura	Turbidez	Potencial Hidrógeno	Oxígeno Disuelto	Fosfatos	Nitritos	Sulfatos	Arsénico	Cadmio	Cobre	Hierro	Coliformes Fecales	Coliformes Totales
Demanda Bioquímica Oxígeno (mg/l)	1,000	,972	,999	-,952	-,302	-,074	-,294	-,103	-,479	,971	-,629	-,274	-,994	-,626	-,152	-,232
Sólidos Disueltos Totales(mg/l)	,972	1,000	,961	-,997	-,096	-,181	-,350	-,296	-,271	1,000	-,787	-,330	-,956	-,536	-,342	-,441
Demanda Química Oxígeno	,999	,961	1,000	-,940	-,349	-,027	-,252	-,052	-,500	,961	-,592	-,293	-,991	-,661	-,102	-,185
Temperatura	-,952	-,997	-,940	1,000	,039	,199	,350	,342	,201	-,998	,822	,356	,932	,513	,387	,494
Turbidez NTU	-,302	-,096	-,349	,039	1,000	-,768	-,557	-,914	,695	-,104	-,537	,277	,290	,774	-,892	-,847
Potencial Hidrógeno	-,074	-,181	-,027	,199	-,768	1,000	,957	,878	-,076	-,165	,594	-,669	,142	-,729	,878	,736
Oxígeno Disuelto (mg/l)	-,294	-,350	-,252	,350	-,557	,957	1,000	,760	,196	-,332	,598	-,692	,373	-,556	,773	,624
Fosfatos (mg/l)	-,103	-,296	-,052	,342	-,914	,878	,760	1,000	-,465	-,286	,807	-,256	,122	-,585	,999	,965
Nitritos (mg/l)	-,479	-,271	-,500	,201	,695	-,076	,196	-,465	1,000	-,264	-,239	-,374	,542	,336	-,434	-,538
Sulfatos (mg/l)	,971	1,000	,961	-,998	-,104	-,165	-,332	-,286	-,264	1,000	-,783	-,347	-,953	-,549	-,332	-,435
Arsénico (mg/l)	-,629	-,787	-,592	,822	-,537	,594	,598	,807	-,239	-,783	1,000	,160	,617	,000	,832	,901
Cadmio(mg/l)	-,274	-,330	-,293	,356	,277	-,669	-,692	-,256	-,374	-,347	,160	1,000	,167	,748	-,245	,001
Cobre(mg/l)	-,994	-,956	-,991	,932	,290	,142	,373	,122	,542	-,953	,617	,167	1,000	,564	,172	,225
Hierro(mg/l)	-,626	-,536	-,661	,513	,774	-,729	-,556	-,585	,336	-,549	,000	,748	,564	1,000	-,552	-,377
Coliformes Fecales NPM/100 ml	-,152	-,342	-,102	,387	-,892	,878	,773	,999	-,434	-,332	,832	-,245	,172	-,552	1,000	,969
Coliformes Totales NPM/100 ml	-,232	-,441	-,185	,494	-,847	,736	,624	,965	-,538	-,435	,901	,001	,225	-,377	,969	1,000

a. Determinante = ,000

b. Esta matriz no es cierta positiva.

Se aplicó el método de “Stepwise” o “Por pasos”, para que el software IBM SPSS evalúe y determine un modelo que presenten mayores correlaciones:

La variable SDT explica el 94.4% del comportamiento de la DBO, linealmente se relacionan mediante la ecuación $DBO = 8.394 - 0.001SDT$. Por otra parte, la DQO explica el 99.7% del comportamiento de la DBO, linealmente se relacionan mediante la siguiente ecuación $DBO = -0.631 + 0.497DQO$.

Aunque las correlaciones resultaron en una matriz no cierta positiva (ver tabla 3), niveles altos de correlación en las variables Sólidos Disueltos Totales (,972) y Demanda Química Oxígeno (0,999) permitirían de manera significativa la determinación de la calidad del agua en el Estero El Macho como variables predictoras.

5.2. Parámetros estudiados (data 2019)

Tabla 4: Valores de los parámetros fisicoquímicos y biológicos en el sector 1

INDICADOR	SECTOR 1														
DBO (mg/l)	14,56	11,44	10,32	10,44	10,01	12,01	12,05	13,25	9,28	10,20	13,01	11,80	11,10	9,25	9,75
SDT (mg/l)	1052	1188	1185	1141	1325	1254	1251	1325	1270	1075	1327	1325	1298	1280	1195
Temperatura	28,40	28,80	28,20	28,00	26,20	27,60	27,70	28,00	28,60	28,50	29,00	28,30	27,70	24,80	26,80
Turbidez NTU	41,52	45,07	50,05	46,48	45,08	50,78	50,47	54,05	50,45	42,98	49,78	44,98	51,75	49,75	50,75
Aceites grasas (mg/l)	1,20	0,75	1,42	1,50	1,40	1,60	1,42	1,27	1,25	1,20	1,20	1,40	1,40	1,40	1,40
Coliformes fecales NPM/100 ml	15000	20000	21000	27500	17000	18000	18000	18000	17000	17000	17000	17000	17000	17000	17000
Coliformes totales NPM/100 ml	160000	200000	210000	270000	170000	170000	1600000	170000	180000	180000	170000	180000	180000	170000	170000
PH	7,60	7,40	7,70	7,29	7,30	7,40	7,40	7,40	7,50	7,40	7,40	7,50	7,40	7,40	7,40
DQO	27,50	21,70	18,80	14,23	12,80	17,56	15,62	19,27	13,80	13,07	18,94	15,27	16,31	13,21	14,44
OD (mg/l)	6,50	4,25	4,15	4,25	4,05	3,95	4,52	4,15	3,85	4,20	4,20	3,75	4,25	4,15	4,75
Fosfatos (mg/l)	0,95	1,11	1,34	1,21	1,05	1,01	1,15	1,05	1,48	1,15	1,01	1,60	1,65	1,50	1,75
Nitritos (mg/l)	0,73	0,85	0,63	0,71	0,82	0,65	0,85	0,75	0,81	0,58	0,81	0,72	0,68	0,71	0,61
Nitratos (mg/l)	1,55	1,61	1,65	1,78	1,75	0,68	1,65	1,71	1,55	1,68	1,65	1,72	1,75	1,72	1,75
Sulfatos (mg/l)	182,23	127,25	133,86	154,78	145,75	172,64	168,54	165,64	168,24	161,73	105,38	178,08	172,73	164,02	182,04
Hierro (mg/l)	0,22	0,28	0,20	0,20	0,25	0,22	0,28	0,20	0,20	0,25	0,20	0,22	0,25	0,20	0,20

Fuente: Aguas Machala EP (2019)

Tabla 5: Valores de los parámetros fisicoquímicos y biológicos en el sector 2

INDICADOR	SECTOR 2														
DBO (mg/l)	54,76	55,44	50,21	40,44	44,06	48,04	49,01	53,25	48,75	50,24	53,78	47,65	48,09	45,07	48,91
SDT (mg/l)	17638	18469	17354	15425	16378	15987	15489	16015	16364	15948	16870	14050	15643	15214	14054
Temperatura	28,40	28,80	28,20	28,00	26,20	27,60	27,70	28,00	28,60	28,50	29,00	28,30	27,70	24,80	26,80
Turbidez NTU	51,00	54,00	50,00	56,00	50,00	50,00	50,00	44,00	50,00	52,00	52,00	54,00	55,20	49,00	50,00
Aceites grasas (mg/l)	3,70	3,25	3,85	2,75	2,89	3,15	2,75	2,85	2,95	3,25	3,15	2,95	2,05	1,90	1,70
Coliformes fecales NPM/100 ml	340000	300000	290000	300000	300000	290000	290000	300000	300000	300000	310000	300000	310000	300000	300000
Coliformes totales NPM/100 ml	340000 0	330000 0	320000 0	340000 0	410000 0	360000 0	410000 0	350000 0	310000 0	330000 0	390000 0	310000 0	315000 0	305000 0	300000 0
PH	7,90	7,80	7,60	7,70	7,60	7,60	7,60	7,70	7,60	7,60	7,60	7,60	7,60	7,60	7,60
DQO	72,24	69,54	68,21	54,65	66,24	67,24	58,75	70,05	66,48	66,51	72,48	65,46	67,28	59,48	62,47
OD (mg/l)	2,18	2,23	2,92	2,50	2,07	2,05	2,75	2,15	2,35	3,01	2,85	3,05	3,05	2,25	2,45
Fosfatos (mg/l)	1,34	1,45	1,13	1,25	1,15	0,98	1,65	1,55	1,75	1,35	1,30	1,75	1,45	1,52	1,63
Nitritos (mg/l)	0,57	0,71	0,62	0,73	0,58	0,57	0,56	0,61	0,67	0,69	0,64	0,70	0,65	0,67	0,65
Nitratos (mg/l)	1,27	1,43	1,54	1,65	1,43	1,52	1,41	1,67	1,57	1,34	1,46	1,62	1,44	1,39	1,51
Sulfatos (mg/l)	525,72	479,24	541,05	530,07	495,78	457,06	478,64	466,45	480,05	491,08	501,24	475,62	501,24	471,02	458,42
Hierro (mg/l)	0,23	0,28	0,30	0,30	0,34	0,42	0,37	0,42	0,40	0,45	0,40	0,40	0,41	0,45	0,43

Fuente: Aguas Machala EP (2019).

Tabla 6: Valores de los parámetros fisicoquímicos y biológicos en el sector 3

INDICADOR	SECTOR 3														
DBO (mg/l)	97,45	95,25	85,92	95,85	85,75	98,25	102,35	105,65	97,85	107,12	110,25	97,85	95,25	95,15	99,50
SDT (mg/l)	24350	23786	23548	22780	22550	23950	23780	24100	24350	24100	24050	23680	23570	24690	23900
Temperatura	28,40	28,80	28,20	28,00	26,20	27,60	27,70	28,00	28,60	28,50	29,00	28,30	27,70	24,80	26,80
Turbidez NTU	92,50	98,00	100,00	108,00	103,00	98,00	105,00	102,00	104,00	103,00	106,00	107,00	107,70	104,50	103,50
Aceites grasas (mg/l)	4,50	4,75	5,15	3,75	3,75	4,10	4,05	3,95	4,10	4,00	4,25	4,15	4,25	4,25	4,60
Coliformes fecales NPM/100 ml	250000	400000	300000	320000	420000	320000	400000	350000	390000	310000	350000	370000	450000	410000	460000
Coliformes totales NPM/100 ml	241970 0	332520 0	322630 0	337970 0	413620 0	368010 0	405510 0	346190 0	308500 0	325520 0	386640 0	303910 0	413628 2	416070 0	462840 0
PH	7,80	7,50	7,60	7,50	7,50	7,50	7,50	7,50	7,60	7,50	7,60	7,60	7,60	7,50	7,60
DQO	135,20	135,35	120,35	115,76	118,54	138,15	137,25	142,25	128,95	135,75	154,26	135,50	140,15	132,75	141,75
OD (mg/l)	1,78	1,53	1,92	1,51	1,24	1,35	1,85	1,65	1,87	2,01	1,95	2,01	1,90	1,75	1,85
Fosfatos (mg/l)	1,70	1,01	1,25	1,07	0,95	0,85	1,15	1,05	1,55	1,25	1,01	1,60	1,65	1,50	1,75
Nitritos (mg/l)	0,52	0,65	0,58	0,65	0,62	0,65	0,65	0,55	0,52	0,57	0,62	0,60	0,58	0,65	0,62
Nitratos (mg/l)	0,55	0,60	0,65	0,70	0,75	0,68	0,65	0,70	0,55	0,70	0,65	0,72	0,75	0,72	0,75
Sulfatos (mg/l)	482,53	329,93	333,69	352,21	365,91	327,46	386,69	356,75	388,42	351,37	305,83	387,80	327,54	360,35	328,20
Hierro (mg/l)	0,32	0,28	0,30	0,30	0,35	0,32	0,28	0,30	0,30	0,35	0,30	0,32	0,35	0,30	0,30

Fuente: Aguas Machala EP (2019)

Tabla 7: Valores de los parámetros fisicoquímicos y biológicos en el sector 4

INDICADOR	SECTOR 4														
DBO (mg/l)	23,30	18,50	15,70	20,50	27,30	17,40	22,40	28,10	30,20	27,40	24,50	24,75	35,20	30,14	24,25
SDT (mg/l)	19000	19500	13676	18668	14554	17350	18668	16540	17000	17800	14540	18500	17200	17750	16800
Temperatura	28,50	29,50	28,10	27,40	26,10	27,50	27,80	28,20	28,40	28,75	29,20	28,20	27,50	24,80	26,80
Turbidez NTU	21,20	23,30	24,50	26,45	28,45	26,75	24,56	23,15	20,14	21,14	22,67	23,45	23,01	20,05	21,05
Aceites grasas (mg/l)	4,25	5,25	4,75	2,61	3,14	2,70	3,75	3,25	4,12	3,85	5,10	3,75	4,85	5,60	5,05
Coliformes fecales NPM/100 ml	68000	61000	70000	65000	58000	75000	58000	62000	60500	57000	68000	64500	62400	60200	60000
Coliformes totales NPM/100 ml	108800	99125	112000	104000	92400	135000	110000	125000	115000	125000	124000	130000	135000	120000	125000
PH	8,30	8,30	8,00	7,29	7,38	7,40	7,40	7,40	8,00	7,35	8,10	7,35	7,75	8,15	8,02
DQO	70,20	80,24	50,50	75,70	100,10	65,70	100,24	80,20	100,70	85,23	82,75	72,50	95,40	80,75	75,23
OD (mg/l)	8,24	8,15	8,12	12,60	6,14	9,50	10,00	10,25	12,00	9,75	12,50	8,25	10,15	9,25	9,50
Fosfatos (mg/l)	0,43	1,05	1,33	0,51	0,79	1,36	0,51	1,25	0,75	2,00	1,50	1,50	0,75	0,85	0,91
Nitritos (mg/l)	0,25	0,15	0,12	0,12	0,20	0,22	0,12	0,27	0,22	0,21	0,18	0,13	0,14	0,28	0,20
Nitratos (mg/l)	0,45	0,47	0,40	0,48	0,42	0,51	0,60	0,62	0,48	0,62	0,65	0,65	0,62	0,62	0,67
Sulfatos (mg/l)	320,52	270,54	300,28	220,51	315,78	340,25	299,60	300,57	350,24	270,45	254,62	302,25	285,47	284,02	270,54
Hierro (mg/l)	0,20	0,20	0,35	0,45	0,30	0,30	0,25	0,30	0,28	0,25	0,30	0,28	0,30	0,30	0,30

Fuente: Aguas Machala EP (2019)

La base de datos fue extraída de los registros de un monitoreo realizado por la Empresa Pública Aguas Machala en el año 2019, en cuatro sectores (ver anexo 2).

A continuación de analiza el comportamiento de cada uno de los parámetros:

a) Parámetros biológicos

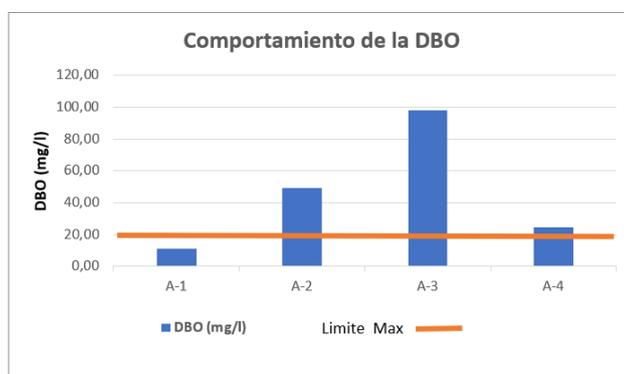


Figura 12: Demanda Bioquímica Oxígeno

Fuente: Aguas Machala EP (2019).

Dentro del parámetro de demanda biológica de oxígeno, el cual mide la cantidad de oxígeno requerido para oxidar la materia orgánica en una muestra de agua, se evidencia que en el sector A-2, A-3 y A-4, en el año 2019, ha sobrepasado los límites permisibles; el sector A-1 es susceptible de llegar a un punto máximo permitido.

b) Parámetros físicos

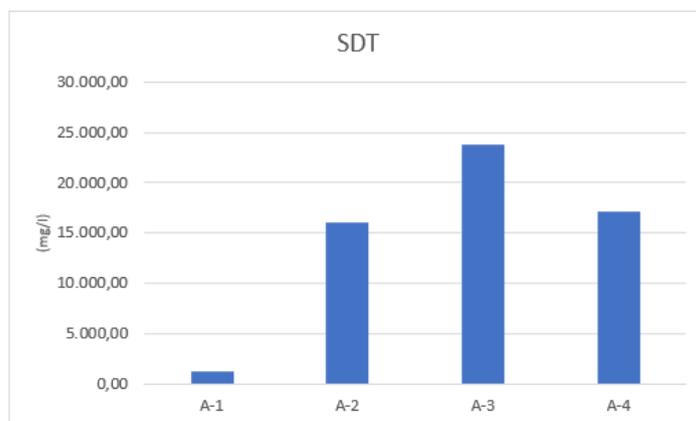


Figura 13: Sólidos disueltos totales

Fuente: Aguas Machala EP (2019)

Aunque la cantidad SDT en sistemas estuarinos es alta, el cambio que experimenta el agua desde el sector A-1 hasta el sector A-2, es significativo, delatando procesos de contaminación.

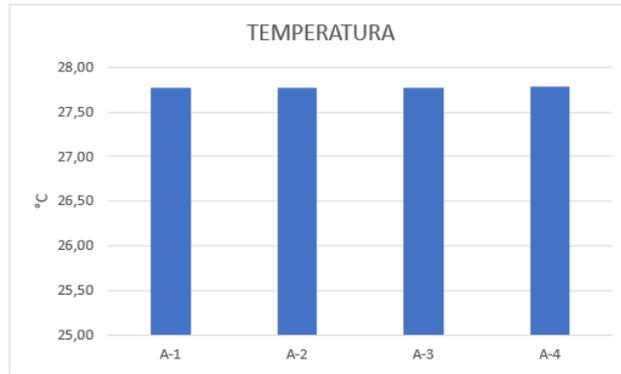


Figura 14: Temperatura del agua
Fuente: Aguas Machala EP (2019)

La temperatura del agua esta ligeramente elevada, pero se considera normal por la temporada en que se tomaron las muestras, esto es invierno.

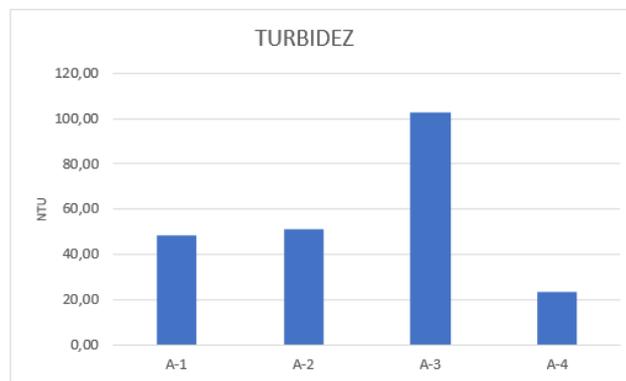


Figura 15: Turbidez
Fuente: Aguas Machala EP (2019).

El agua presenta una turbiedad notoria, se atribuye a la contaminación producto de las aguas residuales y a los procesos biológicos que se dan.

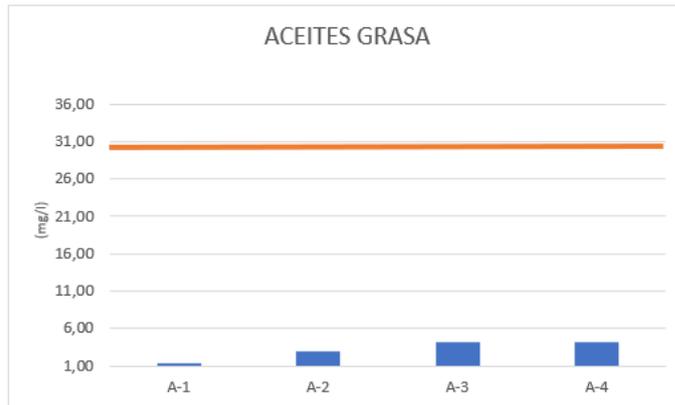


Figura 16: Aceites Grasa
Fuente: Aguas Machala EP (2019).

Aunque la presencia de grasas está muy por debajo del límite, es de tomar en cuenta por su incidencia acumulativa a lo largo del cuerpo de agua.

c) Parámetros microbiológicos

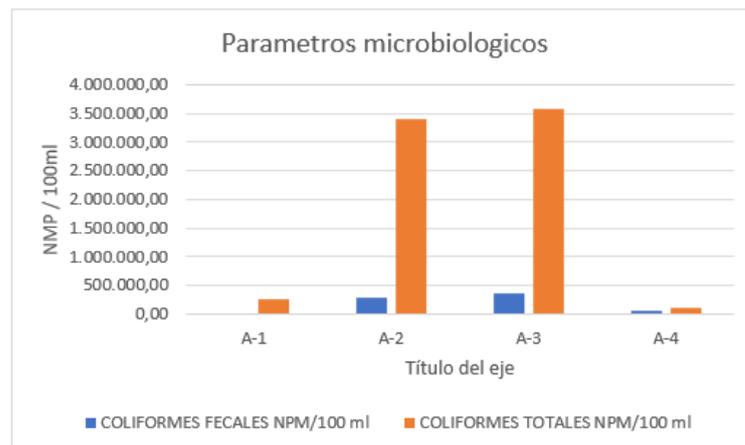


Figura 17: Coliformes fecales y coliformes totales
Fuente: Aguas Machala EP (2019).

Respecto a estos parámetros, se evidencia que la zona A-2 y A-3 presentan niveles altos de coliformes fecales y totales. Su presencia indica que existe una vía de contaminación de bacterias en las aguas superficiales, ya sea debido al sistema séptico, desechos animales, entre otros.

d) Parámetros químicos

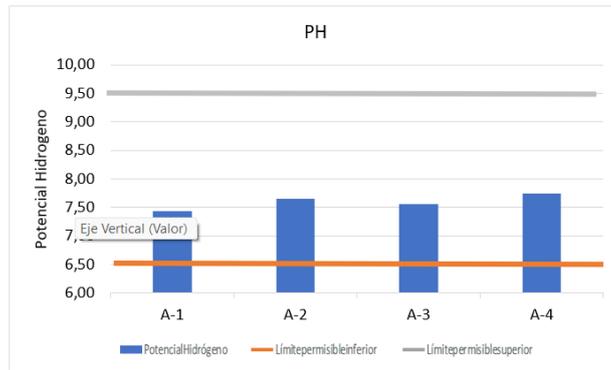


Figura 18: Potencial de hidrógeno pH

Fuente: Aguas Machala EP (2019).

En relación con el potencial de hidrógeno, los límites permitidos se encuentran entre 6,5 y 9,5 según la norma de calidad ambiental del ministerio de ambiente, por tanto, se considera que las aguas del estero El Macho aún se mantiene dentro de los rangos permitido.

La presencia de Nitritos y Nitratos dan muestras de procesos de Nitrificación, en el cuerpo de agua tal como se indica en las Figura 19 y 20.

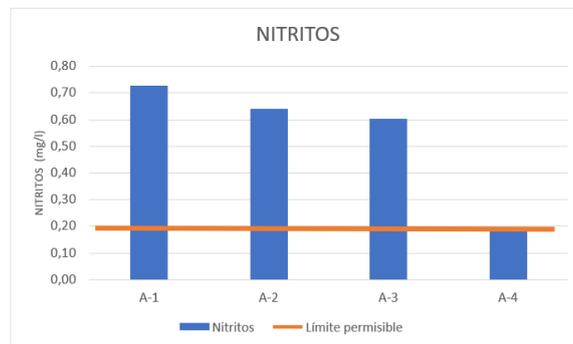


Figura 19: Nitritos

Fuente: Aguas Machala EP (2019).

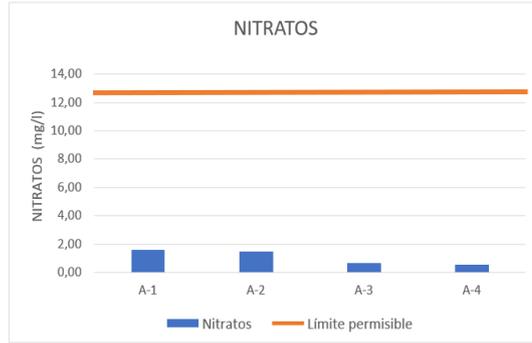


Figura 20: Nitratos
Fuente: Aguas Machala EP (2019).

Las concentraciones de hierro están por encima de los límites permisibles en los sectores A-2 y A-3, ocasionando cambio en las características físicas del agua (ver figura 21).

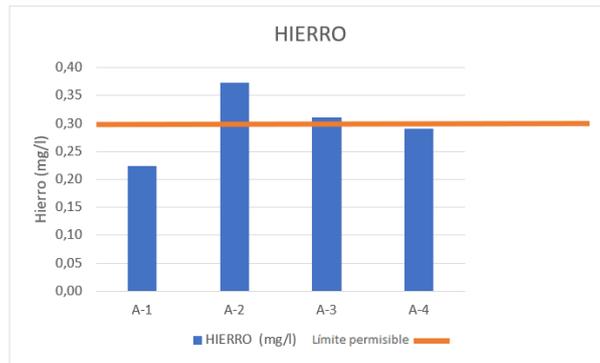


Figura 21: Hierro
Fuente: Aguas Machala EP (2019).

La cantidad de sulfatos contenidas en las aguas está bajo el límite permitido, sin embargo, en el sector A2, su presencia es significativa y pone de manifiestos procesos de nitrificación.

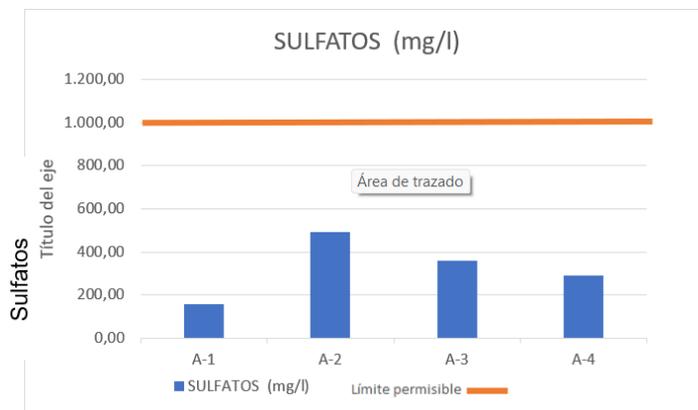


Figura 22: Sulfatos
Fuente: Aguas Machala EP (2019)

5.3. Análisis de Componentes Principales (ACP)

El ACP es una técnica de reducción de dimensiones, consiste en agrupar las variables en grupos que comparten características en común, en una primera instancia se testearon todas las variables independientes, y dado que existieron variables que tenían saturaciones compartidas, se optó por excluirlas del modelo y volver a testearlas, quedando de la siguiente manera:

Tabla 8: Matrix de correlacion^a

	DBO_DQO	SDT	TEM	TUR	A_G	pH	OD	FOSF	NITRI	NITRA	SUL	Fe	CF	CT
Correlation DBO_DQO	1,000	-,069	-,037	,654	-,327	-,325	-,908	,397	,873	,559	,271	,169	,541	,640
SDT	-,069	1,000	,022	,457	,846	,303	-,119	-,039	-,375	-,708	,658	,469	,742	,636
TEM	-,037	,022	1,000	-,009	,067	,142	,057	,038	-,020	-,027	,003	-,149	-,039	-,049
TUR	,654	,457	-,009	1,000	,212	-,229	-,767	,207	,498	-,136	,216	,126	,752	,726
A_G	-,327	,846	,067	,212	1,000	,576	,166	-,141	-,579	-,787	,409	,201	,436	,316
pH	-,325	,303	,142	-,229	,576	1,000	,260	-,143	-,395	-,309	,281	,014	,054	,004
OD	-,908	-,119	,057	-,767	,166	,260	1,000	-,356	-,794	-,380	-,361	-,181	-,679	-,752
FOSF	,397	-,039	,038	,207	-,141	-,143	-,356	1,000	,305	,318	,220	,154	,260	,274
NITRI	,873	-,375	-,020	,498	-,579	-,395	-,794	,305	1,000	,718	-,016	-,043	,269	,383
NITRA	,559	-,708	-,027	-,136	-,787	-,309	-,380	,318	,718	1,000	-,089	-,062	-,165	-,008
SUL	,271	,658	,003	,216	,409	,281	-,361	,220	-,016	-,089	1,000	,670	,750	,756
Fe	,169	,469	-,149	,126	,201	,014	-,181	,154	-,043	-,062	,670	1,000	,556	,561
CF	,541	,742	-,039	,752	,436	,054	-,679	,260	,269	-,165	,750	,556	1,000	,973
CT	,640	,636	-,049	,726	,316	,004	-,752	,274	,383	-,008	,756	,561	,973	1,000

a. Determinant = 1,425E-9

La matriz de correlación, tabla 8, muestra las relaciones que existen entre las variables, su determinante al ser muy pequeño nos indica que existen variables con Inter correlaciones muy altas, y entonces es factible continuar con el análisis de componentes principales. Sin embargo, el determinante no debe ser igual a cero, pues en este caso los datos no serían válidos. (Montoya, 2007)

Para el caso de este estudio se obtuvo un determinante igual a 1,425E-9. Esto nos indica que dicho determinante es muy próximo a cero, por lo que es factible continuar con el análisis factorial.

Que el determinante sea cercano a cero nos indica que las variables están fuertemente correlacionadas o que tienen un alto grado de asociación.

En la prueba de Kaiser-Meyer-Olkin tabla 9, el coeficiente KMO es mayor a 0.7 lo que nos indica que los datos son adecuados para realizar un Análisis de Componentes Principales. (Montoya, 2007).

Tabla 9: Prueba KMO y Bartlett

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,797
Bartlett's Test of Sphericity	Approx. Chi-Square	1089.745
	df	91
	Sig.	,000

Por otra parte, la prueba de esfericidad de Bartlett establece como hipótesis que:

Ho: Las variables son no correlacionadas

H1: Las variables son correlacionadas

Criterio de decisión

Si p-valor < 0.05 se rechaza Ho y se acepta H1

Si p valor \geq 0.05 se acepta H0 y se rechaza H1 (Montoya, 2007)

En nuestro caso p-valor es menor que 0.05 por lo que se rechaza Ho y se acepta que las variables no son no correlacionadas es decir se acepta H1, y que el análisis de componentes Principales para nuestro caso es adecuado realizarlo.

Revisando la tabla 10 de Comunalidades y la tabla 11 de Varianza total explicadas nos damos cuenta que se han retenido 4 componentes y que con ellos se explica el 84.647% de las variables de manera global, de manera particular con estos cuatro componentes se puede explicar el 92.6% de la DBO_DQO, el 96.5 de la SDT, el 76.4% de la TEMP, el 95.4% de la TURD y así sucesivamente.

Tabla 10: Comunalidades

	Initial	Extraction
DBO_DQO	1,000	,926
SDT	1,000	,965
TEM	1,000	,764
TUR	1,000	,954
A_G	1,000	,905
pH	1,000	,581
OD	1,000	,927
FOSF	1,000	,400
NITRI	1,000	,918
NITRA	1,000	,913
SUL	1,000	,885
Fe	1,000	,778
CF	1,000	,973
CT	1,000	,962

Extraction Method: Principal

Component Analysis.

Tabla 11: Varianza total explicada

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5,291	37,794	37,794	5,291	37,794	37,794
2	4,150	29,640	67,434	4,150	29,640	67,434
3	1,288	9,200	76,634	1,288	9,200	76,634
4	1,122	8,013	84,647	1,122	8,013	84,647
5	,783	5,590	90,236			
6	,715	5,105	95,341			
7	,263	1,877	97,217			
8	,110	,786	98,004			
9	,096	,685	98,689			
10	,064	,460	99,149			
11	,052	,370	99,519			
12	,034	,243	99,761			
13	,022	,154	99,915			
14	,012	,085	100,000			

Extraction Method: Principal Component Analysis.

Si vemos la figura de sedimentación (fig. 23), notamos que existen 4 puntos, cuyos valores son mayores que uno, lo que refuerza la tabla anterior.

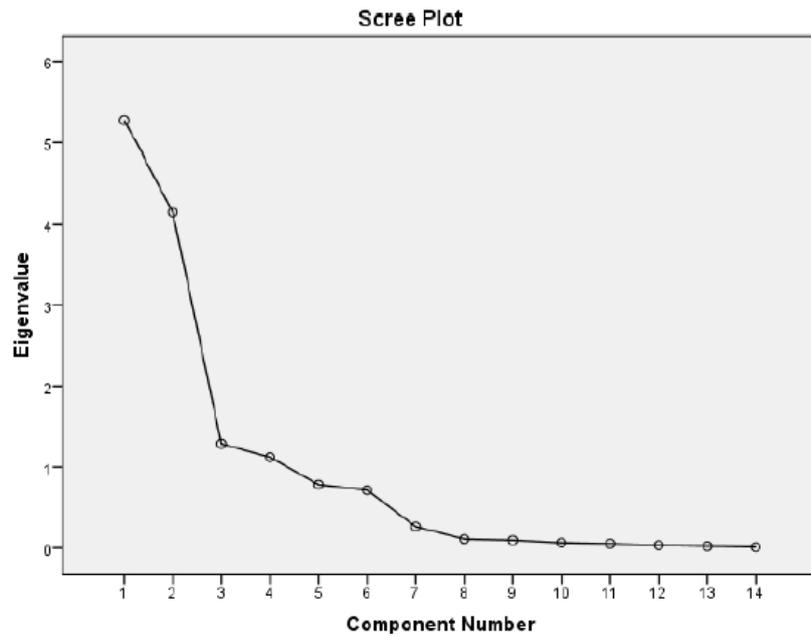


Figura 23: Figura De Sedimentación

El cuarto punto está ligeramente sobre el uno, lo que nos indica que existe poca compatibilidad de las variables que forman ese componente.

Analizando la matriz de componente tabla 12, nos damos cuenta que el primer componente tiene una correlación positiva muy alta con C_F y C_T, una correlación positiva alta con DBO_DQO y turbiedad y una correlación positiva moderada con la variable sulfatos, una correlacion positiva baja con la variable fosforo y una correlacion negactiva alta con la variable oxigeno disuelto; el segundo componente tiene una correlación positiva muy alta con A_G, positiva alta con SDT, negativa alta con NITRI Y NITRA, una correlación positiva moderada con pH, el tercer componente tiene una correlación positiva moderada con el hierro y el cuarto componente tiene una correlacionado positiva alta con la temperatura.

Tabla 12: Matrix de componente ^a

	Component			
	1	2	3	4
DBO_DQO	,768	-,572	-,065	,061
SDT	,532	,815	-,112	-,064
TEM	-,054	,055	-,228	,840
TUR	,797	-,042	-,553	-,108
A_G	,195	,908	-,192	,079
pH	-,099	,583	,138	,461
OD	-,854	,410	,167	-,022
FOSF	,389	-,279	,295	,289
NITRI	,524	-,791	-,114	,069
NITRA	,091	-,859	,371	,173
SUL	,687	,422	,466	,137
Fe	,522	,306	,610	-,199
CF	,943	,288	-,030	-,020
CT	,967	,158	,035	-,006

Extraction Method: Principal Component Analysis.

a. 4 components extracted.

Discriminando se puede eliminar a las variables TEMP, pH, SDT, A_G, Nitratos y Hierro y extrayendo un solo factor tendríamos:

Tabla 13: Comunalidades

	Initial	Extraction
DBO_DQO	1,000	,803
TUR	1,000	,734
OD	1,000	,910
NITRI	1,000	,541
CF	1,000	,679
CT	1,000	,763

Extraction Method: Principal Component Analysis.

Revisando la tabla 13 de Comunalidades y la tabla 14 de la Varianza total explicadas nos damos cuenta de que se han retenido un componentes y que con el se explica el 73.815% de las variables de manera global, de manera particular con este

componente se puede explicar el 80.30% de la DBO_DQO, 73.40% de la TUR, el 91.00% del OD, el 54.10% de los NITRI y así sucesivamente.

Tabla 14: Varianza total explicada

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4,429	73,815	73,815	4,429	73,815	73,815
2	1,102	18,366	92,181			
3	,301	5,021	97,202			
4	,084	1,400	98,602			
5	,070	1,171	99,773			
6	,014	,227	100,000			

Extraction Method: Principal Component Analysis.

Si vemos la figura 24 notamos que existen 2 puntos, cuyos auto valores son mayores que uno, sin embargo hemos utilizado la opción de SPSS para extraer un solo componente, este componente tiene un auto valor mayor que 4, lo que refuerza la tabla anterior.

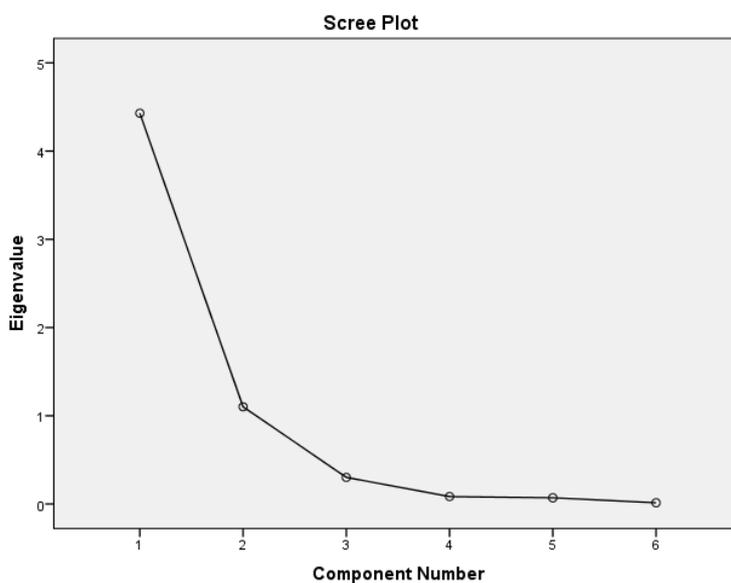


Figura 24: ACP2 Sedimentación

Analizando la matriz de componente, nos damos cuenta el componente extraído tiene una correlación positiva muy alta con OD, una correlación alta con coliformes fecales, coliformes totales, turbiedad, con DBO_DQO y nitritos.

Tabla 15: *Matrix de componentes* ^a

	Component
	1
DBO_DQO	,896
TUR	,857
OD	-,954
NITRI	,735
CF	,824
CT	,874

Extraction Method: PCA

a. 1 components extracted.

Con el ACP se tiene una idea del comportamiento de las variables para formar nuestro modelo estadístico.

5.4. Modelo de regresión lineal múltiple

5.4.1. Prueba de Normalidad.

Aplicamos esta prueba para validar la NORMALIDAD de las variables, para definir si usamos un enfoque paramétrico o no paramétrico para establecer la correlación de las variables.

Para la prueba de hipótesis consideramos:

Ho: Las variables tiene normalidad

H1: Las variables no tienen normalidad

Criterio de decisión:

Si $p\text{-valor} < 0.05$ se rechaza H_0 y se acepta H_1

Si $p\text{-valor} \geq 0.05$ se acepta H_0 y se rechaza H_1

La data tiene 14 variables y sesenta datos, con ayuda del SPSS tenemos:

Tabla 16: Resumen de casos

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
DBO_DQ	60	100,0%	0	0,0%	60	100,0%
O	60	100,0%	0	0,0%	60	100,0%
SDT	60	100,0%	0	0,0%	60	100,0%
TEMP	60	100,0%	0	0,0%	60	100,0%
TURB	60	100,0%	0	0,0%	60	100,0%
A_G	60	100,0%	0	0,0%	60	100,0%
pH	60	100,0%	0	0,0%	60	100,0%
OD	60	100,0%	0	0,0%	60	100,0%
FOSF	60	100,0%	0	0,0%	60	100,0%
NITRI	60	100,0%	0	0,0%	60	100,0%
NITRA	60	100,0%	0	0,0%	60	100,0%
SULF	60	100,0%	0	0,0%	60	100,0%
Fe	60	100,0%	0	0,0%	60	100,0%
C_F	60	100,0%	0	0,0%	60	100,0%
C_T	60	100,0%	0	0,0%	60	100,0%

De los datos de la tabla validamos que no existen datos perdidos.

Como son sesenta datos usamos el estadístico Kolmogórov-Smirnov para el nivel de significancia (p valor):

Tabla 17: Prueba de normalidad

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
DBO_DQO	,300	60	,000	,780	60	,000
SDT	,209	60	,000	,820	60	,000
TEMP	,186	60	,000	,864	60	,000
TURB	,255	60	,000	,833	60	,000
A_G	,129	60	,014	,936	60	,004
pH	,260	60	,000	,852	60	,000
OD	,238	60	,000	,822	60	,000
FOSF	,085	60	,200*	,980	60	,428
NITRI	,239	60	,000	,844	60	,000
NITRA	,262	60	,000	,824	60	,000
SULF	,125	60	,021	,942	60	,007
Fe	,183	60	,000	,927	60	,001
C_F	,268	60	,000	,823	60	,000
C_T	,317	60	,000	,754	60	,000

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Como observamos en la tabla 17 el p-valor es menor a 0.05 con excepción de la variable Fosfatos, por lo cual se rechaza H_0 y se acepta H_1 , es decir las variables no tienen normalidad, por cual vamos a usar un enfoque no paramétrico, para determinar la correlación entre las variables.

5.4.2. Cálculo de las correlaciones.

Con ayuda el SPSS encontramos las correlaciones:

Tabla 18: Correlaciones

		DBO_DQO	SDT	TEMP	TURB	A_G	pH	OD	FOSF	NITRI	NITRA	SULF	Fe	C_F	C_T
Spearman's rho	DBO_DQO Correlation Coefficient	1,000	,038	,034	,561**	-,213	,007	-,656**	,323*	,511**	,488**	,410**	,253	,350**	,643**
	Sig. (2-tailed)	.	,770	,797	,000	,103	,959	,000	,012	,000	,000	,001	,051	,006	,000
	N	60	60	60	60	60	60	60	60	60	60	60	60	60	60
SDT	Correlation Coefficient	,038	1,000	,063	,406**	,775**	,230	-,487**	-,067	-,446**	-,641**	,486**	,329*	,756**	,352**
	Sig. (2-tailed)	,770	.	,635	,001	,000	,077	,000	,611	,000	,000	,000	,010	,000	,006
	N	60	60	60	60	60	60	60	60	60	60	60	60	60	60
TEMP	Correlation Coefficient	,034	,063	1,000	,011	,090	,213	,038	,089	,040	-,077	,005	-,211	-,008	-,057
	Sig. (2-tailed)	,797	,635	.	,936	,496	,102	,774	,498	,759	,559	,972	,105	,953	,665
	N	60	60	60	60	60	60	60	60	60	60	60	60	60	60
TURB	Correlation Coefficient	,561**	,406**	,011	1,000	,077	-,042	-,865**	,269*	,402**	,182	,384**	,214	,646**	,784**
	Sig. (2-tailed)	,000	,001	,936	.	,559	,748	,000	,037	,001	,164	,002	,100	,000	,000
	N	60	60	60	60	60	60	60	60	60	60	60	60	60	60
A_G	Correlation Coefficient	-,213	,775**	,090	,077	1,000	,544**	-,173	-,124	-,618**	-,739**	,340**	,270*	,568**	,073
	Sig. (2-tailed)	,103	,000	,496	,559	.	,000	,188	,345	,000	,000	,008	,037	,000	,581
	N	60	60	60	60	60	60	60	60	60	60	60	60	60	60

pH	Correlation Coefficient	,007	,230	,213	-,042	,544**	1,000	-,111	,030	-,250	-,248	,465**	,264*	,334**	,126
	Sig. (2-tailed)	,959	,077	,102	,748	,000	.	,399	,817	,054	,056	,000	,042	,009	,339
	N	60	60	60	60	60	60	60	60	60	60	60	60	60	60
OD	Correlation Coefficient	-,656**	-,487**	,038	-,865**	-,173	-,111	1,000	-,231	-,325*	-,136	-,532**	-,347**	-,734**	-,885**
	Sig. (2-tailed)	,000	,000	,774	,000	,188	,399	.	,075	,011	,301	,000	,007	,000	,000
	N	60	60	60	60	60	60	60	60	60	60	60	60	60	60
FOSF	Correlation Coefficient	,323*	-,067	,089	,269*	-,124	,030	-,231	1,000	,174	,327*	,249	,180	,163	,296*
	Sig. (2-tailed)	,012	,611	,498	,037	,345	,817	,075	.	,184	,011	,055	,170	,214	,022
	N	60	60	60	60	60	60	60	60	60	60	60	60	60	60
NITRI	Correlation Coefficient	,511**	-,446**	,040	,402**	-,618**	-,250	-,325*	,174	1,000	,763**	-,120	-,202	-,140	,349**
	Sig. (2-tailed)	,000	,000	,759	,001	,000	,054	,011	,184	.	,000	,363	,121	,287	,006
	N	60	60	60	60	60	60	60	60	60	60	60	60	60	60
NITRA	Correlation Coefficient	,488**	-,641**	-,077	,182	-,739**	-,248	-,136	,327*	,763**	1,000	-,171	-,188	-,316*	,297*
	Sig. (2-tailed)	,000	,000	,559	,164	,000	,056	,301	,011	,000	.	,192	,150	,014	,021
	N	60	60	60	60	60	60	60	60	60	60	60	60	60	60
SULF	Correlation Coefficient	,410**	,486**	,005	,384**	,340**	,465**	-,532**	,249	-,120	-,171	1,000	,675**	,747**	,580**
	Sig. (2-tailed)	,001	,000	,972	,002	,008	,000	,000	,055	,363	,192	.	,000	,000	,000
	N	60	60	60	60	60	60	60	60	60	60	60	60	60	60

Fe	Correlation Coefficient	,253	,329*	-,211	,214	,270*	,264*	-,347**	,180	-,202	-,188	,675**	1,000	,615**	,412**
	Sig. (2-tailed)	,051	,010	,105	,100	,037	,042	,007	,170	,121	,150	,000	.	,000	,001
	N	60	60	60	60	60	60	60	60	60	60	60	60	60	60
C_F	Correlation Coefficient	,350**	,756**	-,008	,646**	,568**	,334**	-,734**	,163	-,140	-,316*	,747**	,615**	1,000	,724**
	Sig. (2-tailed)	,006	,000	,953	,000	,000	,009	,000	,214	,287	,014	,000	,000	.	,000
	N	60	60	60	60	60	60	60	60	60	60	60	60	60	60
C_T	Correlation Coefficient	,643**	,352**	-,057	,784**	,073	,126	-,885**	,296*	,349**	,297*	,580**	,412**	,724**	1,000
	Sig. (2-tailed)	,000	,006	,665	,000	,581	,339	,000	,022	,006	,021	,000	,001	,000	.
	N	60	60	60	60	60	60	60	60	60	60	60	60	60	60

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

La tabla 18 permite determinar si existe relación entre la variable dependiente $\frac{DBO}{DQO}$ y sus predictoras y también saber si existe multicolinealidad entre las predictoras; lo recomendable es que las variables predictoras no tengan relación entre ellas para que de esta manera el comportamiento de la variable dependiente sea mejor explicado por sus predictoras. En el caso que existe multicolinealidad se procederá a eliminar las variables que presenten este comportamiento. Para esto procedemos a realizar la prueba de hipótesis, de la siguiente manera:

1. Ho: No existe relación entre la DBO_DQO y la SDT
H1: Existe relación entre la DBO_DQO y la SDT
2. Ho: No existe relación entre la DBO_DQO y la TEMPERATURA
H1: Existe relación entre la DBO_DQO y la TEMPERATURA
3. Ho: No existe relación entre la DBO_DQO y la TURBIEDAD
H1: Existe relación entre la DBO_DQO y la TURBIEDAD
4. Ho: No existe relación entre la DBO_DQO y la ACEITES Y GRASA
H1: Existe relación entre la DBO_DQO y la ACEITES Y GRASA
5. Ho: No existe relación entre la DBO_DQO y la pH
H1: Existe relación entre la DBO_DQO y la pH
6. Ho: No existe relación entre la DBO_DQO y la OD
H1: Existe relación entre la DBO_DQO y la OD
7. Ho: No existe relación entre la DBO_DQO y la FOSFATOS
H1: Existe relación entre la DBO_DQO la FOSFATOS
8. Ho: No existe relación entre la DBO_DQO y la NO2
H1: Existe relación entre la DBO_DQO y la NO2
9. Ho: No existe relación entre la DBO_DQO y la NO3
H1: Existe relación entre la DBO_DQO y la NO3
10. Ho: No existe relación entre la DBO_DQO y la SO4
H1: Existe relación entre la DBO_DQO y la SO4
11. Ho: No existe relación entre la DBO_DQO y la Fe
H1: Existe relación entre la DBO_DQO y la Fe
12. Ho: No existe relación entre la DBO_DQO y la CF
H1: Existe relación entre la DBO_DQO y la CF
13. Ho: No existe relación entre la DBO_DQO y la CT

H1: Existe relación entre la DBO_DQO y la CT

Criterio de decisión:

Si p-valor < 0.05 se rechaza Ho y se acepta H1

Si p-valor es ≥ 0.05 se acepta Ho y se rechaza H1

Analizando tenemos:

1. El p-valor es menor a 0.05 por lo que se puede decir que se acepta la hipótesis H1, es decir existe correlación entre $\frac{DBO}{DQO}$ y TURBIEDAD, OD, FOSFATOS, NITRITOS, NITRATOS, SULFATOS, HIERRO, COLIFORMES_TOTALES Y COLIFORMES_FECALES; es decir, estas variables explican el comportamiento de la variable dependiente DBO. Si analizamos el coeficiente de correlación nos damos cuenta de que:
 - Las variables predictoras TURBIEDAD, NITRITOS Y COLIFORMES TOTALES tienen una correlación positiva moderada con la variable dependiente.
 - Las variables predictoras FOSFATO, NITRATO, SULFATO, HIERRO, COLIFORMES FECALES tienen una correlación positiva de moderada a baja con la variable dependiente.
 - La variable OD tiene una correlación negativa moderada con la variable dependiente
2. Las variables SDT, TEMPERATURA, ACEITES Y GRASAS, pH no influyen o explican el comportamiento de la variable dependiente (DBO), situación que se refleja en el coeficiente de correlación.

5.4.3. Determinación del modelo de regresión lineal

Descartamos las siguientes variables SDT, temperatura, aceites y grasas, pH para generar nuestro primer modelo "M1".

Para eso partimos de la siguiente prueba de hipótesis:

Ho: $\beta_1 = \beta_2 = \beta_3 \dots = 0$

H1: $\beta_1 \neq \beta_2 \neq \beta_3 \dots \neq 0$

Criterio de decisión:

Si p-valor < 0.05 se rechaza Ho y se acepta H1

Si p-valor es ≥ 0.05 se acepta H_0 y se rechaza H_1

La tabla que a continuación se detalla se indican predictoras y la variable dependiente.

Tabla 19: Variables ingresadas/eliminadas ^a

Model	Variables Entered	Variables Removed	Method
1	C_T, NITRA, FOSF, Fe, SULF, OD, NITRI, TURB, C_F ^b		Enter

a. Dependent Variable: DBO_DQO

b. All requested variables entered.

La tabla No. 20, nos define R ajustado (prueba de bondad de ajuste o coeficiente de determinación), nos indica que el 90.00% de los casos pueden ser predeterminados por el modelo y el R nos indica la fortaleza de la relación de las variables seleccionadas 95.60%.

Tabla 20: Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,956 ^a	,915	,900	,06051

a. Predictors: (Constant), C_T, NITRA, FOSF, Fe, SULF, OD, NITRI, TURB, C_F

La tabla Anova (tabla 21), es el análisis de la Varianza nos indica si es factible construir un modelo de regresión lineal múltiple a partir de la variable dependiente y de sus predictoras para que ello ocurra el p-valor debe ser menor a 0.05, situación que en nuestro caso si se cumple.

Tabla 21: ANOVA ^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1,967	9	,219	59,694	,000 ^b
	Residual	,183	50	,004		
	Total	2,150	59			

a. Dependent Variable: DBO_DQO

b. Predictors: (Constant), C_T, NITRA, FOSF, Fe, SULF, OD, NITRI, TURB, C_F

La tabla 22 de coeficientes nos indica los p-valores para que los coeficientes sean aceptados, estos valores de p-valor deben ser menores a 0.05, lo que no se está cumpliendo para las variables fosfatos, nitritos, sulfatos, hierro, coliformes fecales y coliformes torales, por lo que se procederá a realizar un nuevo modelo.

Tabla 22: Coeficientes ^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	,195	,130		1,504	,139	-,066	,456
	TURB	,003	,001	,417	2,652	,011	,001	,005
	OD	-,017	,008	-,289	-2,172	,035	-,033	-,001
	FOSF	,019	,027	,035	,720	,475	-,035	,074
	NITRI	,223	,114	,256	1,951	,057	-,007	,453
	NITRA	,103	,046	,270	2,248	,029	,011	,195
	SULF	,000	,000	,182	1,599	,116	,000	,001
	Fe	,146	,161	,055	,912	,366	-,176	,469
	C_F	-4,727E-7	,000	-,380	-1,374	,176	,000	,000
	C_T	2,406E-8	,000	,217	,818	,417	,000	,000

a. Dependent Variable: DBO_DQO

Para el nuevo modelo M2, vamos a eliminar las variables turbiedad, fosfatos, nitritos, sulfatos, hierro, coliformes totales, coliformes fecales, luego, el modelo quedaría definido por:

Tabla 23: Variables ingresadas/eliminadas ^a

Model	Variables Entered	Variables Removed	Method
1	NITRI, OD ^b	.	Enter

a. Dependent Variable: DBO_DQO

b. All requested variables entered.

En este modelo, el 88.2% de los casos pueden ser predeterminados por el modelo y el R nos indica la fortaleza de la relación de las variables seleccionadas 94.1% (tabla 24).

Tabla 24 : Resumen del modelo

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,941 ^a	,886	,882	,06549

a. Predictors: (Constant), NITRI, OD

El análisis de la Varianza (tabla 25), nos indica si es factible construir un modelo de regresión lineal múltiple a partir de la variable dependiente y de sus predictoras para que ello ocurra el p-valor debe ser menor a 0.05, situación que en nuestro caso si se cumple.

Tabla 25: ANOVA ^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1,906	2	,953	222,183	,000 ^b
	Residual	,244	57	,004		
	Total	2,150	59			

a. Dependent Variable: DBO_DQO

b. Predictors: (Constant), NITRI, OD

La tabla 26 de coeficientes nos indica los valores de B y para que sean aceptados los valores de p-valor deben ser menor a 0.05, lo que se está cumpliendo; es decir, rechazamos la hipótesis nula y aceptamos la hipótesis alternativa.

Tabla 26: Coeficientes ^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	,579	,052		11,081	,000	,474	,684
	OD	-,034	,004	-,581	-7,911	,000	-,043	-,026
	NITRI	,359	,064	,412	5,606	,000	,231	,487

a. Dependent Variable: DBO_DQO

Quedando nuestro MODELO DE REGRESIÓN LINEAL MÚLTIPLE de la siguiente manera:

$$\text{DBO_DQO} = 0.579 - 0.034(\text{OD}) + 0.359(\text{NITRITOS})$$

Utilizando el software R tenemos:

Call:

lm(formula = DBO_DQO ~ OD + Nitritos, data = base1)

Residuals:

Min	1Q	Median	3Q	Max
-0.211614	-0.036115	0.002758	0.040100	0.148104

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.579081	0.052259	11.081	7.64e-16 ***
OD	-0.034213	0.004325	-7.911	9.67e-11 ***
Nitritos	0.358751	0.063988	5.606	6.28e-07 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06549 on 57 degrees of freedom

Multiple R-squared: 0.8863, Adjusted R-squared: 0.8823

F-statistic: 222.2 on 2 and 57 DF, p-value: < 2.2e-16

5.5. Validación del modelo

5.5.1. Usando un modelo de prueba

Se tomó aleatoriamente el 80% de los datos de la data y se generó nuevamente el modelo y con este modelo se pronosticó los valores excluidos de la data los cuales deben de ser muy cercanos a los valores excluidos.

En este modelo, el 86.90% de los casos pueden ser predeterminados por el modelo y el R nos indica la fortaleza de la relación de las variables seleccionadas (93.5%). El nuevo R ajustado es muy similar al calculado anteriormente en el M1, lo que nos indica que vamos bien (tabla 27).

Tabla 27: Resumen del modelo

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,935 ^a	,874	,869	,06671

a. Predictors: (Constant), NITRI, OD

El análisis de la Varianza (tabla 28), nos indica si es factible construir un modelo de regresión lineal múltiple a partir de la variable dependiente y de sus predictoras para que ello ocurra el p-valor debe ser menor a 0.05, situación que en nuestro caso si se cumple.

Tabla 28: ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1,568	2	,784	176,225	,000 ^b
	Residual	,227	51	,004		
	Total	1,795	53			

a. Dependent Variable: DBO_DQO

b. Predictors: (Constant), NITRI, OD

La tabla de coeficientes (tabla 29), nos indica los valores de B y para que sean aceptados los valores de p-valor deben ser menor a 0.05, lo que se está cumpliendo; es decir, rechazamos la hipótesis nula y aceptamos la hipótesis alternativa. Los coeficientes obtenidos son muy similares a los del Modelo M2.

Tabla 29: Coeficientes^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
							B	Std. Error
1	(Constant)	,576	,055		10,552	,000	,466	,685
	OD	-,034	,004	-,580	-7,548	,000	-,043	-,025
	NITRI	,360	,067	,414	5,380	,000	,226	,494

a. Dependent Variable: DBO_DQO

El modelo de validación queda así:

$$DBO_DQO = 0.576 - 0.034(OD) + 0.36(NITRITOS)$$

Con este modelo (M2) que es el de prueba o validación (MP) calculamos los valores que fueron excluidos de la data de manera aleatoria y nos damos cuenta que los valores encontrados están dentro del margen de error lo que nos indica que el modelo encontrado es correcto.

$$DBO_DQO = 0.576 - 0.034(OD) + 0.36(NITRITOS)$$

PARA DATO No. 16

$$DBO_DQO = 0,76 \text{ VALOR A PRONOSTICAR}$$

$$OD = 2,18$$

$$NITRITOS = 0,57$$

DBO_DQO = 0,71 VALOR MUY CERCANO AL MEDIDO EN SITIO

PARA DATO No. 26

DBO_DQO = 0,75 VALOR A PRONOSTICAR

OD = 2,85

NITRITOS = 0,64

DBO_DQO = 0,71 VALOR MUY CERCANO AL MEDIDO EN SITIO

5.5.2. Encontrando los supuestos de normalidad con SPSS y comparándolos con los valores obtenidos en R.

Ahora probamos Linealidad, Independencia, Homocedasticidad, Normalidad y no Colinealidad, para esto calculamos las puntuaciones residuales analizando la diferencia entre las observaciones realizadas y las predichas por el modelo de regresión

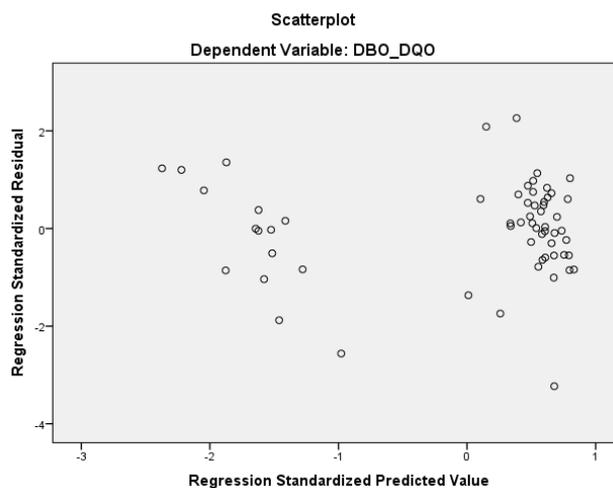
Tabla 30: Residuales ^a

	Minimu m	Maximu m	Mean	Std. Deviation	N
Predicted Value	,1903	,7661	,6167	,17973	60
Residual	-,21161	,14810	,00000	,06437	60
Std. Predicted Value	-2,372	,831	,000	1,000	60
Std. Residual	-3,231	2,261	,000	,983	60

a. Dependent Variable: DBO_DQO

1. Linealidad. – vamos a verificar que la media de los errores esta llamada a ser cero (o casi cero), que la relación entra cada variable independiente (x) de la regresión con la media POBLACIONAL de la variable dependiente (y) es lineal; de no cumplirse podría desencadenar problemas de multicolinealidad.

En la figura 25 y 26 se puede apreciar una baja relación positiva entre las dos variables seleccionadas.



**Figura 25: Poca relación positiva entre las dos variables.
Software IBM SPSS**

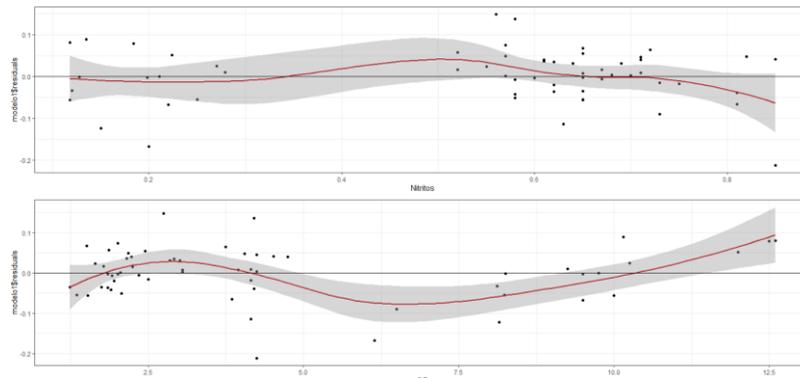


Figura 26: Poca relación positiva entre las dos variables. Software R

Para validar este supuesto con mayor certeza, con ayuda del software R calculamos la media de los residuos del modelo “mean(modelo1\$residuals)”, obteniéndose un valor $7.775268e-19$ que es prácticamente cero con lo que este supuesto queda validado.

2. Independencia. - Que los errores en la medición de las variables explicativas sean independientes entre sí, es decir las covarianzas de los errores deben ser cero.

Para que las variables tengan independencia el coeficiente de Durbin-Watson debe estar en el intervalo de 1.5 y 2.5, en nuestro caso si se encuentra, aunque pasa topando el límite inferior. El valor del estadístico Durbin Watson encontrado tanto con el software IBM SPSS y el software R fue de 1.507 Tabla 31 y 32 respectivamente.

Tabla 31: Resumen del modelo. Software IBM SPSS ^b

Modelo	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,941 ^a	,886	,882	,06549	1,507

a. Predictors: (Constant), NITRI, OD

b. Dependent Variable: DBO_DQO

Con el software R realizamos la prueba de hipótesis donde para aceptar la existencia de independencia el p-value debe de ser >0.05 ; en el caso que que el p-value sea <0.5 existiría autocorrelación de orden 1.

Tabla 32: Resumen del modelo. Software R DW

lag	Autocorrelation	D-W Statistic	p-value
1	0.2299742	1.507184	0.04

Alternative hypothesis: rho! = 0

Al ser el $p < 0.05$ hay evidencia estadística que los errores no son independientes, sin embargo, el coeficiente de Durbin Watson está dentro del intervalo en el que se puede asumir independencia entre los residuos. Statologos, (2021, 7 mayo).

Como la prueba de hipótesis realizada con el coeficiente Durbin Watson no es clara usamos las pruebas de hipótesis de Breusch-Gorfrey y la prueba de hipótesis de Ljung-Box que son mas robustas, obteniéndose el $p\text{-value} > 0.05$ por lo que existe evidencia estadística de la independencia de los residuos tabla 33 – 34

Tabla 33: Prueba Breusch-Godfrey

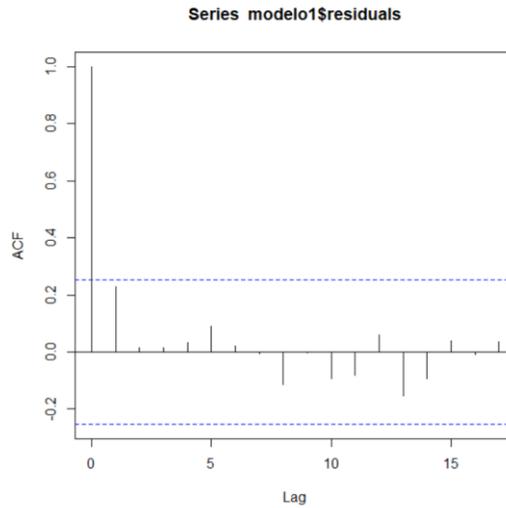
LM test	df	p-value
3.2614,	1	0.07093

Tabla 34: Prueba de Ljung-Box

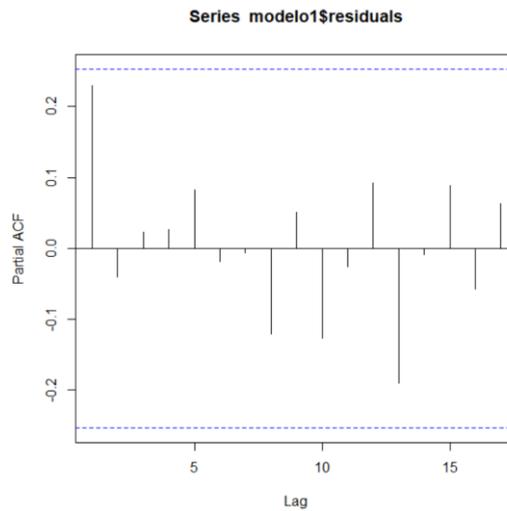
X-squared	df	p-value
3.3346	1	0.06784

Para corroborar la independencia calculamos las autocorrelaciones usando la función `acf()` del software R que calcula la función de autocorrelación simple de una serie temporal, y `pacf()` la función de autocorrelación parcial.

En ambos casos, por defecto se muestra el gráfico con bandas de confianza al 95%. Figura 27 y 28 respectivamente en los que se indica que no hay autocorrelación.



**Figura 27: Poca relación positiva entre las dos variables.
Software R**



**Figura 28: Poca relación positiva entre las dos variables.
Software R**

3. Homocedasticidad. – Que la varianza de los errores es la misma para cualquier combinación de valores de las variables independientes (que la varianza de los errores sea constante).

La figura 29 muestra agrupamientos débiles por lo que no se puede concluir que las varianzas no sean constantes.

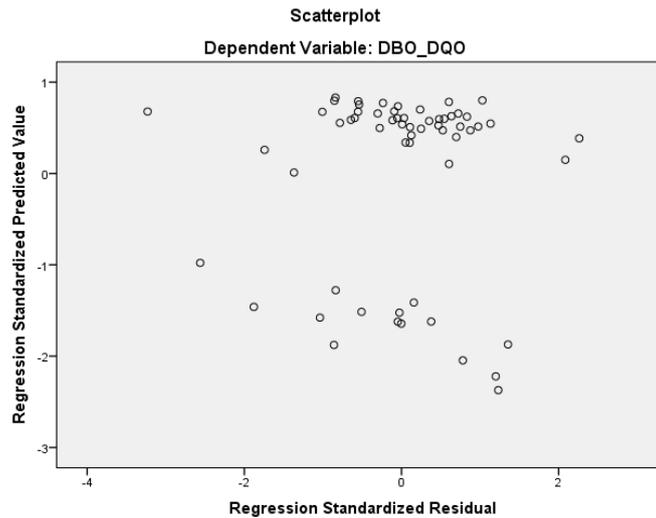


Figura 29: Agrupamientos débiles. Software IBM SPSS

Con ayuda del software R usando la función “ncvTest(modelo1)” se tiene que por la evidencia estadística $p > 0.05$ las varianzas de los errores son constantes.

Tabla 35: Validación de la varianza. Software R

Chisquare	df	p-value
0.7242346	1	0.39476

Usando “plot(modelo1, 3)” del software R, nos damos cuenta que si existe homocedasticidad.

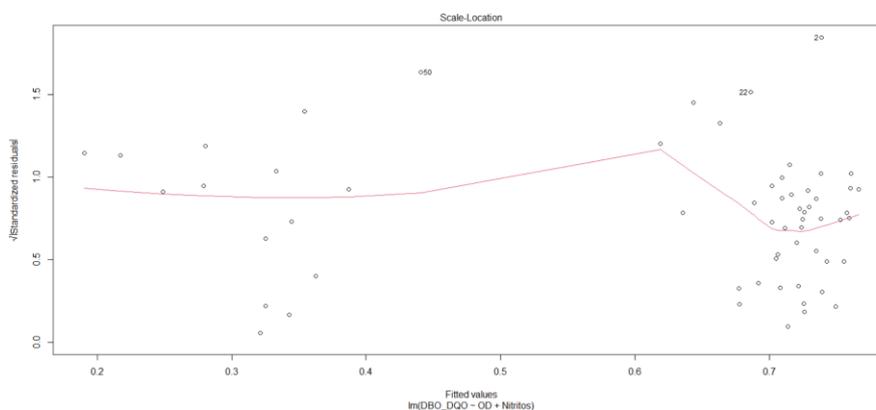


Figura 30: Verificación de la homocedasticidad del modelo

4. Normalidad. – Que los errores sigan la Ley Normal.

De acuerdo con los gráficos adjuntos los errores parecen no cumplen el principio de normalidad.

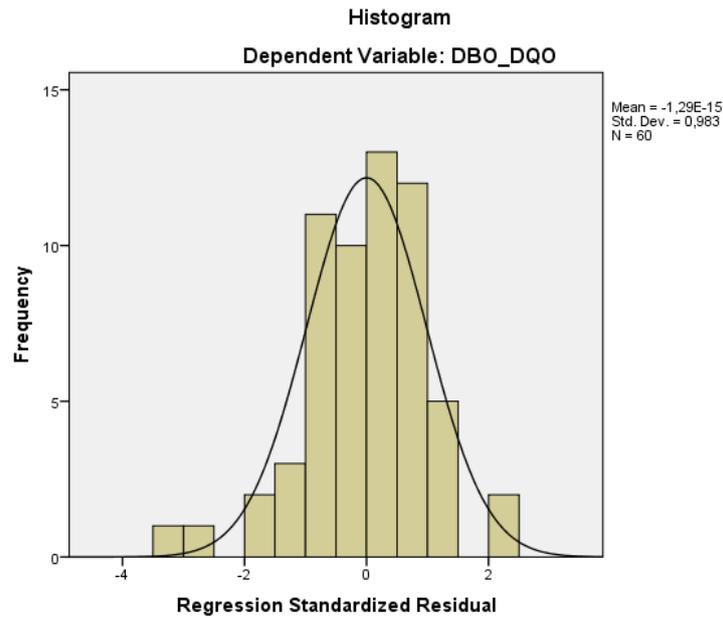


Figura 31: Variables cumplen principio de normalidad

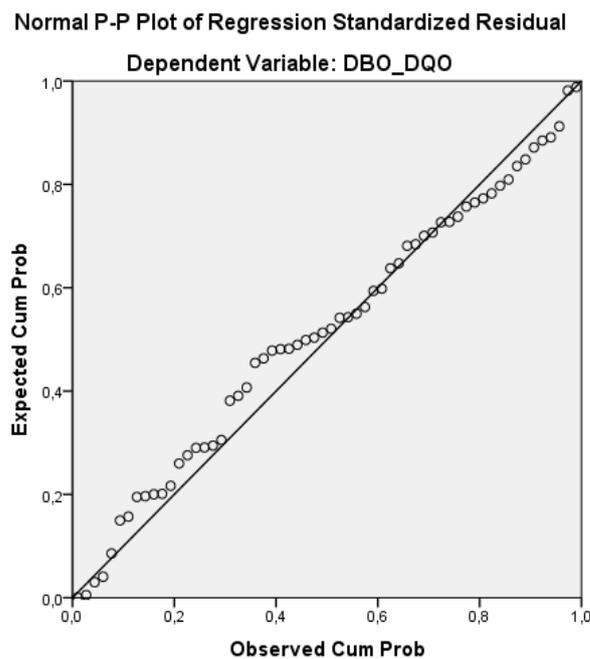


Figura 32: Variables cumplen principio de normalidad

Por lo que se realizan pruebas mas robustas como la prueba Kolmogorov-Smirnov (tabla 36), para que las variables sigan la Ley Normal el nivel de significancia asintótica (bilateral) debe ser mayor a 0.05, lo que en nuestro caso se cumple (0.171)

Tabla 36: Prueba de Kolmogorov-Smirnov

		Unstandardized Residual
N		60
Normal Parameters ^{a,b}	Mean	,0000000
	Std. Deviation	,06437015
Most Extreme Differences	Absolute	,104
	Positive	,055
	Negative	-,104
Test Statistic		,104
Asymp. Sig. (2-tailed)		,171 ^c

a. Test distribution is Normal.

b. Calculated from data.

c. Lilliefors Significance Correction.

Por otra parte, existe evidencia estadística para decir que las variables siguen una distribución normal $p > 0.05$, tabla 37.

Tabla 37: Prueba de normalidad Lilliefors (Kolmogorov-Smirnov). R

D	p-value
0.10382	0.1112

5. No colinealidad. – Que las variables independientes no estén correlacionadas entre ellas.

El valor de la varianza inflada (VIF) debe ser menor a 10 para que se cumpla el principio de no Colinealidad, en nuestro caso el VIF es de 2.703 (tabla 38).

Por otra parte, el estadístico de tolerancia está basado en la correlación múltiple de una determinada variable independiente con las restantes, es decir que para que no haya multicolinealidad la tolerancia debe ser alta, una tolerancia menor a 0.1 indica grave problemas de colinealidad, en nuestro caso el estadístico de tolerancia tiene un valor de 0.37 por lo que se puede decir que se cumple el principio de no colinealidad.

Tabla 38: Coeficientes ^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
1 (Constant)	,579	,052		11,081	,000		
OD	-,034	,004	-,581	-7,911	,000	,370	2,703
NITRI	,359	,064	,412	5,606	,000	,370	2,703

a. Dependent Variable: DBO_DQO

Igual resultado se encontró aplicando el software R (tabla 39).

Tabla 39: Valores de la varianza inflada. Software R

OD	Nitritos
2.702634	2.702634

En función de la evidencia estadística no existe multicolinealidad al ser VIF menor a 10.

CAPITULO VI DISCUSIÓN DE RESULTADOS

El análisis de los parámetros físicos, químicos, biológicos y microbiológicos se desarrolló en cuatro sectores o puntos del estero El Macho de la Ciudad de Machala; este cuerpo de agua atraviesa a la ciudad de este a oeste, el punto A1 está en la parte superior del estero en donde la ciudad empieza, los puntos A2 y A3 están en la zona media y el punto A4 en la desembocadura del estero en el Mar, lo indicado explica porque el Sector A1 es el menos contaminado al tener poco contacto con las aguas residuales de la ciudad y la forma como sube el nivel de contaminación en los sectores A2 y A3 al aumentar las descargas de aguas residuales y como baja el nivel de contaminación en el sector A4 al diluirse con el agua del Mar.

En general, analizando los parámetros físicos, químicos, biológicos y microbiológicos de las aguas del Estero hemos encontrado que estos indican un grave problema de contaminación de acuerdo a los parámetros indicados en el Texto Unificado de Legislación Secundaria de Medio Ambiente de la Republica del Ecuador y que a la fecha de la presente investigación está vigente; por otra parte Samboni, Reyes y Carvajal (2011) consideran que un adecuado procedimiento para conocer los niveles de contaminación consiste en integrar parámetros de tipo fisicoquímicas y biológicas.

Sotomayor (2016) en su trabajo investigativo para diagnosticar la calidad de las aguas superficiales de un segmento del Río Paute, implementó el análisis de componentes principales (PCA) de datos multivariados para garantizar un nivel alto de rigurosidad científica en la evaluación de la calidad de agua, y, de esta forma contar con datos que consoliden planes apropiados para la gestión y remediación de los ecosistemas ribereños, en este trabajo se usó el PCA para establecer agrupaciones de variables predictoras para obtener un modelo con mayor asertividad, tal es así que el modelo PCA encontrado describe a la variable predictora en el 73.815% de los casos.

López y Palací (2014), en su investigación sobre el estudio multivariante de la calidad del agua, indican que es necesario efectuar un diagnóstico de tipo longitudinal y

transversal. Para ello, que aplicó en primer lugar el Análisis de Componentes Principales (PCA) con la finalidad de reducir los datos de las variables de investigación.

Carrera, Guevara, Tamayo y Guallichico (2015) en su análisis multivariado de calidad de agua para uso agrícola, tuvo por propósito conocer las concentraciones de elementos inorgánicos a través de la aplicación de la técnica de componentes Principales (PCA), utilizando como parámetros de comparación la norma INEN 2169 y realizando 26 tomas de muestras, los cuales permitieron determinar las características de las aguas y su incidencia en los suelos de cultivo.

Conociendo el comportamiento de las variables se aplicó regresión lineal múltiple para encontrar el modelo que pronostique la calidad del agua del Estero El Macho basado en la relación DBO_DQO Becerra et al. (2014).

Gil Marin (2020) en su investigación desarrolla un modelo de calidad de agua subterránea mediante la combinación de técnicas estadísticas de análisis de componentes principales con la finalidad de simplificar las correlaciones entre los parámetros de calidad de agua, así como a través de regresiones lineales múltiples para el pronóstico de las cualidades del agua.

Del modelo encontrado mediante regresión lineal múltiple la variable dependiente (DBO_DQO) es pronosticada por sus predictoras (Nitritos y Oxígeno disuelto) con un R ajustado de 88.200%.

El modelo encontrado cumple los supuestos de la siguiente manera:

1. Linealidad; ya que los residuos siguen una distribución normal cuya media encontrada es igual a $7.775268e-19$ equivalente a cero (Ramsey, 1969).
2. Independencia; usamos las pruebas de hipótesis de Breusch-Gorfrey que es más robusta que la prueba de DW, obteniéndose significancia estadística ($p\text{-value} = 0.07093$) por lo que los errores son independientes.
3. Homocedasticidad; la prueba varianza no constante con la función `ncvTest()` estableció significancia estadística ($p = 0.39476$) por lo que la varianza de los errores es la misma para cualquier combinación de valores de las variables independientes, por lo que el supuesto de homocedasticidad se cumple.

4. Normalidad; de acuerdo con la prueba Kolmogorov-Smirnov (tabla 32), para que las variables sigan la Ley Normal el nivel de significancia asintótica (bilateral) debe ser mayor a 0.05, lo que en nuestro caso se cumple (0.171) por lo que se cumple el supuesto de normalidad.
5. Colinealidad; para que se cumpla este supuesto el valor de la varianza inflada (VIF) debe ser menor a 10 para que se cumpla el principio de no Colinealidad, en nuestro caso el VIF es de 2.703 (tabla 36), cumpliéndose por lo tanto este supuesto

CAPITULO VII

CONCLUSIONES

1. Analizados los parámetros físico, químicos y biológicos que caracterizan las aguas del Estero El Macho, se tiene:
 - Que los parámetros como coliformes fecales, demanda bioquímica de oxígeno, demanda química de oxígeno, hierro y oxígeno disuelto se encuentran fuera de los límites permitidos de calidad de agua.
 - La contaminación del agua también se debe a la presencia de un número considerable de sólidos disueltos y turbidez que afecta las características físicas del agua. En este sentido, el agua del estero El Macho no cumple con los criterios de calidad del agua.
 - Según Becerra et al. (2014), de la relación $\frac{DBO}{DQO}$ podemos decir que las aguas del estero están contaminadas por materia orgánica biodegradable situación que se evidencia por las descargas de aguas residuales que soporta el canal las cuales eliminan aguas domesticas al ser la ciudad de Machala una ciudad no Industrializada.
2. Mediante el análisis de componentes principales, se logró agrupar las variables predictoras lo que facilito el camino para la obtención del modelo que determina la calidad del agua del estero El Macho.
3. El modelo encontrado permite explicar la calidad del agua del estero El Macho con un R ajustado del 88.2% en función de los Nitritos y el Oxígeno disuelto.
4. El modelo encontrado cumple los supuestos de linealidad, independencia, homocedasticidad, normalidad y colinealidad, además fue validado con un modelo de prueba el cual genero datos dentro de los límites permitidos.
5. El modelo encontrado permitirá la toma de decisiones de manera rápida, al tener que realizar dos exámenes (oxígeno disuelto y nitratos) que duran muy poco tiempo y tienen un costo bajo, para conocer el estado de las aguas del estero.

CAPITULO VIII RECOMENDACIONES

En función a los resultados obtenidos se recomienda:

1. Establecer acciones de remediación para la recuperación de la calidad de agua del estero El Macho, siendo una medida viable la eliminación completa de los desagües de aguas residuales a lo largo del estero.
2. Aplicar el modelo encontrado para el análisis de los resultados de calidad de agua, a fin de establecer correlaciones significativas que contribuyan a la identificación de áreas de mayor afectación, así como desarrollar medidas anticipada en base al modelo predictivo.
3. Desarrollar un monitoreo permanente de los parámetros químicos y físicos para comprobar la eficacia de las acciones de remediación para la recuperación de la calidad de agua del estero El Macho

CAPITULO IX REFERENCIAS BIBLIOGRÁFICAS

- Abarca, F. J. (2007). Técnicas para evaluación y monitoreo del estado de los humedales y otros ecosistemas acuáticos. En *Perspectivas sobre conservación de ecosistemas acuáticos en México* (Segunda, pp. 113-144). Secretaría de Medio Ambiente y Recursos Naturales.
- Aguas Machala EP. (2019). *Informe de rendición de cuentas 2018* (p. 19) [Rendición de cuentas]. Empresa Pública Municipal de Agua Potable y Alcantarillado del Cantón Machala. <https://aguasmachala.gob.ec/wp-content/uploads/2019/04/INFORME-DE-RENDICION-DE-CUENTAS-2018-AGUAS-MACHALA.pdf>
- Aguas Machala EP. (2022). *Transparencia*. Ley Orgánica de Transparencia y Acceso a la Información Pública. <https://www.aguasmachala.gob.ec/transparencia>
- Aldas Manzano, J., & Uriel Jimenez, E. (2017). *Análisis multivariante aplicado con R* (Segunda). Ediciones Paraninfo, SA.
- Amat Rodrigo, J. (2017, junio). *Análisis de Componentes Principales (Principal Component Analysis, PCA) y t-SNE*. https://www.cienciadedatos.net/documentos/35_principal_component_analysis
- Aparicio, R. F., Espinoza, M. F., Milla, V. C., & Reyes, R. E. (2013). Modelo estadístico para predecir la calidad del agua de consumo humano en el ámbito rural del Callejón de Huaylas. *Aporte Santiaguino*, 6(2), 25-34. <https://doi.org/10.32911/as.2013.v6.n2.501>

- Arroyave Gómez, D. M., Moreno Tovar, A. A., Toro Botero, F. M., Gallego Suárez, D. de J., & Carvajal Serna, L. F. (2013). Estudio del modelamiento de la calidad del agua del Río Sinú, Colombia. *Revista Universidad de Medellín*, 12(22), 33-44.
- Baeza-Serrato, R., & Vázquez-López, J. A. (2014). Transición de un modelo de regresión lineal múltiple predictivo, a un modelo de regresión no lineal simple explicativo con mejor nivel de predicción: Un enfoque de dinámica de sistemas. *Revista Facultad de Ingeniería Universidad de Antioquia*, 71, 59-71.
- Baldiris Navarro, I., Sánchez Aponte, J., Cabeza, D., & Acosta, J. C. (2017). Evaluación de la calidad de agua en la bahía de barbacoas Cartagena utilizando métodos estadísticos multivariados. *Teknos revista científica*, 17(1), 59-67.
- Balzarini, M., Bruno, C., Córdoba, M., & Teich, I. (2015). *Herramientas en el análisis estadístico multivariado* [Tesis de Maestría]. Universidad Nacional.
- Banco Mundial. (2011). *Diagnóstico de la gestión de los recursos hídricos* [Artículo]. Departamento de Medio Ambiente y Desarrollo Sostenible. https://www.academia.edu/download/62341620/Diagnostico_gestion_de_recursos_hidricos_en_Chile_Banco_Mundial20200311-25794-1vn2e5i.pdf
- Barreto-Villanueva, A. (2012). El progreso de la Estadística y su utilidad en la evaluación del desarrollo. *Papeles de población*, 18(73), 241-271.
- Bécher Quinodóz, F. N., & Blarasin, M. (2017). Evaluación de relaciones geomorfología-calidad de agua subterránea mediante técnicas estadísticas y modelación numérica en la planicie arenosa del sur de Córdoba, Argentina. *Geoacta*, 41(2), 1-16.
- Benavides Alarcón, C. (2023). *Análisis tipológico de lagos nortapagónicos chilenos para la elaboración de normas de calidad de aguas superficiales*. [Tesis de Pregrado, Universidad de Concepción]. <http://repositorio.udec.cl/jspui/handle/11594/10831>
- Brousett-Minaya, M., Chambi Rodríguez, A., Mollocondo Turpo, M., Aguilar Atamari, L., & Lujano Laura, E. (2018). Evaluación físico-química y microbiológica de agua para consumo humano Puno-Perú. *Fides et Ratio-Revista de Difusión cultural y científica*

de la Universidad La Salle en Bolivia, 15(15), 47-68.

Caballero, A. L. (2009). Valoración del impacto de efluentes minero industriales del carbón en el arroyo San José de Río Turbio en Santa Cruz. *Informes Científicos Técnicos-UNPA*, 1(1), 85-112.

Caho-Rodríguez, C. A., & López-Barrera, E. A. (2017). Determinación del Índice de Calidad de Agua para el sector occidental del humedal Torca-Guaymaral empleando las metodologías UWQI y CWQI. *Producción + Limpia*, 12(2), 35-49. <https://doi.org/10.22507/pml.v12n2a3>

Carrera, D., Guevara, P., Tamayo, L., & Guallichico, D. (2015). *Análisis multivariado de las aguas de la Subcuenca del Río Ambi en época de estiaje y su relación con la calidad desde el punto de vista agrícola*. 10(1), 123-129.

Carrillo, J. K., & Duran, C. M. (2018). Metodología para determinar la calidad del agua a partir de métodos analíticos sensoriales. *BISTUA Revista de la Facultad de Ciencias Básicas*, 16(1), 185-192.

Casares, M. V., & De Cabo, L. I. (2018). Análisis de tendencias de variables indicadoras de calidad de agua para el Riachuelo (Cuenca Matanza-Riachuelo, Argentina). *Revista internacional de contaminación ambiental*, 34(4), 651-665.

Closas, A. H., Arriola, E. A., Kuc Zening, C. I., Amarilla, M. R., & Jovanovich, E. C. (2013). Análisis multivariante, conceptos y aplicaciones en Psicología Educativa y Psicometría. *Enfoques*, 25(1), 65-92.

Código Orgánico del Ambiente [CODA]. Registro Oficial Suplemento Número 983 del 2017. 12 de abril del 2017, Ecuador.

Corradi, E., Agostini, M., Greco, G., Massidda, D., Santi, M., Calderisi, M., Signore, G., & Cecchini, M. (2020). An objective, principal-component-analysis (PCA) based, method which improves the quartz-crystal-microbalance (QCM) sensing performance. *Sensors and Actuators A: Physical*, 315, 112323. <https://doi.org/10.1016/j.sna.2020.112323>

Delgado Zambrano, D. M., Zambrano García, D. P., Alcívar Cedeño, U. E., Palacios Revelo,

- J. G., Santos Falcones, M. C., Cevallos Cedeño, R. E., & Burgos Briones, G. A. (2020). Toxicidad del agua en el estuario del río Burro en la ciudad de Manta. *Polo del Conocimiento: Revista científico-profesional*, 5(1), 305-316.
- Doménech, J. (2002). Sanidad ambiental.-Control de la calidad del agua. *Offarm: farmacia y sociedad*, 21(10), 138-146.
- Escobar Jaramillo, L. A. (2008). Indicadores ambientales sintéticos: Una aproximación conceptual desde la estadística multivariante. *Gestión y Ambiente*, 11(1), 121-140.
- Flores Suárez, Á. G., & Pozo Merejildo, E. J. (2023). *Evaluación estadística de los parámetros de análisis físico, químico y biológico de los afluentes y efluentes del sistema de tratamiento de aguas residuales regulado por la Empresa Pública Mancomunada Aguapen-EP, ubicado en la cabecera cantonal del cantón Santa Elena, provincia de Santa Elena.*
- Forero Gómez, G., & Martínez Lozano, J. A. (2020). *Modelo de regresión lineal múltiple para el pronóstico de ventas de bolsas ecológicas para la empresa Boleco SA, en la ciudad de Bogotá DC.*
- García Merino, L. S., Galvez Diaz, N. del C., Villanueva Calderon, J. A., Olivos Romero, F. G., Guerrero Millones, A. M., & Marin del Aguila, W. T. (2023). Estadística aplicada a la investigación científica. *Revista Científica Emprendimiento Científico Tecnológico*, 5, 144-144.
- Gil Gómez, J. A. (2014). *Determinación de la calidad del agua mediante variables físico químicas, y la comunidad de macroinvertebrados como bioindicadores de calidad del agua en la cuenca del río Garagoa [Universidad de Manizales].*
<https://ridum.umanizales.edu.co/handle/20.500.12746/1803>
- Gil Marin, J. A. (2020). Modelo de calidad del agua subterránea mediante el uso combinado del análisis de componentes principales (ACP) y regresiones lineales múltiples (RLM): Caso de estudio: Acuíferos de Maturín, Monagas, Venezuela. *Innotec*, 20 jul-dic, 67-88.

- Gil Marín, J. A., Vizcaino González, C. del V., & Montañó Mata, N. J. (2018). Evaluación de la calidad del agua superficial utilizando el índice de calidad del agua (ICA). Caso de estudio: Cuenca del Río. *Anales Científicos*, 79(1), 111-119.
- Giraud Herrera, L. M., & Morantes Quintana, G. R. (2017). Aplicación del análisis multivariante para la sostenibilidad ambiental urbana. *Bitácora urbano territorial*, 27(1), 89-100.
- Gómez, I. N. & Peñuela, G. A. (). *Revisión de los métodos estadísticos multivariados usados en el análisis de calidad de aguas*. Recuperado de: <http://hdl.handle.net/20.500.12010/647>.
- Gómez Miranda, I. N., & Peñuela Mesa, G. A. (2016). Revisión de los métodos estadísticos multivariados usados en el análisis de calidad de aguas. *Revista Mutis*, 6(1), 54-63. <https://doi.org/10.21789/22561498.1112>
- Guano Avila, M. E., & Velasteguí Carrillo, V. A. (2014). *Manual básico para la modelación numérica de la calidad del agua en ríos: Programa computacional HEC-Ras 4.0. Ejemplo de aplicación a un tramo del río Machángara dentro de la zona de influencia de la ciudad de Quito*.
- Guerrero Martínez, C. D. (2019). *Evaluación temporal y espacial de la calidad del agua de la quebrada Las Delicias (cerros orientales de Bogotá) mediante un análisis estadístico multivariado* [Tesis de Pregrado, Universidad Cooperativa de Colombia]. <http://hdl.handle.net/20.500.12494/15130>
- Guzmán, B. L., Nava, G., & Díaz Bevilacqua, P. (2015). La calidad del agua para consumo humano y su asociación con la morbimortalidad en Colombia, 2008-2012. *Biomédica*, 35(2), 90-177. <https://doi.org/10.7705/biomedica.v35i0.2511>
- Hernandez-Sampieri, R. y Mendoza (2018). *Metodología de la Investigación. Las rutas cuantitativas, cualitativa y mixta*. McGraw-Hill Interamericana Editores
- Ledesma, C., Bonansea, M., Rodríguez, C. M., & Sánchez Delgado, A. R. (2013). Determinación de indicadores de eutrofización en el embalse Río Tercero, Córdoba (Argentina). *Revista Ciência Agronômica*, 44(3), 419-425.

- López, M. (2015). Análisis Físico-Químico de la calidad del agua en el canal El Macho de la ciudad de Machala. *Universidad Técnica de Machala*.
- López Rodríguez, M. I., & Palací López, D. G. (2014). Estudio multivariante de la calidad del agua: Aplicación al río Jucar en el período 1990-2013. *M+ A Revista Electronica de Medioambiente*, 15(1), 37-53.
- López-Aguado, M., & Gutiérrez-Provecho, L. (2019). Cómo realizar e interpretar un análisis factorial exploratorio utilizando SPSS. *REIRE Revista d'Innovació i Recerca en Educació*, 12(2), 1-14. <https://doi.org/10.1344/reire2019.12.227057>
- Lorbes Medina, J., García Orellana, Y., Ohep, C., & Milla Pino, M. (2014). Discriminación de tipos de calidad de agua de riego según atributos químicos utilizando una técnica multivariada. *Revista Mexicana de Ciencias Agrícolas*, 5(1), 29-36.
- Méndez Zambrano, P. V., Arcos Logroño, J. P., & Cazorla Vinueza, X. R. (2020). Determinación del índice de calidad del agua (NSF) del río Copueno ubicado en Cantón Morona. *Dominio de las Ciencias*, 6(3), 734-746.
- Mendoza, K. S., Acosta, D. S., Torres, L. F., & Parra, J. M. (2023, mayo 2). Modelo de regresión lineal múltiple sobre el PIB. *RPubs*. https://rpubs.com/Karen_Mendoza/1038030
- Montoya, O. 2007 Aplicación del análisis factorial a la investigación de mercados. *Scientia et Technica Año XIII*
- Moreno Tovar, A. A., Toro Botero, M., & Carvajal, L. F. (2008). Revisión de criterios y metodologías de diseño de redes para el monitoreo de la calidad del agua en ríos. *Avances en Recursos Hidráulicos*, 18, 57-68. Redalyc.
- Osorio Angarita, M. A., & Suárez Parra, A. B. (2014). Importancia de la probabilidad y la estadística en la formación del Ingeniero. *I3+*, 1(2), Article 2. <https://doi.org/10.24267/23462329.63>
- Otero, M. R., Greca, I. M., & Silveira, F. L. da. (2003). Imágenes visuales en el aula y rendimiento escolar en Física: Un estudio comparativo. *Revista electrónica de*

enseñanza de las ciencias. Ourense. Vol. 2, no. 1 (2003), p. 1-30.

Pauta Calle, G., & Chang Gómez, J. (2016). Índices de calidad del agua de fuentes superficiales y aspectos toxicológicos, evaluación del Río Burgay. *Maskana*, 5, 165-176.

Pereyra, L. E. (2022). *Metodología de la investigación*. Klik.

Pérez, J. I., Nardini, A. G., & Galindo, A. A. (2018). Análisis Comparativo de Índices de Calidad del Agua Aplicados al Río Ranchería, La Guajira-Colombia. *Información tecnológica*, 29(3), 47-58. <https://doi.org/10.4067/S0718-07642018000300047>

Pérez Petitón, J., López Seijas, T., Hernández Cuello, G., & González Robaina, F. (2009). Modelo matemático para determinar la calidad del agua en dos puntos del arroyo Guachinango. *Revista Ciencias Técnicas Agropecuarias*, 18(3), 19-24.

Poza Lara, C. (2008). Técnicas estadísticas multivariantes para la generación de variables latentes. *Revista Escuela de Administración de Negocios*, 64, 89-99. <https://doi.org/10.21158/01208160.n64.2008.454>

Ramírez Ríos, A., & Polack Peña, A. M. (2020). Estadística inferencial. Elección de una prueba estadística no paramétrica en investigación científica. *Horizonte de la Ciencia*, 10(19), 191-208.

Recabarren Santibáñez, O. (2016). El estándar del derecho de aguas desde la perspectiva del derecho internacional de los derechos humanos y del medio ambiente. *Estudios constitucionales*, 14(2), 305-346.

Rodríguez, J. G., Giménez Lega, E., Rodríguez, J. G., & Giménez Lega, E. (2017). Aplicación del análisis de componentes principales en la investigación de aguas de pozo para el consumo humano. *Población y Desarrollo*, 23(45), 38-52. [https://doi.org/10.18004/pdfce/2076-054x/2017.023\(45\)038-052](https://doi.org/10.18004/pdfce/2076-054x/2017.023(45)038-052)

Rubio Arias, H. O., Ortiz Delgado, R. C., Quintana Martínez, R. M., Saucedo Terán, R. A., Ochoa Rivero, J. M., & Rey Burciaga, N. I. (2014). Índice de calidad de agua (ICA) en la presa la boquilla en Chihuahua, México. *Ecosistemas y recursos agropecuarios*,

1(2), 139-150.

- Samboni, N., Trujillo, A. R., & Carvajal, Y. (2011). Aplicación de los indicadores de calidad y contaminación del agua en la determinación de la oferta hídrica neta. *Ingeniería y competitividad*, 13(2), 49-60.
- Sarie Palas, A. D. (2020). Análisis de la vulnerabilidad de las familias en el sector denominado Manglar del Afro de las Riberas del Canal El Macho de la Ciudad de Machala. *Repositorio Digital FLACSO Ecuador*, 50. <http://hdl.handle.net/10469/16543>
- Sotomayor, G. (2016). *Evaluación de la calidad de las aguas superficiales mediante técnicas de estadística multivariante: Un estudio de caso en la cuenca del Río Paute, al sur de Ecuador* [Tesis de Maestría, Universidad Nacional de La Plata]. <http://sedici.unlp.edu.ar/handle/10915/53959>
- Suárez-Medina, I. (2020). Calidad del agua de consumo humano influenciada por aguas servidas: Quality of water for human consumption influenced by sewage. *Revista Ciencia Nor@ndina*, 3(1), 80-89. <https://doi.org/10.37518/2663-6360X2020v3n1p80>
- Statologos. (2021, 7 mayo). La prueba de Durbin-Watson: definición y ejemplo. <https://statologos.com/prueba-de-durbin-watson/>
- Torres, P., Cruz, C. H., & Patiño, P. J. (2009). Índices de calidad de agua en fuentes superficiales utilizadas en la producción de agua para consumo humano: Una revisión crítica. *Revista Ingenierías Universidad de Medellín*, 8(15), 79-94.
- Valencia Delfa, J. L. (2007). *Estudio estadístico de la calidad de las aguas en la cuenca hidrográfica del río Ebro* [[Http://purl.org/dc/dcmitype/Text](http://purl.org/dc/dcmitype/Text), Universidad Politécnica de Madrid]. <https://dialnet.unirioja.es/servlet/tesis?codigo=2783&orden=0&info=link&info=link>
- Vigueras Velázquez, M. E., Carbajal Hernández, J. J., Sánchez Fernández, L. P., & Hernández Martínez, M. (2016). Modelo computacional para la evaluación de la calidad del agua en el cultivo del pez blanco. *Pistas Educativas*, 38(120), Article 120.
- Yamashita, T., Hasegawa, T., Hayashida, Y., Ninomiya, K., Shibata, S., Ito, K., Mizuguchi, H.,

& Yokoyama, H. (2022). Energy savings with a biochemical oxygen demand (BOD)- and pH-based intermittent aeration control system using a BOD biosensor for swine wastewater treatment. *Biochemical Engineering Journal*, 177, 108266. <https://doi.org/10.1016/j.bej.2021.108266>

Zhinda, R., (2024) Aplicación de la estadística multivariante para el desarrollo de un modelo de la calidad del agua del estero El Macho, Machala, El Oro. <https://produccioncientificaluz.org/index.php/rluz/article/view/41368>

I. ANEXOS

1.) Matriz de consistencia

Modelo estadístico de calidad del agua del Estero El Macho, Machala-El Oro

Problema	Objetivo	Hipótesis	Variables	Objeto de estudio	Metodología
¿Qué modelo es factible para determinar en forma significativa la calidad del agua del estero El Macho de la ciudad de Machala, provincia de El Oro?	Determinar el modelo estadístico para conocer la calidad del agua del estero El Macho de la ciudad de Machala, provincia de El Oro.	Un modelo estadístico determina significativamente la calidad del agua del estero El Macho de la ciudad de Machala, provincia de El Oro, conforme a los parámetros físico-químicos y biológicos.	Variable dependiente: Y = Modelo estadístico Variables independientes: X ₁ = Parámetros físico-químicos de calidad del agua X ₂ = Parámetros biológicos de calidad del agua	Cuatro puntos de análisis del Estero El Macho de la ciudad de Machala.	Tipo de investigación según el objetivo: Básica y aplicada Nivel de profundización del objeto de estudio: Descriptiva Método según datos empleados: Cuantitativo Diseño de la investigación: Investigación correlacional bajo un modelo estadístico multivariante de regresión lineal múltiple y métrica de componentes principales

2.) Mapa de la ubicación del Estero El Macho respecto de la ciudad de Machala. Ubicación de sitios de toma de muestras.

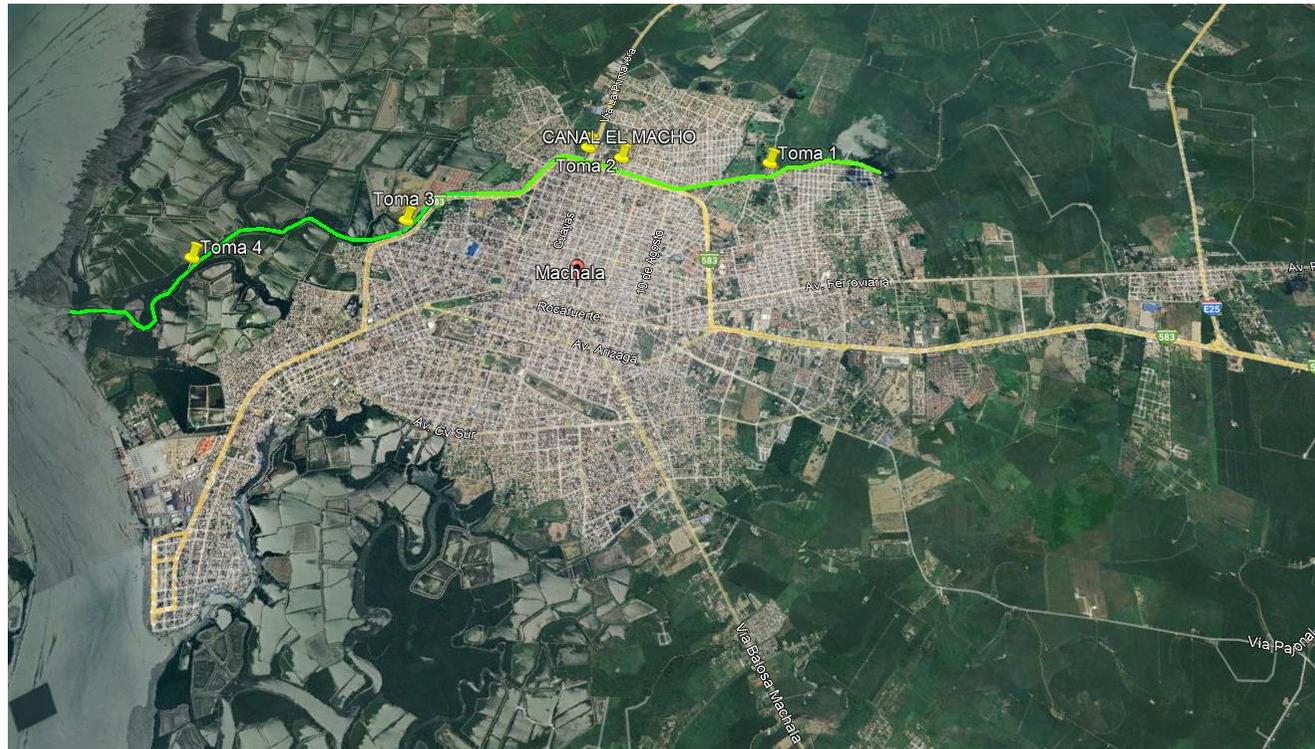


Figura 32. Ubicación del canal y punto de toma de muestras.

Fuente: Google 2019

3.) Normativa de calidad de agua

Tabla 40: Criterios de calidad admisibles para la preservación de la vida acuática y silvestre en aguas dulces, marinas y de estuarios.

Fuente: TULSMA (2017)

PARÁMETROS	Expresados como	Unidad	Criterio de calidad	
			Agua dulce	Agua marina y de estuario
Aluminio ⁽¹⁾	Al	mg/l	0,1	1,5
Amoniaco Total ⁽²⁾	NH ₃	mg/l	-	0,4
Arsénico	As	mg/l	0,05	0,05
Bario	Ba	mg/l	1,0	1,0
Berilio	Be	mg/l	0,1	1,5
Bifenilos Policlorados	Concentración de PCBs totales	µg/l	1,0	1,0
Boro	B	mg/l	0,75	5,0
Cadmio	Cd	mg/l	0,001	0,005
Cianuros	CN ⁻	mg/l	0,01	0,01
Cinc	Zn	mg/l	0,03	0,015
Cloro residual total	Cl ₂	mg/l	0,01	0,01
Clorofenoles ⁽³⁾		mg/l	0,05	0,05
Cobalto	Co	mg/l	0,2	0,2
Cobre	Cu	mg/l	0,005	0,005
Cromo total	Cr	mg/l	0,032	0,05
Estaño	Sn	mg/l		2,00
Fenoles monohídricos	Expresado como fenoles	mg/l	0,001	0,001
Aceites y grasas	Sustancias solubles en hexano	mg/l	0,3	0,3
Hidrocarburos Totales de Petróleo	TPH	mg/l	0,5	0,5
Hierro	Fe	mg/l	0,3	0,3
Manganeso	Mn	mg/l	0,1	0,1
Materia flotante de origen antrópico	visible		Ausencia	Ausencia
Mercurio	Hg	mg/l	0,0002	0,0001
Níquel	Ni	mg/l	0,025	0,1
Oxígeno Disuelto	OD	% de saturación	> 80	> 60
Piretroides	Concentración de piretroides totales	mg/l	0,05	0,05
Plaguicidas organoclorados totales	Organoclorados totales	µg/l	10,0	10,0
Plaguicidas organofosforados totales	Organofosforados totales	µg/l	10,0	10,0
Plata	Ag	mg/l	0,01	0,005
Plomo	Pb	mg/l	0,001	0,001
Potencial de Hidrógeno	pH	unidades de pH	6,5 – 9	6,5 – 9,5
Selenio	Se	mg/l	0,001	0,001
Tensoactivos	Sustancias activas al azul de metileno	mg/l	0,5	0,5
Nitritos	NO ₂ ⁻	mg/l	0,2	
Nitratos	NO ₃ ⁻	mg/l	13	200
DQO	DQO	mg/l	40	-
DBO ₅	DBO ₅	mg/l	20	-
Sólidos Suspendidos Totales	SST	mg/l	max incremento de 10% de la condición natural	-

⁽¹⁾ Aluminio: Si el pH es menor a 6,5 el criterio de calidad será 0,005 mg/l

⁽²⁾ Aplicar la Tabla 2a como criterio de calidad para agua dulce

⁽³⁾ Si sobrepasa el criterio de calidad se debe analizar el diclorofenol cuyo criterio de calidad es 0,2 µg/l

Relación $\frac{DBO}{DQO}$, según Becerra et al. (2014):

- a. Cuando la relación $\frac{DBO}{DQO}$ es inferior a 0.2 indica que el agua esta contaminada por una carga orgánica no biodegradable
- b. Cuando la relación $\frac{DBO}{DQO}$ esta comprendida entre 0.2 – 0.4 indica que el agua está contaminada por una carga orgánica biodegradable.
- c. Cuando la relación $\frac{DBO}{DQO}$ es superior 0.4 indica que el agua está contaminada por una carga orgánica muy biodegradable.

4.) Cuadro de coeficientes de correlación

-1,00	Correlacion negativa perfecta. (" A mayor X, menor Y", de manera proporcional Es decir, cada vez que X aumenta una unidad, Y disminuye siempre una cantidad constante). Esto tambien se aplica "a menor X, mayor Y".
- 0,90 a - 0,99	Correlacion negativa muy alta
-0,70 a - 0,89	Correlacion negativa alta
- 0,40 a - 0,69	Correlacion negativa moderada
- 0,20 a - 0,39	Correlacion negativa baja
- 0,01 a - 0,19	Correlacion negativa muy baja
0	Correlacion nula
0,01 a 0,19	Correlacion positiva muy baja
0,20 a 0,39	Correlacion positiva baja
0,40 a 0,69	Correlacion positiva moderada
0,70 a 0,89	Correlacion positiva alta
0,90 a 0,99	Correlacion positiva muy alta
1,00	Correlacion positiva perfecta. " A mayor X, mayor Y" o a "menorX, menor Y" de manera proporcional. Cada vez que X aumenta, Y aumenta siempre una cantidad constante. Igual cuando X disminuye).

Fuente: Hernandez-Sampieri, R. y Mendoza (2018).