



UNIVERSIDAD POLITÉCNICA SALESIANA
SEDE CUENCA
CARRERA DE INGENIERÍA AMBIENTAL

ESTIMACIÓN DE LA CALIDAD DEL AGUA MEDIANTE TÉCNICAS DE
MODELIZACIÓN CON ÁRBOLES DE REGRESIÓN Y EL ÍNDICE DE LA FUNDACIÓN
NACIONAL DE SANEAMIENTO PARA LA CUENCA HIDROGRÁFICA DEL RÍO
YANUNCAY

Trabajo de titulación previo a la obtención del
título de Ingeniera(o) Ambiental

AUTORES: CRISTINA MARIBEL CÁRDENAS PATIÑO

BRYAM ALEXANDER CRESPO CUZCO

TUTORA: ING. PAOLA JACKELINE DUQUE SARANGO, MSc.

COTUTOR: ING. DIEGO ROMÁN CABRERA MENDIETA, PhD.

Cuenca - Ecuador

2024

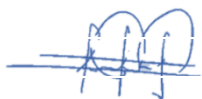
CERTIFICADO DE RESPONSABILIDAD Y AUTORÍA DEL TRABAJO DE TITULACIÓN

Nosotros, Cristina Maribel Cárdenas Patiño con documento de identificación N° 0302994249 y Bryam Alexander Crespo Cuzco con documento de identificación N° 0302708177; manifestamos que:

Somos los autores y responsables del presente trabajo; y, autorizamos a que sin fines de lucro la Universidad Politécnica Salesiana pueda usar, difundir, reproducir o publicar de manera total o parcial el presente trabajo de titulación.

Cuenca, 04 de marzo del 2024

Atentamente,



Cristina Maribel Cárdenas Patiño

0302994249



Bryam Alexander Crespo Cuzco

0302708177

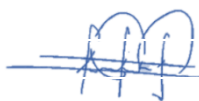
CERTIFICADO DE CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO DE TITULACIÓN A LA UNIVERSIDAD POLITÉCNICA SALESIANA

Nosotros, Cristina Maribel Cárdenas Patiño con documento de identificación N° 0302994249 y Bryam Alexander Crespo Cuzco con documento de identificación N° 0302708177, expresamos nuestra voluntad y por medio del presente documento cedemos a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que somos autores del Trabajo experimental: “Estimación de la calidad del agua mediante técnicas de modelización con árboles de regresión y el Índice de la Fundación Nacional de Saneamiento para la cuenca hidrográfica del Río Yanuncay”, el cual ha sido desarrollado para optar por el título de: Ingeniera(o) Ambiental, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En concordancia con lo manifestado, suscribimos este documento en el momento que hacemos la entrega del trabajo final en formato digital a la Biblioteca de la Universidad Politécnica Salesiana.

Cuenca, 04 de marzo del 2024

Atentamente,



Cristina Maribel Cárdenas Patiño

0302994249



Bryam Alexander Crespo Cuzco

0302708177

CERTIFICADO DE DIRECCIÓN DEL TRABAJO DE TITULACIÓN

Yo, Paola Jackeline Duque Sarango con documento de identificación N° 1104257835, docente de la Universidad Politécnica Salesiana, declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: ESTIMACIÓN DE LA CALIDAD DEL AGUA MEDIANTE TÉCNICAS DE MODELIZACIÓN CON ÁRBOLES DE REGRESIÓN Y EL ÍNDICE DE LA FUNDACIÓN NACIONAL DE SANEAMIENTO PARA LA CUENCA HIDROGRÁFICA DEL RÍO YANUNCAY, realizado por Cristina Maribel Cárdenas Patiño con documento de identificación N° 0302994249 y por Bryam Alexander Crespo Cuzco con documento de identificación N° 0302708177, obteniendo como resultado final el trabajo de titulación bajo la opción Trabajo experimental que cumple con todos los requisitos determinados por la Universidad Politécnica Salesiana.

Cuenca, 04 de marzo del 2024

Atentamente,



Ing. Paola Jackeline Duque Sarango, MSc.

1104257835

DEDICATORIA

Dedico el resultado de este trabajo de titulación primeramente a Dios por guiarme por el camino correcto y por permitirme llegar a donde estoy, asimismo dedico a toda mi familia, especialmente a mis queridos padres. Su apoyo ha sido fundamental en el desarrollo y aprendizaje de esta gran experiencia universitaria. Gracias a sus enseñanzas y consejos, puedo enfrentar las dificultades que se me presenten. Aprecio y valoro el soporte que me brindaron y me permitieron haber vivido esa maravillosa experiencia no solo en el ámbito académico, sino también en el deportivo. También quiero dedicar este trabajo de titulación a mis hermanos, quienes siempre me han brindado cuidado y han sido un refugio de calor humano para mí, permitiéndome nunca sentirme sola y seguir luchando por mis sueños anhelados.

Cristina Cárdenas

Dedico el resultado de este trabajo de titulación a mis adorados padres y hermano. Su apoyo incondicional ha sido el pilar fundamental de mi desarrollo académico y personal. Desde el momento en que decidí embarcarme en este desafío, ustedes han estado a mi lado, alentándome en cada paso del camino. Han sido mi roca, brindándome palabras de aliento cuando me sentía abrumado y compartiendo mi alegría en cada logro alcanzado. Su constante confianza en mí ha sido mi mayor motivación y ha alimentado mi determinación para seguir adelante.

Bryam Crespo

AGRADECIMIENTO

Queremos expresar nuestro más sincero agradecimiento a los distinguidos miembros de la Empresa ETAPA E. P., quienes generosamente brindaron los datos necesarios para llevar a cabo esta investigación. Su contribución ha sido invaluable para la calidad y relevancia de este trabajo. Asimismo, queremos extender nuestra gratitud a nuestros respetados tutores, cuya dedicación, paciencia y sabias orientaciones han sido fundamentales en esta etapa académica. Su guía experta ha sido fundamental para nuestro crecimiento y desarrollo profesional. Por último, pero no menos importante, deseamos expresar nuestro más profundo agradecimiento a todas las personas que participaron en esta investigación. Su disposición y contribución han sido invaluable para la calidad y relevancia de este trabajo.

A todos ustedes, les agradecemos de corazón por su apoyo, confianza y por haber sido parte esencial en este importante capítulo de nuestra vida académica.

Cristina Cárdenas y Bryam Crespo

Índice de Contenido

1	Introducción	1
1.1	Antecedentes	1
1.2	Problema	2
1.3	Delimitación	3
1.3.1	Características morfométricas	5
1.4	Pregunta de investigación	7
1.5	Objetivos	7
1.5.1	Objetivo general	7
1.5.2	Objetivos específicos	8
1.6	Hipótesis	8
2	Fundamentación teórica	9
2.1	Cuenca hidrográfica	9
2.2	Campañas de monitoreo de calidad del agua	11
2.3	Carga contaminante	12
2.4	Calidad del agua	14
2.5	Índice de Calidad del Agua	15
2.6	Índice de la Fundación Nacional de Saneamiento (ICA NSF)	16
2.7	Cálculo del Índice de la Fundación Nacional de Saneamiento	25
2.8	Software ArcMap	28
2.8.1	Modelo digital del terreno (MDT)	28
2.8.2	Delimitación de una cuenca hidrográfica utilizando un MDT	28
2.8.3	Interpolación por el método Kriging	29
2.9	Lenguaje de programación Python	30
2.10	Aprendizaje automático	32

2.11 Creación de modelos basados en datos	33
2.12 Árboles de decisión	34
2.12.1 Árboles de clasificación	34
2.12.2 Árboles de regresión	35
2.13 Bosque aleatorio.....	37
2.13.1 Bosque aleatorio para clasificación.....	37
2.13.2 Bosque aleatorio por regresión.....	39
2.14 Prueba de hipótesis (Prueba t).....	41
2.15 Diagrama de flujo de datos	42
3 Materiales y Métodos	45
3.1 Recopilación de datos correspondientes a la cuenca hidrográfica del Río Yanuncay ...	45
3.2 Cálculo del Índice de la Fundación Nacional de Saneamiento (ICA NSF)	46
3.2.1 Determinación de los factores de escala Q_i	46
3.2.2 Rangos de clasificación para el ICA NSF.....	55
3.3 Diseño de los modelos basados en datos.....	55
3.3.1 Creación del conjunto de datos	56
3.3.2 Definición del modelo base.....	57
3.3.3 Creación y modelado de subconjuntos de parámetros	59
3.3.4 Determinación de la similitud con el modelo base	61
4 Resultados	62
4.1 Calidad del agua obtenida en cada estación de monitoreo correspondientes a la cuenca del Río Yanuncay.....	62
4.2 Deterioro del ICA NSF en la cuenca del Río Yanuncay.....	86
4.3 Interpolación de la calidad del agua por el método Kriging	88
4.4 Rendimiento de los modelos base por Árboles de regresión	89

4.5 Modelos de Bosque aleatorio por regresión entrenados con 3 parámetros	91
4.6 Modelos de Bosque aleatorio por regresión entrenados con 4 parámetros	94
4.7 Elección del modelo más apropiado con la capacidad similar al ICA NSF.....	100
4.7.1 Modelo de Bosque aleatorio entrenado con 3 parámetros	101
4.7.2 Modelo de Bosque aleatorio entrenado con 4 parámetros	102
5 Conclusiones	103
6 Recomendaciones.....	104
7 Bibliografía	105
8 Anexos.....	113
Anexo 1: Información general de las estaciones de monitoreo de la cuenca del Río Yanuncay.	113
Anexo 2: Gráfico del factor de escala Q correspondiente al Oxígeno Disuelto.	114
Anexo 3: Gráfico del factor de escala Q correspondiente a Coliformes Fecales.....	114
Anexo 4: Gráfico del factor de escala Q correspondiente al pH.....	115
Anexo 5: Gráfico del factor de escala Q correspondiente a la DBO ₅	115
Anexo 6: Gráfico del factor de escala Q correspondiente a los Nitratos.	116
Anexo 7: Gráfico del factor de escala Q correspondiente a los Fosfatos.....	116
Anexo 8: Gráfico del factor de escala Q correspondiente a la Temperatura.	117
Anexo 9: Gráfico del factor de escala Q correspondiente a la Turbiedad.	117
Anexo 10: Gráfico del factor de escala Q correspondiente a los Sólidos Totales.	118
Anexo 11: Gráfico de la calidad del agua en cada estación de monitoreo.....	118
Anexo 12: Resumen del promedio y desviación estándar de la calidad del agua en cada estación de monitoreo.	119
Anexo 13: Calidad del agua en función del uso del suelo.	120
Anexo 14: Resultados de modelos con un rendimiento mayor a 0.90.	121

LISTADO DE FIGURAS

Figura 1: Cuenca hidrográfica del Río Yanuncay.	4
Figura 2: Uso del suelo en la cuenca del Río Yanuncay.	6
Figura 3: Función de Calidad NSF (Oxígeno Disuelto).	17
Figura 4: Función de Calidad NSF (Coliformes Fecales).....	18
Figura 5: Función de Calidad NSF (Potencial de Hidrógeno).....	19
Figura 6: Función de Calidad NSF (Demanda Bioquímica de Oxígeno).	20
Figura 7: Función de Calidad NSF (Nitratos).....	21
Figura 8: Función de Calidad NSF (Fosfatos).	22
Figura 9: Función de Calidad NSF (Temperatura).	23
Figura 10: Función de Calidad NSF (Turbiedad).	24
Figura 11: Función de Calidad NSF (Sólidos Totales).....	25
Figura 12: Estructura de un Árbol de clasificación.	35
Figura 13: Estructura de un Árbol de regresión.....	37
Figura 14: Estructura de un Bosque aleatorio por clasificación.	39
Figura 15: Estructura de un Bosque aleatorio por regresión.	41
Figura 16: Diagrama de delimitación de una cuenca hidrográfica.	45
Figura 17: Conjunto de datos para el cálculo del ICA NSF.	46
Figura 18: Diagrama de flujo para el cálculo del factor de escala Q_{OD}	47
Figura 19: Diagrama de flujo para el cálculo del factor de escala Q_{CF}	48
Figura 20: Diagrama de flujo para el cálculo del factor de escala Q_{pH}	49
Figura 21: Diagrama de flujo para el cálculo del factor de escala Q_{DBO5}	50
Figura 22: Diagrama de flujo para el cálculo del factor de escala Q_N	51
Figura 23: Diagrama de flujo para el cálculo del factor de escala Q_p	52
Figura 24: Diagrama de flujo para el cálculo del factor de escala Q_{CT}	53
Figura 25: Diagrama de flujo para el cálculo del factor de escala Q_T	54
Figura 26: Diagrama de flujo para el cálculo del factor de escala Q_{ST}	55
Figura 27: Etapas para la creación de modelos basados en árboles de regresión.	56
Figura 28: Ejemplificación del conjunto de datos.	57
Figura 29: Ejemplificación de la elección del modelo base.	57
Figura 30: Ejemplificación del modelado de subconjuntos de parámetros.	60

Figura 31: Red monitoreada en la cuenca del Río Yanuncay.....	62
Figura 32: Comportamiento del ICA NSF en la cuenca del Río Yanuncay.	87
Figura 33: Interpolación del ICA NSF.	88

LISTADO DE TABLAS

Tabla 1: Ubicación geográfica de la cuenca del Río Yanuncay.....	5
Tabla 2: Características generales de la cuenca del Río Yanuncay.	5
Tabla 3: Valoración del uso del suelo en la cuenca del Río Yanuncay.	6
Tabla 4: Solubilidad del Oxígeno en Agua Dulce.	17
Tabla 5: Pesos relativos para cada parámetro del ICA.	26
Tabla 6: Clasificación del ICA propuesto por Brown.....	26
Tabla 7: Simbología para un diagrama de flujo de datos.....	43
Tabla 8: Rendimiento de los modelos base.....	90
Tabla 9: Rendimiento de los modelos entrenados con 3 parámetros.....	93
Tabla 10: Rendimiento de los modelos entrenados con 4 parámetros.....	96
Tabla 11: Calificación cualitativa para los 9 parámetros requeridos por ICA NSF.	101

LISTADO DE GRÁFICOS

Gráfico 1: ICA NSF en la estación E-1.....	63
Gráfico 2: ICA NSF en la estación E-2.....	64
Gráfico 3: ICA NSF en la estación E-3.....	65
Gráfico 4: ICA NSF en la estación E-4.....	66
Gráfico 5: ICA NSF en la estación E-5.....	67
Gráfico 6: ICA NSF en la estación E-6.....	68
Gráfico 7: ICA NSF en la estación E-7.....	69
Gráfico 8: ICA NSF en la estación E-8.....	70
Gráfico 9: ICA NSF en la estación E-9.....	71
Gráfico 10: ICA NSF en la estación E-10.....	72
Gráfico 11: ICA NSF en la estación E-11.....	73
Gráfico 12: ICA NSF en la estación E-12.....	74

Gráfico 13: ICA NSF en la estación E-13.....	75
Gráfico 14: ICA NSF en la estación E-14.....	76
Gráfico 15: ICA NSF en la estación E-15.....	77
Gráfico 16: ICA NSF en la estación E-16.....	78
Gráfico 17: ICA NSF en la estación E-17.....	79
Gráfico 18: ICA NSF en la estación E-18.....	80
Gráfico 19: ICA NSF en la estación E-19.....	81
Gráfico 20: ICA NSF en la estación E-20.....	82
Gráfico 21: ICA NSF en la estación E-21.....	83
Gráfico 22: ICA NSF en la estación E-22.....	84
Gráfico 23: ICA NSF en la estación E-23.....	85
Gráfico 24: Comportamiento del ICA NSF en función de la variable altitud.	86

LISTADO DE ECUACIONES

Ecuación 1: Ecuación para el cálculo del ICA NSF.	25
Ecuación 2: Ecuación para el cálculo del factor de escala Q_{OD}	47
Ecuación 3: Ecuación para el cálculo del factor de escala Q_{CF}	47
Ecuación 4: Ecuación para el cálculo del factor de escala Q_{pH}	48
Ecuación 5: Ecuación para el cálculo del factor de escala Q_{pH}	48
Ecuación 6: Ecuación para el cálculo del factor de escala Q_{DBO5}	49
Ecuación 7: Ecuación para el cálculo del factor de escala Q_N	50
Ecuación 8: Ecuación para el cálculo del factor de escala Q_P	51
Ecuación 9: Ecuación para el cálculo del factor de escala Q_{CT}	52
Ecuación 10: Ecuación para el cálculo del factor de escala Q_T	53
Ecuación 11: Ecuación para el cálculo del factor de escala Q_{ST}	54

ENLACE DEL PROYECTO

<https://github.com/BryamCrespo/Trabajo-de-Titulaci-n.git>

1 Introducción

1.1 Antecedentes

La importancia crucial del agua se extiende tanto a los ecosistemas acuáticos como a la salud humana. A lo largo de la historia, se ha reconocido la necesidad de mantener un suministro de agua limpio y seguro para salvaguardar el bienestar de las comunidades y preservar la diversidad biológica. No obstante, las actividades humanas, como la industrialización, la agricultura intensiva y la urbanización, han incrementado la presión sobre los recursos hídricos a nivel mundial, resultando en la contaminación y deterioro de la calidad del agua. La presencia de contaminantes, que incluyen sustancias químicas, patógenos y nutrientes en exceso, puede ocasionar efectos perjudiciales en los ecosistemas acuáticos y constituir un riesgo para la salud pública.

La supervisión y la valoración de la calidad del agua se han convertido en una prioridad a nivel mundial para asegurar la sostenibilidad y una gestión apropiada de este recurso esencial. Se emplea el análisis de diversos parámetros físicos, químicos y biológicos con el fin de comprender y prevenir la contaminación, así como implementar medidas para proteger y restaurar los cuerpos de agua. En este contexto, los avances en tecnología y ciencia han desempeñado un papel esencial en el desarrollo de metodologías y herramientas que posibilitan una evaluación precisa y eficiente de la calidad del agua, facilitando de esta manera la toma de decisiones informadas para su conservación y uso sostenible.

Según el estudio realizado por Koranga et al. (2022) titulado “Modelos eficientes de predicción de calidad del agua basados en algoritmos de aprendizaje automático para el lago Nainital, Uttarakhand”, se concluyó que el algoritmo Random Forest demostró ser la opción más eficaz en términos de aprendizaje automático al utilizar tres parámetros fisicoquímicos clave, como el pH, la turbidez y los sólidos disueltos totales, para prever la calidad del agua en el lago Nainital.

Además, es relevante destacar la novedosa aproximación desarrollada por Cañar Uyaguari (2023), que consistió en la utilización de bibliotecas de visión por computadora, como OpenCV, y la implementación de una red neuronal secuencial. Estas técnicas permitieron acelerar la evaluación de las condiciones actuales del agua en el Río Sinincay utilizando macroinvertebrados acuáticos como indicadores confiables de la calidad del agua. Este enfoque logró ofrecer

información instantánea y eficaz, al mismo tiempo que se utilizaron métodos no perjudiciales para el ecosistema.

Las estaciones de monitoreo de calidad del agua establecidas por ETAPA E. P. en diversas áreas de la cuenca del Río Yanuncay han sido una herramienta esencial para evaluar y supervisar de manera cercana la salud de los recursos hídricos en esa región. Estratégicamente ubicadas a lo largo del río y sus afluentes, estas estaciones han facilitado la recopilación de datos hidrológicos y diversos parámetros físicos, químicos y microbiológicos. Gracias a este monitoreo constante y exhaustivo, se ha construido una base de datos completa y actualizada que proporciona información valiosa sobre la calidad del agua en cada zona de la cuenca del Río Yanuncay.

Estos datos han sido empleados en diversos estudios e investigaciones para examinar la evolución de la calidad del agua a lo largo del tiempo, identificar posibles fuentes de contaminación y evaluar el impacto de las actividades humanas en el ecosistema acuático. Además, estas estaciones de monitoreo han permitido la detección de cambios y tendencias en la calidad del agua, siendo crucial para la toma de decisiones fundamentadas y la implementación de medidas apropiadas de protección y conservación.

1.2 Problema

La calidad del agua en la cuenca del Río Yanuncay se ve amenazada por diversos factores que la afectan negativamente. La contaminación proveniente de actividades industriales y domésticas, así como el uso inadecuado de la tierra y la disposición inapropiada de desechos sólidos, son elementos clave que contribuyen a la degradación de la calidad del agua en esta área. Estos problemas se ven agravados por la variabilidad geográfica, el crecimiento urbano y los cambios en los patrones climáticos (Gao et al., 2021).

Por esta razón, resulta esencial llevar a cabo un monitoreo constante y una evaluación de la calidad del agua en la cuenca del Río Yanuncay. Esto se debe a que el agua representa un recurso vital tanto para los ecosistemas acuáticos como para las comunidades humanas que dependen de ella. Además, la calidad del agua sirve como un indicador significativo de la salud ambiental general de la cuenca. Este enfoque proporciona una base sólida para la toma de decisiones informadas en términos de gestión y conservación del agua.

Las herramientas tradicionales para evaluar la calidad del agua, como los Índices de Calidad del Agua (ICAs), han sido utilizadas convencionalmente y abarcan una amplia variedad de parámetros físicos, químicos y biológicos. Aunque estos índices ofrecen una visión completa de la calidad del agua, su aplicación puede ser costosa y logísticamente desafiante debido a la necesidad de medir y mantener múltiples parámetros en una cuenca hidrográfica (Boyle et al., 2000).

En contraste, los modelos basados en datos utilizan técnicas de aprendizaje automático y estadísticas para prever la calidad del agua utilizando datos previamente recopilados en la cuenca hidrográfica (Bui et al., 2020). En la actualidad, surge la pregunta sobre si estos modelos basados en datos pueden superar la limitación inherente de los ICAs al considerar relaciones no lineales y complejas entre los parámetros de calidad del agua (Rahat et al., 2023).

Para abordar este problema, en la presente investigación se reconoció la necesidad de mejorar la evaluación de la calidad del agua en la cuenca del Río Yanuncay mediante un enfoque innovador que implica un análisis comparativo entre modelos diseñados para estimar el Índice de la Fundación Nacional de Saneamiento (ICA NSF), utilizando datos específicos de parámetros físicos, químicos y biológicos del Río Yanuncay.

La propuesta de reducir los parámetros utilizados por el ICA NSF y desarrollar modelos se justifica por dos razones principales:

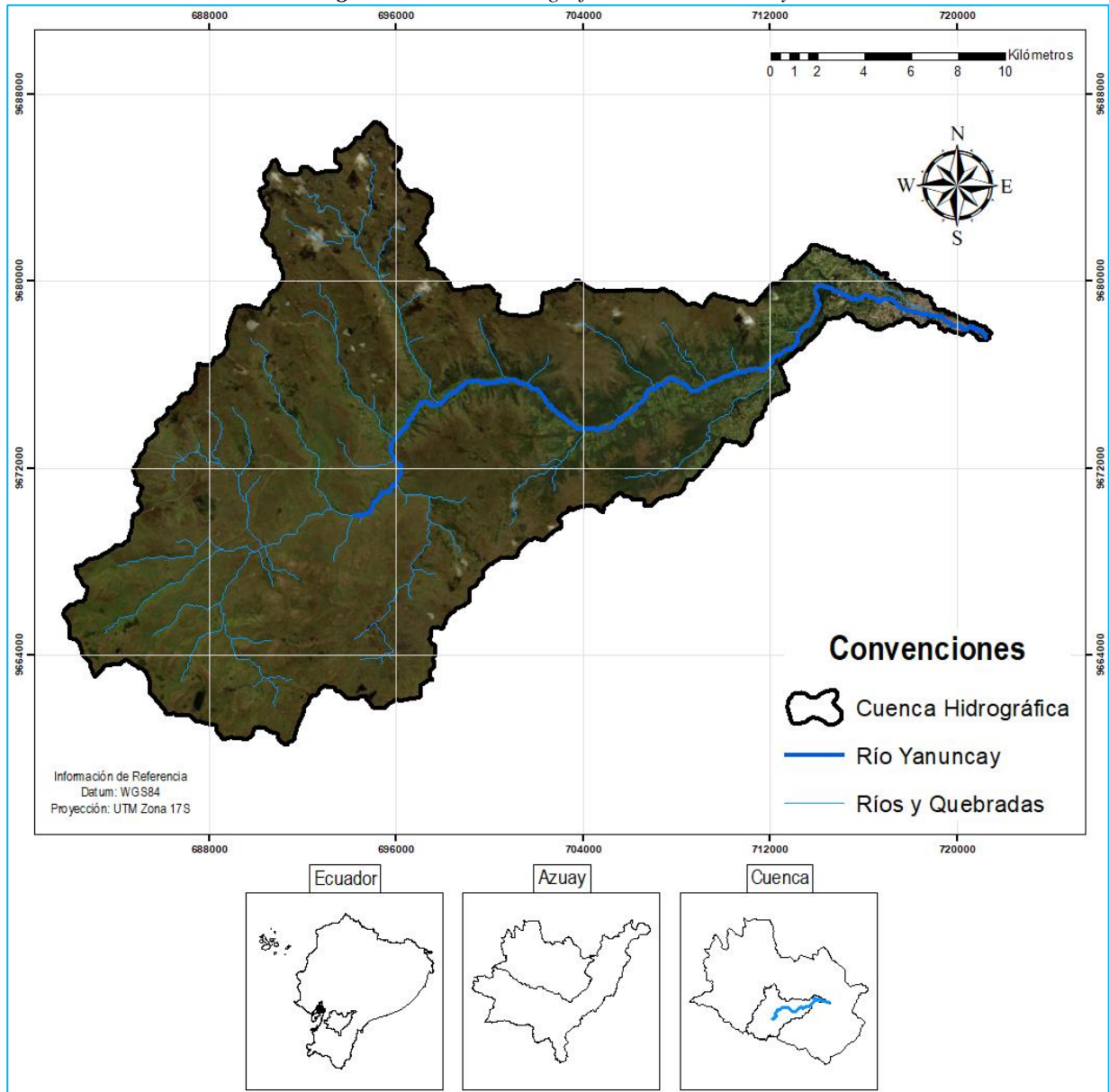
- En primer lugar, se parte de la premisa de que un conjunto más reducido, pero aún representativo, de parámetros clave del ICA, vinculado a las características específicas de la cuenca, podría ofrecer una evaluación más precisa y contextualizada de la calidad del agua en comparación con los ICAs actuales (Li et al., 2012; Fu et al., 2020).
- Además, esta herramienta podría simplificar el proceso de evaluación de la calidad del agua, lo que facilitaría una implementación más ágil de medidas de gestión y mitigación de la contaminación, posibilitando la toma de decisiones informadas para la conservación y la gestión sostenible de los recursos hídricos (Bekele et al., 2018).

1.3 Delimitación

La cuenca del Río Yanuncay abarca una extensión que comprende las localidades de Cuenca y Girón, junto con varias comunidades y áreas rurales circundantes. El Río Yanuncay desempeña un

papel crucial como fuente de agua para la provincia de Azuay y la ciudad de Cuenca, sirviendo tanto para el abastecimiento de agua potable, como para la agricultura y la generación de energía hidroeléctrica. Además, el río constituye un elemento significativo del sistema hídrico regional y contribuye al caudal del Río Tomebamba (ver Figura 1).

Figura 1: Cuenca hidrográfica del Río Yanuncay.



Fuente: Elaboración propia.

La subcuenca del Río Yanuncay está ubicada en la región montañosa de Ecuador y forma parte integral de la cuenca del Río Paute. Al norte, limita con la cuenca del Río Cañar y la subcuenca del Río Tomebamba; al sur, colinda con la cuenca del Río Jubones y la subcuenca del Río Tarqui;

al este, comparte límites con la subcuenca del Río Tomebamba y Tarqui; y al oeste, se encuentra delimitada por la cuenca del Río Balao. En su totalidad, esta subcuenca está situada en la provincia de Azuay, específicamente en el cantón Cuenca, y sus coordenadas geográficas se detallan en la siguiente tabla (ver Tabla 1).

Tabla 1: Ubicación geográfica de la cuenca del Río Yanuncay.

Coordenada X del centroide	Coordenada Y del centroide
696687	9672900

Fuente: Elaboración propia.

1.3.1 Características morfométricas

- **Parámetros generales**

La cuenca del Río Yanuncay, ubicada en la provincia de Azuay en la región sur de Ecuador, cubre un área de 413.87 km² y tiene un perímetro de 135.61 km. Su topografía montañosa y diversa se despliega a lo largo de estrechos valles y quebradas que definen su curso. El Río Yanuncay, con una extensión de 39.44 km, se origina en las elevadas montañas de la Cordillera Occidental y fluye hacia el este, atravesando una serie de paisajes pintorescos y variados.

Tabla 2: Características generales de la cuenca del Río Yanuncay.

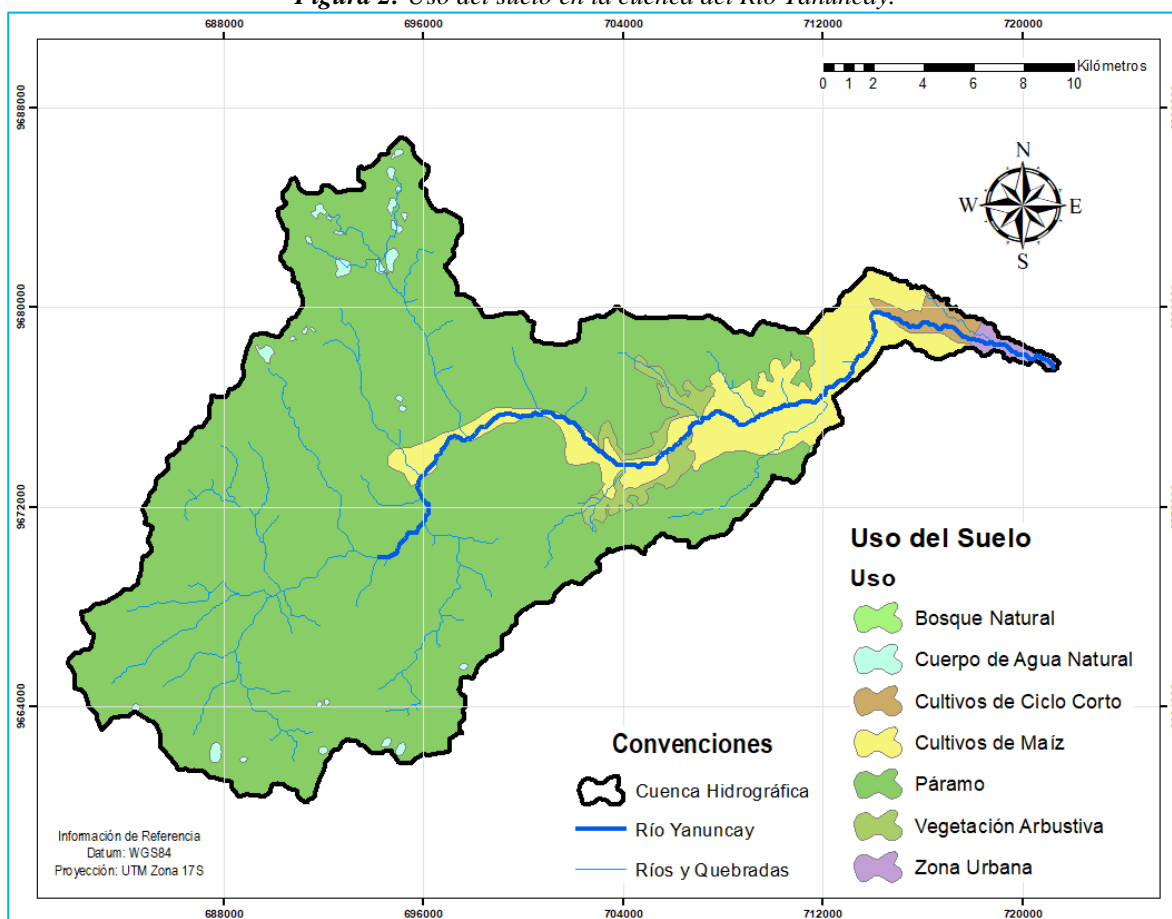
Área (km²)	413.87
Perímetro (km)	135.61
Longitud (km)	39.44

Fuente: Elaboración propia.

- **Uso del suelo**

En las zonas altas y medias de la cuenca del Río Yanuncay, el uso del suelo está predominantemente caracterizado por el ecosistema de páramo, que alberga cuerpos de agua naturales, bosques autóctonos y una vegetación arbustiva distintiva. Además, se identifica una pequeña extensión dedicada al cultivo de maíz. En contraste, en la parte baja de la cuenca, se observan predominantemente cultivos de ciclo corto y maíz, así como áreas urbanizadas, tal como se muestra en la Figura 2.

Figura 2: Uso del suelo en la cuenca del Río Yanuncay.



Fuente: Elaboración propia.

Tabla 3: Valoración del uso del suelo en la cuenca del Río Yanuncay.

Uso	Área (km ²)	Porcentaje (%)
Bosque Natural	0.13	0.03
Cuerpo de Agua Natural	3.17	0.77
Cultivos de Ciclo Corto	4.29	1.04
Cultivos de Maíz	41.28	9.97
Páramo	354.28	85.58
Vegetación Arbustiva	8.14	1.97
Zona Urbana	2.57	0.62

Fuente: Elaboración propia.

La variabilidad en la utilización del suelo a diferentes altitudes en la cuenca refleja la interacción entre la actividad agrícola, la preservación de los ecosistemas naturales y el desarrollo urbano en la región.

- **Páramo:** En las zonas elevadas de la cuenca, se localiza el ecosistema de páramo, caracterizado por su vegetación herbácea, arbustiva y musgosa. Este ecosistema juega un papel esencial en la regulación del flujo de agua, actuando como una especie de esponja natural que almacena y libera gradualmente el agua a lo largo del año.
- **Bosques naturales:** En ciertas áreas de la cuenca, se preservan bosques naturales que albergan una amplia diversidad de especies vegetales y animales. Estos bosques contribuyen a regular los ciclos hidrológicos y a conservar la biodiversidad.
- **Cultivos agrícolas:** En las zonas de altitudes intermedias y bajas de la cuenca, se encuentran áreas destinadas a la actividad agrícola. Los cultivos más comunes comprenden maíz, papa, hortalizas y frutas. La agricultura desempeña un papel económico crucial en la región, y el uso del suelo agrícola se orienta tanto a la subsistencia como a la producción comercial.
- **Zonas urbanas:** En las áreas más bajas de la cuenca, se encuentran asentamientos urbanos como la ciudad de Cuenca y otras localidades. Estas zonas urbanas están destinadas a la vivienda, el comercio, la infraestructura y los servicios públicos.

1.4 Pregunta de investigación

La cuenca del Río Yanuncay ha sido objeto de un monitoreo continuo, generando una valiosa base de datos que contiene información sobre diversos parámetros relacionados con la calidad del agua. Al utilizar estos datos históricos, ¿es posible entrenar modelos basados en Árboles de regresión, que sean capaces de capturar las complejas relaciones y patrones existentes de un reducido número de variables que influyen en la calidad del agua?

1.5 Objetivos

1.5.1 Objetivo general

- ✓ Comparar modelos basados en datos desarrollados con un conjunto reducido de parámetros en relación con el Índice de la Fundación Nacional de Saneamiento (ICA NSF), para la estimación de la calidad del agua en la cuenca hidrográfica del Río Yanuncay.

1.5.2 Objetivos específicos

- ✓ Recopilar, con ayuda de software, una data de información hidrológica, meteorológica y de parámetros físicos, químicos y microbiológicos de la cuenca hidrográfica del Río Yanuncay.
- ✓ Calcular el ICA NSF en la cuenca hidrográfica del Río Yanuncay, por medio de las ecuaciones obtenidas en los ajustes de curvas aplicadas en un lenguaje de programación para la obtención de los factores de escala Q_i .
- ✓ Desarrollar modelos basados en Árboles de calidad del agua con un número reducido de parámetros que los requeridos por el ICA NSF, mediante técnicas de análisis de datos y aprendizaje automático para la optimización de los requerimientos de monitoreo.
- ✓ Comparar los resultados obtenidos entre los modelos de predicción desarrollados para la estimación del ICA NSF, a través de pruebas de hipótesis para la evaluación de su precisión en concordancia y las diferencias en la detección de cambios en la calidad del agua.

1.6 Hipótesis

Los modelos basados en Árboles de regresión entrenados con datos históricos de la cuenca del Río Yanuncay pueden capturar de manera consistente las complejas relaciones y patrones de un número reducido de variables con una alternativa ICA NSF para la predicción de la calidad del agua con menos recursos.

2 Fundamentación teórica

2.1 Cuenca hidrográfica

La cuenca hidrográfica representa una unidad geográfica fundamental en la gestión integral de los recursos hídricos, donde todas las aguas de precipitación convergen hacia un único punto de salida, ya sea un río principal o un lago (Wani & Garg, 2009). Esta región, delimitada por las divisorias de aguas, forma una intrincada red de ríos, arroyos y afluentes, creando un sistema hidrológico interconectado. En resumen, la cuenca hidrográfica constituye un ecosistema complejo que abarca una amplia gama de procesos hidrológicos, biogeoquímicos y ecológicos (Truchy et al., 2019).

Al evaluar la salud de una cuenca hidrográfica, es esencial considerar diversos aspectos, como la calidad del agua, la cantidad de precipitación, la capacidad de infiltración del suelo y la interacción entre la vegetación y el sustrato. Estos elementos se entrelazan en un entramado complejo de procesos que afectan la funcionalidad y resiliencia de los ecosistemas de la cuenca (Pinto & Maheshwari, 2014).

- La calidad del agua es un factor crítico a considerar, ya que impacta tanto a los organismos acuáticos como a los seres humanos que dependen de ella. Evaluar la calidad del agua implica analizar parámetros como los niveles de contaminantes, la presencia de nutrientes y la salud de los ecosistemas acuáticos.
- La cantidad de precipitación es otro aspecto crucial, ya que determina la disponibilidad de agua en la cuenca. Un análisis detallado de los patrones de precipitación a lo largo del tiempo puede ayudar a comprender la variabilidad hidrológica y a tomar decisiones informadas sobre la gestión de los recursos hídricos.
- La infiltración del suelo es esencial para el ciclo del agua en una cuenca hidrográfica. La capacidad del suelo para absorber y retener el agua afecta la recarga de los acuíferos subterráneos, la disponibilidad de agua para las plantas y la regulación del flujo de agua en los ríos y arroyos.
- Además, la interacción entre la vegetación y el sustrato desempeña un papel crucial en la salud de la cuenca hidrográfica. La vegetación actúa como una barrera natural, contribuyendo a prevenir la erosión del suelo, mejorando la infiltración y filtración del agua, y proporcionando hábitats para una diversidad de especies.

Comprender estos procesos y su interconexión es esencial para implementar estrategias de manejo efectivas que conserven la funcionalidad y resiliencia de los ecosistemas de la cuenca hidrográfica. Al tener en cuenta estos aspectos, se pueden tomar medidas adecuadas para proteger y preservar estos valiosos recursos hídricos (Pinto & Maheshwari, 2014).

Río

Dentro de la amplia red de una cuenca hidrográfica, los ríos desempeñan un papel esencial al actuar como las arterias que transportan agua y sedimentos, conectando de manera serpenteante y dinámica diversos paisajes y ecosistemas a lo largo de su curso (Golet et al., 2006). La importancia de los ríos va más allá de su función de transporte, ya que también desempeñan un papel fundamental al servir como hábitats críticos para numerosas especies acuáticas, proporcionando refugio, alimento y condiciones propicias para su supervivencia y reproducción. Además, los ríos son verdaderos proveedores de servicios ecosistémicos valiosos, que incluyen la regulación del clima y la recarga de acuíferos, lo que implica que su salud y funcionamiento óptimo son esenciales para el bienestar de los ecosistemas y las comunidades humanas que dependen de ellos (Martínez-Dalmau et al., 2023).

- En su función como corredores de vida, los ríos albergan una diversidad de especies acuáticas que han evolucionado y adaptado a las condiciones específicas de cada segmento fluvial, creando así ecosistemas únicos y altamente especializados. Estos hábitats fluviales no solo sostienen la biodiversidad local, sino que también actúan como puntos de conexión entre diferentes ecosistemas terrestres y acuáticos, fomentando la dispersión de especies y facilitando la migración de animales en busca de recursos o durante sus ciclos reproductivos.
- Además de su importancia biológica, los ríos cumplen un papel crucial en la prestación de servicios ecosistémicos que benefician directamente a las sociedades humanas. Por ejemplo, su capacidad para regular el clima local y regional contribuye a moderar las temperaturas, reducir la intensidad de las inundaciones y mitigar los efectos del cambio climático. Asimismo, los ríos actúan como reservorios naturales de agua dulce, recargando los acuíferos subterráneos y garantizando el suministro de agua potable para comunidades cercanas.

2.2 Campañas de monitoreo de calidad del agua

Las campañas de monitoreo de calidad del agua constituyen estrategias planificadas y sistemáticas cuyo propósito es recopilar datos exhaustivos y representativos sobre las condiciones físicas, químicas y biológicas del agua en una región específica. Estas iniciativas desempeñan un papel esencial en la evaluación de la salud de los cuerpos de agua, en la identificación de fuentes de contaminación y en el desarrollo de estrategias efectivas para la gestión del recurso hídrico (Vienna & Spearing, 2012).

Durante estas campañas, se aplican métodos rigurosos de muestreo en diversas estaciones de monitoreo ubicadas estratégicamente a lo largo de los cuerpos de agua en una cuenca hidrográfica. La frecuencia y duración de estas campañas se ajustan a las características específicas de la cuenca y a los factores que pueden influir en la calidad del agua. Estos factores pueden abarcar la variabilidad estacional, los eventos climáticos extremos y las actividades humanas presentes en la zona. Es crucial tener en cuenta que la calidad del agua puede verse afectada por cambios estacionales, como las variaciones en los flujos de agua, las temperaturas y los patrones de precipitación. Además, las actividades humanas, como la agricultura, la industria y las prácticas de manejo de residuos, pueden introducir contaminantes al agua y afectar su calidad (Bartram & Ballance, 1996).

Estaciones de monitoreo de calidad del agua

Las estaciones de monitoreo desempeñan un papel esencial en la recopilación sistemática de muestras de agua y en la evaluación de su calidad. Estos puntos geográficos estratégicos cuentan con instrumentos especializados diseñados para medir una amplia gama de parámetros que indican la salud y la condición del agua (Lazorchak et al., 1998).

En estas estaciones de monitoreo, se realiza una minuciosa recolección de datos, utilizando tecnología de vanguardia para medir diversos factores clave como la temperatura del agua, la concentración de oxígeno disuelto, la turbidez y la presencia de contaminantes específicos. Estos parámetros ofrecen información valiosa sobre la calidad física, química y biológica del agua, permitiendo evaluar su aptitud para diversos usos y su estado general de salud.

La ubicación estratégica de las estaciones de monitoreo es una consideración fundamental. Se seleccionan cuidadosamente para representar una variedad de condiciones ambientales y posibles

fuentes de contaminación presentes en la cuenca hidrográfica. Esto asegura que los datos recopilados sean representativos de la calidad del agua en diferentes áreas y situaciones. Además, las estaciones de monitoreo pueden situarse aguas arriba y aguas abajo de actividades humanas significativas, como áreas urbanas, industrias o explotaciones agrícolas intensivas, permitiendo evaluar el impacto de estas actividades en la calidad del agua y su posible contaminación (Bergamasco et al., 2021).

La implementación de un diseño cuidadoso de las estaciones de monitoreo garantiza una cobertura adecuada de la cuenca hidrográfica y proporciona una visión más completa de la calidad del agua en la región. Además, permite identificar patrones espaciales y temporales, así como tendencias a lo largo del tiempo, en los parámetros que se están monitoreando. Esto es esencial para evaluar los cambios en la calidad del agua, identificar posibles fuentes de contaminación y tomar decisiones informadas en la gestión y protección de los recursos hídricos.

2.3 Carga contaminante

La carga contaminante en el agua se refiere a la cantidad de sustancias, partículas y microorganismos presentes en cuerpos de agua, ya sean naturales o artificiales, que tienen el potencial de afectar negativamente su calidad. Estas cargas contaminantes provienen de diversas fuentes, como vertidos industriales, agrícolas, urbanos y atmosféricos. Evaluar y comprender la carga contaminante es esencial para la gestión de la salud de los ecosistemas acuáticos y para garantizar la disponibilidad de agua segura para el consumo humano.

La contaminación del agua puede originarse en múltiples fuentes. Por ejemplo, las descargas industriales a menudo introducen sustancias químicas tóxicas, metales pesados y compuestos orgánicos persistentes en los cuerpos de agua, representando un peligro para su calidad (Freeman, 1995). De manera similar, la actividad agrícola contribuye con nutrientes como el nitrógeno y el fósforo, así como con pesticidas y herbicidas que pueden filtrarse hacia fuentes de agua superficiales y subterráneas (Merrington et al., 2003). Además, las actividades urbanas liberan contaminantes como aceites, sedimentos y productos químicos domésticos en las aguas pluviales y los sistemas de alcantarillado (Khatib et al., 2018).

La carga contaminante puede clasificarse en dos categorías principales: cargas puntuales y cargas no puntuales (EPA, 2003).

- Las cargas puntuales provienen de fuentes específicas y bien definidas, como vertidos industriales o descargas de plantas de tratamiento de aguas residuales.
- En cambio, las cargas no puntuales se originan de manera difusa, como la escorrentía superficial que arrastra contaminantes desde áreas agrícolas o urbanas hacia los cuerpos de agua.

La medición y evaluación de la carga contaminante en el agua son procesos fundamentales para comprender la magnitud y la naturaleza de la contaminación. Se utilizan diversos parámetros para cuantificar la carga contaminante, incluyendo los niveles de oxígeno disuelto, la concentración de nutrientes, la presencia de patógenos y la detección de sustancias químicas específicas mediante el uso de técnicas analíticas avanzadas (EPA, 2017).

Carga máxima permisible

La carga máxima permisible (CMP) hace referencia a la cantidad máxima de contaminantes que un cuerpo de agua puede recibir sin comprometer su integridad ambiental ni tener un impacto negativo en la salud humana. Establecer la CMP implica considerar factores clave como la capacidad de autodepuración del agua, la sensibilidad del ecosistema acuático y los estándares de calidad del agua establecidos por las autoridades competentes (Behmel et al., 2016).

- Determinar la CMP es un proceso complejo y multidisciplinario que requiere una evaluación exhaustiva de varios factores. Un elemento fundamental es comprender la capacidad de autodepuración del cuerpo de agua en cuestión. Cada ecosistema acuático tiene la capacidad de eliminar una cierta cantidad de contaminantes de forma natural a través de procesos biológicos, químicos y físicos. Estos procesos incluyen la degradación biológica de sustancias, la sedimentación de partículas y la dilución de contaminantes mediante la mezcla del agua. Al considerar la capacidad de autodepuración, se busca determinar cuánta carga contaminante puede ser asimilada y eliminada sin comprometer la salud del ecosistema.
- Además, la sensibilidad del ecosistema acuático es un aspecto crucial a tener en cuenta. Algunos cuerpos de agua son más frágiles y susceptibles a los impactos de los contaminantes, mientras que otros pueden ser más resistentes. La evaluación de la sensibilidad implica considerar factores como la diversidad biológica, la estructura del

hábitat, los flujos de agua y la capacidad de recuperación del ecosistema. Estos elementos ayudan a determinar cuál es la carga máxima permisible que el ecosistema puede tolerar sin sufrir daños significativos.

- Por último, es esencial tener en cuenta los estándares de calidad del agua establecidos por las autoridades competentes. Estos estándares definen los límites aceptables para la presencia de contaminantes en el agua, con el objetivo de proteger la salud humana y el medio ambiente. La CMP se establece en función de estos estándares, asegurando que la calidad del agua se mantenga dentro de los límites establecidos.

2.4 Calidad del agua

La evaluación de la calidad del agua es un aspecto crucial tanto para salvaguardar la salud humana como para mantener la armonía de los ecosistemas acuáticos. Esta evaluación se realiza a través de la consideración de diversos parámetros físicos, químicos y biológicos que ofrecen información sobre la pureza y seguridad del agua para su uso previsto. Es fundamental destacar que la calidad del agua puede variar según su origen y puede ser influida tanto por factores naturales como por actividades humanas.

Los parámetros físicos engloban características físicas del agua, como la temperatura, que puede afectar las reacciones químicas y la vida acuática; la turbidez, relacionada con la presencia de partículas suspendidas y que incide en la transparencia del agua; el color, indicativo de sustancias disueltas; y la conductividad eléctrica, vinculada a la capacidad del agua para conducir corriente eléctrica y que puede ofrecer información sobre la presencia de sales y minerales disueltos (Bartram & Ballance, 1996).

En contraste, los parámetros químicos se centran en la concentración de sustancias químicas en el agua. Estos abarcan nutrientes como nitrógeno y fósforo, esenciales para el crecimiento de plantas acuáticas, pero problemáticos en concentraciones elevadas debido a la eutrofización; metales, potencialmente tóxicos para la vida acuática y humana en altas concentraciones; compuestos orgánicos, como pesticidas y productos químicos industriales, con posibles efectos adversos en la salud y el entorno; y contaminantes, una amplia gama de sustancias perjudiciales para organismos acuáticos y la salud humana (Fewtrell & Bartram, 2001).

En lo que respecta a los parámetros biológicos, se refieren a la presencia de microorganismos, algas y organismos vivos en el agua. La detección de estos organismos puede ofrecer información valiosa sobre la calidad del agua y reflejar su estado ecológico. Por ejemplo, la presencia de bacterias coliformes puede indicar contaminación fecal y la posible presencia de patógenos en el agua. Además, la abundancia y diversidad de algas y otros organismos acuáticos pueden servir como indicadores del estado de salud del ecosistema acuático (Warren, 1971).

2.5 Índice de Calidad del Agua

El Índice de Calidad del Agua (ICA) se presenta como una herramienta valiosa para evaluar y resumir la calidad general del agua mediante la integración de diversos parámetros relevantes en un solo valor. El cálculo del ICA se fundamenta en estándares y límites establecidos por organismos gubernamentales y ambientales, brindando un marco objetivo para la evaluación de la calidad del agua (Ward, 2001).

Existen diversas metodologías para calcular el ICA, cada una con sus propias ventajas y enfoques. El Método de Índices Ponderados es una de estas metodologías, donde se asignan pesos a los diferentes parámetros de calidad del agua en función de su importancia relativa. Posteriormente, se calcula un valor ponderado que refleja la calidad general del agua. Al asignar pesos, se puede otorgar mayor énfasis a los parámetros con un impacto más significativo, posibilitando una evaluación más precisa y equilibrada (Bartram & Ballance, 1996).

Otra metodología empleada es el Método de Umbrales de Calidad, que establece límites o umbrales para cada parámetro de calidad del agua basándose en estándares y directrices definidos. Se asigna una puntuación a cada parámetro según el grado de cumplimiento de los límites establecidos y, finalmente, se genera un puntaje total que refleja la calidad general del agua. Este método ofrece una evaluación clara y comparativa de la calidad del agua al establecer estándares objetivos que deben cumplirse (Fewtrell & Bartram, 2001).

Adicionalmente, se recurre al Método Multimétrico, el cual utiliza diversas métricas biológicas, físicas y químicas para evaluar la salud y el estado del ecosistema acuático. Este enfoque considera una amplia gama de factores, como la diversidad y abundancia de especies acuáticas, la presencia de contaminantes químicos y los parámetros físicos del agua. Al integrar estas métricas variadas,

se obtiene una evaluación más completa y holística de la calidad del agua, reflejando así la salud general del ecosistema acuático (Descy & Coste, 1991).

2.6 Índice de la Fundación Nacional de Saneamiento (ICA NSF)

Se trata de un Índice de Calidad del Agua creado en 1970 por la National Sanitation Foundation (NSF) de Estados Unidos, utilizando la técnica de investigación Delphi de la Rand Corporation (Fernández & Solano, 2007). Este ICA es ampliamente utilizado entre los diversos índices de calidad de agua disponibles. Su aplicación se extiende para evaluar las variaciones en la calidad del agua en segmentos específicos de los ríos a lo largo del tiempo. Asimismo, facilita la comparación de la calidad del agua entre distintos tramos del mismo río y permite realizar comparativas con la calidad del agua de diferentes ríos a nivel mundial. Los resultados obtenidos son valiosos para determinar la salud de un tramo específico de dicho río (SNET, 2007).

Dentro de este método, se trabajan con nueve parámetros de respuesta, siendo esencial contar con sus mediciones para el cálculo del Índice de Calidad del Agua (ICA) en una ubicación específica (Fernández & Solano, 2007).

- **Oxígeno Disuelto – OD (% de Saturación)**

El oxígeno disuelto desempeña una función crucial en la vida acuática y se considera un indicador fundamental de la calidad del agua. Hace referencia a la cantidad de oxígeno molecular (O₂) presente en el agua en forma disuelta. La presencia adecuada de oxígeno disuelto es esencial para sostener los ecosistemas acuáticos, ya que respalda la vida de una amplia gama de organismos acuáticos, que incluyen desde peces y crustáceos hasta bacterias y algas. La disolución del oxígeno en el agua está influenciada por varios factores. La temperatura del agua desempeña un papel crucial, ya que el oxígeno tiene una mayor capacidad de disolverse a temperaturas más bajas. Por lo tanto, las aguas más frías tienden a contener mayores niveles de oxígeno disuelto en comparación con las aguas más cálidas. Además, la presión atmosférica también puede influir en la disolución del oxígeno, ya que una presión más baja puede dificultar su disolución. Por último, la presencia de contaminantes en el agua puede reducir la cantidad de oxígeno disuelto, ya que los contaminantes pueden consumir oxígeno durante los procesos de descomposición y degradación, lo que resulta en niveles más bajos de oxígeno disponible para los organismos acuáticos (Bhardwaj et al., 2015).

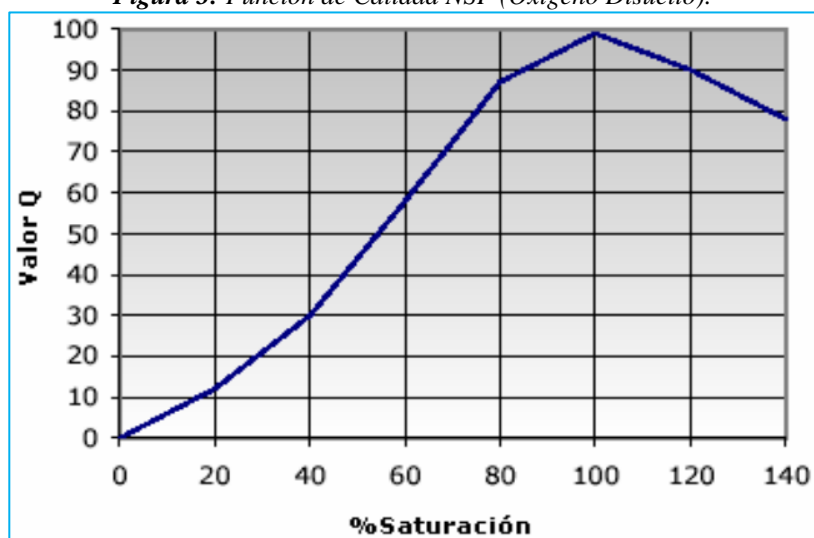
Para el caso del parámetro de Oxígeno Disuelto, el primer paso implica calcular el porcentaje de saturación del OD en el agua. Este cálculo se realiza identificando el valor de saturación del OD de acuerdo con la temperatura del agua (SNET, 2007), tal como se muestra en la Tabla 4.

Tabla 4: Solubilidad del Oxígeno en Agua Dulce.

T (°C)	OD (mg/L)	T (°C)	OD (mg/L)	T (°C)	OD (mg/L)	T (°C)	OD (mg/L)
1	14.19	12	10.76	23	8.56	34	7.05
2	13.81	13	10.52	24	8.40	35	6.93
3	13.44	14	10.29	25	8.24	36	6.82
4	13.09	15	10.07	26	8.09	37	6.71
5	12.75	16	9.85	27	7.95	38	6.61
6	12.43	17	9.65	28	7.81	39	6.51
7	12.12	18	9.45	29	7.67	40	6.41
8	11.83	19	9.26	30	7.54	41	6.31
9	11.55	20	9.07	31	7.41	42	6.22
10	11.27	21	8.90	32	7.28	43	6.13
11	11.01	22	8.72	33	7.16	44	6.04

Fuente: Adaptado de SNET (2007).

Figura 3: Función de Calidad NSF (Oxígeno Disuelto).



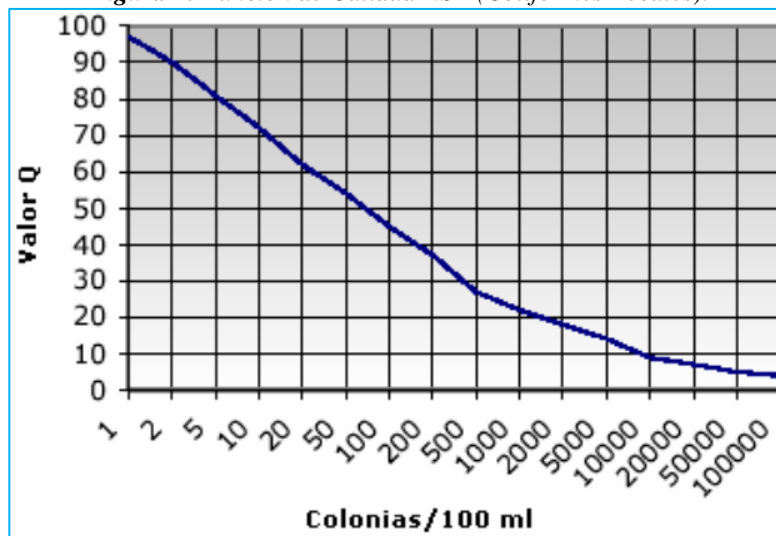
Fuente: Adaptado de Fernández & Solano (2007).

Cuando el porcentaje de Saturación del Oxígeno Disuelto es superior al 140%, se establece que el Q-OD sea igual a 47. En caso de que el valor obtenido sea menor al 140% de Saturación de Oxígeno Disuelto, se busca el valor correspondiente en el eje X de la Figura 3 y se realiza una interpolación para determinar el valor en el eje Y. Este resultado constituirá el Q-OD del Oxígeno Disuelto. Para concluir, se multiplica el valor de Q-OD por el peso W-OD (SNET, 2007).

- **Coliformes Fecales – CF (Colonias/100 mL)**

Los coliformes fecales constituyen un conjunto de bacterias que señalan la posible existencia de contaminación fecal en el agua, sugiriendo la presencia potencial de patógenos perjudiciales para la salud humana. Es relevante destacar que, aunque los coliformes fecales en sí mismos no son patógenos, actúan como indicadores de la eventual presencia de microorganismos fecales que podrían representar un riesgo para la salud. La detección de coliformes fecales en el agua puede originarse por diversas fuentes de contaminación, tales como aguas residuales no tratadas, desbordamientos de sistemas sépticos, infiltraciones de aguas pluviales y la proximidad de animales de granja o vida silvestre a las fuentes de agua (OMS, 2008).

Figura 4: Función de Calidad NSF (Coliformes Fecales).



Fuente: Adaptado de Fernández & Solano (2007).

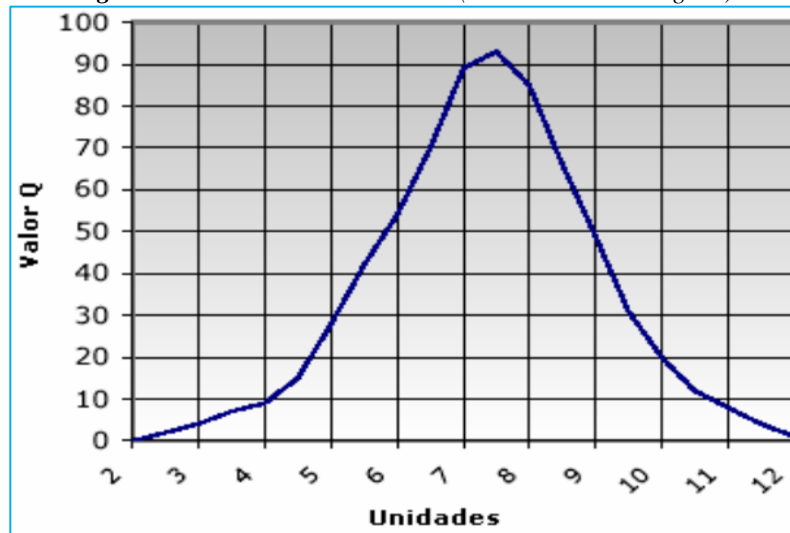
Cuando la cantidad de Coliformes Fecales supera las 100000 colonias por cada 100 mL, el Q-CF se establece en 3. En caso de que la cifra de Coliformes Fecales sea inferior a 100000 colonias por cada 100 mL, se busca el valor correspondiente en el eje X de la Figura 4 y se lleva a cabo una

interpolación para determinar el valor en el eje Y. Este resultado constituirá el Q-CF de Coliformes Fecales. Para concluir, se multiplica el valor de Q-CF por el peso W-CF (SNET, 2007).

- **Potencial de Hidrógeno – pH (Unidades de pH)**

El pH, esencial para evaluar la acidez o alcalinidad del agua, se mide en una escala que abarca de 0 a 14, donde valores inferiores a 7 denotan acidez, mientras que valores superiores a 7 indican alcalinidad. Este parámetro, crítico para los ecosistemas acuáticos, impacta la solubilidad de nutrientes esenciales y la disponibilidad de metales, ejerciendo así una influencia directa sobre la salud y el bienestar de los organismos acuáticos (Gribbin, 2013).

Figura 5: Función de Calidad NSF (Potencial de Hidrógeno).



Fuente: Adaptado de Fernández & Solano (2007).

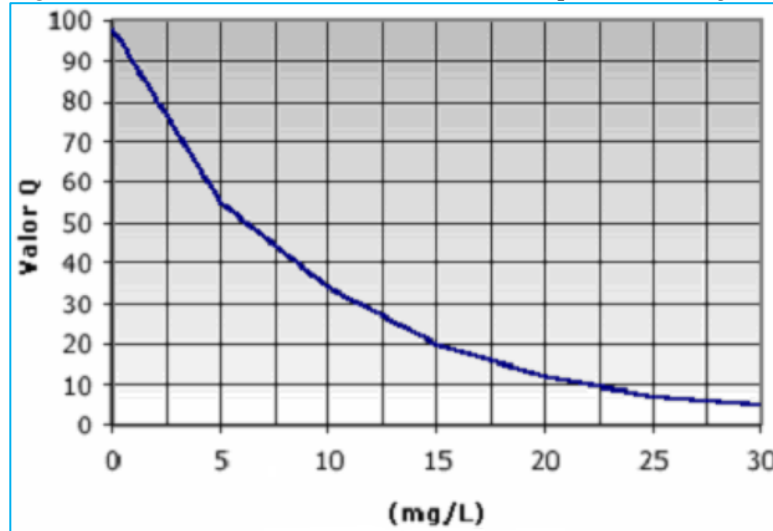
Cuando el valor del Potencial de Hidrógeno es igual o inferior a 2 unidades, se establece que el Q-pH sea 2. En el caso de que el valor del Potencial de Hidrógeno sea igual o superior a 10 unidades, el Q-pH se fija en 3. Si el pH se encuentra en el rango entre 2 y 10 unidades, se busca el valor correspondiente en el eje X de la Figura 5 y se realiza una interpolación para obtener el valor en el eje Y. Este resultado representa el Q-pH del Potencial de Hidrógeno. Para concluir, se multiplica el valor de Q-pH por el peso W-pH (SNET, 2007).

- **Demanda Bioquímica de Oxígeno en 5 días – DBO₅ (mg/L)**

La Demanda Bioquímica de Oxígeno a los cinco días, también conocida como DBO₅, mide cuánto oxígeno se requiere para que los microorganismos descompongan la materia orgánica en el

agua durante un lapso de cinco días. Esta medición proporciona una evaluación precisa de la contaminación orgánica en el agua al cuantificar la cantidad de oxígeno necesaria para la descomposición de la materia orgánica. Niveles más elevados de DBO5 indican una carga orgánica más significativa y una mayor demanda de oxígeno por parte de los microorganismos descomponedores. Este fenómeno puede atribuirse a la descarga de aguas residuales domésticas, industriales o agrícolas no tratadas o parcialmente tratadas (Safoniuk, 2004).

Figura 6: Función de Calidad NSF (Demanda Bioquímica de Oxígeno).



Fuente: Adaptado de Fernández & Solano (2007).

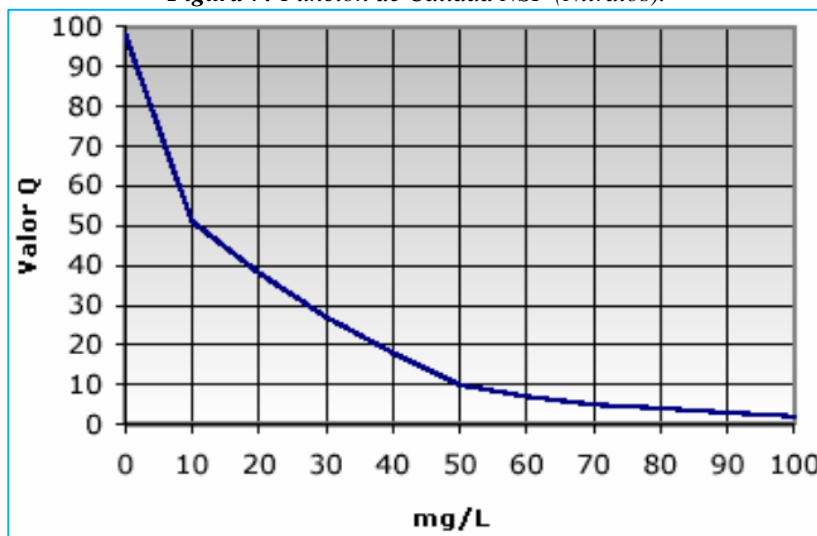
Cuando la Demanda Bioquímica de Oxígeno en 5 días supera los 30 mg/L, el Q-DBO5 se establece en 2. En el caso de que la Demanda Bioquímica de Oxígeno en 5 días sea menor a 30 mg/L, se busca el valor correspondiente en el eje X de la Figura 6 y se realiza una interpolación para determinar el valor en el eje Y. Este resultado representa el Q-DBO5 de la DBO5. Para concluir, se multiplica el valor de Q-DBO5 por el peso W-DBO5 (SNET, 2007).

- **Nitratos – N (mg/L)**

Los nitratos son compuestos químicos formados por nitrógeno y oxígeno que pueden contaminar el agua al entrar por la escorrentía agrícola o por el uso excesivo de fertilizantes. La detección de nitratos en el agua es un indicador crucial de contaminación y puede afectar considerablemente la calidad del agua potable. La principal fuente de nitratos en el agua proviene de la actividad agrícola, donde los fertilizantes ricos en nitrógeno se utilizan para mejorar el crecimiento de los cultivos. No obstante, cuando estos fertilizantes no se aplican correctamente o

se utilizan en exceso, los nitratos pueden infiltrarse en el suelo y, eventualmente, llegar a las fuentes de agua subterránea o superficial a través de la escorrentía. Además, la actividad ganadera también puede contribuir a la presencia de nitratos en el agua, ya que los residuos animales pueden liberar nitratos al medio ambiente (Biswas, 1991).

Figura 7: Función de Calidad NSF (Nitratos).



Fuente: Adaptado de Fernández & Solano (2007).

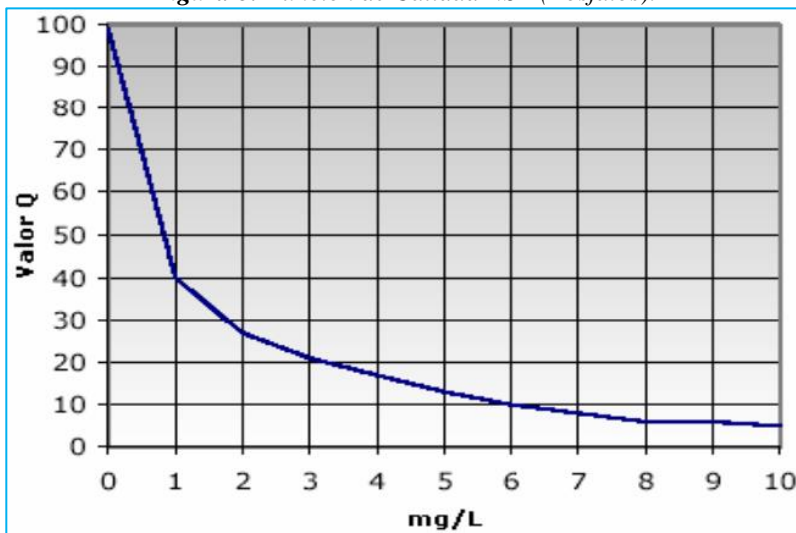
Cuando los Nitratos superan los 100 mg/L, el Q-N se establece en 2. En el caso de que los Nitratos sean menores a 100 mg/L, se busca el valor correspondiente en el eje X de la Figura 7 y se realiza una interpolación para determinar el valor en el eje Y. Este resultado representa el Q-N de los Nitratos. Para concluir, se multiplica el valor de Q-N por el peso W-N (SNET, 2007).

- **Fosfatos – P (mg/L)**

Los fosfatos son nutrientes esenciales para el desarrollo saludable de las plantas, pero su exceso puede provocar problemas ecológicos, como la eutrofización en cuerpos de agua. Cuando los fosfatos se acumulan en exceso en estos cuerpos de agua, suministran nutrientes adicionales para el crecimiento descontrolado de algas y otras plantas acuáticas. Esto conduce a una proliferación abundante de algas, formando capas densas en la superficie del agua. A medida que estas algas mueren y se descomponen, los microorganismos descomponedores consumen grandes cantidades de oxígeno, resultando en una reducción de la disponibilidad de este elemento vital para otros organismos acuáticos, como peces y crustáceos. En consecuencia, los organismos acuáticos

pueden experimentar una escasez de oxígeno y, en algunos casos, sufrir la pérdida de vida (Valsami-Jones, 2004).

Figura 8: Función de Calidad NSF (Fosfatos).



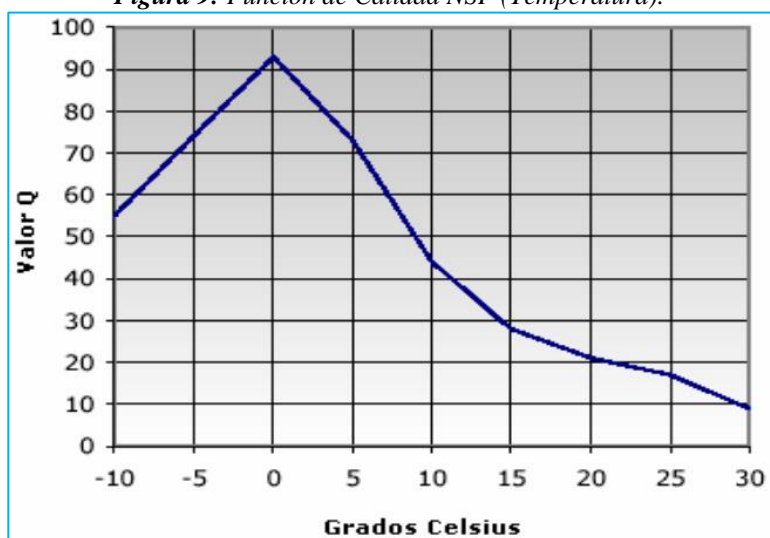
Fuente: Adaptado de Fernández & Solano (2007).

Cuando los Fosfatos superan los 10 mg/L, el Q-P se establece en 5. En el caso de que los Fosfatos sean menores a 10 mg/L, se busca el valor correspondiente en el eje X de la Figura 8 y se realiza una interpolación para determinar el valor en el eje Y. Este resultado representa el Q-P de los Fosfatos. Para concluir, se multiplica el valor de Q-P por el peso W-P (SNET, 2007).

- **Cambio de Temperatura – CT (°C)**

La temperatura del agua juega un papel fundamental en la solubilidad del oxígeno y tiene consecuencias significativas en los procesos biológicos de los ecosistemas acuáticos. Las variaciones extremas en la temperatura pueden resultar perjudiciales para la flora y fauna que dependen de un equilibrio en el entorno acuático. A medida que la temperatura del agua aumenta, su capacidad para retener oxígeno disminuye, ya que las moléculas de agua se mueven más rápidamente, lo que reduce su propensión a retener moléculas de oxígeno. En contraste, a temperaturas más bajas, las moléculas de agua se vuelven más densas y tienen una mayor capacidad para retener mayores cantidades de oxígeno (Baird et al., 2012).

Figura 9: Función de Calidad NSF (Temperatura).



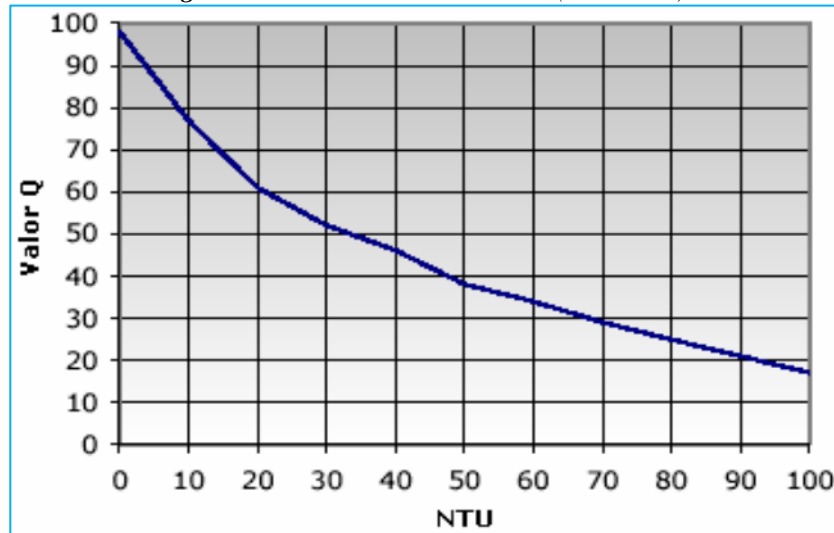
Fuente: Adaptado de Fernández & Solano (2007).

En el caso del parámetro de Temperatura, se inicia calculando la diferencia entre la temperatura ambiente y la temperatura de la muestra. Posteriormente, se busca el valor correspondiente en el eje X de la Figura 9 y se realiza una interpolación para determinar el valor en el eje Y. Este resultado representa el Q-CT de la Temperatura. Finalmente, para concluir, se multiplica el valor de Q-CT por el peso W-CT (SNET, 2007).

- **Turbiedad – T (NTU)**

La turbidez del agua se refiere a la presencia de partículas suspendidas que generan opacidad en el entorno acuático. Cuando los niveles de turbidez son elevados, estas partículas pueden obstaculizar la penetración de la luz solar en el agua, afectando negativamente los procesos de fotosíntesis en los ecosistemas acuáticos. Factores como sedimentos, arcilla, materia orgánica en descomposición, algas u otros contaminantes suspendidos en el agua pueden contribuir a la turbidez. Estas partículas pueden originarse tanto de manera natural, como la erosión del suelo o la actividad biológica, como de actividades humanas, incluyendo la agricultura intensiva, la construcción o el vertido de aguas residuales sin tratar (Boyd, 2019).

Figura 10: Función de Calidad NSF (Turbiedad).



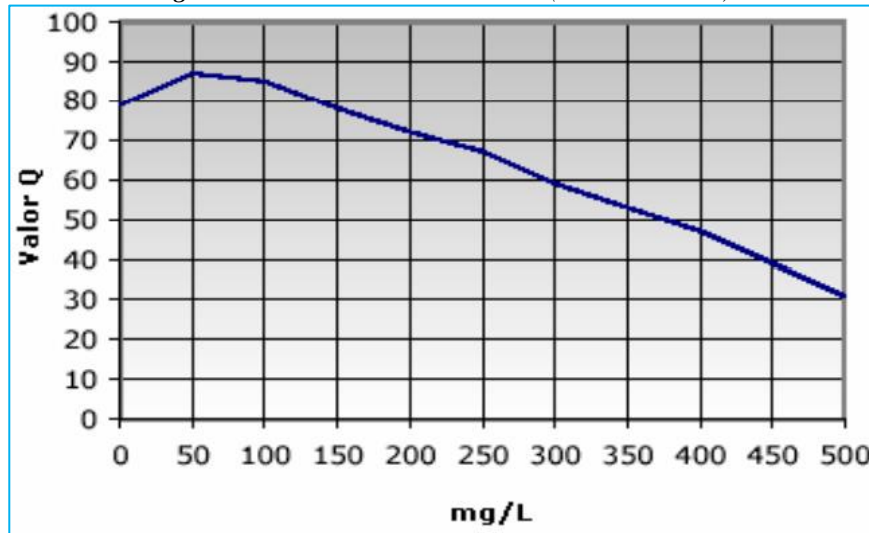
Fuente: Adaptado de Fernández & Solano (2007).

Cuando el valor de la Turbiedad es superior a 100 NTU, se establece que el Q-T sea 5. En el caso de que el valor de la Turbiedad sea inferior a 100 NTU, se busca el valor correspondiente en el eje X de la Figura 10 y se realiza una interpolación para obtener el valor en el eje Y. Este resultado representa el Q-T de la Turbiedad. Para concluir, se multiplica el valor de Q-T por el peso W-T (SNET, 2007).

- **Sólidos Totales – ST (mg/L)**

Los sólidos totales en el agua comprenden la suma de los sólidos suspendidos y los sólidos disueltos presentes en el entorno acuático. Estos sólidos pueden consistir en diversos materiales, tanto orgánicos como inorgánicos, y su existencia puede proporcionar información esencial sobre la calidad y la composición del agua. Los sólidos totales pueden originarse de diversas fuentes. Los sólidos suspendidos engloban partículas visibles en el agua, como sedimentos, arcilla, materia orgánica en descomposición, residuos industriales o agrícolas, entre otros. Estas partículas pueden ser transportadas por la erosión del suelo, el escurrimiento de aguas pluviales o la actividad humana, y pueden afectar la transparencia del agua y el hábitat de los organismos acuáticos. En contraste, los sólidos disueltos refieren a sustancias químicas en forma de iones, como sales minerales, nutrientes y compuestos orgánicos disueltos. Aunque no son visibles a simple vista, estos sólidos pueden influir en la calidad del agua y tener efectos tanto beneficiosos como perjudiciales en los ecosistemas acuáticos (Bear, 2007).

Figura 11: Función de Calidad NSF (Sólidos Totales).



Fuente: Adaptado de Fernández & Solano (2007).

Cuando los Sólidos Totales superan los 500 mg/L, el Q-ST se establece en 3. En el caso de que los Sólidos Totales sean menores a 500 mg/L, se busca el valor correspondiente en el eje X de la Figura 11 y se realiza una interpolación para determinar el valor en el eje Y. Este resultado representa el Q-ST de los Sólidos Totales. Para concluir, se multiplica el valor de Q-ST por el peso W-ST (SNET, 2007).

2.7 Cálculo del Índice de la Fundación Nacional de Saneamiento

Brown propuso la evaluación cuantitativa del Índice de la Fundación Nacional de Saneamiento (ICA NSF) a través de métodos multiplicativos y ponderados, haciendo uso de asignaciones de pesos específicos (SNET, 2007). Por ende, el cálculo del ICA puede llevarse a cabo mediante una combinación ponderada lineal de los subíndices, expresada de la siguiente manera:

Ecuación 1: Ecuación para el cálculo del ICA NSF.

$$ICA = \sum_{i=1}^n Q_i \cdot W_i$$

Fuente: Brown, RM et al. (1970).

Donde:

- i : Parámetro considerado
- Q_i : Valor del parámetro i
- n : Número de parámetros
- W_i : Factor de ponderación para el parámetro i

Los factores relativos asignados a cada parámetro, y ajustados en un rango de 0 a 1, de manera que la suma total sea igual a uno, son:

Tabla 5: Pesos relativos para cada parámetro del ICA.

n	Parámetro i	Unidades	W _i
1	Oxígeno disuelto	% de saturación	0.17
2	Coliformes fecales	NMP/100mL	0.15
3	pH	pH	0.12
4	DBO ₅	mg/L	0.10
5	Nitratos	mg/L	0.10
6	Fosfatos	mg/L	0.10
7	Temperatura	°C	0.10
8	Turbiedad	NTU	0.08
9	Sólidos totales	mg/L	0.08

Fuente: Brown et al. (1970).

Después de realizar el cálculo del ICA, se procede a la clasificación con base en el resultado numérico que debería oscilar entre 0 y 100:

Tabla 6: Clasificación del ICA propuesto por Brown.

Calidad del agua	Color	Valor
Excelente		91 a 100
Buena		71 a 90
Regular		51 a 70
Mala		26 a 50
Pésima		0 a 25

Fuente: Brown et al. (1970).

- **Excelente (91 a 100):** Cuando nos referimos a la calidad del agua como “excelente”, estamos hablando de un estado en el que los niveles de varios parámetros clave son óptimos, indicando pureza y la ausencia de contaminantes dañinos. Estos parámetros incluyen, pero no se limitan a, la concentración de oxígeno disuelto, el pH, la turbidez y la

presencia de sustancias químicas específicas. Estos indicadores son esenciales para evaluar la salud y la calidad general del agua (Svirčev et al., 2014).

- **Buena (71 a 90):** Cuando nos referimos a aguas clasificadas como “buena”, estamos hablando de aquellas que cumplen con los estándares aceptables establecidos para asegurar la salud humana y la sostenibilidad de los ecosistemas acuáticos. Aunque puede haber ciertos niveles de contaminantes presentes, estos se mantienen por debajo de los límites críticos establecidos. Numerosas investigaciones respaldan la afirmación de que la calidad del agua clasificada como buena sigue siendo apropiada tanto para el consumo humano como para el respaldo de la vida acuática (Ibekwe et al., 2011).
- **Regular (51 a 70):** Cuando nos referimos a la clasificación de la calidad del agua como “regular”, estamos hablando de niveles de ciertos parámetros que se sitúan cerca de los límites aceptables o han experimentado fluctuaciones temporales. En este contexto, es esencial destacar que estas aguas pueden requerir una supervisión continua y la implementación de medidas correctivas para prevenir una degradación mayor. La investigación científica ha señalado que las aguas de calidad regular pueden estar vinculadas con ciertos riesgos tanto para la salud humana como para los ecosistemas acuáticos (Espinosa Ramírez, 2018).
- **Mala (26 a 50):** Cuando la calidad del agua se califica como “mala”, esto indica la presencia de contaminantes en niveles que superan los límites aceptables establecidos. Esta condición tiene repercusiones negativas tanto para la salud humana como para la vida acuática. La investigación científica ha confirmado de manera concluyente que las aguas de baja calidad están directamente vinculadas a una variedad de problemas de salud, como enfermedades gastrointestinales, y también pueden resultar en una disminución significativa de la biodiversidad acuática (Lin et al., 2022).
- **Pésima (0 a 25):** Cuando hacemos referencia a la categorización de la calidad del agua como “pésima”, estamos hablando de la peor condición posible en términos de calidad del agua. En este escenario, los niveles de contaminantes en el agua son excepcionalmente elevados, representando una amenaza significativa tanto para la salud humana como para la supervivencia de los ecosistemas acuáticos. La investigación científica ha confirmado de manera concluyente que la exposición prolongada a aguas de calidad pésima puede conllevar consecuencias graves para la salud humana, incluyendo el desarrollo de

enfermedades crónicas, así como la degradación irreversible de los ecosistemas acuáticos (Palaniappan et al., 2011).

2.8 Software ArcMap

ArcMap, desarrollado por Environmental Systems Research Institute (ESRI), constituye un componente esencial de la suite ArcGIS, que proporciona soluciones avanzadas en el ámbito de los sistemas de información geográfica (SIG). Distinguido por ofrecer a los usuarios un entorno integral y resistente, ArcMap se destaca en la creación, análisis y visualización de datos espaciales. Dotado de una interfaz intuitiva y una amplia variedad de herramientas para el análisis y la cartografía, ArcMap se ha convertido en una herramienta indispensable para profesionales en diversos campos, tales como la cartografía, la planificación del uso del suelo y la gestión de recursos naturales (ESRI, 2021).

2.8.1 Modelo digital del terreno (MDT)

El Modelo Digital del Terreno (MDT) representa de manera numérica y tridimensional la superficie terrestre con una alta precisión, capturando minuciosamente las variaciones en la elevación del terreno. Mediante el empleo de tecnologías avanzadas como la fotogrametría, el láser escáner y los sistemas de información geográfica, se generan modelos detallados que ofrecen información exacta y exhaustiva acerca de la topografía de una determinada área. Estos modelos posibilitan realizar análisis detallados del relieve y explorar en profundidad las características geográficas de un lugar (Inegi, 2012).

2.8.2 Delimitación de una cuenca hidrográfica utilizando un MDT

1. En el inicio del proceso de demarcación de una cuenca hidrográfica, se realiza la importación del Modelo Digital del Terreno (MDT) en el software. ArcMap facilita la carga y representación visual de los datos del MDT como una representación gráfica de la topografía del área en estudio (TYC-GIS, 2015).
2. Después de importar el MDT, se llevan a cabo tareas de preparación para asegurar la calidad y precisión de los datos. Estas tareas pueden abarcar desde la eliminación de ruido o artefactos hasta la corrección de errores sistemáticos, interpolación de datos faltantes y, si es necesario, reproyección. Estas actividades son esenciales para garantizar la

confiabilidad de los datos del MDT y su idoneidad para análisis posteriores (TYC-GIS, 2015).

3. A continuación, se realiza el análisis hidrológico utilizando las herramientas y funciones proporcionadas por ArcMap. Estas herramientas permiten determinar la dirección del flujo de agua basándose en las elevaciones del MDT. Se identifican las áreas de acumulación de agua y los flujos hacia los cuerpos de agua, estableciendo así los límites de la cuenca hidrográfica (TYC-GIS, 2015).
4. Después de completar el análisis del flujo de agua, se procede a la demarcación de la cuenca hidrográfica. Este proceso implica trazar límites divisorios de aguas, que separan las áreas de drenaje que fluyen en direcciones diferentes. Dichos límites divisorios se generan utilizando los resultados del análisis hidrológico y se emplean para definir los límites precisos de la cuenca hidrográfica (TYC-GIS, 2015).
5. Además de la delimitación de la cuenca hidrográfica, ArcMap proporciona herramientas para calcular diversos parámetros hidrológicos. Estos incluyen la superficie de la cuenca, el caudal, la longitud del curso principal de agua, la pendiente media y otros atributos relevantes. Estos cálculos suministran información esencial para comprender el comportamiento hidrológico de la cuenca y son fundamentales para la gestión adecuada del recurso hídrico (TYC-GIS, 2015).
6. Al completar el análisis y la delimitación de la cuenca hidrográfica, ArcMap también permite la visualización de los resultados mediante la creación de mapas temáticos y cartografía. Este aspecto resulta especialmente útil para presentar y comunicar datos de manera clara y comprensible a través de representaciones visuales (TYC-GIS, 2015).

2.8.3 Interpolación por el método Kriging

El método Kriging, desarrollado por el geólogo sudafricano Danie G. Krige en la década de 1950, se utiliza en diversas disciplinas, como hidrología y gestión de recursos hídricos, para llevar a cabo interpolación espacial. Esta técnica, fundamentada en la teoría de estadísticas espaciales, destaca por su eficacia en la estimación de valores en ubicaciones no muestreadas dentro de una región espacial específica. Su aplicación resulta valiosa en situaciones donde la distribución espacial de los datos es heterogénea y se busca minimizar la incertidumbre en las predicciones (Krige, 1951).

El Kriging se basa en el concepto de semivarianza, que cuantifica la variabilidad espacial de los datos en función de la distancia entre los puntos de muestreo. Al hacerlo, el Kriging calcula los pesos óptimos para combinar las observaciones conocidas y realizar predicciones en ubicaciones no muestreadas. Los pasos clave de este método comprenden la definición de un modelo de semivarianza, la estimación de los parámetros del modelo y la predicción de valores en ubicaciones específicas (Bilonick, 1991).

En el ámbito de la calidad del agua en una cuenca hidrográfica, la interpolación de datos desempeña un papel crucial para comprender la distribución espacial de parámetros fundamentales, como la concentración de contaminantes. Este proceso facilita la identificación de áreas críticas, la evaluación del impacto ambiental y la toma de decisiones informadas en la gestión de los recursos hídricos. Dado que la calidad del agua puede variar significativamente en una cuenca debido a factores geográficos, climáticos y antropogénicos, se destaca la importancia de utilizar técnicas de interpolación avanzadas (Goovaerts, 1997).

2.9 Lenguaje de programación Python

Python, ideado por Guido van Rossum y presentado en 1991, es un lenguaje de programación de alto nivel y propósito general conocido por su sencillez y legibilidad. Su sintaxis clara y concisa facilita su aprendizaje y comprensión. La elegancia de su diseño y su filosofía de “baterías incluidas” se traducen en la inclusión de numerosas funcionalidades útiles en la biblioteca estándar de Python. Además, su extensa comunidad de desarrolladores contribuye con numerosas bibliotecas de terceros, simplificando la búsqueda de soluciones para una amplia variedad de desafíos (Van Rossum, 2007).

Bibliotecas

Una biblioteca de programación consta de funciones o capacidades desarrolladas por programadores con el objetivo de llevar a cabo tareas que previamente resultaban difíciles. Estas bibliotecas proporcionan soluciones y herramientas predefinidas que optimizan el tiempo y el esfuerzo, agilizando así el proceso de desarrollo y mejorando la eficacia y calidad del código (Flores, 2021). Ejemplos de bibliotecas importantes en Python son math, pandas, scikit-learn y SciPy.

- La biblioteca “math”, integrada en la biblioteca estándar de Python, ofrece un conjunto exhaustivo de funciones y constantes matemáticas para realizar operaciones fundamentales. Los programadores tienen la capacidad de realizar diversas operaciones matemáticas, incluyendo funciones trigonométricas, exponenciales, logarítmicas y de redondeo. Asimismo, proporciona constantes matemáticas esenciales, como π (π) y e (número de Euler), que desempeñan un papel crucial en una variedad de cálculos científicos e ingenieriles (Python Software Foundation, 2023).
- La biblioteca “pandas” es una herramienta de manipulación y análisis de datos que presenta estructuras de datos eficientes, como el DataFrame, una tabla bidimensional que facilita la organización y análisis de datos. Pandas proporciona diversas funciones para cargar y guardar datos en distintos formatos, realizar tareas de limpieza y transformación de datos, ejecutar operaciones de filtrado y selección, calcular estadísticas descriptivas y crear visualizaciones. Debido a su versatilidad y potencia, se utiliza extensamente en campos como la ciencia de datos, el análisis financiero, la investigación académica y otras disciplinas que demandan un manejo eficaz de datos tabulares (McKinney, 2010).
- La biblioteca “scikit-learn” abarca un amplio espectro de algoritmos y herramientas destinadas a tareas de aprendizaje, ya sea supervisado o no supervisado. En su repertorio se incluyen algoritmos de clasificación, regresión, agrupamiento, reducción de dimensionalidad, selección de características, entre otros. Además de estos algoritmos, scikit-learn proporciona utilidades para la preparación de datos, evaluación de modelos y ajuste de hiperparámetros. Su uso es prevalente en el desarrollo de modelos predictivos y análisis de datos en diversas aplicaciones, como detección de fraudes, procesamiento del lenguaje natural, visión por computadora, entre otras (Pedregosa et al., 2011).
- La biblioteca “SciPy” representa un recurso de código abierto que amplía las funcionalidades de la biblioteca math, ofreciendo capacidades más avanzadas y especializadas para aplicaciones científicas e ingenieriles. Aborda una diversidad de áreas, que incluyen estadísticas, procesamiento de señales, álgebra lineal, optimización, interpolación, entre otras. Dentro de su submódulo “stats”, SciPy proporciona herramientas estadísticas y, en el contexto de la prueba t, la función “ttest_ind” permite llevar a cabo la prueba t para muestras independientes. Esta prueba facilita la comparación de las medias

de dos conjuntos de datos, ayudando a determinar si existen diferencias significativas entre ellas (Virtanen et al., 2020).

2.10 Aprendizaje automático

El aprendizaje automático, también denominado Machine Learning, constituye una rama de la inteligencia artificial que se enfoca en la creación de algoritmos y modelos con la capacidad de aprender de manera autónoma. Este proceso implica identificar patrones y llevar a cabo tareas específicas sin depender de una programación explícita. La característica distintiva de esta disciplina es la habilidad de las máquinas para mejorar su rendimiento a medida que se les suministra más información y datos, lo que resulta en una adaptación y optimización continua de los modelos (Cabrera Mendieta, 2017). El aprendizaje automático encuentra aplicaciones en diversos ámbitos, tales como el reconocimiento de voz, la visión por computadora, el procesamiento de lenguaje natural, el diagnóstico médico, las finanzas, entre otros.

Los 3 tipos de aprendizaje automático

- **Aprendizaje Supervisado:** En el aprendizaje supervisado, los algoritmos son entrenados mediante un conjunto de datos etiquetado, en el cual cada entrada está asociada con una etiqueta o resultado conocido. El modelo adquiere la capacidad de realizar predicciones o clasificaciones basadas en ejemplos previamente etiquetados. El objetivo es que el algoritmo pueda generalizar de manera efectiva a nuevas instancias que no hayan sido vistas durante la fase de entrenamiento. Algunos ejemplos de algoritmos comúnmente utilizados en el aprendizaje supervisado son la Regresión Lineal, las Máquinas de Soporte Vectorial (SVM) y las Redes Neuronales (Hastie et al., 2001).
- **Aprendizaje No Supervisado:** A diferencia del aprendizaje supervisado, el aprendizaje no supervisado trabaja con conjuntos de datos que carecen de etiquetas. En este enfoque, los algoritmos buscan identificar patrones, estructuras o relaciones inherentes en los datos sin contar con información previa sobre las salidas esperadas. Algunas técnicas comunes de aprendizaje no supervisado comprenden el clustering (agrupamiento), la reducción de dimensionalidad y la asociación. Entre los algoritmos destacados se encuentran el K-Means, el Análisis de Componentes Principales (PCA) y los Algoritmos de Asociación (Bishop, 2006).

- **Aprendizaje por Refuerzo:** En el aprendizaje por refuerzo, un agente interactúa con su entorno tomando decisiones y recibiendo retroalimentación en forma de recompensas o castigos. El objetivo es que el agente aprenda a realizar acciones que maximicen la recompensa acumulada a lo largo del tiempo. Este enfoque es esencial en la inteligencia artificial aplicada a la toma de decisiones, como en juegos y el control de robots. Algoritmos notables en este contexto incluyen Q-learning y Redes Neuronales Profundas para Q-learning (DQN) (Sutton & Barto, 2018).

2.11 Creación de modelos basados en datos

La construcción de modelos basados en datos es un procedimiento esencial en el ámbito de la ciencia de datos, cuyo objetivo primordial consiste en descubrir patrones y conocimientos valiosos a partir de conjuntos de datos. Este enfoque emplea una combinación de técnicas estadísticas, algoritmos de aprendizaje automático y métodos de análisis exploratorio de datos para desarrollar modelos que puedan prever y describir fenómenos con precisión. El proceso de creación de modelos basados en datos abarca diversas etapas clave (Hastie et al., 2001).

1. En una primera fase, se lleva a cabo una exploración exhaustiva de los datos disponibles para comprender su estructura, características y posibles relaciones. Esto incluye actividades como la visualización de datos, la detección de valores atípicos y la evaluación general de la calidad de los datos.
2. Tras completar la exploración inicial, se eligen las variables pertinentes y se preparan los datos para su análisis. Esto involucra la limpieza de datos, la normalización y transformación de variables, y la división del conjunto de datos en conjuntos de entrenamiento y prueba.
3. Posterior a la preparación de los datos, se selecciona el algoritmo de modelado más adecuado para abordar el problema específico. Esto puede incluir el uso de técnicas de regresión, clasificación, agrupamiento u otras estrategias de aprendizaje automático, dependiendo de la naturaleza de los datos y del objetivo del modelo.
4. Una vez seleccionado el algoritmo, se procede a entrenar el modelo utilizando el conjunto de datos de entrenamiento. Durante este proceso, el modelo ajusta sus parámetros para encontrar la representación óptima de los datos y capturar las relaciones subyacentes.

5. Después del entrenamiento, se evalúa el rendimiento del modelo utilizando el conjunto de datos de prueba. Esto implica comparar las predicciones del modelo con los valores reales y calcular métricas de rendimiento, como la precisión, el error cuadrático medio o el coeficiente de determinación.

Si el modelo exhibe un rendimiento satisfactorio, se puede implementar en aplicaciones prácticas para realizar predicciones o extraer conocimientos. No obstante, el proceso de creación de modelos es iterativo y puede requerir ajustes adicionales, como la optimización de hiperparámetros o la incorporación de nuevas variables, para mejorar aún más el rendimiento del modelo (Hastie et al., 2001).

2.12 Árboles de decisión

Los árboles de decisión son modelos de aprendizaje supervisado ampliamente empleados en el ámbito de la inteligencia artificial y la estadística, siendo utilizados tanto para problemas de clasificación como para problemas de regresión. Estos modelos se fundamentan en la estructura de un árbol, en el que cada nodo representa una decisión tomada en base a la información proporcionada por una característica específica del conjunto de datos (Hastie et al., 2001).

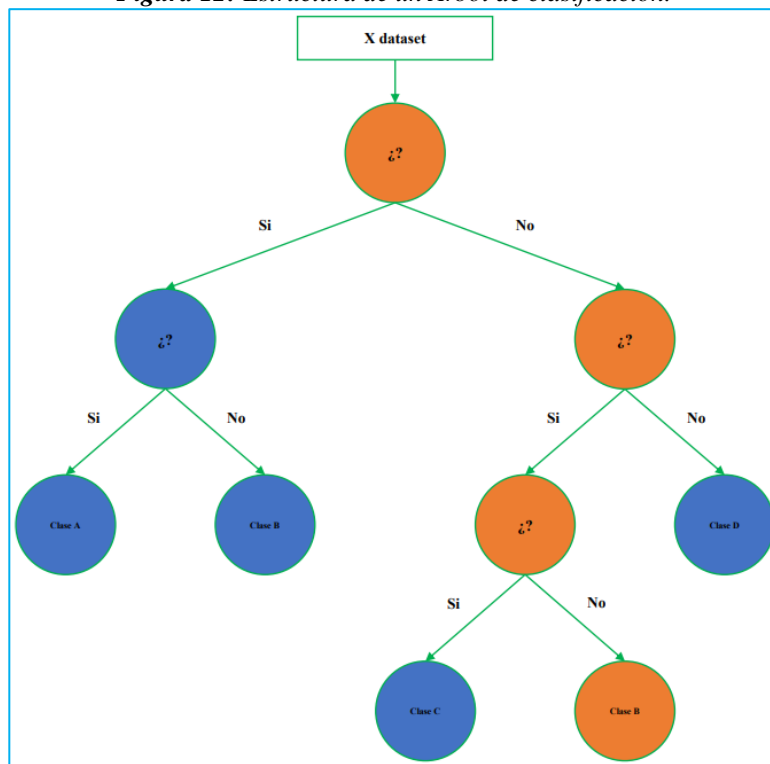
2.12.1 Árboles de clasificación

Los árboles de clasificación representan una variante especializada de los árboles de decisión diseñada específicamente para abordar problemas de clasificación, en los cuales la variable objetivo es categórica. Similar a los árboles de decisión, los árboles de clasificación adoptan una estructura en forma de árbol, en la que los nodos establecen condiciones que dividen el conjunto de datos en subconjuntos más pequeños, y las hojas del árbol representan las categorías de clasificación asignadas a las instancias (Witten & James, 2013).

Entre los algoritmos destacados para construir árboles de clasificación, se encuentra el algoritmo CART (Classification and Regression Trees). Este algoritmo incorpora varios hiperparámetros que posibilitan ajustar el comportamiento y la estructura del árbol resultante para adaptarse a las particularidades específicas del problema (Witten & James, 2013). Algunos de los hiperparámetros más relevantes son:

- **Profundidad del Árbol:** Este parámetro de ajuste regula la extensión máxima permitida para los nodos en el árbol. Restringir la profundidad contribuye a prevenir el sobreajuste, manteniendo una estructura más sencilla y fácil de comprender (Witten & James, 2013).
- **Número Mínimo de Muestras en una Hoja:** Este hiperparámetro establece la cantidad mínima de instancias necesarias en una hoja del árbol. Al fijar un límite mínimo, se previene la formación de hojas con un número insuficiente de ejemplos, evitando así predicciones poco confiables o inestables (Witten & James, 2013).
- **Número Mínimo de Muestras en un Nodo:** Este parámetro determina la cantidad mínima de instancias requeridas en un nodo para realizar una división adicional. Establecer un umbral mínimo evita divisiones que carecen de información significativa, reduciendo la complejidad innecesaria de la estructura del árbol (Witten & James, 2013).

Figura 12: Estructura de un Árbol de clasificación.



Fuente: Adaptado de Alvaro (2023).

2.12.2 Árboles de regresión

A diferencia de los árboles de clasificación, los árboles de regresión son específicamente empleados para abordar problemas de regresión, donde la variable objetivo es continua en lugar

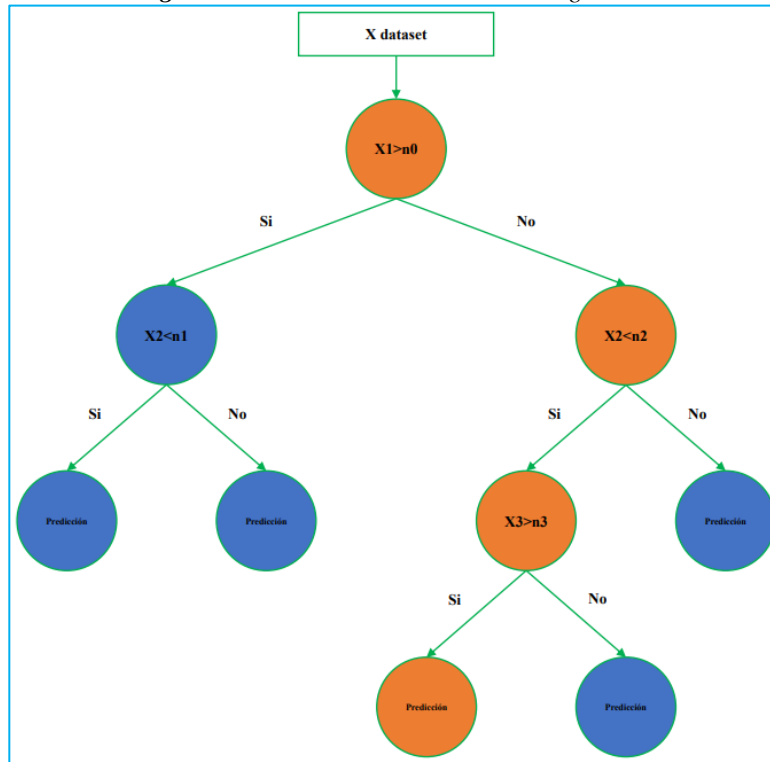
de categórica. Estos árboles están diseñados para prever un valor numérico para cada hoja del árbol, convirtiéndolos en una herramienta efectiva para estimar relaciones y tendencias en conjuntos de datos con variables continuas (Witten & James, 2013).

Aunque el algoritmo CART se utiliza tanto en árboles de clasificación como en árboles de regresión, difiere en el enfoque cuando se trata de problemas de regresión. En este caso, se ajusta para minimizar la varianza de las predicciones en lugar de maximizar la homogeneidad de las clases, como ocurre en los árboles de clasificación. Esto implica que el objetivo principal es reducir la dispersión de los valores estimados y lograr la máxima precisión posible en la predicción de valores continuos (Witten & James, 2013).

Cuando se construyen árboles de regresión, es esencial tener en cuenta los hiperparámetros pertinentes que permiten ajustar la estructura del árbol para obtener resultados óptimos (Witten & James, 2013). Algunos de los hiperparámetros más destacados son los siguientes:

- **Profundidad del Árbol:** Este parámetro controla la máxima profundidad que pueden alcanzar los nodos en el árbol de regresión. Restringir esta profundidad contribuye a prevenir la creación de modelos excesivamente intrincados que podrían sobreajustarse a los datos de entrenamiento y enfrentar dificultades para generalizar a nuevos conjuntos de datos (Witten & James, 2013).
- **Número Mínimo de Muestras en una Hoja:** Este hiperparámetro determina el número mínimo de instancias necesario en una hoja del árbol de regresión. Al imponer un límite mínimo, se evita la formación de hojas con un número insuficiente de ejemplos, lo que podría resultar en predicciones poco fiables y menos estables (Witten & James, 2013).
- **Número Mínimo de Muestras en un Nodo:** Este parámetro especifica el número mínimo de instancias requerido en un nodo para realizar una nueva división en el árbol de regresión. Establecer un límite mínimo ayuda a evitar divisiones que no proporcionan información significativa y podrían conducir a una estructura innecesariamente compleja del árbol (Witten & James, 2013).

Figura 13: Estructura de un Árbol de regresión.



Fuente: Adaptado de Alvaro (2023).

2.13 Bosque aleatorio

El bosque aleatorio, o Random Forest (RF), es una técnica robusta de aprendizaje automático ampliamente reconocida por su eficacia en diversas aplicaciones, ya sea en tareas de clasificación o regresión. Su principio se fundamenta en la creación de numerosos árboles de decisión y la combinación de sus resultados para generar un modelo más sólido y adaptable (Breiman, 2001).

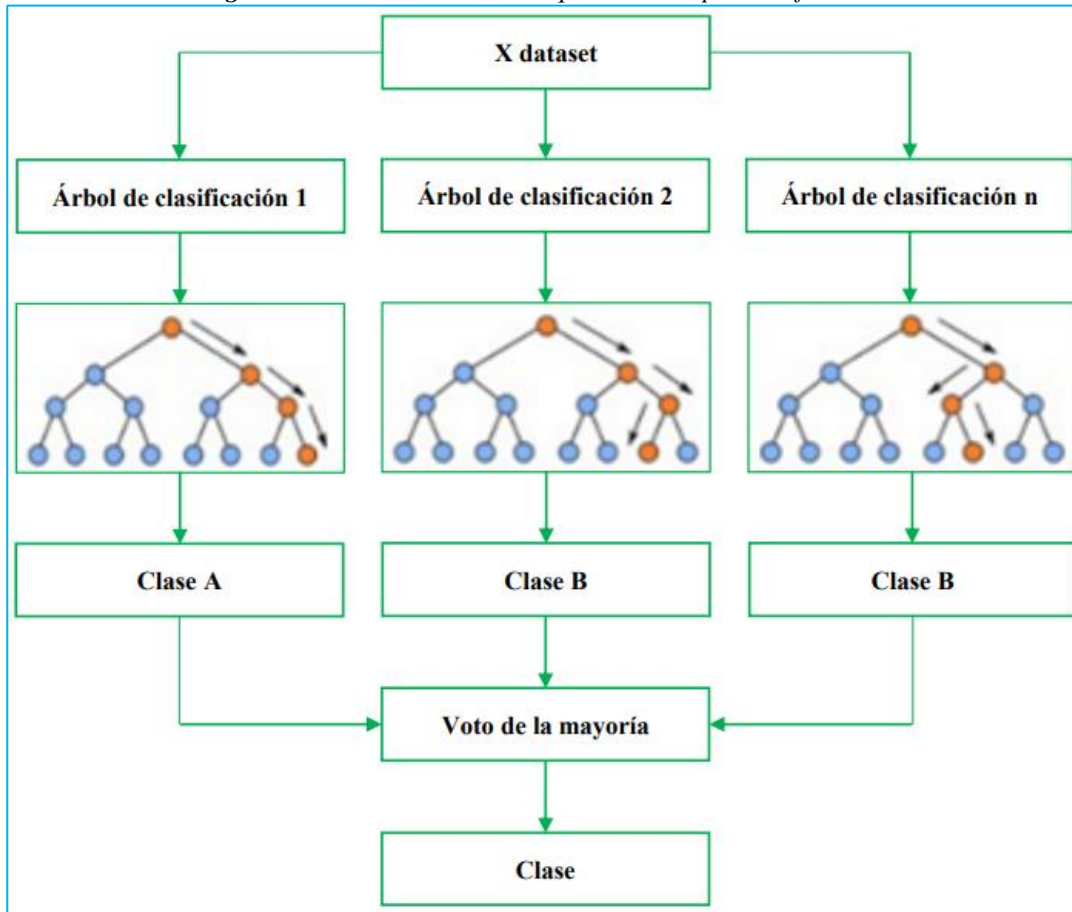
2.13.1 Bosque aleatorio para clasificación (Liaw & Wiener, 2002)

El Bosque Aleatorio para problemas de clasificación representa una extensión del algoritmo original diseñada especialmente para afrontar tareas de clasificación. Esta estrategia se ha evidenciado como particularmente ventajosa al tratar con conjuntos de datos complejos y de alta dimensionalidad, donde otros métodos podrían tener dificultades para lograr resultados precisos y confiables (Liaw & Wiener, 2002).

Al emplear el Bosque Aleatorio para clasificación, resulta crucial considerar y ajustar de manera cuidadosa los hiperparámetros clave que influyen en su rendimiento (Liaw & Wiener, 2002). Algunos de los hiperparámetros más destacados son los siguientes:

- **Cantidad de Árboles:** Este parámetro clave determina la cantidad de árboles de decisión que se construirán en el conjunto del Bosque Aleatorio. Un incremento en el número de árboles suele mejorar la estabilidad y precisión del modelo, ya que se aprovecha la sabiduría colectiva de múltiples árboles. No obstante, es esencial tener en cuenta que un mayor número de árboles también puede aumentar el costo computacional asociado con el entrenamiento y la predicción (Liaw & Wiener, 2002).
- **Profundidad Máxima de los Árboles:** Este parámetro limita la profundidad máxima permitida para cada árbol en el Bosque Aleatorio. El control de la profundidad es fundamental para prevenir el sobreajuste, donde los árboles se ajustan excesivamente a los datos de entrenamiento y tienen dificultades para generalizar a nuevos datos. Valores más altos de `max_depth` pueden propiciar el sobreajuste, mientras que valores bajos pueden resultar en un modelo excesivamente simplificado que no capta la complejidad de los datos (Liaw & Wiener, 2002).
- **Número Mínimo de Muestras en las Hojas:** Este hiperparámetro establece el número mínimo de muestras requeridas para que se forme un nodo hoja durante la construcción de los árboles. Ajustar este parámetro puede afectar la capacidad del modelo para generalizar o especializarse en ciertos patrones de los datos. Un valor más alto de `min_samples_leaf` puede ayudar a prevenir el sobreajuste, ya que se necesita un mayor número de muestras para formar una hoja, fomentando así la generalización (Liaw & Wiener, 2002).
- **Número Mínimo de Muestras para Dividir un Nodo:** Este hiperparámetro define el número mínimo de muestras necesario en un nodo interno para que se realice una nueva división. Ajustar este parámetro permite controlar la complejidad de los árboles y la capacidad del modelo para adaptarse a los datos. Un valor más alto de `min_samples_split` puede resultar en árboles más simples y menos complejos, siendo útil para conjuntos de datos más pequeños o con menos variabilidad (Liaw & Wiener, 2002).

Figura 14: Estructura de un Bosque aleatorio por clasificación.



Fuente: Adaptado de Puma (2020).

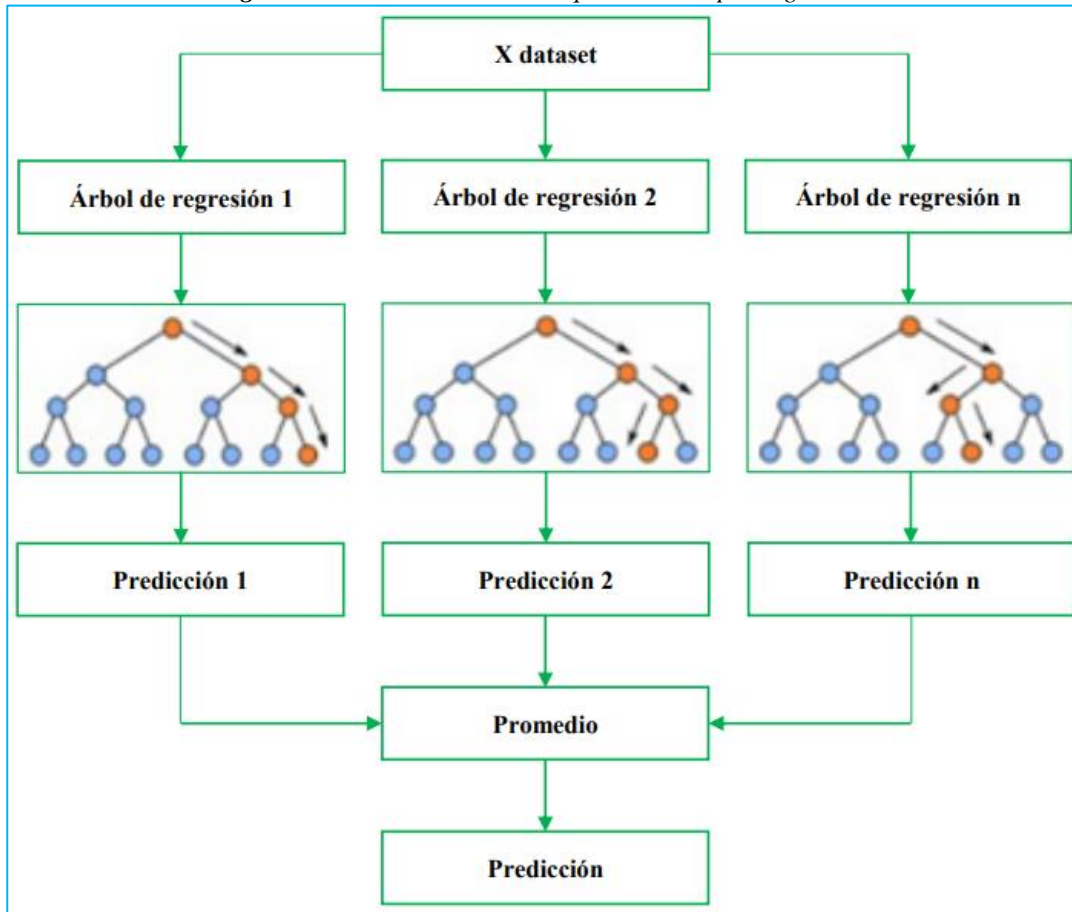
2.13.2 Bosque aleatorio por regresión

En el contexto de la regresión, el Bosque Aleatorio se emplea para realizar predicciones de valores continuos en lugar de clasificar en categorías discretas. Esta estrategia se ha mostrado sumamente eficaz en problemas de regresión, ya que posibilita la modelización de relaciones complejas entre variables y la obtención de predicciones precisas. Al igual que en la clasificación, diversos hiperparámetros fundamentales afectan el rendimiento del Bosque Aleatorio en el ámbito de la regresión (Liaw & Wiener, 2002).

- **Cantidad de Árboles:** Determina cuántos árboles se incluirán en el conjunto del Bosque Aleatorio. Aumentar este número amplía la capacidad del modelo para identificar diversos patrones en los datos y mejora su estabilidad. No obstante, es crucial considerar que un número excesivamente alto de árboles puede incrementar el costo computacional asociado con el entrenamiento y la predicción (Liaw & Wiener, 2002).

- **Profundidad Máxima de los Árboles:** Limitar la profundidad de los árboles resulta crucial para prevenir el sobreajuste en problemas de regresión. Un árbol demasiado profundo tiende a ajustarse en exceso a los datos de entrenamiento y enfrenta dificultades al generalizar a nuevas instancias. El control de la profundidad máxima regula la complejidad del modelo y mejora su capacidad de generalización (Liaw & Wiener, 2002).
- **Número Mínimo de Muestras en las Hojas:** Este parámetro establece el número mínimo de ejemplos necesario en una hoja para su formación durante la construcción de los árboles. Ajustar este valor influye en la capacidad del modelo para identificar patrones más generales o específicos en los datos de regresión. Un valor más alto de `min_samples_leaf` ayuda a prevenir el sobreajuste y promueve la generalización del modelo (Liaw & Wiener, 2002).
- **Número Mínimo de Muestras para Dividir un Nodo:** Este parámetro determina el número mínimo de ejemplos requeridos en un nodo interno para realizar una nueva división. Ajustar este hiperparámetro impacta en la complejidad de los árboles y la capacidad del modelo para adaptarse a los datos de regresión (Liaw & Wiener, 2002).

Figura 15: Estructura de un Bosque aleatorio por regresión.



Fuente: Adaptado de Puma (2020).

2.14 Prueba de hipótesis (Prueba t)

La prueba de hipótesis, una herramienta crucial en la estadística inferencial, posibilita la toma de decisiones sobre afirmaciones relacionadas con una población, basándose en información recopilada de una muestra representativa. Este enfoque implica la formulación de dos hipótesis: la hipótesis nula (H_0) y la hipótesis alternativa (H_1). La hipótesis nula es una afirmación inicial considerada verdadera hasta que haya suficiente evidencia para refutarla, mientras que la hipótesis alternativa es la afirmación que se busca respaldar mediante la evidencia recopilada (Montgomery, 2019).








El procedimiento de la prueba de hipótesis sigue una secuencia de pasos generales (Montgomery, 2019):



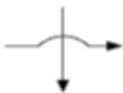
1. En primer lugar, se formulan las hipótesis nula y alternativa, donde H_0 representa la afirmación inicial considerada verdadera y H_1 es la afirmación que se busca respaldar. Posteriormente, se elige un nivel de significancia (α), que indica la probabilidad de cometer un error tipo I al rechazar H_0 cuando es verdadera. Comúnmente, se selecciona un nivel de significancia de $\alpha=0.05$ o $\alpha=0.01$.
2. A continuación, se recopilan los datos de la muestra y se calcula el estadístico de prueba apropiado. Por ejemplo, al comparar medias, se puede emplear la prueba t de Student, desarrollada por William Sealy Gosset en 1908. Esta prueba se utiliza cuando no se cumplen las condiciones para la prueba Z, como un tamaño de muestra pequeño o una desviación estándar poblacional desconocida (STUDENT, 1908).
3. Una vez obtenido el estadístico de prueba, se toma una decisión al compararlo con un valor crítico o utilizando un p-valor. Si el valor de prueba es inferior al valor crítico o el p-valor es menor que el nivel de significancia seleccionado, se rechaza la hipótesis nula, respaldando la hipótesis alternativa. En caso contrario, no se rechaza la hipótesis nula.
4. Finalmente, se emite una conclusión basada en los resultados de la prueba de hipótesis. Si se rechaza la hipótesis nula, se concluye que hay evidencia suficiente para respaldar la hipótesis alternativa. Por el contrario, si no se rechaza la hipótesis nula, no se dispone de suficiente evidencia para respaldar la hipótesis alternativa.

2.15 Diagrama de flujo de datos

Un diagrama de flujo de datos es una representación gráfica que ofrece una visión visual de un procedimiento utilizado para resolver un problema. Estos diagramas son ampliamente empleados en el ámbito de la informática y la programación para ilustrar algoritmos y programas de computadora de manera clara y precisa. La estructura de un diagrama de flujo de datos implica la interconexión de diversas figuras mediante flechas. Cada figura simboliza una acción o proceso, mientras que las flechas señalan la dirección del flujo de datos o información entre las distintas etapas del procedimiento (consultar Tabla 7). Este formato facilita la visualización secuencial de cómo los datos se transforman y procesan a medida que se avanza en el diagrama (Yourdon & Constantine, 1979).

Tabla 7: Simbología para un diagrama de flujo de datos.

Nombre	Símbolo	Función
Inicio/Final		<p>Se utiliza para marcar tanto el inicio como el final de un diagrama de flujo. Es importante destacar que solo puede haber una línea de flujo que salga del símbolo de Inicio y solo una línea de flujo que llegue al símbolo de Final.</p>
Acción/Proceso		<p>Representa una acción o instrucción específica que debe realizarse durante el procedimiento. Esto puede incluir operaciones aritméticas, asignaciones de valores a variables, cálculos matemáticos u otras tareas similares.</p>
Decisión		<p>Se utiliza cuando es necesario tomar una decisión basada en una comparación de dos datos. Dependiendo del resultado lógico (falso o verdadero) de esta comparación, se tomará un camino u otro en el diagrama de flujo.</p>
Impresora/Documento		<p>Indica la presentación de resultados o información en forma impresa. Esto puede incluir la generación de informes, listados o cualquier otro tipo de salida impresa que sea necesaria en el procedimiento.</p>
Pantalla		<p>Se utiliza para indicar la presentación de mensajes o resultados en una pantalla de visualización. Esto puede incluir la visualización de datos, mensajes de confirmación o cualquier otra información relevante que deba mostrarse al usuario.</p>
Entrada/Salida		<p>Se utiliza para representar la entrada o salida de datos a través de cualquier dispositivo externo, como un escáner, lector de códigos de barras, micrófono, parlantes, entre otros. Este símbolo indica la interacción del sistema con el entorno externo.</p>
Entrada por teclado		<p>Se utiliza cuando el procedimiento requiere que el usuario ingrese datos a través del teclado. El computador espera a que el usuario proporcione un</p>

		dato mediante la escritura, y ese dato se guarda en una variable o constante dentro del programa.
Conector interno		Se utilizan para enlazar diferentes partes de un diagrama de flujo dentro de la misma página. Estos conectores permiten una mejor organización y comprensión del diagrama al evitar la repetición innecesaria de ciertas secciones.
Conector externo		Se utilizan para enlazar diferentes partes de un diagrama de flujo en páginas diferentes. Esto es especialmente útil cuando el diagrama de flujo es extenso y no se puede acomodar en una sola página. Los conectores externos permiten mantener la continuidad y la coherencia del diagrama a través de múltiples páginas.
Flechas de dirección/flujo		Indican el seguimiento lógico del diagrama de flujo. Estas flechas establecen el orden de ejecución de las operaciones y muestran la dirección en la que se avanza a través del procedimiento. Siguiendo el flujo de las flechas, se pueden identificar claramente las etapas y el recorrido que se debe seguir en el diagrama de flujo.

Fuente: Adaptado de WordPress (n.d.).

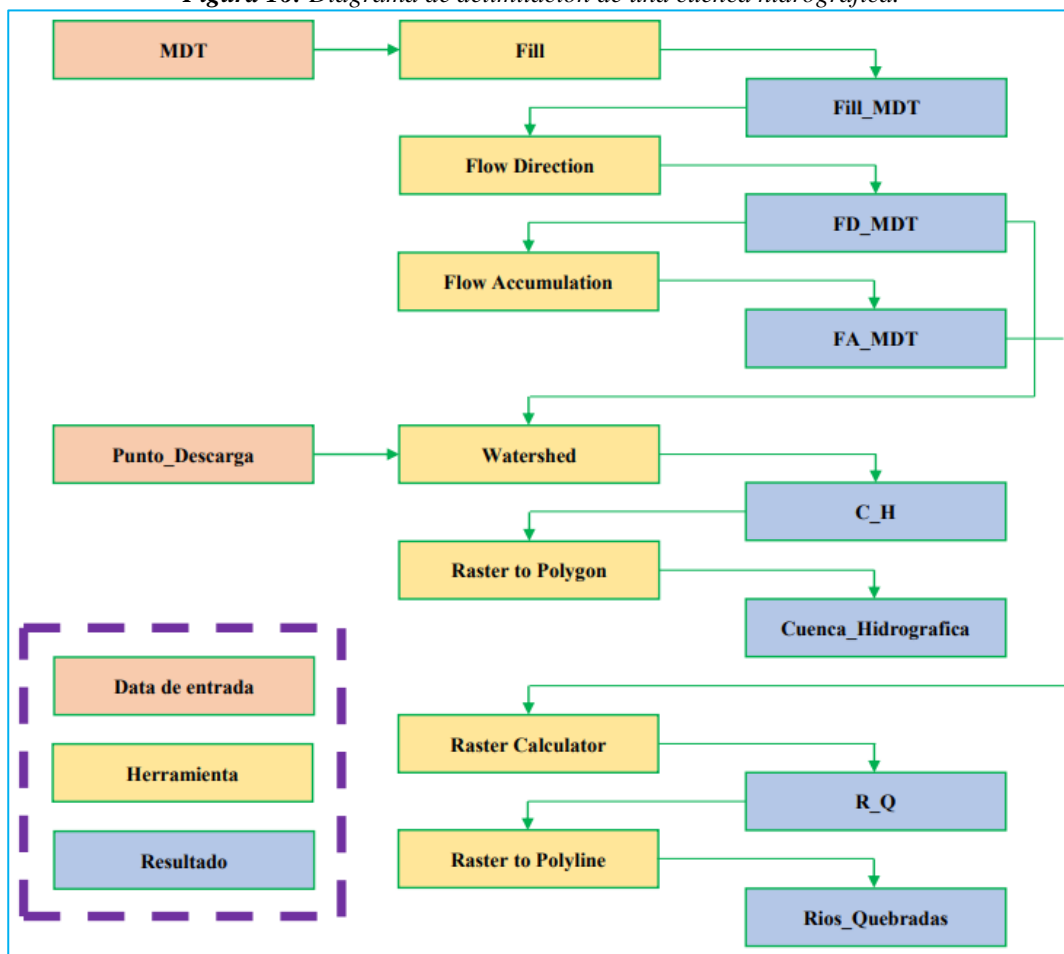
3 Materiales y Métodos

3.1 Recopilación de datos correspondientes a la cuenca hidrográfica del Río Yanuncay

El primer paso consistió en realizar la delimitación de la cuenca del Río Yanuncay, partiendo de un Modelo de Elevación Digital (MDT), el cual fue adquirido del Servicio Geológico de los Estados Unidos (USGS) (ver Figura 16). Este modelo proporcionó datos altamente detallados sobre la topografía de la zona, permitiendo así una delimitación efectiva de la cuenca hidrográfica.

Una vez obtenido el MDT, se empleó el software ArcMap 10.5 para llevar a cabo el proceso de depuración de la zona de estudio. Para lograrlo, se hicieron uso de diversas herramientas disponibles en el programa, tales como Hydrology y Map Algebra, tal como se muestra en la Figura 16. Estas herramientas, diseñadas específicamente para el análisis hidrológico, desempeñaron un papel clave en la limpieza y preparación de los datos geoespaciales.

Figura 16: Diagrama de delimitación de una cuenca hidrográfica.



Fuente: Elaboración propia.

3.2 Cálculo del Índice de la Fundación Nacional de Saneamiento (ICA NSF)

Para realizar el cálculo del ICA NSF en la cuenca del Río Yanuncay, se llevó a cabo un proceso de tratamiento de los datos proporcionados por la empresa ETAPA E.P. Se aplicaron técnicas de limpieza y filtrado utilizando macros en el lenguaje de programación de Excel, con el fin de eliminar valores nulos y atípicos que pudieran afectar la precisión de los resultados.

Durante esta fase de procesamiento, se llevó a cabo una minuciosa identificación de las fechas, estaciones de monitoreo y los distintos parámetros físicos, químicos y biológicos que eran relevantes para el análisis de la calidad del agua en la cuenca (Figura 17). Entre los parámetros seleccionados se incluyeron el oxígeno disuelto, los coliformes fecales, el potencial de hidrógeno, la demanda bioquímica de oxígeno en 5 días, los nitratos, los fosfatos, la temperatura, la turbiedad y los sólidos totales.

3.2.1 Determinación de los factores de escala Q_i

El conjunto de datos utilizado para calcular los factores de escala Q_i para cada parámetro se presenta en forma de una matriz. En esta estructura, las columnas representan las distintas características que corresponden a los parámetros, mientras que cada fila contiene la información de un monitoreo específico. Tener en cuenta que N°, Estación y Fecha se les consideran como meta atributos dentro de la matriz. Este formato organizado se ilustra claramente en la Figura 17.

Figura 17: Conjunto de datos para el cálculo del ICA NSF.

N°	Estación	Fecha	OD	CF	pH	DBO5	N	P	CT	T	ST
1											
...											
292											

Fuente: Elaboración propia.

La metodología de utilizar ecuaciones en lugar de curvas de estandarización asociadas a cada parámetro se basa en el trabajo realizado por Jiménez y Vélez (2006). Estos investigadores llevaron a cabo ajustes polinómicos en las curvas de estandarización con el fin de agilizar el proceso. Esta decisión se tomó debido a la complejidad que implica obtener directamente los factores de escala Q_i a partir de los diagramas.

Como parte de la metodología utilizada, las ecuaciones se implementaron en el lenguaje de programación Python para calcular los valores de los factores de escala correspondientes. Esto permitió automatizar los cálculos y obtener resultados precisos de manera eficiente.

- **Oxígeno Disuelto (% de saturación)**

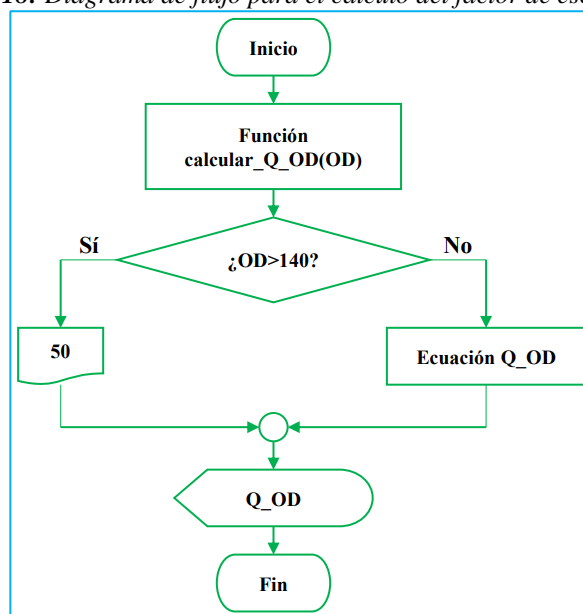
Ecuación 2: Ecuación para el cálculo del factor de escala Q_{OD} .

$$Q_{OD} = 3.1615 \cdot 10^{-8} \cdot (OD)^5 - 1.0304 \cdot 10^{-5} \cdot (OD)^4 + 1.0076 \cdot 10^{-3} \cdot (OD)^3 - 2.7883 \cdot 10^{-2} \cdot (OD)^2 + 8.4068 \cdot 10^{-1} \cdot (OD) - 1.612 \cdot 10^{-1}$$

Fuente: Jiménez & Vélez (2006).

En el cálculo del factor de escala Q_{OD} , se establece una función específica para determinar su valor basándose en la condición $OD > 140$. Si esta condición es verdadera, es decir, si el valor de OD es mayor a 140, el factor Q_{OD} se establece en 50 de manera directa. Sin embargo, si la respuesta es negativa, lo cual implica que el valor de OD es menor o igual a 140, entonces se utiliza la Ecuación 2 para calcular el valor de Q_{OD} (ver Figura 18).

Figura 18: Diagrama de flujo para el cálculo del factor de escala Q_{OD} .



Fuente: Elaboración propia.

- **Coliformes Fecales (NMP/100mL)**

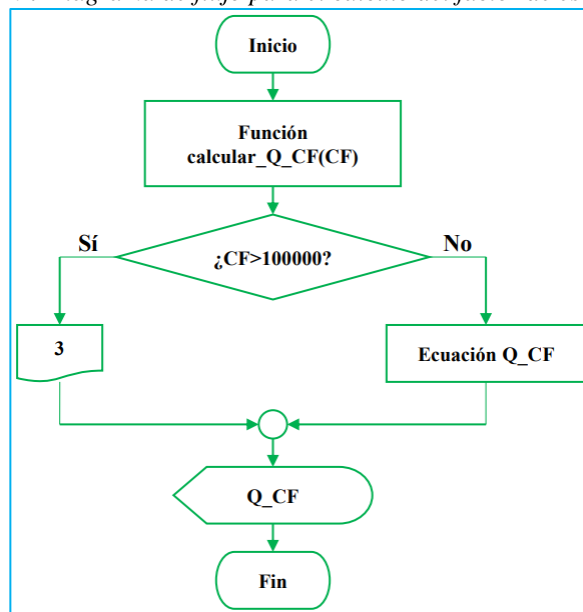
Ecuación 3: Ecuación para el cálculo del factor de escala Q_{CF} .

$$Q_{CF} = e^{-0.0152 \cdot (\ln(CF))^2 - 0.1063 \cdot (\ln(CF)) + 4.5922}$$

Fuente: Adaptado de Jiménez & Vélez (2006).

En el proceso de cálculo del factor de escala Q_{CF} , se establece una función específica para determinar su valor basándose en la condición $CF > 100000$. Si esta condición es verdadera, lo que indica que el valor de CF es mayor a 100000, el factor Q_{CF} se establece directamente en 3. Sin embargo, si la respuesta es negativa, es decir, si el valor de CF es menor o igual a 100000, entonces se recurre al uso de la Ecuación 3 para calcular el valor de Q_{CF} (ver Figura 19).

Figura 19: Diagrama de flujo para el cálculo del factor de escala Q_{CF} .



Fuente: Elaboración propia.

- **Potencial de Hidrógeno (Unidades)**

Ecuación 4: Ecuación para el cálculo del factor de escala Q_{pH} .

$$Q_{pH > 7.5} = -1.11429 \cdot (pH)^4 + 44.50952 \cdot (pH)^3 - 656.6 \cdot (pH)^2 + 4215.34762 \cdot (pH) - 9840.14286$$

Fuente: Jiménez & Vélez (2006).

Ecuación 5: Ecuación para el cálculo del factor de escala Q_{pH} .

$$Q_{pH:7.5} = -0.1789 \cdot (pH)^5 + 3.7932 \cdot (pH)^4 - 30.517 \cdot (pH)^3 + 119.75 \cdot (pH)^2 - 224.58 \cdot (pH) + 159.46$$

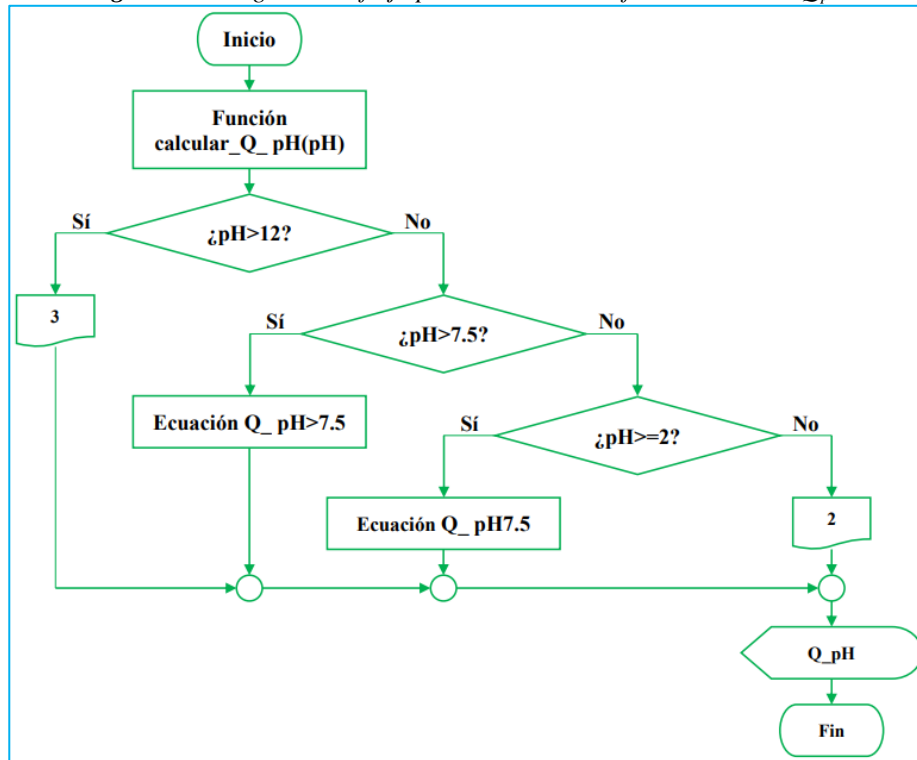
Fuente: Jiménez & Vélez (2006).

Para calcular el factor de escala Q_{pH} , se sigue una metodología que utiliza tres condicionantes (ver Figura 20). Estos son los siguientes:

1. Si el valor de pH es mayor a 12, se establece directamente el valor de Q_{pH} en 3. En caso contrario, se pasa al siguiente condicionante.

2. Si el pH es mayor a 7.5, se utiliza la Ecuación 4 para calcular el valor de Q_{pH} . Si el pH es menor o igual a 7.5, se pasa al siguiente condicionante.
3. Si el pH es igual o mayor a 2, se emplea la Ecuación 5 para determinar el valor de Q_{pH} . En caso contrario, el valor de Q_{pH} se establece en 2.

Figura 20: Diagrama de flujo para el cálculo del factor de escala Q_{pH} .



Fuente: Elaboración propia.

- **Demanda Bioquímica de Oxígeno en 5 días (mg/L)**

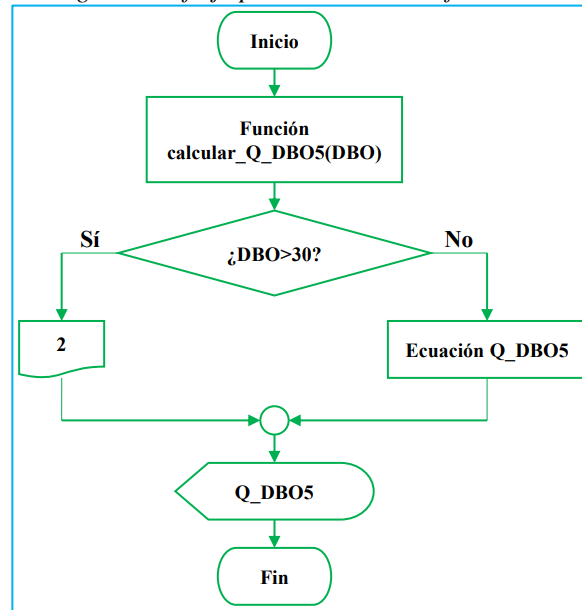
Ecuación 6: Ecuación para el cálculo del factor de escala Q_{DBO5} .

$$Q_{DBO5} = 1.8677 \cdot 10^{-4} \cdot (DBO)^4 - 1.6615 \cdot 10^{-2} \cdot (DBO)^3 + 5.9636 \cdot 10^{-1} \cdot (DBO)^2 - 1.1152 \cdot 10^1 \cdot (DBO) + 1.0019 \cdot 10^2$$

Fuente: Jiménez & Vélez (2006).

En el cálculo del factor de escala Q_{DBO5} , se establece una función específica para determinar su valor basándose en el condicionante $DBO > 30$. Si esta condición es verdadera, es decir, si el valor de DBO es mayor a 30, el factor Q_{DBO5} se establece directamente en 2. Sin embargo, si la respuesta es negativa, es decir, si el valor de DBO es menor o igual a 30, entonces se recurre al uso de la Ecuación 6 para calcular el valor de Q_{DBO5} (ver Figura 21).

Figura 21: Diagrama de flujo para el cálculo del factor de escala Q_{DBO5} .



Fuente: Elaboración propia.

- **Nitratos (mg/L)**

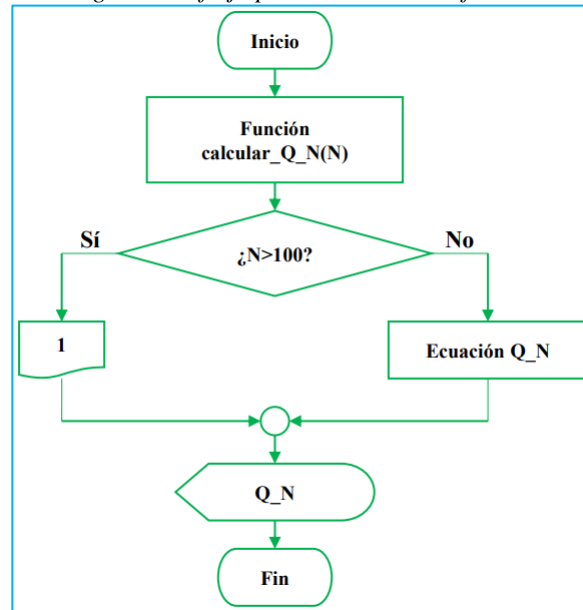
Ecuación 7: Ecuación para el cálculo del factor de escala Q_N .

$$Q_N = 3.5603 \cdot 10^{-9} \cdot (N)^6 - 1.2183 \cdot 10^{-6} \cdot (N)^5 + 1.6238 \cdot 10^{-4} \cdot (N)^4 - 1.0693 \cdot 10^{-2} \cdot (N)^3 + 3.7304 \cdot 10^{-1} \cdot (N)^2 - 7.521 \cdot (N) + 1.0095 \cdot 10^2$$

Fuente: Jiménez & Vélez (2006).

En el cálculo del factor de escala Q_N , se utiliza una función específica que determina su valor en base al condicionante $N > 100$. Si esta condición es verdadera, es decir, si el valor de N es mayor a 100, el factor Q_N se establece directamente en 1. Sin embargo, si la respuesta es negativa, es decir, si el valor de N es menor o igual a 100, se procede al uso de la Ecuación 7 para calcular el valor de Q_N (ver Figura 22).

Figura 22: Diagrama de flujo para el cálculo del factor de escala Q_N .



Fuente: Elaboración propia.

- **Fosfatos (mg/L)**

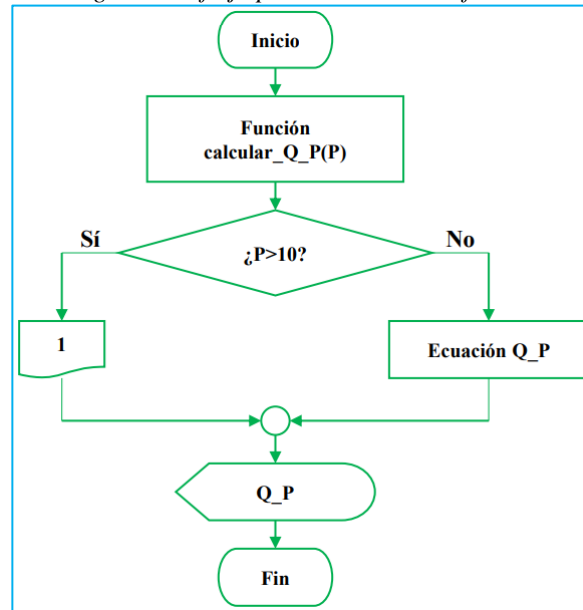
Ecuación 8: Ecuación para el cálculo del factor de escala Q_P .

$$Q_P = 4.6732 \cdot 10^{-3} \cdot (P)^6 - 1.6167 \cdot 10^{-1} \cdot (P)^5 + 2.20595 \cdot (P)^4 - 1.50504 \cdot 10^1 \cdot (P)^3 + 5.38893 \cdot 10^1 \cdot (P)^2 - 9.98933 \cdot 10^1 \cdot (P) + 9.98311 \cdot 10^1$$

Fuente: Jiménez & Vélez (2006).

En el cálculo del factor de escala Q_P , se implementa una función que determina su valor en función del condicionante $P > 10$. Si esta condición es verdadera, es decir, si el valor de P es mayor a 10, el factor Q_P se establece directamente en 1. Sin embargo, si la respuesta es negativa, es decir, si el valor de P es menor o igual a 10, se recurre al uso de la Ecuación 8 para calcular el valor de Q_P (ver Figura 23).

Figura 23: Diagrama de flujo para el cálculo del factor de escala Q_P .



Fuente: Elaboración propia.

- **Cambio de Temperatura (°C)**

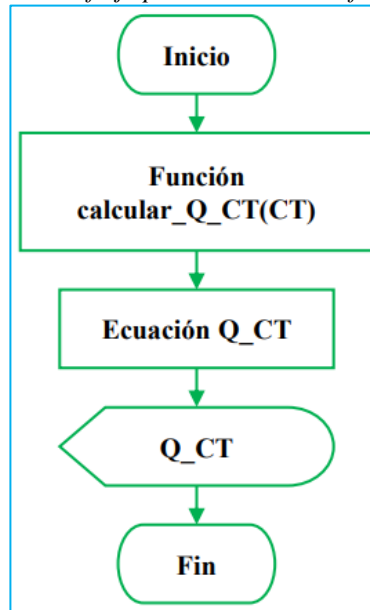
Ecuación 9: Ecuación para el cálculo del factor de escala Q_{CT} .

$$Q_{\Delta T} = 1.9619 \cdot 10^{-6} \cdot (\Delta T)^6 - 1.3964 \cdot 10^{-4} \cdot (\Delta T)^5 + 2.5908 \cdot 10^{-3} \cdot (\Delta T)^4 + 1.5398 \cdot 10^{-2} \cdot (\Delta T)^3 - 6.7952 \cdot 10^{-1} \cdot (\Delta T)^2 - 6.7204 \cdot 10^{-1} \cdot (\Delta T) + 9.0392 \cdot 10^1$$

Fuente: Jiménez & Vélez (2006).

En el cálculo del factor de escala Q_{CT} , se utiliza una función específica que se basa en la Ecuación 9 para determinar su valor. Esta ecuación es aplicada directamente en el proceso de cálculo del factor Q_{CT} (ver Figura 24). Esta metodología garantiza una evaluación precisa del factor de escala Q_{CT} , ya que se utiliza una fórmula específica diseñada para este propósito.

Figura 24: Diagrama de flujo para el cálculo del factor de escala Q_{CT} .



Fuente: Elaboración propia.

- **Turbiedad (NTU)**

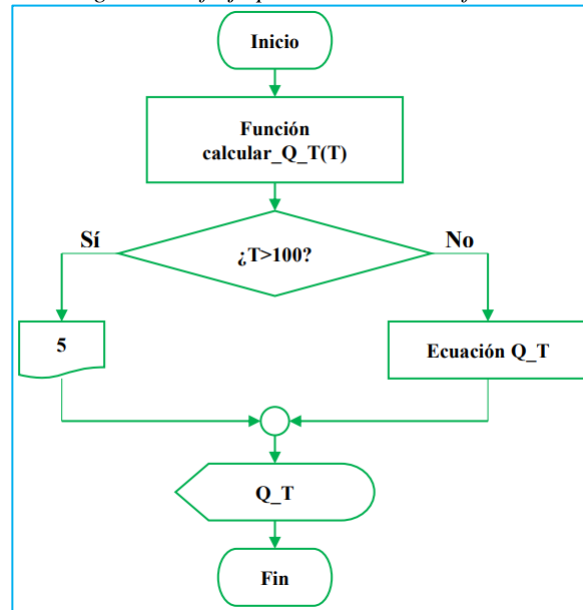
Ecuación 10: Ecuación para el cálculo del factor de escala Q_T .

$$Q_T = 1.8939 \cdot 10^{-6} \cdot (T)^4 - 4.9942 \cdot 10^{-4} \cdot (T)^3 + 4.9181 \cdot 10^{-2} \cdot (T)^2 - 2.6284 \cdot (T) + 9.8098 \cdot 10^1$$

Fuente: Jiménez & Vélez (2006).

En el cálculo del factor de escala Q_T , se utiliza una función específica que determina su valor en base al condicionante $T > 100$. Si esta condición es verdadera, es decir, si el valor de T es mayor a 100, el factor Q_T se establece directamente en 5. Sin embargo, si la respuesta es negativa, es decir, si el valor de T es menor o igual a 100, se procede al uso de la Ecuación 10 para calcular el valor de Q_T (ver Figura 25).

Figura 25: Diagrama de flujo para el cálculo del factor de escala Q_T .



Fuente: Elaboración propia.

- **Sólidos Totales (mg/L)**

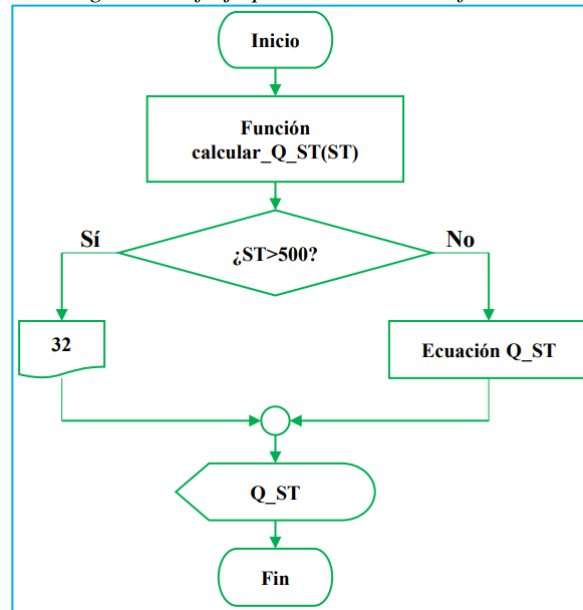
Ecuación 11: Ecuación para el cálculo del factor de escala Q_{ST} .

$$Q_{ST} = -4.4289 \cdot 10^{-9} \cdot (ST)^4 + 4.65 \cdot 10^{-6} \cdot (ST)^3 - 1.9591 \cdot 10^{-3} \cdot (ST)^2 + 1.8973 \cdot 10^{-1} \cdot (ST) + 8.0608 \cdot 10^1$$

Fuente: Jiménez & Vélez (2006).

En el proceso de cálculo del factor de escala Q_{ST} , se establece una función específica para determinar su valor basándose en el condicionante $ST > 500$. Si esta condición es verdadera, lo que indica que el valor de ST es mayor a 500, el factor Q_{ST} se establece directamente en 32. Sin embargo, si la respuesta es negativa, es decir, si el valor de ST es menor o igual a 500, entonces se recurre al uso de la Ecuación 11 para calcular el valor de Q_{ST} (ver Figura 26).

Figura 26: Diagrama de flujo para el cálculo del factor de escala Q_{ST} .



Fuente: Elaboración propia.

3.2.2 Rangos de clasificación para el ICA NSF

Una vez completados los cálculos de los factores de escala (Q_i) utilizando la Ecuación 1 y considerando los pesos relativos (W_i) asociados a cada parámetro, se procede al cálculo del ICA NSF. Sin embargo, es importante destacar que, para realizar la clasificación del ICA, es necesario definir una función que asignará niveles a los valores obtenidos.

Esta función de clasificación permitirá asignar una categoría específica al ICA NSF basándose en los valores numéricos obtenidos. La metodología para la asignación de categorías se encuentra detallada en la sección 3.7 del capítulo de Fundamentación Teórica.

Al aplicar esta función a cada valor numérico del ICA NSF calculado, se asignará la categoría correspondiente según los criterios establecidos. Esto facilitará la interpretación y la comunicación de los resultados obtenidos, ya que se podrán identificar diferentes niveles de calidad del agua en función de las categorías asignadas.

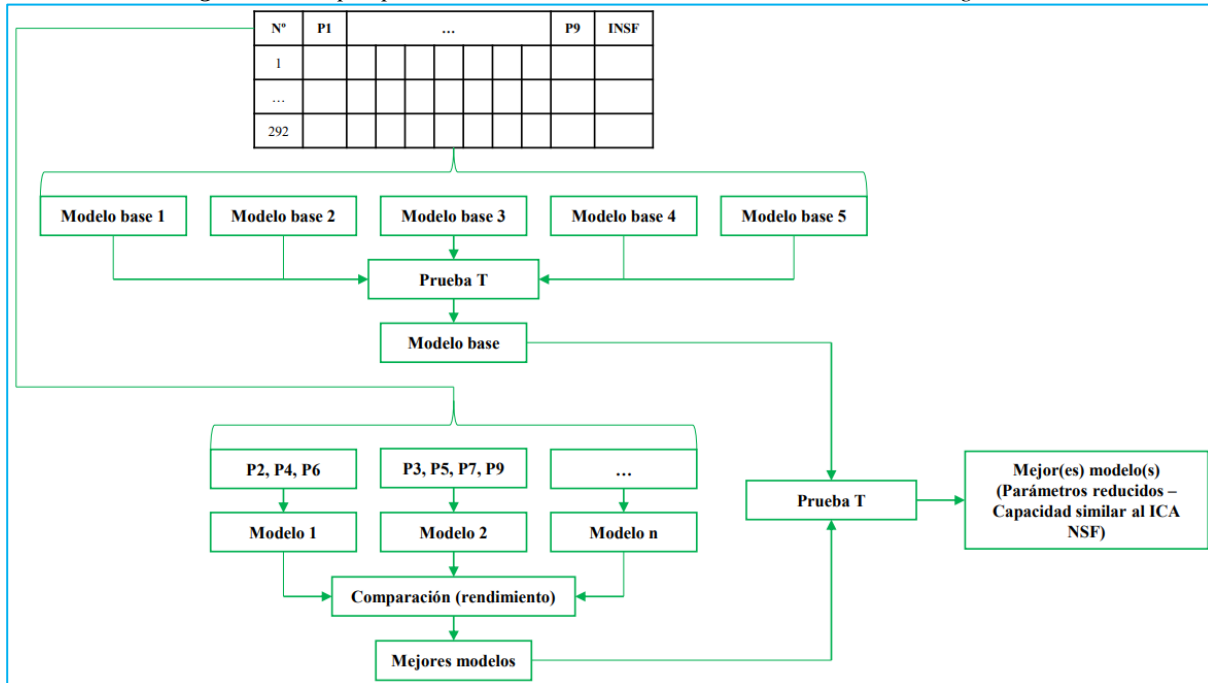
3.3 Diseño de los modelos basados en datos

Con el ICA NSF calculado, se procederá a crear modelos basados en Árboles de regresión. Para esto se cuenta con 4 etapas (ver Figura 27), siendo las siguientes:

1. Creación del conjunto de datos

2. Definición del modelo base
3. Creación y modelado de subconjuntos de parámetros
4. Determinación de la similitud con el modelo base

Figura 27: Etapas para la creación de modelos basados en árboles de regresión.



Fuente: Elaboración propia.

A continuación, se detallan las etapas para la creación de los modelos basados en datos.

3.3.1 Creación del conjunto de datos

El conjunto de datos utilizado para crear los modelos basados en Árboles de regresión se construyó a partir de información recopilada de las 23 estaciones de monitoreo. Esta información se estructuró en forma de matriz, donde cada fila representa un monitoreo específico y las columnas contienen las características, que son los parámetros y el ICA NSF. De igual manera, Nº, Estación y Fecha se les consideran como meta tributos dentro de la matriz. Un ejemplo visual de esta estructura se puede observar en la Figura 28.

Figura 28: Ejemplificación del conjunto de datos.

N°	Estación	Fecha	OD	CF	pH	DBO5	N	P	CT	T	ST	INSF
1												
...												
292												

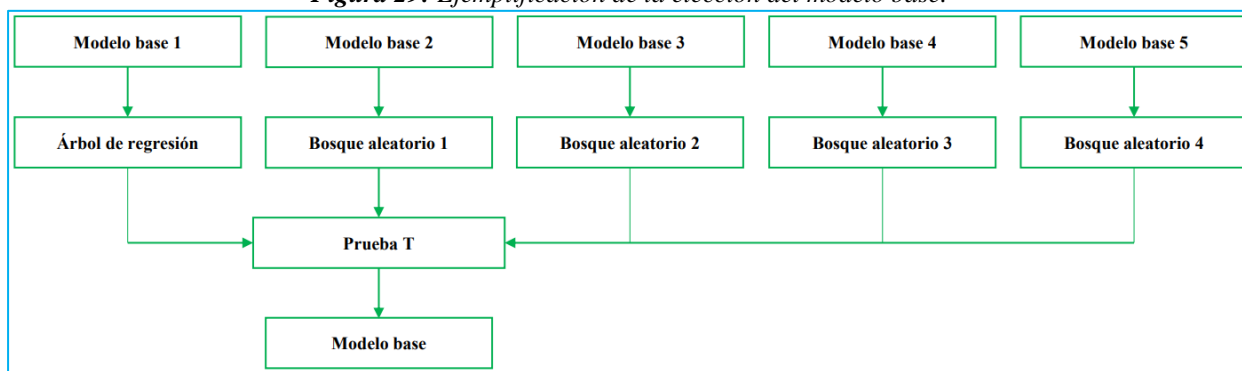
Fuente: Elaboración propia.

Es importante destacar que los valores del ICA NSF obtenidos en el cálculo anterior deben interpretarse como los valores reales de cada ejemplo en el conjunto de datos. Este enfoque garantiza que la muestra sea representativa de las condiciones reales y permita un análisis adecuado de la calidad del agua.

3.3.2 Definición del modelo base

Con el fin de establecer un modelo base, se desarrollaron cinco modelos en total. Uno de ellos corresponde a un modelo de Árbol de regresión, mientras que los otros cuatro son modelos de Bosque aleatorio para regresión. El propósito de esta creación de modelos es identificar al mejor modelo que se ajuste de manera precisa a la estimación de la calidad del agua utilizando el ICA NSF como variable principal. Esta información se encuentra ilustrada en la Figura 29, proporcionando una visualización clara de los modelos y su evaluación comparativa.

Figura 29: Ejemplificación de la elección del modelo base.



Fuente: Elaboración propia.

La utilización de diferentes enfoques de modelado, como los modelos de Árbol de regresión y los modelos de Bosque aleatorio para regresión, permite explorar diferentes técnicas y enfoques para lograr una estimación precisa de la calidad del agua. Cada uno de estos modelos tiene sus propias fortalezas y características particulares, y la comparación entre ellos nos permite identificar el modelo más adecuado para el propósito específico de la estimación del ICA NSF:

1. **Modelo base 1 (Árbol de regresión):** Este modelo de Árbol de regresión se construye utilizando el algoritmo CART, el cual se basa en el concepto de impureza para tomar decisiones sobre qué variables y puntos de corte utilizar en cada división del árbol. Las medidas de impureza comunes para los árboles de regresión son la varianza o el error cuadrático medio. Los parámetros, como la profundidad del árbol se configuran en su valor máximo, mientras que el número de muestras en una hoja y el número de muestras para dividir un nodo se configuran en su valor mínimo. Esto posibilita que el árbol crezca hasta ajustarse perfectamente a los datos de entrenamiento.
2. **Modelo base 2 (Bosque aleatorio 1):** Para crear modelos de Bosque aleatorio mediante regresión, se genera cada árbol del conjunto utilizando una muestra extraída con reemplazo del conjunto de entrenamiento, lo que se conoce como muestra de arranque. En este caso, se emplean 100 árboles en el bosque y se consideran todas las características y ejemplos disponibles. Además, se configura la profundidad máxima de los árboles para permitir su expansión hasta que todas las hojas sean puras, es decir, contengan solo instancias de una única clase o valor.
3. **Modelo base 3 (Bosque aleatorio 2):** Para crear el segundo modelo de Bosque aleatorio mediante regresión, se construye cada árbol del conjunto utilizando una muestra extraída con reemplazo del conjunto de entrenamiento, conocida como muestra de arranque. En esta ocasión, se utilizan 100 árboles en el bosque y se consideran todos los ejemplos disponibles. El número de características seleccionadas para cada árbol se elige al azar, siendo este igual a la raíz cuadrada del número total de características. Además, se configura la profundidad máxima de los árboles para permitir su crecimiento hasta que todas las hojas sean puras, es decir, solo contengan instancias de una única clase o valor.
4. **Modelo base 4 (Bosque aleatorio 3):** Para crear el tercer modelo de Bosque aleatorio mediante regresión, se genera cada árbol del conjunto utilizando una muestra extraída con reemplazo del conjunto de entrenamiento, conocida como muestra de arranque. En este caso, se emplean 100 árboles en el bosque y se consideran todos los ejemplos disponibles. El número de muestras seleccionadas para entrenar cada árbol se elige al azar, utilizando un subconjunto de datos que contiene el 70% de las muestras. Además, se configura la profundidad máxima de los árboles para permitir su crecimiento hasta que todas las hojas sean puras, es decir, contengan solo instancias de una única clase o valor.

5. **Modelo base 5 (Bosque aleatorio 4):** Para crear el cuarto modelo de Bosque Aleatorio mediante regresión, se construye cada árbol del conjunto utilizando una muestra extraída con reemplazo del conjunto de entrenamiento, conocida como muestra de arranque. En esta ocasión, se utilizan 100 árboles en el bosque. El número de características seleccionadas para cada árbol se elige al azar, siendo este igual a la raíz cuadrada del número total de características. Asimismo, el número de ejemplos seleccionados para entrenar cada árbol se decide de manera aleatoria, utilizando un subconjunto de datos que contiene el 70% de las muestras disponibles. Además, se configura la profundidad máxima de los árboles para permitir su crecimiento hasta que todas las hojas sean puras, es decir, solo contengan instancias de una única clase o valor.

Para evaluar el rendimiento de cada modelo de manera objetiva, se emplea la validación cruzada con 10 divisiones. Esto implica dividir el conjunto de datos en 10 partes y realizar 10 iteraciones, utilizando una parte diferente como conjunto de prueba en cada iteración. Se calculan estadísticas resumidas, como el promedio y la desviación estándar de los puntajes obtenidos en cada iteración, lo cual proporciona una medida del rendimiento del modelo de manera generalizada. Estas estadísticas resumidas permiten una comparación objetiva del desempeño de cada modelo.

Se realizarán pruebas de hipótesis entre todos los modelos de regresión creados para determinar el modelo base. El objetivo es evaluar si existe evidencia de diferencias significativas entre los modelos y seleccionar el Bosque aleatorio 1 como modelo base debido a la simplicidad en la aplicación de sus valores de parámetros. Los resultados revelarán que los modelos de Bosque aleatorio por regresión son similares, ya que no se encontrará evidencia de diferencias significativas entre ellos. Sin embargo, es importante destacar que el modelo de Árbol de regresión puede diferir, ya que podrían encontrarse indicios de diferencias significativas con respecto al Bosque aleatorio 1.

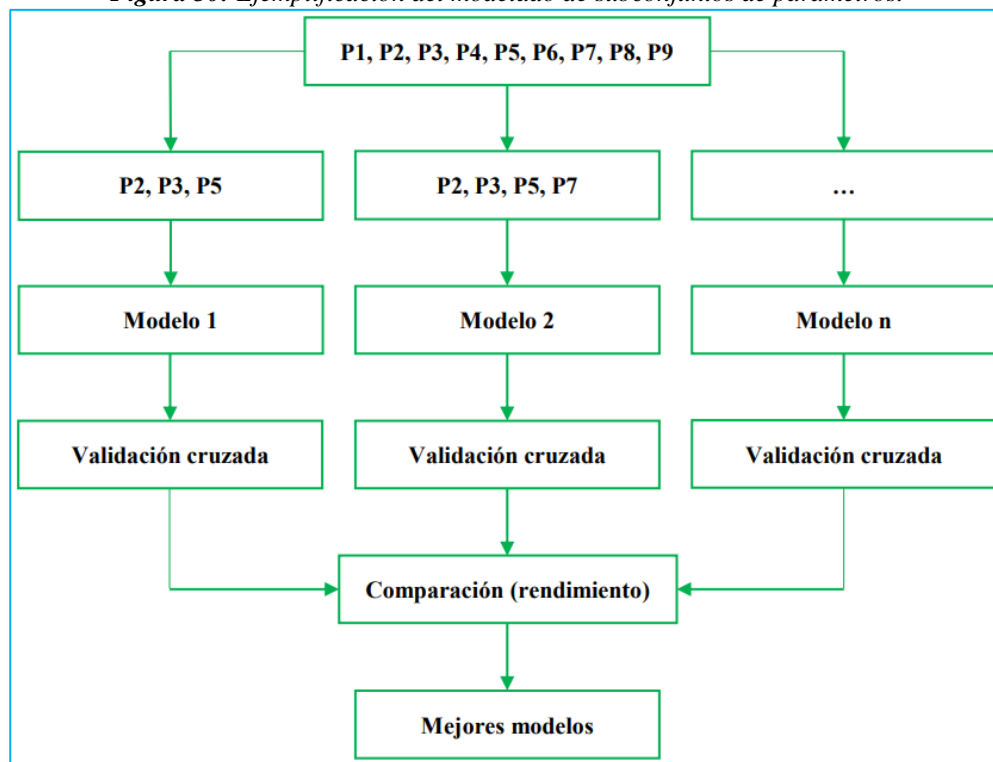
3.3.3 Creación y modelado de subconjuntos de parámetros

Para generar modelos con subconjuntos de parámetros reducidos, se empleó el mismo conjunto de datos utilizado para crear los modelos base (ver Figura 28). A partir de los 9 parámetros disponibles, se generaron subconjuntos que abarcaban todas las combinaciones posibles. Esto implica considerar diferentes combinaciones de los parámetros, seleccionando algunos y excluyendo otros en cada subconjunto. Al crear estos subconjuntos, se exploraron todas las

posibles configuraciones de parámetros para evaluar su impacto en el desempeño de los modelos. Esto permitirá determinar qué combinaciones de parámetros son más relevantes o influyentes en la estimación del modelo y cuáles podrían ser omitidas sin afectar significativamente su rendimiento.

La cantidad de subconjuntos que se pueden formar a partir de todas las posibles combinaciones se determinó utilizando la expresión 2 elevado a la 9, lo cual da como resultado 512. El código binario, que utiliza los dígitos 0 y 1, es fundamental en este proceso. Cada dígito binario representa un estado específico, donde el 0 indica “ausencia de características” y el 1 indica “presencia de características”. En el caso de formar subconjuntos, se consideran todas las combinaciones posibles excepto la combinación de todos los dígitos en 0, ya que en ese caso no habría ninguna característica seleccionada en el subconjunto. Por lo tanto, el número total de subconjuntos utilizable en este contexto es de 511, ya que se excluye el conjunto vacío.

Figura 30: Ejemplificación del modelado de subconjuntos de parámetros.



Fuente: Elaboración propia.

En la Figura 30, se presenta el proceso de creación de modelos de Bosque aleatorio por regresión para cada uno de los subconjuntos. Para construir estos modelos, se siguió el

procedimiento establecido para el Bosque Aleatorio 1, el cual se caracteriza por su simplicidad en la aplicación de los valores de sus parámetros. En cada iteración de construcción del modelo, se evaluó su rendimiento utilizando la validación cruzada con 10 divisiones. Posteriormente, se calcularon estadísticas resumidas, como el promedio y la desviación estándar, de los puntajes obtenidos en cada iteración.

3.3.4 Determinación de la similitud con el modelo base

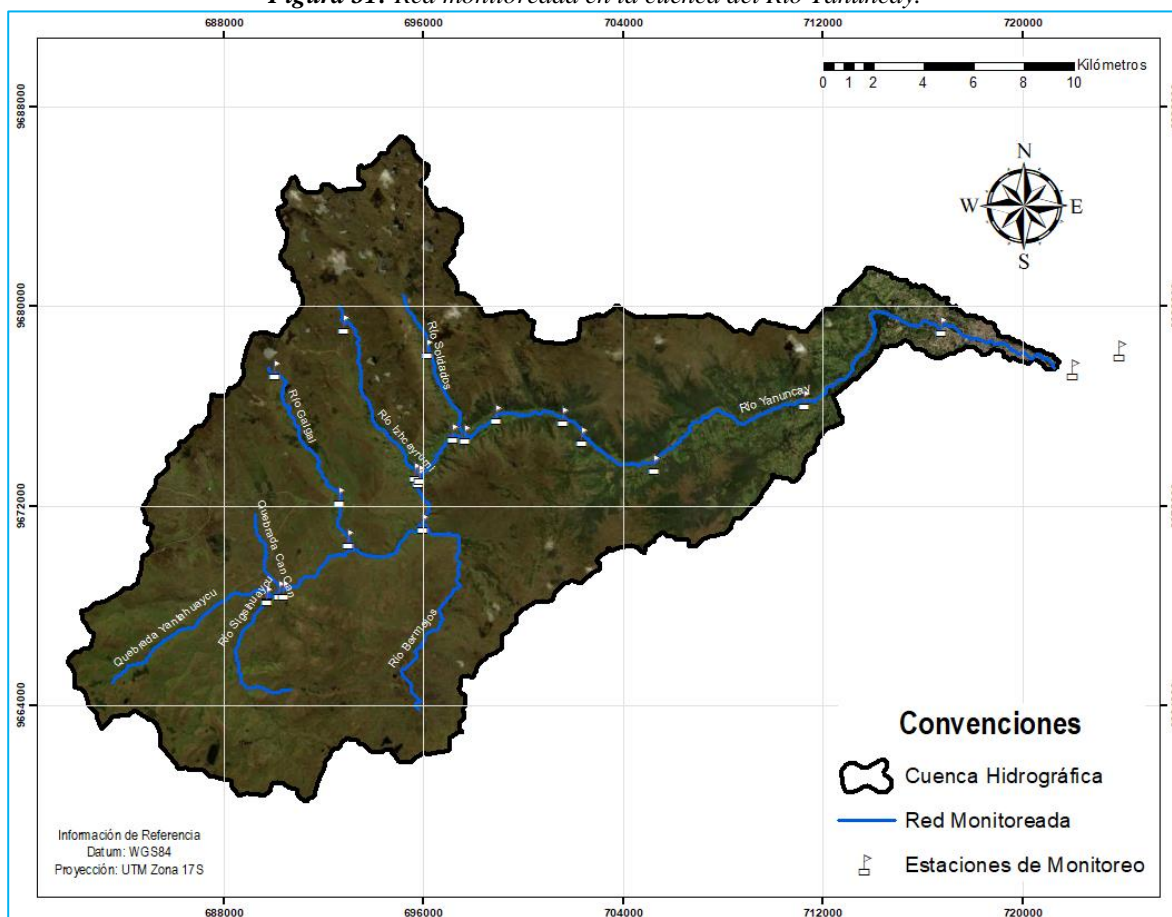
Para finalizar, se realizarán comparaciones y pruebas de hipótesis, donde se evaluará si existen evidencias de diferencias estadísticamente significativas entre los modelos propuestos y el modelo base. Esto nos brindará una comprensión más profunda de la efectividad y el potencial de los modelos basados en Árboles de regresión en la evaluación y estimación de la calidad del agua en la cuenca del Río Yanuncay. Con esta información, podremos tomar decisiones fundamentadas y establecer conclusiones sólidas que nos permitan identificar el modelo más apropiado para nuestras necesidades, considerando tanto el número de parámetros como la capacidad de evaluación y estimación en comparación con el ICA NSF.

4 Resultados

4.1 Calidad del agua obtenida en cada estación de monitoreo correspondientes a la cuenca del Río Yanuncay

En el Anexo 1 se encuentra disponible información detallada sobre las 23 estaciones de monitoreo de la cuenca del Río Yanuncay. Cada estación posee un código único que corresponde a su respectivo río o quebrada, así como también se proporcionan las coordenadas geográficas y la altitud asociada a cada ubicación. Además, como identificación adicional, se ha asignado un nombre a cada estación utilizando la letra “E” seguida de un número del 1 al 23.

Figura 31: Red monitoreada en la cuenca del Río Yanuncay.



Fuente: Elaboración propia.

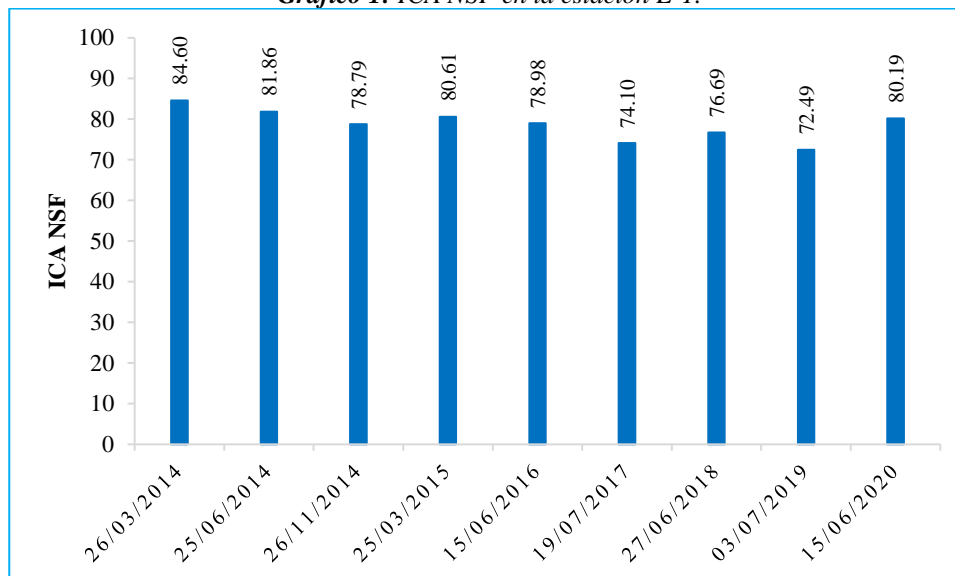
La Figura 31 proporciona una representación visual clara de la ubicación espacial de las estaciones de monitoreo en la cuenca. Tomando en cuenta la altitud, se puede observar que las estaciones están estratégicamente distribuidas a lo largo de los ríos y quebradas que forman parte

de esta cuenca. Esto garantiza una cobertura adecuada de la red fluvial y proporciona una base sólida para el monitoreo y la comprensión de los recursos hídricos en la zona.

1. Estación 1

La estación de monitoreo E-1, identificada con el código YAN-YH-CONTROL y ubicada en la Quebrada Yantahuaycu, ha sido objeto de un exhaustivo seguimiento a lo largo de un periodo de seis años, desde 2014 hasta 2020. Durante este tiempo, se han llevado a cabo un total de nueve monitoreos en dicha estación, con el objetivo de evaluar la calidad del agua en la zona y obtener una comprensión clara de las condiciones hídricas en la Quebrada Yantahuaycu a lo largo del tiempo.

Gráfico 1: ICA NSF en la estación E-1.



Fuente: Elaboración propia.

Al analizar el Gráfico 1, se observa que la estación E-1 ha registrado valores notables en cuanto a la calidad del agua. En particular, se ha identificado que el punto de mayor calidad del agua se registró el 26 de marzo de 2014, alcanzando un valor de 84.60. Por otro lado, el punto de menor calidad del agua fue registrado el 3 de julio de 2019, con un valor de 72.49. Ambos valores se clasifican cualitativamente como “Buena” calidad del agua.

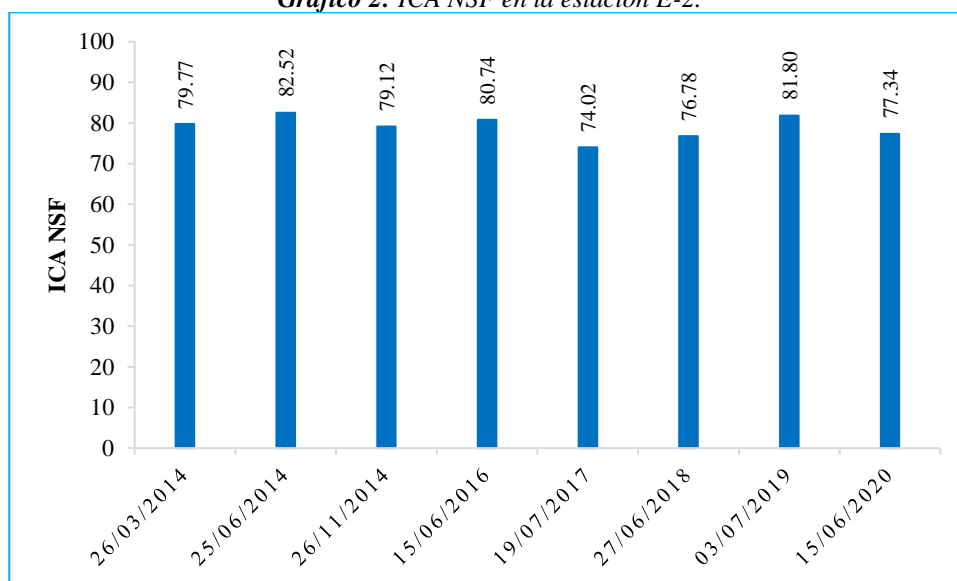
Estos resultados indican que, a lo largo de los seis años de monitoreo, la calidad del agua en la estación E-1 se ha mantenido relativamente constante. Los valores obtenidos, tanto el más alto como el más bajo, indican una calidad del agua dentro de los parámetros considerados como

buenos. Esto sugiere una estabilidad en las condiciones hídricas de la Quebrada Yantahuaycu durante el período de estudio.

2. Estación 2

La estación de monitoreo E-2, identificada con el código YAN-SH-CONTROL y ubicada en el Río Sigsihuaycu, ha sido objeto de un seguimiento detallado durante un periodo de seis años, desde 2014 hasta 2020. Durante este tiempo, se han llevado a cabo un total de ocho monitoreos en dicha estación, con el propósito de evaluar la calidad del agua en la zona y obtener una comprensión clara de las condiciones hídricas en el Río Sigsihuaycu a lo largo del tiempo.

Gráfico 2: ICA NSF en la estación E-2.



Fuente: Elaboración propia.

Al examinar el Gráfico 2, se destaca que la estación E-2 ha registrado valores significativos en relación a la calidad del agua. En particular, se ha identificado que el punto de mayor calidad del agua se registró el 25 de junio de 2014, alcanzando un valor de 82.52. Por otro lado, el punto de menor calidad del agua se registró el 19 de julio de 2017, con un valor de 74.02. Ambos valores se clasifican cualitativamente como “Buena” calidad del agua.

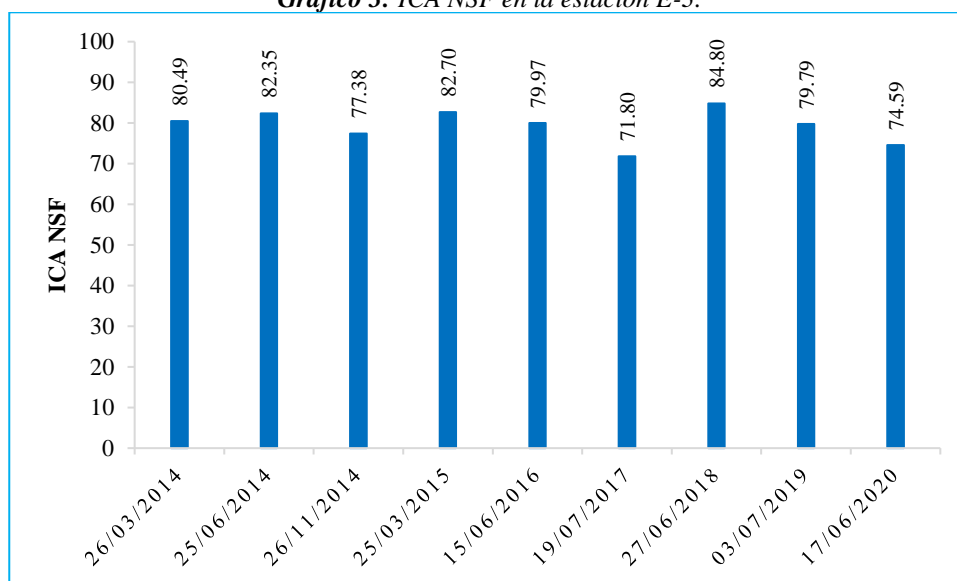
Estos resultados indican que, a lo largo de los seis años de monitoreo, la calidad del agua en la estación E-2 se ha mantenido relativamente constante. Los valores obtenidos, tanto el más alto como el más bajo, se encuentran dentro de los parámetros considerados como buena calidad del

agua. Esto sugiere una estabilidad en las condiciones hídricas del Río Sigsihuaycu durante el período de estudio.

3. Estación 3

La estación de monitoreo E-3, identificada con el código YAN-CAN-CONTROL y ubicada en la Quebrada Can Can, ha sido objeto de un seguimiento exhaustivo durante seis años, desde 2014 hasta 2020. Durante este periodo, se han llevado a cabo un total de nueve monitoreos en dicha estación, con el objetivo de evaluar la calidad del agua en la zona y obtener una comprensión clara de las condiciones hídricas en la Quebrada Can Can a lo largo del tiempo.

Gráfico 3: ICA NSF en la estación E-3.



Fuente: Elaboración propia.

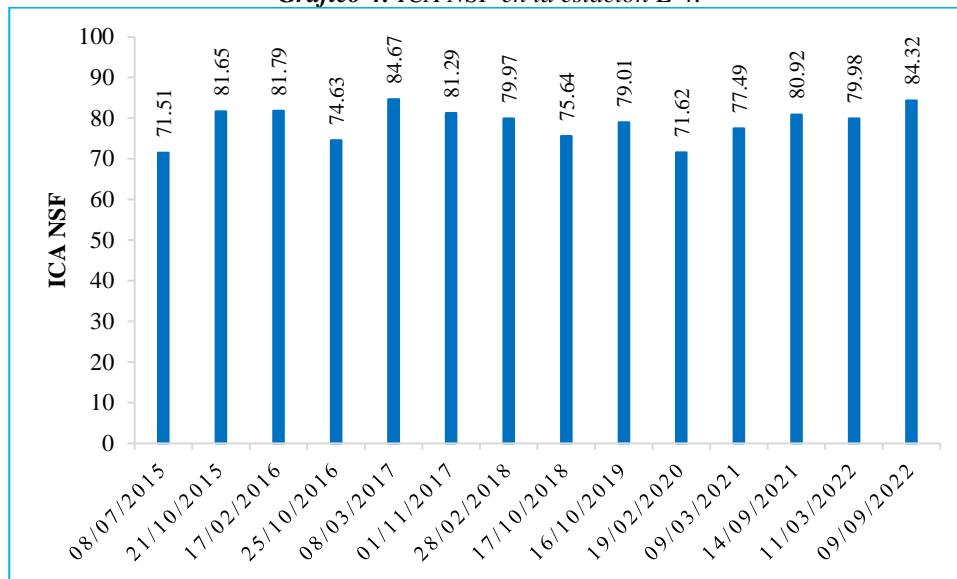
Al analizar el Gráfico 3, se destaca que la estación E-3 ha registrado valores notables en cuanto a la calidad del agua. En particular, se ha identificado que el punto de mayor calidad del agua se registró el 27 de junio de 2018, alcanzando un valor de 84.80. Por otro lado, el punto de menor calidad del agua fue registrado el 19 de julio de 2017, con un valor de 71.80. Ambos valores se clasifican cualitativamente como “Buena” calidad del agua.

Estos resultados indican que, a lo largo de los seis años de monitoreo, la calidad del agua en la estación E-3 se ha mantenido relativamente constante. Tanto el valor más alto como el más bajo se encuentran dentro de los parámetros considerados como buena calidad del agua. Esto sugiere una estabilidad en las condiciones hídricas de la Quebrada Can Can durante el periodo de estudio.

4. Estación 4

La estación de monitoreo E-4, identificada con el código YAN-QUIN-040 y ubicada en el Río Quingoyacu, ha sido objeto de un seguimiento exhaustivo durante un periodo de siete años, desde 2015 hasta 2022. Durante este tiempo, se han llevado a cabo un total de catorce monitoreos en dicha estación, con el objetivo de evaluar la calidad del agua en la zona y obtener una comprensión clara de las condiciones hídricas en el Río Quingoyacu a lo largo del tiempo.

Gráfico 4: ICA NSF en la estación E-4.



Fuente: Elaboración propia.

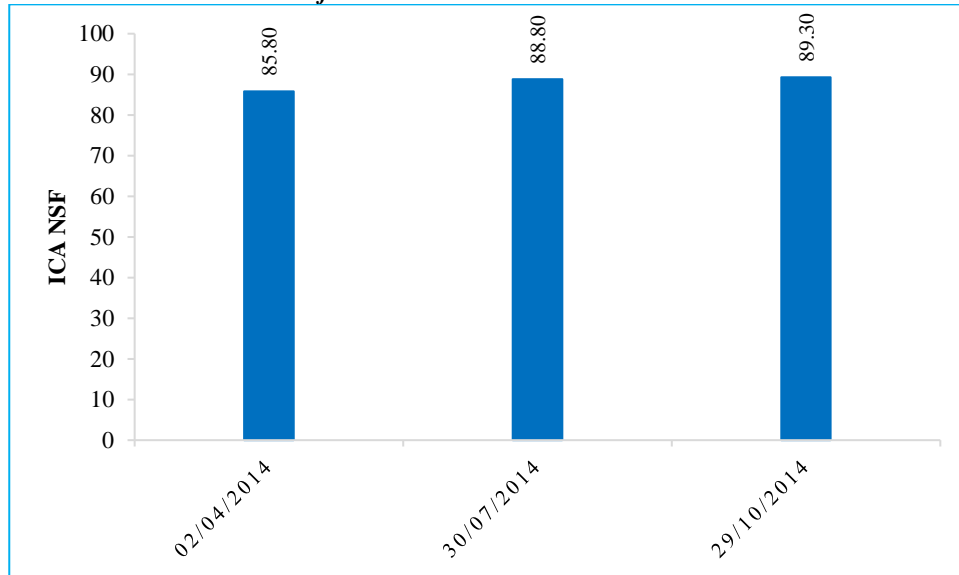
Al examinar el Gráfico 4, se observa que la estación E-4 ha registrado valores significativos en relación a la calidad del agua. En particular, se ha identificado que el punto de mayor calidad del agua se registró el 8 de marzo de 2017, alcanzando un valor de 84.67. Por otro lado, el punto de menor calidad del agua fue registrado el 8 de julio de 2015, con un valor de 71.51. Ambos valores se clasifican cualitativamente como “Buena” calidad del agua.

Estos resultados indican que, a lo largo de los siete años de monitoreo, la calidad del agua en la estación E-4 se ha mantenido relativamente constante. Tanto el valor más alto como el más bajo se encuentran dentro de los parámetros considerados como buena calidad del agua. Esto sugiere una estabilidad en las condiciones hídricas del Río Quingoyacu durante el periodo de estudio.

5. Estación 5

La estación de monitoreo E-5, identificada con el código YAN-GAL-020 y ubicada en el Río Galgal, ha sido objeto de un seguimiento en el año 2014, con un total de tres monitoreos. El objetivo de estos monitoreos fue evaluar la calidad del agua en la zona y obtener una comprensión clara de las condiciones hídricas en el Río Galgal a lo largo del tiempo.

Gráfico 5: ICA NSF en la estación E-5.



Fuente: Elaboración propia.

Al analizar el Gráfico 5, se destaca que la estación E-5 ha registrado valores significativos en cuanto a la calidad del agua. En particular, se ha identificado que el valor más alto de calidad del agua se registró en el mes de octubre, alcanzando un nivel de 89.30. Por otro lado, el valor más bajo de calidad del agua fue registrado en el mes de abril, con un nivel de 85.80. Ambos valores se clasifican cualitativamente como “Buena” calidad del agua.

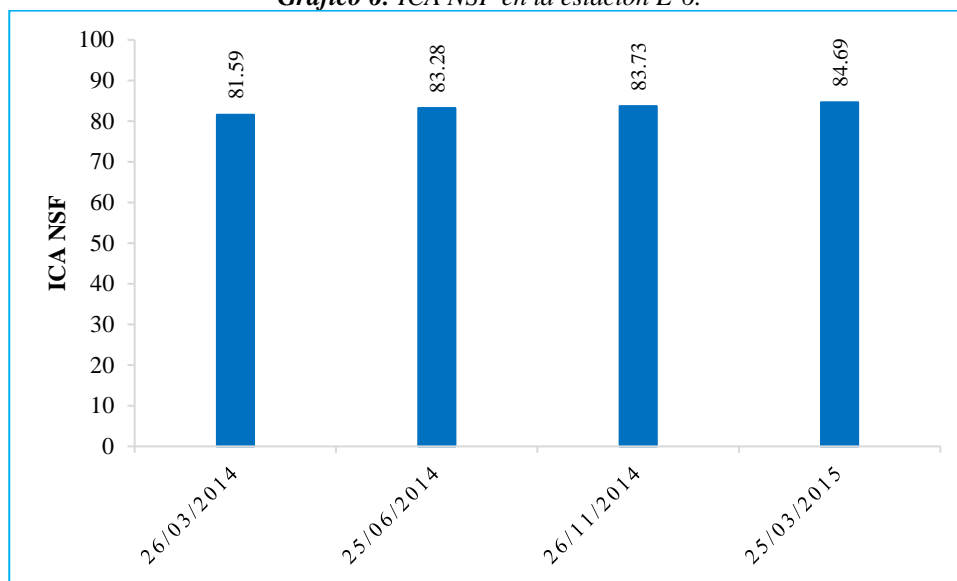
Es importante recalcar que, aunque ambos valores se consideran dentro de la categoría de buena calidad del agua, se observa una mejora progresiva en la calidad del agua a lo largo de los meses. El valor más bajo registrado en abril indica una calidad del agua aceptable, pero el valor más alto alcanzado en octubre indica una mejora en la calidad del agua.

6. Estación 6

La estación de monitoreo E-6, identificada con el código YAN-GAL-005 y ubicada en el Río Galgal, ha sido objeto de un seguimiento exhaustivo durante un periodo de dos años, con tres

monitoreos realizados en el año 2014 y uno en el año 2015. El propósito de estos monitoreos fue evaluar la calidad del agua en la zona y obtener una comprensión clara de las condiciones hídricas en el Río Galgal a lo largo del tiempo.

Gráfico 6: ICA NSF en la estación E-6.



Fuente: Elaboración propia.

Al analizar el Gráfico 6, se destaca que la estación E-6 ha registrado valores significativos en relación a la calidad del agua. El valor más alto de calidad del agua se registró el 25 de marzo de 2015, alcanzando un nivel de 84.69. Por otro lado, el valor más bajo de calidad del agua fue registrado el 26 de marzo de 2014, con un nivel de 81.59. Ambos valores se clasifican cualitativamente como “Buena” calidad del agua.

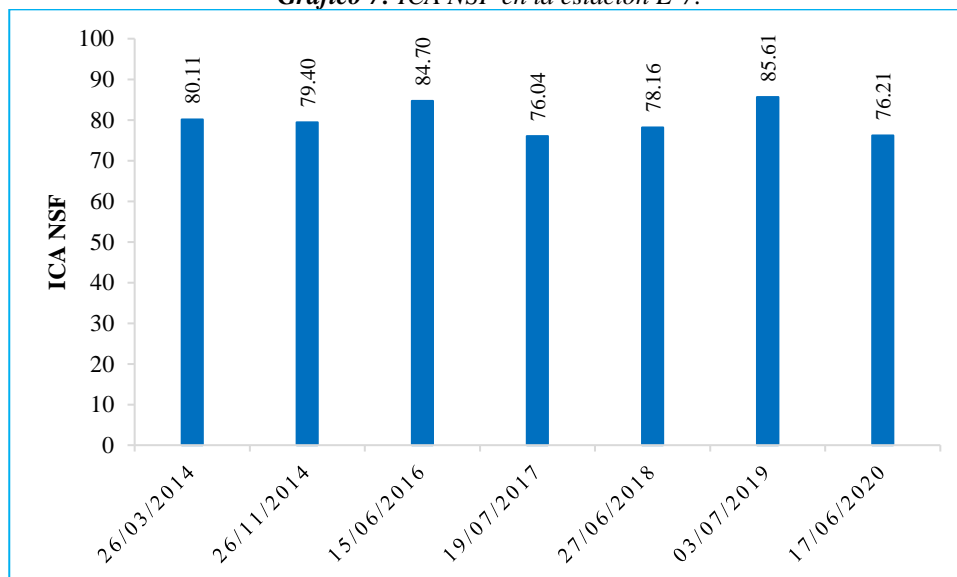
Estos resultados sugieren que la calidad del agua en la estación E-6 se ha mantenido constante a lo largo de los dos años de monitoreo. Tanto el valor más alto registrado en marzo de 2015 como el valor más bajo registrado en marzo de 2014 se encuentran dentro de los parámetros considerados como buena calidad del agua. Esto indica una estabilidad en las condiciones hídricas del Río Galgal durante el periodo de estudio.

7. Estación 7

La estación de monitoreo E-7, identificada con el código YAN-GAL-CONTROL y ubicada en el Río Galgal, ha sido objeto de un seguimiento riguroso durante un período de seis años, desde 2014 hasta 2020. Durante este tiempo, se han llevado a cabo un total de siete monitoreos en dicha

estación, con el objetivo de evaluar la calidad del agua en la zona y obtener una comprensión clara de las condiciones hídricas en el Río Galgal a lo largo del tiempo.

Gráfico 7: ICA NSF en la estación E-7.



Fuente: Elaboración propia.

Al examinar el Gráfico 7, se destaca que la estación E-7 ha registrado valores significativos en relación a la calidad del agua. El valor más alto de calidad del agua se registró el 3 de julio de 2019, alcanzando un nivel de 85.61. Por otro lado, el valor más bajo de calidad del agua fue registrado el 19 de julio de 2017, con un nivel de 76.04. Ambos valores se clasifican cualitativamente como “Buena” calidad del agua.

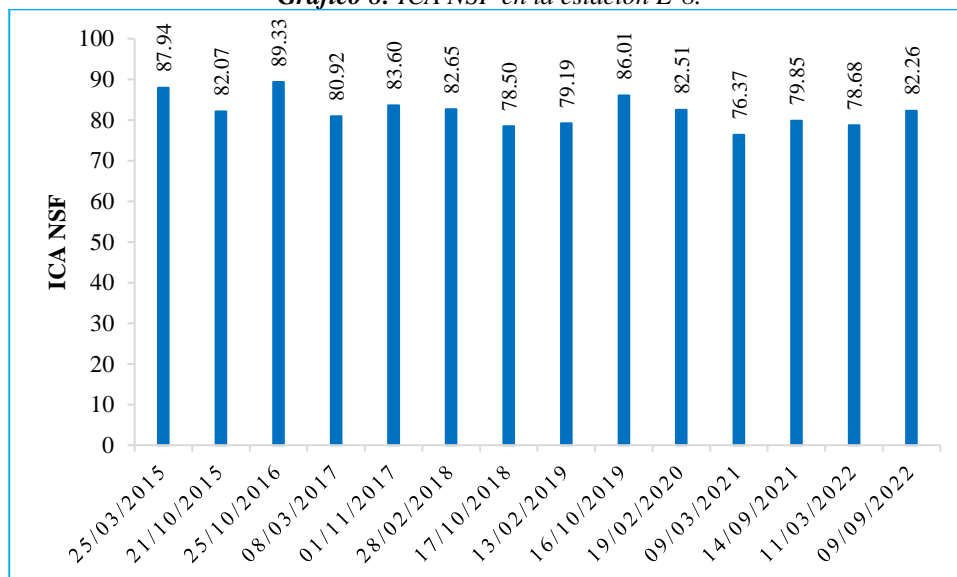
Estos resultados indican que, a lo largo de los seis años de monitoreo, la calidad del agua en la estación E-7 se ha mantenido relativamente constante. Tanto el valor más alto registrado en julio de 2019 como el valor más bajo registrado en julio de 2017 cumplen con los criterios establecidos para una buena calidad del agua. Esto sugiere una estabilidad en las condiciones hídricas del Río Galgal durante el período de estudio.

8. Estación 8

La estación de monitoreo E-8, identificada con el código YAN-BER-CONTROL y ubicada en el Río Bermejos, ha sido objeto de un seguimiento exhaustivo durante un período de siete años, desde 2015 hasta 2022. Durante este tiempo, se han llevado a cabo un total de catorce monitoreos

en dicha estación, con el objetivo de evaluar la calidad del agua en la zona y obtener una comprensión clara de las condiciones hídricas en el Río Bermejós a lo largo del tiempo.

Gráfico 8: ICA NSF en la estación E-8.



Fuente: Elaboración propia.

Al analizar el Gráfico 8, se destaca que la estación E-8 ha registrado valores significativos en relación a la calidad del agua. El valor más alto de calidad del agua se registró el 25 de octubre de 2016, alcanzando un nivel de 89.33. Por otro lado, el valor más bajo de calidad del agua fue registrado el 9 de marzo de 2021, con un nivel de 76.37. Ambos valores se clasifican cualitativamente como “Buena” calidad del agua.

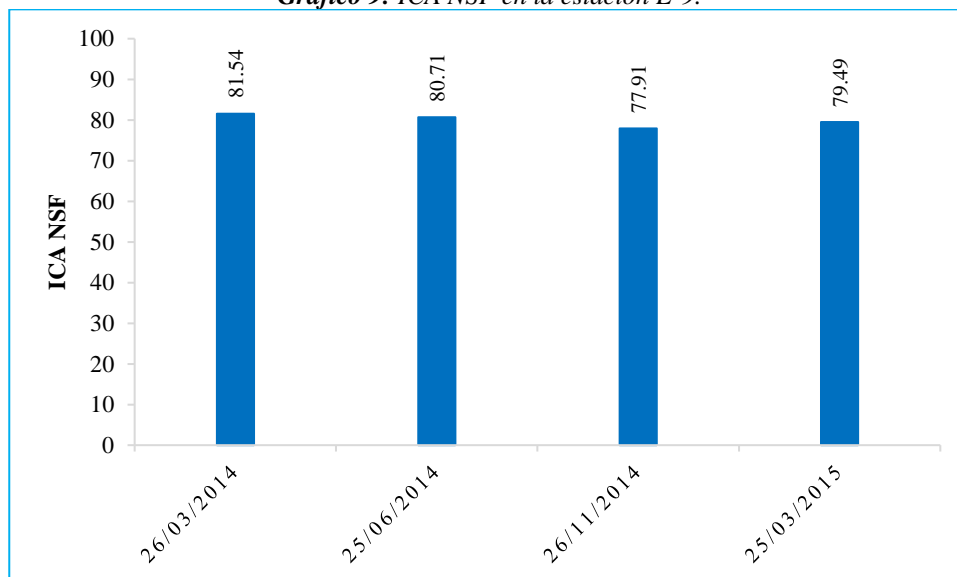
Estos resultados indican que, a lo largo de los siete años de monitoreo, la calidad del agua en la estación E-8 se ha mantenido relativamente constante. Tanto el valor más alto registrado en octubre de 2016 como el valor más bajo registrado en marzo de 2021 cumplen con los criterios establecidos para una buena calidad del agua. Esto sugiere una estabilidad en las condiciones hídricas del Río Bermejós durante el período de estudio.

9. Estación 9

La estación de monitoreo E-9, identificada con el código YAN-YAN-005 y ubicada en el Río Yanuncay, ha sido objeto de un seguimiento riguroso durante un período de dos años, con tres monitoreos realizados en el año 2014 y uno en el año 2015. El propósito de estos monitoreos fue

evaluar la calidad del agua en la zona y obtener una comprensión clara de las condiciones hídricas en el Río Yanuncay a lo largo del tiempo.

Gráfico 9: ICA NSF en la estación E-9.



Fuente: Elaboración propia.

Al analizar el Gráfico 9, se observa que la estación E-9 ha registrado valores significativos en relación a la calidad del agua. El valor más alto de calidad del agua se registró el 26 de marzo de 2014, alcanzando un nivel de 81.54. Por otro lado, el valor más bajo de calidad del agua fue registrado el 26 de noviembre de 2014, con un nivel de 77.91. Ambos valores se clasifican cualitativamente como “Buena” calidad del agua.

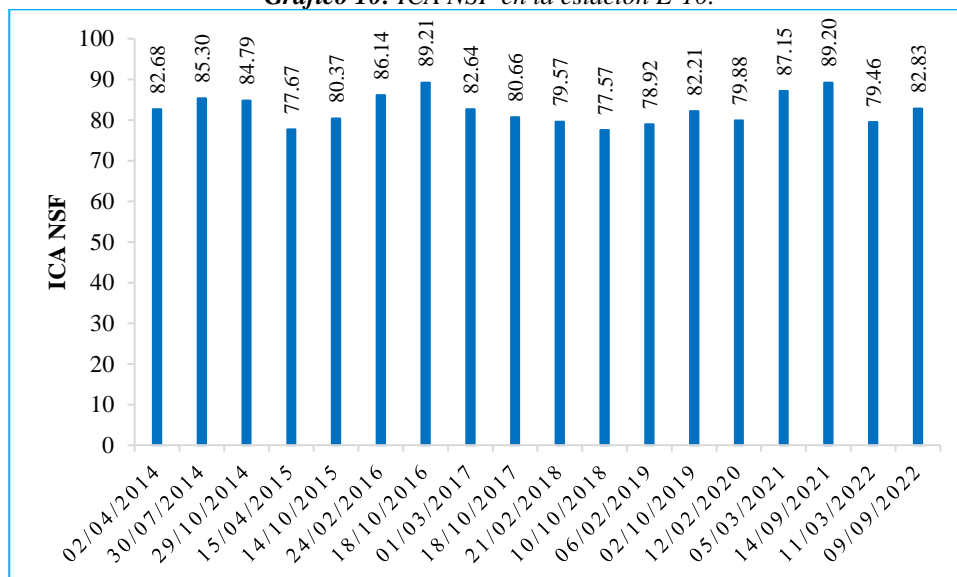
Estos resultados indican que, a lo largo de los dos años de monitoreo, la calidad del agua en la estación E-9 se ha mantenido relativamente constante. Tanto el valor más alto registrado en marzo de 2014 como el valor más bajo registrado en noviembre de 2014 cumplen con los criterios establecidos para una buena calidad del agua. Esto sugiere una estabilidad en las condiciones hídricas del Río Yanuncay durante el período de estudio.

10. Estación 10

La estación de monitoreo E-10, identificada con el código YAN-IZH-020 y ubicada en el Río Izhcayrumi, ha sido sometida a un riguroso seguimiento durante un período de ocho años, desde 2014 hasta 2022. Durante este tiempo, se han llevado a cabo un total de dieciocho monitoreos en

dicha estación, con el objetivo de evaluar la calidad del agua en la zona y obtener una comprensión clara de las condiciones hídricas en el Río Izhcayrumi a lo largo del tiempo.

Gráfico 10: ICA NSF en la estación E-10.



Fuente: Elaboración propia.

Al examinar el Gráfico 10, se destaca que la estación E-10 ha registrado valores significativos en relación a la calidad del agua. El valor más alto de calidad del agua se registró el 18 de octubre de 2016, alcanzando un nivel de 89.21. Por otro lado, el valor más bajo de calidad del agua fue registrado el 10 de octubre de 2018, con un nivel de 77.57. Ambos valores se clasifican cualitativamente como “Buena” calidad del agua.

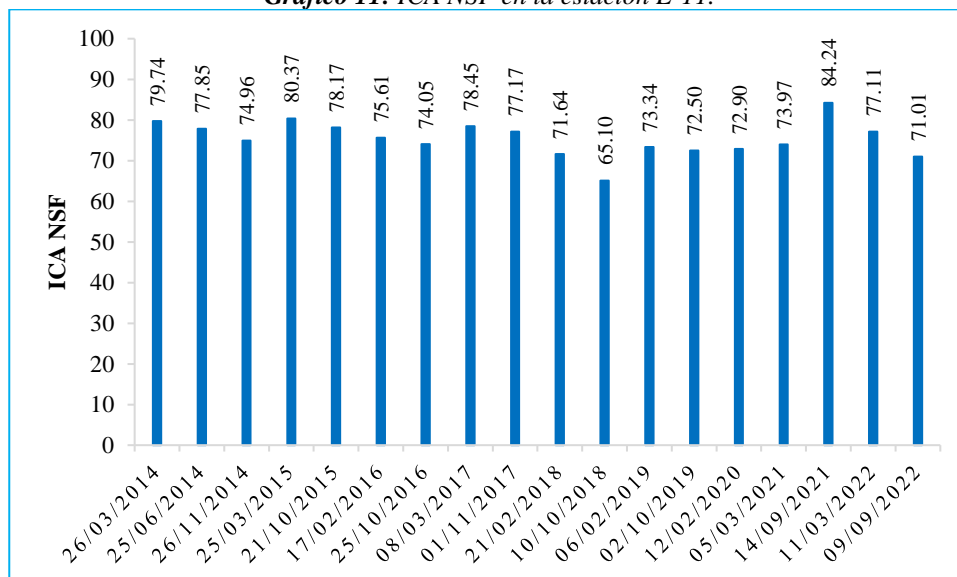
Estos resultados indican que, a lo largo de los ocho años de monitoreo, la calidad del agua en la estación E-10 se ha mantenido relativamente constante. Tanto el valor más alto registrado en octubre de 2016 como el valor más bajo registrado en octubre de 2018 cumplen con los criterios establecidos para una buena calidad del agua. Esto sugiere una estabilidad en las condiciones hídricas del Río Izhcayrumi durante el período de estudio.

11. Estación 11

La estación de monitoreo E-11, identificada con el código YAN-IZH-CONTROL y ubicada en el Río Izhcayrumi, ha sido objeto de un seguimiento exhaustivo durante un período de ocho años, desde 2014 hasta 2022. Durante este tiempo, se han llevado a cabo un total de dieciocho

monitoreos en dicha estación, con el objetivo de evaluar la calidad del agua en la zona y obtener una comprensión clara de las condiciones hídricas en el Río Izhcayrumi a lo largo del tiempo.

Gráfico 11: ICA NSF en la estación E-11.



Fuente: Elaboración propia.

Al analizar el Gráfico 11, se observa que la estación E-11 ha registrado valores significativos en relación a la calidad del agua, pero con una variación notable. El valor más alto de calidad del agua se registró el 14 de septiembre de 2021, alcanzando un nivel de 84.24. Por otro lado, el valor más bajo de calidad del agua fue registrado el 10 de octubre de 2018, con un nivel de 65.10. Estos dos valores se clasifican cualitativamente de manera diferente, donde el valor de 84.24 se considera una calidad “Buena” del agua, mientras que el valor de 65.10 se considera una calidad “Regular”.

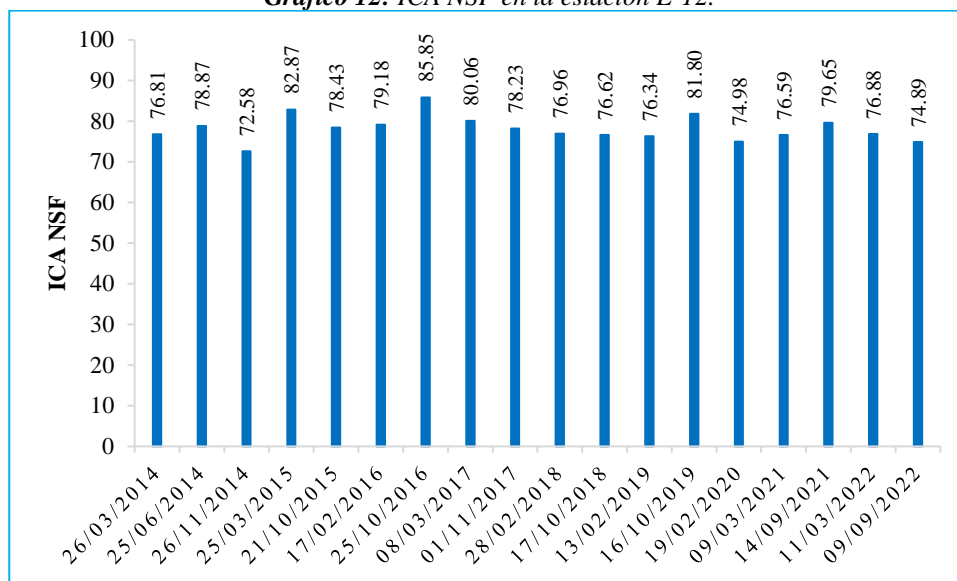
Estos resultados indican que, a lo largo de los ocho años de monitoreo, la calidad del agua en la estación E-11 ha mostrado una variabilidad significativa. El valor más alto registrado en septiembre de 2021 cumple con los criterios para una buena calidad del agua, mientras que el valor más bajo registrado en octubre de 2018 se clasifica como una calidad del agua regular. Esto sugiere que la calidad del agua en el Río Izhcayrumi ha experimentado cambios a lo largo del tiempo y no se ha mantenido constante durante el período de estudio.

12. Estación 12

La estación de monitoreo E-12, identificada con el código YAN-YAN-010 y ubicada en el Río Yanuncay, ha sido objeto de un seguimiento minucioso durante un período de ocho años, desde

2014 hasta 2022. Durante este tiempo, se han llevado a cabo un total de dieciocho monitoreos en dicha estación, con el objetivo de evaluar la calidad del agua en la zona y obtener una comprensión clara de las condiciones hídricas en el Río Yanuncay a lo largo del tiempo.

Gráfico 12: ICA NSF en la estación E-12.



Fuente: Elaboración propia.

Al examinar el Gráfico 12, se destaca que la estación E-12 ha registrado valores significativos en cuanto a la calidad del agua. El valor más alto de calidad del agua se registró el 25 de octubre de 2016, alcanzando un nivel de 85.85. Por otro lado, el valor más bajo de calidad del agua fue registrado el 26 de noviembre de 2014, con un nivel de 72.58. Ambos valores se califican cualitativamente como una calidad del agua “Buena”.

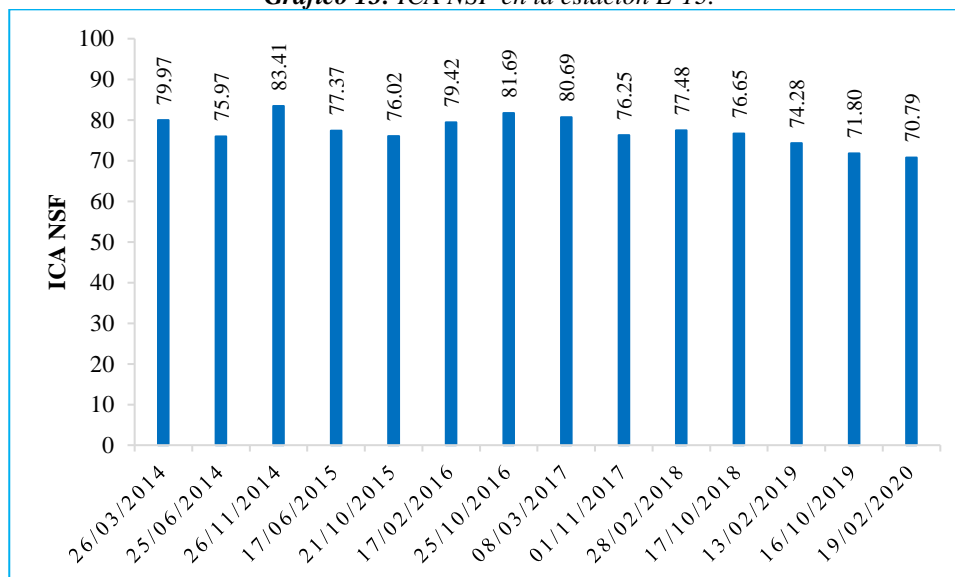
Estos resultados indican que, a lo largo de los ocho años de monitoreo, la calidad del agua en la estación E-12 se ha mantenido relativamente constante. Tanto el valor más alto registrado en octubre de 2016 como el valor más bajo registrado en noviembre de 2014 cumplen con los criterios establecidos para una buena calidad del agua. Esto sugiere una estabilidad en las condiciones hídricas del Río Yanuncay durante el período de estudio.

13. Estación 13

La estación de monitoreo E-13, identificada con el código YAN-YAN-015 y ubicada en el Río Yanuncay, ha sido objeto de un seguimiento detallado durante un período de seis años, desde 2014 hasta 2020. Durante este tiempo, se han llevado a cabo un total de catorce monitoreos en dicha

estación, con el objetivo de evaluar la calidad del agua en la zona y obtener una comprensión clara de las condiciones hídricas en el Río Yanuncay a lo largo del tiempo.

Gráfico 13: ICA NSF en la estación E-13.



Fuente: Elaboración propia.

Al examinar el Gráfico 13, se observa que la estación E-13 ha registrado valores significativos en relación a la calidad del agua, pero con una variación notable. El valor más alto de calidad del agua se registró el 26 de noviembre de 2014, alcanzando un nivel de 83.41. Por otro lado, el valor más bajo de calidad del agua fue registrado el 19 de febrero de 2020, con un nivel de 70.79. Estos dos valores se clasifican cualitativamente de manera diferente, donde el valor de 83.41 se considera una calidad “Buena” del agua, mientras que el valor de 70.79 se considera una calidad “Regular”.

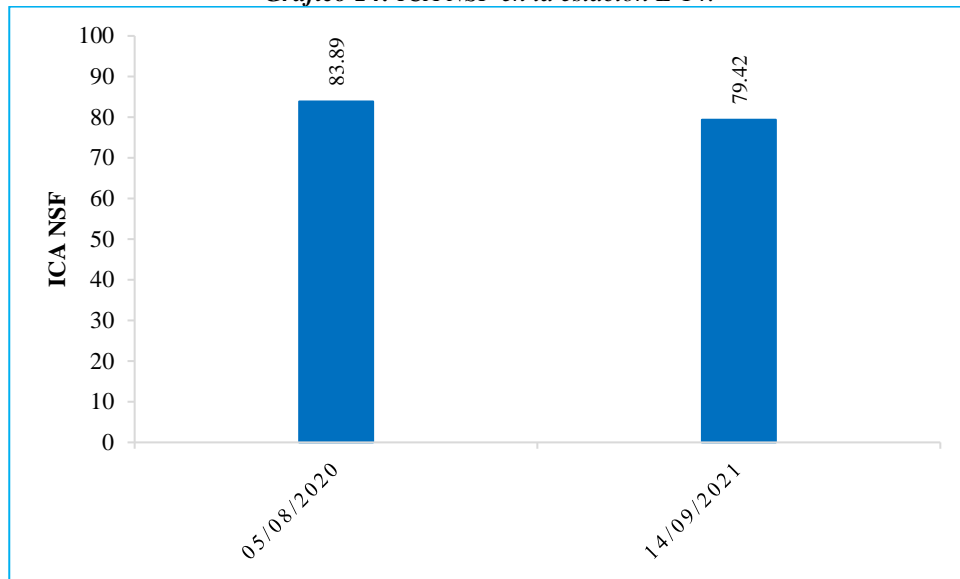
Estos resultados indican que, a lo largo de los seis años de monitoreo, la calidad del agua en la estación E-13 ha mostrado una variabilidad significativa. El valor más alto registrado en noviembre de 2014 cumple con los criterios para una buena calidad del agua, mientras que el valor más bajo registrado en febrero de 2020 se clasifica como una calidad del agua regular. Esto sugiere que la calidad del agua en el Río Yanuncay ha experimentado cambios a lo largo del tiempo y no se ha mantenido constante durante el período de estudio.

14. Estación 14

La estación de monitoreo E-14, identificada con el código YAN-SOL-020 y ubicada en el Río Soldados, ha sido objeto de un seguimiento limitado pero significativo, con solo dos monitoreos

realizados, uno en el año 2020 y otro en el año 2021. El propósito de estos monitoreos ha sido evaluar la calidad del agua en esta ubicación específica y obtener una comprensión inicial de las condiciones hídricas en el Río Soldados a lo largo del tiempo.

Gráfico 14: ICA NSF en la estación E-14.



Fuente: Elaboración propia.

Al examinar el Gráfico 14, se destaca que la estación E-14 ha registrado valores notables en relación a la calidad del agua. El valor más alto de calidad del agua se registró el 5 de agosto de 2020, alcanzando un nivel de 83.89. Por otro lado, el valor más bajo de calidad del agua fue registrado el 14 de septiembre de 2021, con un nivel de 79.42. Ambos valores se califican cualitativamente como una calidad del agua “Buena”.

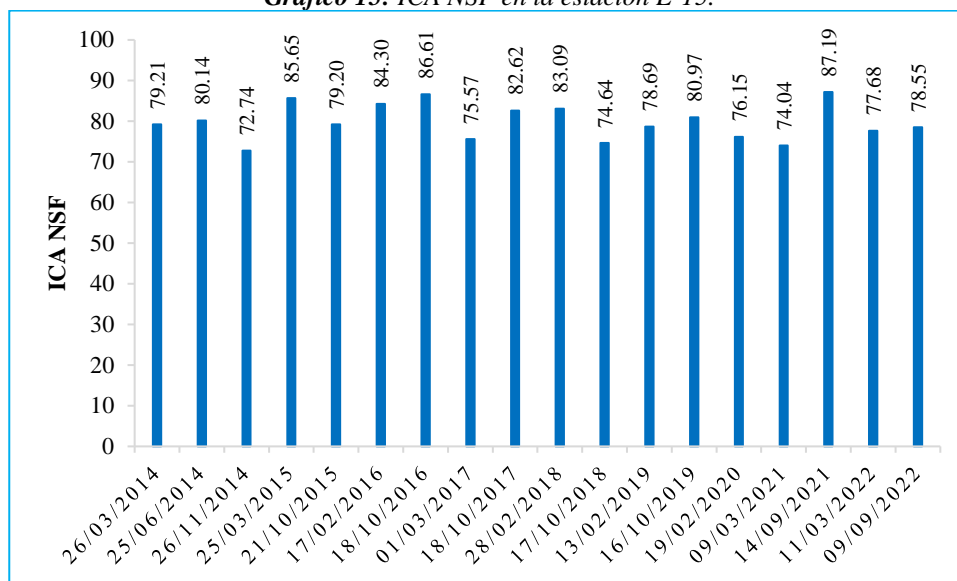
Estos resultados indican que, aunque se cuenta con un número limitado de monitoreos, la calidad del agua en la estación E-14 se ha mantenido relativamente constante durante los dos años de monitoreo. Tanto el valor más alto registrado en agosto de 2020 como el valor más bajo registrado en septiembre de 2021 cumplen con los criterios establecidos para una buena calidad del agua. Esto sugiere una estabilidad en las condiciones hídricas del Río Soldados durante el período de estudio.

15. Estación 15

La estación de monitoreo E-15, identificada con el código YAN-SOL-CONTROL y ubicada en el Río Soldados, ha sido objeto de un seguimiento exhaustivo durante un período de ocho años,

desde 2014 hasta 2022. Durante este tiempo, se han llevado a cabo un total de dieciocho monitoreos en dicha estación, con el objetivo de evaluar la calidad del agua en la zona y obtener una comprensión clara de las condiciones hídricas en el Río Soldados a lo largo del tiempo.

Gráfico 15: ICA NSF en la estación E-15.



Fuente: Elaboración propia.

Al examinar el Gráfico 15, se destaca que la estación E-15 ha registrado valores significativos en cuanto a la calidad del agua. El valor más alto de calidad del agua se registró el 14 de septiembre de 2021, alcanzando un nivel de 87.19. Por otro lado, el valor más bajo de calidad del agua fue registrado el 26 de noviembre de 2014, con un nivel de 74.74. Ambos valores se califican cualitativamente como una calidad del agua “Buena”.

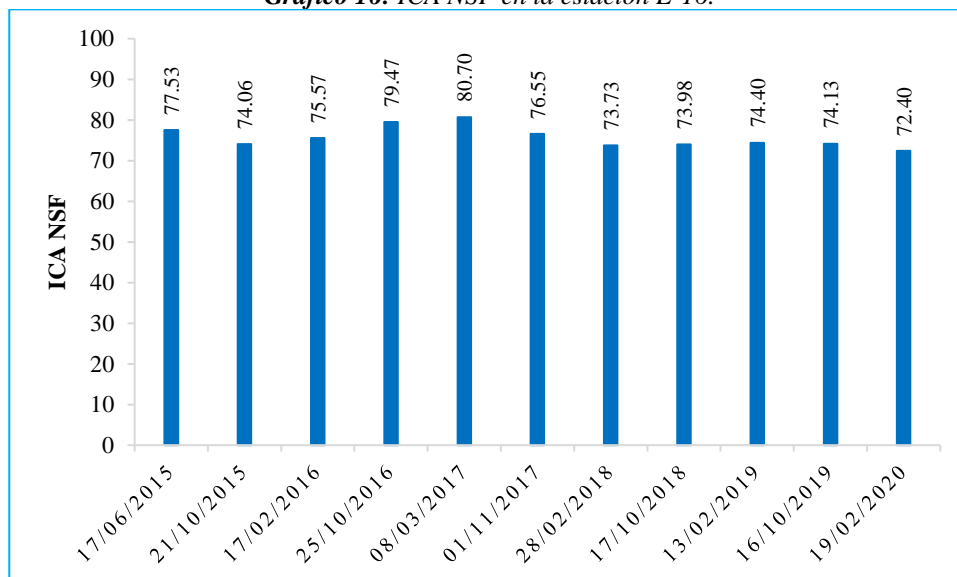
Estos resultados indican que, a lo largo de los ocho años de monitoreo, la calidad del agua en la estación E-15 se ha mantenido constantemente alta. Tanto el valor más alto registrado en septiembre de 2021 como el valor más bajo registrado en noviembre de 2014 cumplen con los criterios establecidos para una buena calidad del agua. Esto sugiere una estabilidad en las condiciones hídricas del Río Soldados durante el período de estudio.

16. Estación 16

La estación de monitoreo E-16, identificada con el código YAN-YAN-018 y ubicada en el Río Yanuncay, ha sido objeto de un seguimiento sistemático durante un período de cinco años, desde 2015 hasta 2020. Durante este tiempo, se han llevado a cabo un total de once monitoreos en dicha

estación, con el objetivo de evaluar la calidad del agua en la zona y obtener una comprensión clara de las condiciones hídricas en el Río Yanuncay a lo largo del tiempo.

Gráfico 16: ICA NSF en la estación E-16.



Fuente: Elaboración propia.

Al examinar el Gráfico 16, se destaca que la estación E-16 ha registrado valores significativos en cuanto a la calidad del agua. El valor más alto de calidad del agua se registró el 8 de marzo de 2017, alcanzando un nivel de 80.70. Por otro lado, el valor más bajo de calidad del agua fue registrado el 19 de febrero de 2020, con un nivel de 72.40. Ambos valores se califican cualitativamente como una calidad del agua “Buena”.

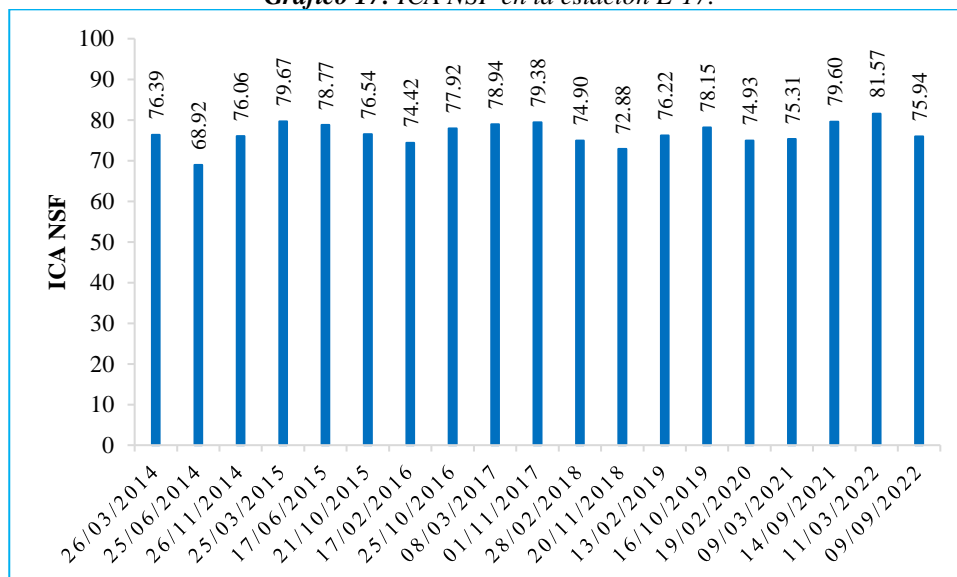
Estos resultados indican que, a lo largo de los cinco años de monitoreo, la calidad del agua en la estación E-16 se ha mantenido constantemente alta. Tanto el valor más alto registrado en marzo de 2017 como el valor más bajo registrado en febrero de 2020 cumplen con los criterios establecidos para una buena calidad del agua. Esto sugiere una estabilidad en las condiciones hídricas del Río Yanuncay durante el período de estudio.

17. Estación 17

La estación de monitoreo E-17, identificada con el código YAN-YAN-020 y ubicada en el Río Yanuncay, ha sido objeto de un seguimiento extenso durante un período de ocho años, desde 2014 hasta 2022. Durante este tiempo, se han llevado a cabo un total de diecinueve monitoreos en dicha

estación, con el objetivo de evaluar la calidad del agua en la zona y obtener una comprensión clara de las condiciones hídricas en el Río Yanuncay a lo largo del tiempo.

Gráfico 17: ICA NSF en la estación E-17.



Fuente: Elaboración propia.

Al examinar el Gráfico 17, se destaca que la estación E-17 ha registrado valores significativos en cuanto a la calidad del agua. El valor más alto de calidad del agua se registró el 11 de marzo de 2022, alcanzando un nivel de 81.57. Por otro lado, el valor más bajo de calidad del agua fue registrado el 25 de junio de 2014, con un nivel de 68.92. Sin embargo, es importante destacar que estos dos valores reciben calificaciones cualitativas diferentes: el valor de 81.57 se considera una calidad del agua “Buena”, mientras que el valor de 68.92 se califica como una calidad del agua “Regular”.

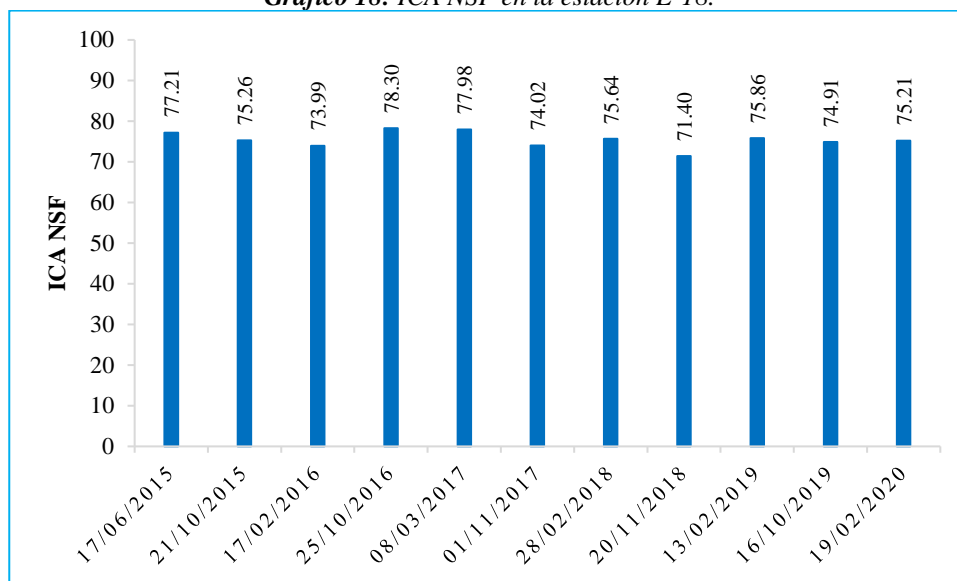
Estos resultados indican que la calidad del agua en la estación E-17 no se ha mantenido constante a lo largo de los ocho años de monitoreo. Mientras que el valor más alto de 81.57 cumple con los criterios establecidos para una buena calidad del agua, el valor más bajo de 68.92 se considera una calidad del agua regular. Esto sugiere fluctuaciones y variaciones en las condiciones hídricas del Río Yanuncay durante el período de estudio.

18. Estación 18

La estación de monitoreo E-18, identificada con el código YAN-YAN-023 y ubicada en el Río Yanuncay, ha sido objeto de un seguimiento sistemático durante un período de cinco años, desde

2015 hasta 2020. Durante este tiempo, se han llevado a cabo un total de once monitoreos en dicha estación, con el objetivo de evaluar la calidad del agua en la zona y obtener una comprensión clara de las condiciones hídricas en el Río Yanuncay a lo largo del tiempo.

Gráfico 18: ICA NSF en la estación E-18.



Fuente: Elaboración propia.

Al examinar el Gráfico 18, se destaca que la estación E-18 ha registrado valores significativos en cuanto a la calidad del agua. El valor más alto de calidad del agua se registró el 25 de octubre de 2016, alcanzando un nivel de 78.30. Por otro lado, el valor más bajo de calidad del agua fue registrado el 20 de noviembre de 2018, con un nivel de 71.40. Ambos valores se califican cualitativamente como una calidad del agua “Buena”.

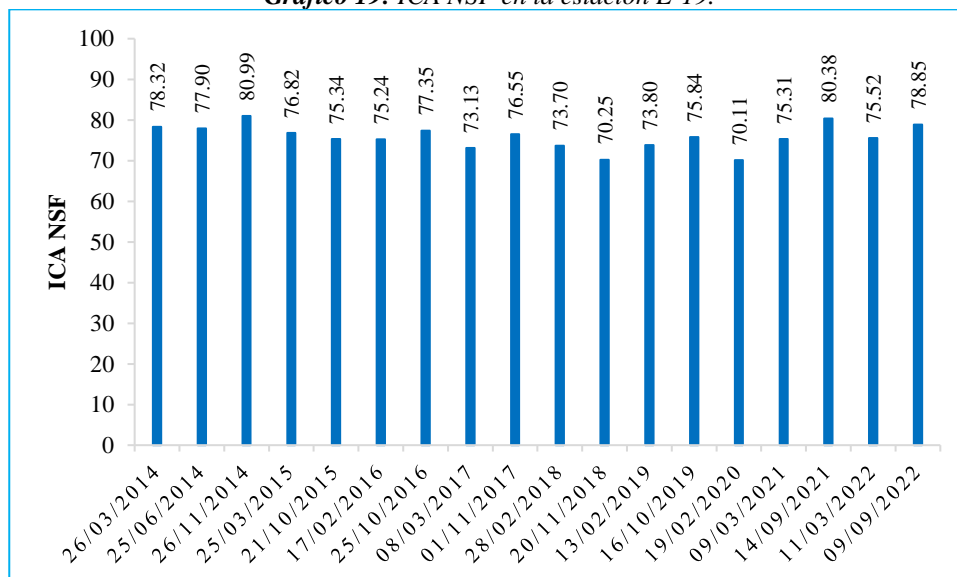
Estos resultados indican que, a lo largo de los cinco años de monitoreo, la calidad del agua en la estación E-18 se ha mantenido constantemente alta. Tanto el valor más alto registrado en octubre de 2016 como el valor más bajo registrado en noviembre de 2018 cumplen con los criterios establecidos para una buena calidad del agua. Esto sugiere una estabilidad en las condiciones hídricas del Río Yanuncay durante el período de estudio.

19. Estación 19

La estación de monitoreo E-19, identificada con el código YAN-YAN-025 y ubicada en el Río Yanuncay, ha sido objeto de un seguimiento exhaustivo durante un período de ocho años, desde 2014 hasta 2022. Durante este tiempo, se han llevado a cabo un total de dieciocho monitoreos en

dicha estación, con el objetivo de evaluar la calidad del agua en la zona y obtener una comprensión clara de las condiciones hídricas en el Río Yanuncay a lo largo del tiempo.

Gráfico 19: ICA NSF en la estación E-19.



Fuente: Elaboración propia.

Al examinar el Gráfico 19, se destaca que la estación E-19 ha registrado valores significativos en cuanto a la calidad del agua. El valor más alto de calidad del agua se registró el 26 de noviembre de 2014, alcanzando un nivel de 80.99. Por otro lado, el valor más bajo de calidad del agua fue registrado el 19 de febrero de 2020, con un nivel de 70.11. Sin embargo, es importante destacar que estos dos valores reciben calificaciones cualitativas diferentes: el valor de 80.99 se considera una calidad del agua “Buena”, mientras que el valor de 70.11 se califica como una calidad del agua “Regular”.

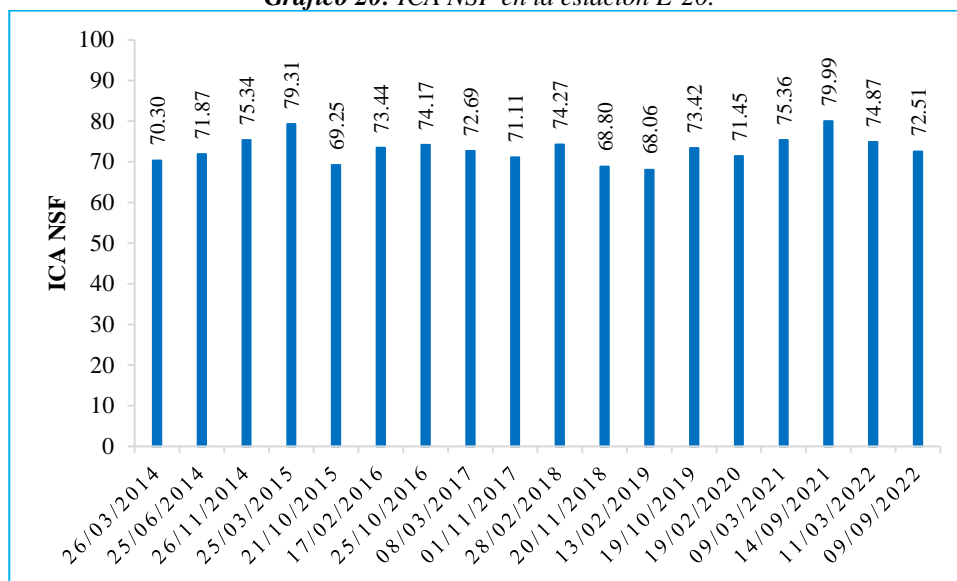
Estos resultados indican que la calidad del agua en la estación E-19 no se ha mantenido constante a lo largo de los ocho años de monitoreo. Mientras que el valor más alto de 80.99 cumple con los criterios establecidos para una buena calidad del agua, el valor más bajo de 70.11 se clasifica como una calidad del agua regular. Esto sugiere fluctuaciones y variaciones en las condiciones hídricas del Río Yanuncay durante el período de estudio.

20. Estación 20

La estación de monitoreo E-20, identificada con el código YAN-YAN-030 y ubicada en el Río Yanuncay, ha sido objeto de un seguimiento exhaustivo durante un período de ocho años, desde

2014 hasta 2022. Durante este tiempo, se han llevado a cabo un total de dieciocho monitoreos en dicha estación, con el objetivo de evaluar la calidad del agua en la zona y obtener una comprensión clara de las condiciones hídricas en el Río Yanuncay a lo largo del tiempo.

Gráfico 20: ICA NSF en la estación E-20.



Fuente: Elaboración propia.

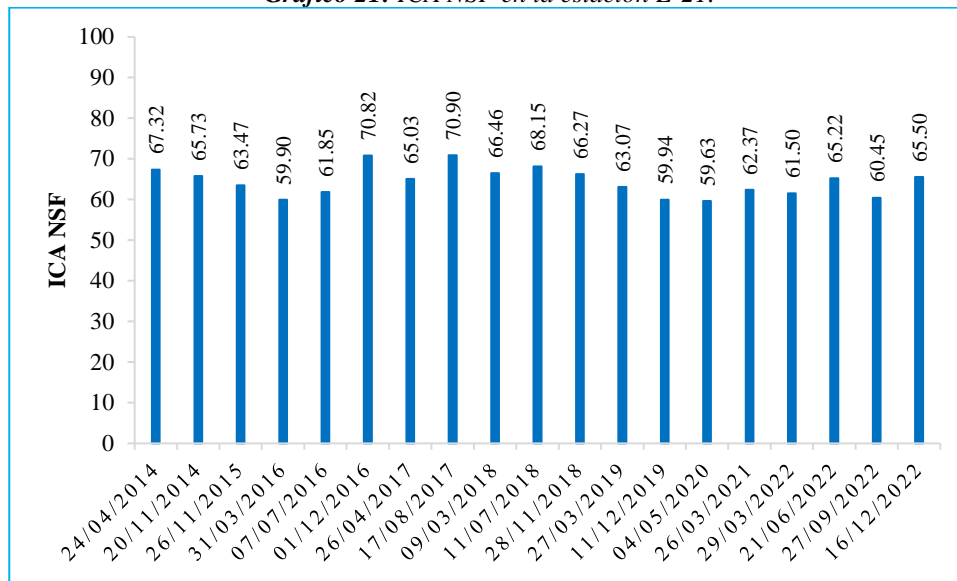
Al analizar el Gráfico 20, se observa que la estación E-20 ha registrado valores significativos en términos de calidad del agua. El valor más alto de calidad del agua se registró el 14 de septiembre de 2021, alcanzando un nivel de 79.99. Por otro lado, el valor más bajo de calidad del agua fue registrado el 13 de febrero de 2019, con un nivel de 68.06. Sin embargo, es importante destacar que estos dos valores reciben calificaciones cualitativas diferentes: el valor de 79.99 se considera una calidad del agua “Buena”, mientras que el valor de 68.06 se califica como una calidad del agua “Regular”.

Estos resultados indican que la calidad del agua en la estación E-20 no se ha mantenido constante a lo largo de los ocho años de monitoreo. Mientras que el valor más alto de 79.99 cumple con los criterios establecidos para una buena calidad del agua, el valor más bajo de 68.06 se clasifica como una calidad del agua regular. Esto sugiere fluctuaciones y variaciones en las condiciones hídricas del Río Yanuncay durante el período de estudio.

21. Estación 21

La estación de monitoreo E-21, identificada con el código YAN-YAN-040 y ubicada en el Río Yanuncay, ha sido objeto de un seguimiento detallado durante un período de ocho años, desde 2014 hasta 2022. Durante este tiempo, se han llevado a cabo un total de diecinueve monitoreos en dicha estación, con el objetivo de evaluar la calidad del agua en la zona y obtener una comprensión clara de las condiciones hídricas en el Río Yanuncay a lo largo del tiempo.

Gráfico 21: ICA NSF en la estación E-21.



Fuente: Elaboración propia.

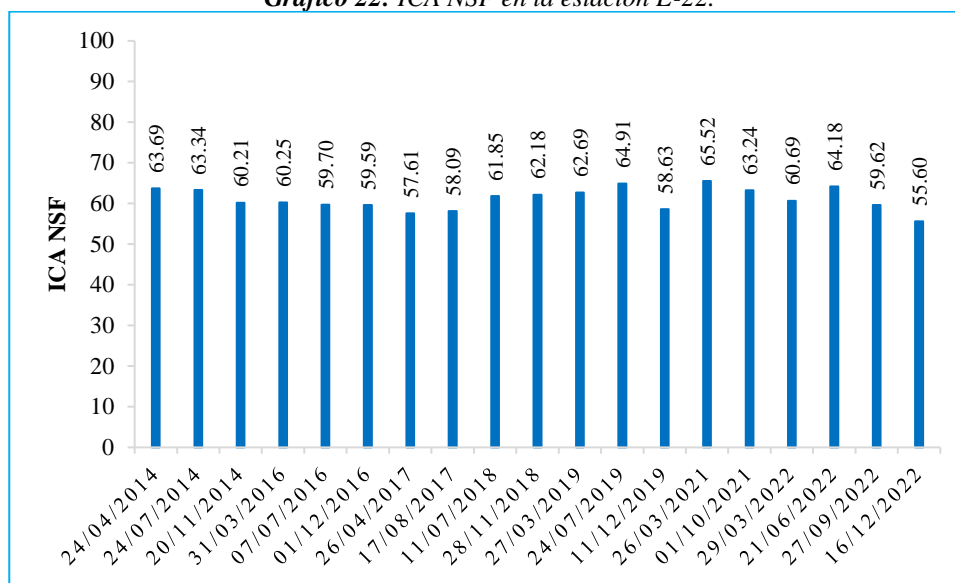
Al revisar el Gráfico 21, se destaca que la estación E-21 ha registrado valores significativos en términos de calidad del agua. El valor más alto de calidad del agua se registró el 17 de agosto de 2017, alcanzando un nivel de 70.90. Por otro lado, el valor más bajo de calidad del agua fue registrado el 4 de mayo de 2020, con un nivel de 59.63. Es importante destacar que ambos valores reciben una calificación cualitativa similar, clasificándose como una calidad del agua “Regular”.

Estos resultados indican que la calidad del agua en la estación E-21 se ha mantenido relativamente constante a lo largo de los ocho años de monitoreo. Tanto el valor más alto de 70.90 como el valor más bajo de 59.63 se califican como una calidad del agua regular. Esto sugiere una estabilidad en las condiciones hídricas del Río Yanuncay durante el período de estudio.

22. Estación 22

La estación de monitoreo E-22, identificada con el código YAN-YAN-060 y ubicada en el Río Yanuncay, ha sido objeto de un seguimiento exhaustivo durante un período de ocho años, desde 2014 hasta 2022. Durante este tiempo, se han llevado a cabo un total de diecinueve monitoreos en dicha estación, con el objetivo de evaluar la calidad del agua en la zona y obtener una comprensión clara de las condiciones hídricas en el Río Yanuncay a lo largo del tiempo.

Gráfico 22: ICA NSF en la estación E-22.



Fuente: Elaboración propia.

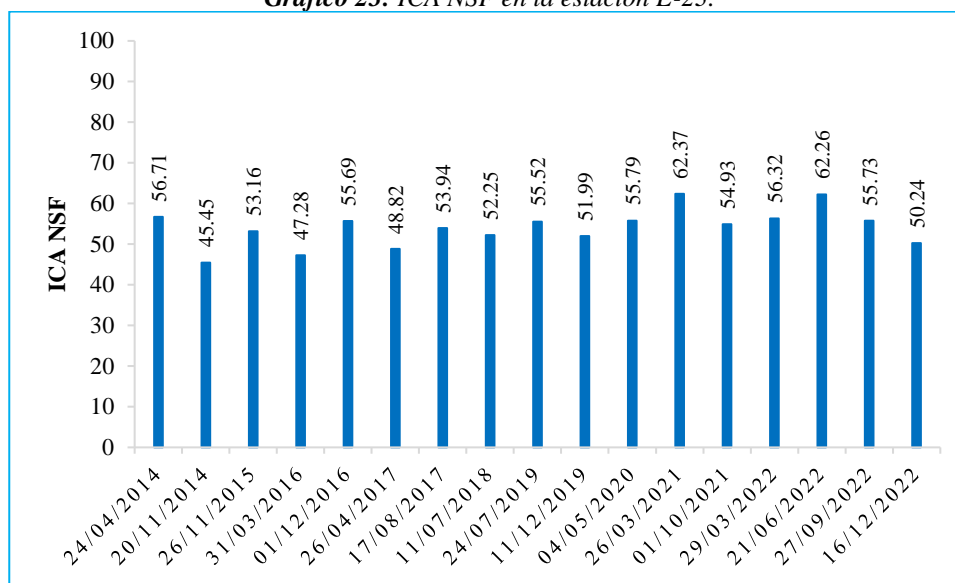
Al examinar el Gráfico 22, se observa que la estación E-22 ha registrado valores significativos en cuanto a la calidad del agua. El valor más alto de calidad del agua se registró el 26 de marzo de 2021, alcanzando un nivel de 65.52. Por otro lado, el valor más bajo de calidad del agua fue registrado el 16 de diciembre de 2022, con un nivel de 55.60. Es importante destacar que ambos valores reciben una calificación cualitativa similar, clasificándose como una calidad del agua “Regular”.

Estos resultados indican que la calidad del agua en la estación E-22 se ha mantenido relativamente constante a lo largo de los ocho años de monitoreo. Tanto el valor más alto de 65.52 como el valor más bajo de 55.60 se califican como una calidad del agua regular. Esto sugiere una estabilidad en las condiciones hídricas del Río Yanuncay durante el período de estudio.

23. Estación 23

La estación de monitoreo E-23, identificada con el código YAN-YAN-CONTROL y ubicada en el Río Yanuncay, ha sido objeto de un seguimiento minucioso durante un período de ocho años, desde 2014 hasta 2022. Durante este tiempo, se han llevado a cabo un total de diecisiete monitoreos en dicha estación, con el objetivo de evaluar la calidad del agua en la zona y obtener una comprensión clara de las condiciones hídricas en el Río Yanuncay a lo largo del tiempo.

Gráfico 23: ICA NSF en la estación E-23.



Fuente: Elaboración propia.

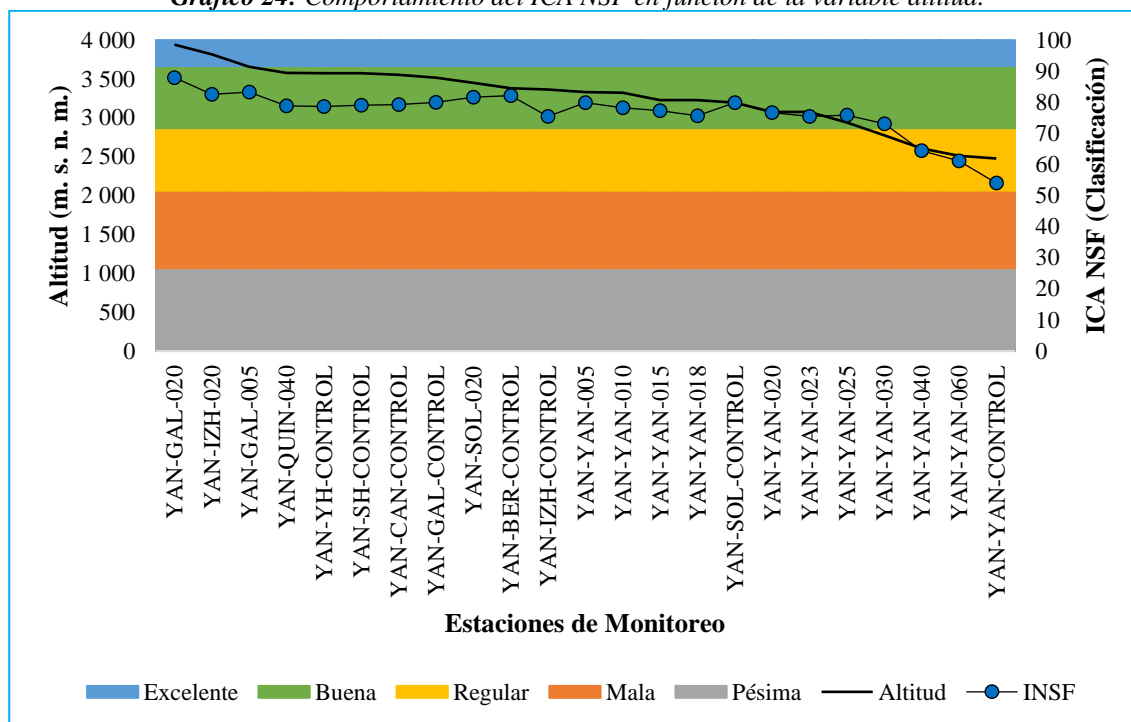
Al analizar el Gráfico 23, se observa que la estación E-23 ha registrado valores significativos en términos de calidad del agua. El valor más alto de calidad del agua se registró el 26 de marzo de 2021, alcanzando un nivel de 62.37. Por otro lado, el valor más bajo de calidad del agua fue registrado el 20 de noviembre de 2014, con un nivel de 45.45. Es importante destacar que estos dos valores reciben calificaciones cualitativas diferentes: el valor de 62.37 se considera una calidad del agua “Regular”, mientras que el valor de 45.45 se clasifica como una calidad del agua “Mala”.

Estos resultados indican que la calidad del agua en la estación E-23 no se ha mantenido constante a lo largo de los ocho años de monitoreo. Mientras que el valor más alto de 62.37 se clasifica como una calidad del agua regular, el valor más bajo de 45.45 se califica como una mala calidad del agua. Esto sugiere fluctuaciones y variaciones significativas en las condiciones hídricas del Río Yanuncay durante el período de estudio.

4.2 Deterioro del ICA NSF en la cuenca del Río Yanuncay

En el Gráfico 24 se presenta un análisis detallado del comportamiento de la calidad del agua en relación con la altitud de las 23 estaciones de monitoreo. Este análisis nos permite visualizar patrones sobre el comportamiento de la calidad del agua en función de la variable altitud.

Gráfico 24: Comportamiento del ICA NSF en función de la variable altitud.



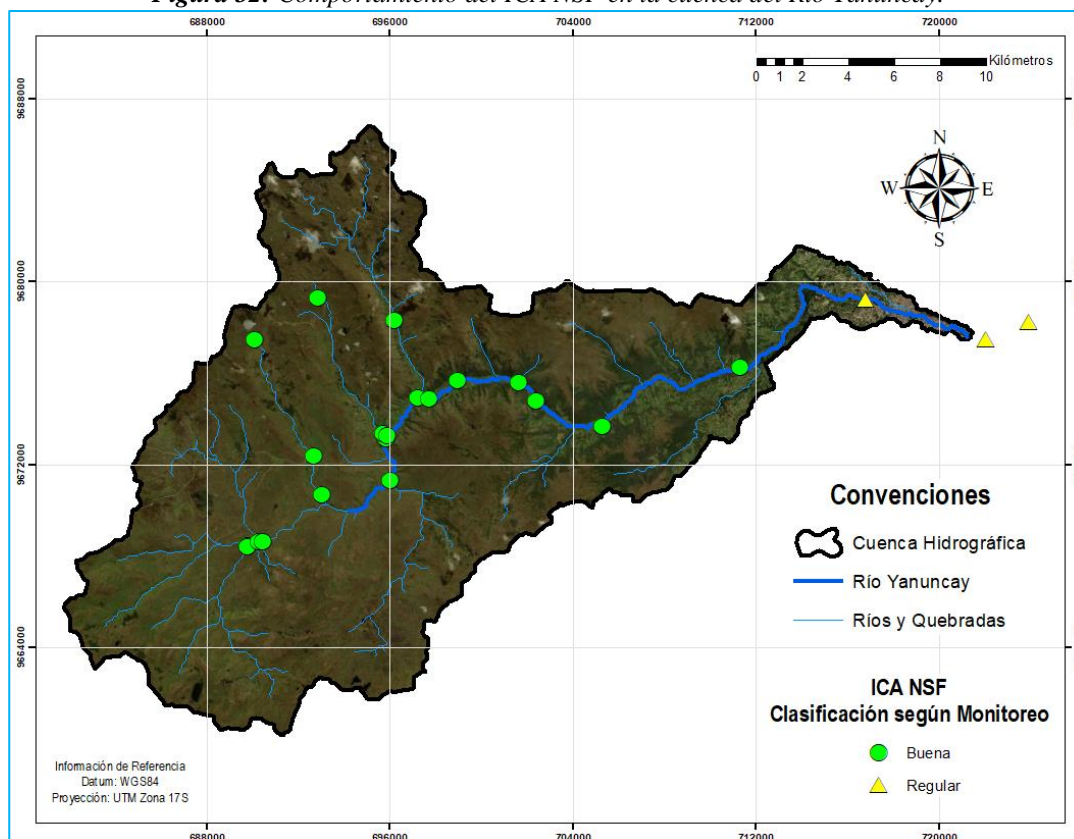
Fuente: Elaboración propia.

En primer lugar, se destaca la estación E-5 (YAN-GAL-020) con coordenadas geográficas 690109X - 9677448Y, la cual se encuentra a una altitud de 3942 metros sobre el nivel del mar. Según los datos recopilados, esta estación muestra la calidad del agua más alta registrada en todas las estaciones de monitoreo, con un valor de 87.97. Esto indica de manera contundente que a medida que se asciende en altitud, la calidad del agua tiende a mejorar en esta área específica.

Por otro lado, se observa que la estación E-23 (YAN-YAN-CONTROL) con coordenadas geográficas 723897X - 9678232Y, se sitúa a una altitud de 2477 metros sobre el nivel del mar. Según los datos recopilados, esta estación muestra la calidad del agua más baja registrada en toda la cuenca hidrográfica, con un valor de 54.03. Esto nos indica que a medida que el agua fluye a través de los ríos y quebradas y desciende en altitud, la calidad del agua tiende a deteriorarse.

Estos resultados respaldan la idea de que la calidad del agua se ve afectada por diversos factores a medida que los ríos y quebradas avanzan a lo largo de la cuenca hidrográfica y finalmente llegan a su punto de descarga. Los procesos naturales y las actividades humanas pueden influir en la calidad del agua a medida que esta se mueve a través del paisaje, interactúa con el suelo, la vegetación y otros elementos del entorno, tal como se puede observar en la Figura 32.

Figura 32: Comportamiento del ICA NSF en la cuenca del Río Yanuncay.



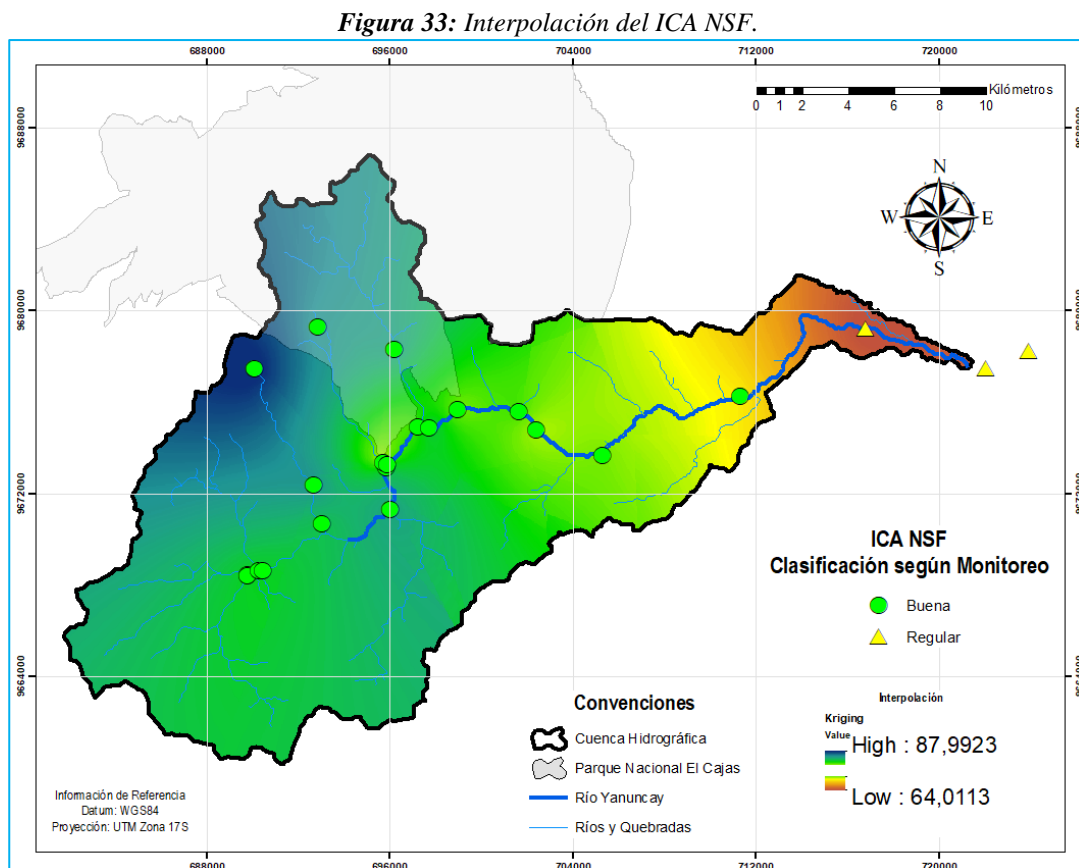
Fuente: Elaboración propia.

Dentro del conjunto de las 23 estaciones de monitoreo analizadas, es importante destacar que la gran mayoría, específicamente 20 estaciones, presentan una clasificación de calidad del agua catalogada como “Buena”. Estas estaciones se encuentran ubicadas en las partes altas de la cuenca hidrográfica, donde predominan las áreas de páramo y la actividad antropogénica es mínima (ver Anexo 13). Esta clasificación positiva indica que el agua en estas zonas se mantiene en un estado de calidad satisfactorio, con bajos niveles de contaminación y un buen equilibrio de los parámetros y componentes que determinan su pureza.

Sin embargo, es importante señalar que existen tres estaciones que presentan una clasificación de calidad del agua catalogada como “Regular”. Estas estaciones se ubican en la parte más baja de la cuenca hidrográfica, donde la actividad antropogénica es más predominante. En esta área, se observa la presencia de áreas de cultivo de maíz y cultivos de ciclo corto, así como zonas urbanas con una mayor concentración de población y actividades industriales (ver Anexo 13). Estos factores contribuyen a un mayor impacto en la calidad del agua, ya que se generan descargas de aguas residuales y existe una mayor probabilidad de contaminación debido a la actividad humana.

4.3 Interpolación de la calidad del agua por el método Kriging

La Figura 33 muestra de manera visual la interpolación realizada mediante el método Kriging, el cual resulta muy útil al proporcionar estimaciones precisas en áreas donde no hay datos medidos disponibles. Esta técnica permite obtener una representación detallada de la calidad del agua en toda el área de la cuenca del Río Yanuncay, teniendo en cuenta tanto las características del uso del suelo como la presencia del Parque Nacional El Cajas.



Fuente: Elaboración propia.

Mediante un análisis visual, se puede observar que, en las partes altas de la cuenca, donde se encuentra ubicado el Parque Nacional El Cajas, se representan áreas con un color azul, lo cual indica valores con una clasificación de calidad del agua considerada como “Buena” (con un valor específico de 87.9923). Esto sugiere que, en esta zona, donde predomina el páramo y la actividad humana es mínima, la calidad del agua se mantiene en niveles satisfactorios, con bajos niveles de contaminación y un equilibrio adecuado de los parámetros que determinan su pureza.

En contraste, en las áreas representadas con colores verde y amarillo, que comienzan desde el punto de descarga del Río Izhcayrumi, se encuentran cultivos de maíz y se observa una leve deterioración de la calidad del agua, aunque todavía se mantiene dentro de la clasificación de calidad “Buena”. Esto podría indicar que la actividad agrícola en esta zona tiene cierto impacto en la calidad del agua, pero aún no ha alcanzado niveles críticos que la clasifiquen como “Regular” o peor.

Finalmente, en las partes bajas de la cuenca, donde se intensifican las actividades agrícolas relacionadas con la zona urbana, se pueden observar áreas con un color rojizo, lo cual indica valores con una clasificación de calidad del agua considerada como “Regular” (con un valor específico de 64.0113). Esto sugiere que, en esta zona, donde hay una mayor influencia de actividades de cultivo, así como descargas de aguas residuales y actividad industrial, la calidad del agua se ve afectada negativamente y alcanza niveles que la clasifican como “Regular”.

4.4 Rendimiento de los modelos base por Árboles de regresión

Los resultados que se presentan en la Tabla 8 corresponden a modelos basados en datos, específicamente Árboles de regresión y Bosques aleatorios de regresión. Se realizaron cambios en los parámetros de los modelos de bosque aleatorio con el objetivo de evaluar si estos cambios mejoran su rendimiento como se indica en la sección 4.3.2 de Materiales y Métodos. Además, se llevaron a cabo pruebas de hipótesis entre estos modelos para determinar si existe evidencia de diferencias significativas entre ellos.

Tabla 8: Rendimiento de los modelos base.

Modelo	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	Promedio* \pm	Valor p
Árbol de regresión	0.847	0.934	0.858	0.927	0.918	0.777	0.907	0.909	0.889	0.911	0.888 \pm 0.045	0.002
Bosque aleatorio-1	0.910	0.960	0.953	0.974	0.976	0.900	0.950	0.968	0.951	0.964	0.951 \pm 0.024	*
Bosque aleatorio-2	0.902	0.942	0.939	0.962	0.978	0.945	0.951	0.969	0.951	0.980	0.952 \pm 0.022	0.675
Bosque aleatorio-3	0.927	0.958	0.962	0.970	0.972	0.916	0.951	0.968	0.946	0.967	0.954 \pm 0.018	0.706
Bosque aleatorio-4	0.897	0.927	0.946	0.955	0.974	0.953	0.954	0.967	0.942	0.981	0.950 \pm 0.023	0.969

* Promedio = R^2

* Modelo base con el que se realizaron las pruebas T (comparación).

Fuente: Elaboración propia.

El modelo base 1, que utiliza Árbol de regresión, exhibió un rendimiento inferior con un promedio de 0.888 y una desviación estándar de 0.045. Esto indica que este modelo tuvo un desempeño más débil en comparación con los otros modelos evaluados. Por otra parte, el modelo base 4, implementado mediante el uso de Bosque aleatorio, demostró un rendimiento destacado con un promedio de 0.954 y una desviación estándar de 0.018. Estos resultados evidencian que este modelo obtuvo el mejor rendimiento en comparación con los demás modelos analizados. La diferencia entre los promedios de rendimiento de estos dos modelos es considerable, con una diferencia de 0.066. Además, la desviación estándar del modelo base 4 es significativamente menor que la del modelo base 1, lo que indica una mayor consistencia y estabilidad en las estimaciones realizadas por el modelo.

Los resultados de las pruebas de hipótesis indicaron que no se encontraron diferencias significativas entre los modelos de Bosque aleatorio por regresión en términos de rendimiento. Esto se debe a que los valores p obtenidos en cada prueba fueron mayores a 0.05, que es el valor crítico establecido. Sin embargo, al realizar una prueba de hipótesis entre el Bosque aleatorio 1 y el Árbol de regresión, se encontró evidencia de diferencias significativas entre los modelos. Esto se debe a que el valor p obtenido en esta prueba fue de 0.002, que es menor a 0.05, el valor crítico establecido.

Dado que no se encontraron diferencias significativas en el rendimiento entre los diferentes modelos de Bosque aleatorio, se puede seleccionar el modelo de Bosque aleatorio-1 como el modelo base. Esta elección se basa en el hecho de que el uso de Bosque aleatorio ha demostrado un rendimiento superior en comparación con el algoritmo de Árbol de regresión. El modelo de Bosque aleatorio-1 ofrece varias ventajas en términos de rendimiento. En primer lugar, ha demostrado ser comparable a los modelos modificados (configuración) en términos de su capacidad para evaluar y estimar la calidad del agua en la cuenca del Río Yanuncay. Esto significa que, a pesar de su simplicidad y la falta de ajustes adicionales, el modelo de Bosque aleatorio-1 sigue siendo efectivo y confiable en su rendimiento.

4.5 Modelos de Bosque aleatorio por regresión entrenados con 3 parámetros

Los resultados que se muestran en la Tabla 9 representan los modelos construidos utilizando Bosques aleatorios de regresión con tres parámetros específicos. Estos modelos fueron sometidos a una evaluación exhaustiva utilizando la técnica de validación cruzada, que proporciona una

estimación confiable del rendimiento del modelo. Para cada modelo, se calculó tanto el promedio como la desviación estándar de las métricas de evaluación. Además, se llevaron a cabo pruebas de hipótesis para comparar estos modelos con el modelo base, conocido como Bosque Aleatorio-1, que incluía los nueve parámetros completos. El propósito de estas pruebas era determinar si los modelos con tres parámetros eran estadísticamente diferentes del modelo base en términos de su rendimiento.

Tabla 9: Rendimiento de los modelos entrenados con 3 parámetros.

Subconjuntos de parámetros	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	Promedio* ±	Valor p
'Oxígeno Disuelto', 'Coliformes Fecales', 'Nitratos'	0.903	0.940	0.976	0.947	0.950	0.870	0.881	0.906	0.900	0.945	0.922 ± 0.032	0.047
'Coliformes Fecales', 'pH', 'Nitratos'	0.923	0.935	0.969	0.957	0.933	0.887	0.895	0.942	0.907	0.956	0.930 ± 0.026	0.101
'Coliformes Fecales', 'DBO5', 'Nitratos'	0.918	0.962	0.975	0.952	0.948	0.845	0.922	0.938	0.951	0.915	0.932 ± 0.034	0.222
'Coliformes Fecales', 'Nitratos', 'Fosfatos'	0.935	0.953	0.913	0.960	0.977	0.878	0.934	0.964	0.935	0.970	0.942 ± 0.028	0.511
'Coliformes Fecales', 'Nitratos', 'Temperatura'	0.901	0.938	0.976	0.954	0.956	0.890	0.882	0.922	0.898	0.945	0.926 ± 0.031	0.077
'Coliformes Fecales', 'Nitratos', 'Turbiedad'	0.923	0.937	0.980	0.954	0.959	0.881	0.908	0.933	0.912	0.952	0.934 ± 0.028	0.192
'Coliformes Fecales', 'Nitratos', 'Sólidos Totales'	0.920	0.947	0.973	0.954	0.967	0.874	0.906	0.944	0.899	0.931	0.931 ± 0.030	0.156

* Promedio = R^2

Fuente: Elaboración propia.

Entre los modelos entrenados, se destaca el rendimiento sobresaliente del modelo que utilizó Coliformes fecales, Nitratos y Fosfatos como variables de entrada. Este modelo logró un promedio de 0.942 y una desviación estándar de 0.028, lo que indica que fue capaz de realizar estimaciones altamente precisas y consistentes en relación a la calidad del agua. Por otro lado, el modelo que empleó Oxígeno disuelto, Coliformes fecales y Nitratos como variables de entrada obtuvo un rendimiento inferior. Este modelo registró un promedio de 0.922 y una desviación estándar de 0.032, lo cual indica que sus predicciones fueron menos precisas y más inconsistentes en comparación con el modelo destacado.

Los resultados de las pruebas de hipótesis revelaron información importante sobre las diferencias de rendimiento entre los modelos. En particular, se encontró evidencia de que el modelo entrenado con Oxígeno disuelto, Coliformes fecales y Nitratos difiere significativamente del modelo base, Bosque aleatorio-1. Esto se debe a que el valor p obtenido en la prueba (0.047) es inferior al valor crítico establecido de 0.05. Por otro lado, en el caso de los demás modelos evaluados, no se encontró evidencia de diferencias significativas en comparación con el modelo base, Bosque aleatorio-1. Esto se debe a que los valores p obtenidos en estas pruebas fueron mayores al valor crítico de 0.05.

Estos resultados subrayan la relevancia de las variables seleccionadas para entrenar los modelos. La exclusión de Oxígeno disuelto en los modelos desempeñó un papel fundamental en su mejor rendimiento. Esto sugiere que la presencia o ausencia de esta variable puede tener un impacto significativo en la capacidad de los modelos. Es importante destacar que las demás variables utilizadas en los modelos pueden desempeñar un papel crucial en la evaluación precisa de la calidad del agua. Al considerar estas variables en conjunto, se logra una representación más completa y precisa, lo que puede conducir a un mejor rendimiento general del modelo.

4.6 Modelos de Bosque aleatorio por regresión entrenados con 4 parámetros

Los resultados que se muestran en la Tabla 10 representan los modelos construidos utilizando Bosques aleatorios de regresión con cuatro parámetros específicos. Estos modelos fueron sometidos a una evaluación exhaustiva utilizando la técnica de validación cruzada, que proporciona una estimación confiable del rendimiento del modelo. Para cada modelo, se calculó tanto el promedio como la desviación estándar de las métricas de evaluación. Además, se llevaron a cabo pruebas de hipótesis para comparar estos modelos con el modelo base, conocido como

Bosque Aleatorio-1, que incluía los nueve parámetros completos. El propósito de estas pruebas era determinar si los modelos con cuatro parámetros eran estadísticamente diferentes del modelo base en términos de su rendimiento.

Tabla 10: Rendimiento de los modelos entrenados con 4 parámetros.

Subconjuntos de parámetros	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	Promedio* ±	Valor p
'Oxígeno Disuelto', 'Coliformes Fecales', 'pH', 'Nitratos'	0.915	0.938	0.974	0.954	0.935	0.894	0.872	0.929	0.896	0.955	0.926 ± 0.030	0.074
'Oxígeno Disuelto', 'Coliformes Fecales', 'DBO5', 'Nitratos'	0.910	0.960	0.972	0.948	0.955	0.863	0.926	0.945	0.949	0.927	0.935 ± 0.030	0.259
'Oxígeno Disuelto', 'Coliformes Fecales', 'Nitratos', 'Fosfatos'	0.904	0.959	0.909	0.954	0.973	0.879	0.930	0.960	0.940	0.964	0.937 ± 0.029	0.312
'Oxígeno Disuelto', 'Coliformes Fecales', 'Nitratos', 'Temperatura'	0.920	0.943	0.981	0.955	0.954	0.900	0.852	0.922	0.917	0.949	0.929 ± 0.034	0.149
'Oxígeno Disuelto', 'Coliformes Fecales', 'Nitratos', 'Turbiedad'	0.924	0.944	0.976	0.956	0.958	0.871	0.892	0.939	0.908	0.954	0.932 ± 0.031	0.186

'Oxígeno Disuelto', 'Coliformes Fecales', 'Nitratos', 'Sólidos Totales'	0.918	0.950	0.975	0.957	0.965	0.879	0.902	0.949	0.906	0.935	0.933 ± 0.030	0.200
'Coliformes Fecales', 'pH', 'DBO5', 'Nitratos'	0.926	0.956	0.972	0.961	0.951	0.873	0.920	0.937	0.953	0.928	0.938 ± 0.027	0.304
'Coliformes Fecales', 'pH', 'Nitratos', 'Fosfatos'	0.934	0.957	0.920	0.965	0.969	0.876	0.932	0.964	0.946	0.973	0.944 ± 0.028	0.611
'Coliformes Fecales', 'pH', 'Nitratos', 'Temperatura'	0.925	0.932	0.977	0.960	0.949	0.904	0.889	0.941	0.916	0.965	0.936 ± 0.026	0.233
'Coliformes Fecales', 'pH', 'Nitratos', 'Turbiedad'	0.928	0.936	0.972	0.961	0.946	0.878	0.908	0.936	0.922	0.961	0.935 ± 0.027	0.209
'Coliformes Fecales', 'pH', 'Nitratos', 'Sólidos Totales'	0.918	0.950	0.975	0.957	0.965	0.879	0.902	0.949	0.906	0.935	0.933 ± 0.027	0.216
'Coliformes Fecales', 'DBO5', 'Nitratos', 'Fosfatos'	0.919	0.961	0.935	0.966	0.968	0.875	0.943	0.964	0.951	0.955	0.944 ± 0.027	0.599
'Coliformes Fecales',	0.931	0.959	0.980	0.957	0.962	0.894	0.926	0.942	0.956	0.930	0.944 ± 0.023	0.578

'DBO5', 'Nitratos', 'Temperatura'												
'Coliformes Fecales', 'DBO5', 'Nitratos', 'Turbiedad'	0.931	0.950	0.975	0.957	0.966	0.858	0.924	0.951	0.946	0.945	0.940 ± 0.031	0.471
'Coliformes Fecales', 'DBO5', 'Nitratos', 'Sólidos Totales'	0.931	0.951	0.978	0.969	0.958	0.888	0.928	0.948	0.946	0.945	0.944 ± 0.024	0.598
'Coliformes Fecales', 'Nitratos', 'Fosfatos', 'Temperatura'	0.938	0.961	0.915	0.964	0.983	0.904	0.923	0.954	0.948	0.971	0.946 ± 0.024	0.733
'Coliformes Fecales', 'Nitratos', 'Fosfatos', 'Turbiedad'	0.923	0.953	0.935	0.966	0.977	0.880	0.928	0.963	0.941	0.967	0.943 ± 0.027	0.580
'Coliformes Fecales', 'Nitratos', 'Fosfatos', 'Sólidos Totales'	0.931	0.956	0.945	0.970	0.976	0.882	0.943	0.964	0.922	0.967	0.946 ± 0.027	0.721
'Coliformes Fecales', 'Nitratos', 'Temperatura', 'Turbiedad'	0.930	0.937	0.982	0.958	0.966	0.894	0.903	0.933	0.935	0.960	0.940 ± 0.026	0.389

'Coliformes Fecales', 'Nitratos', 'Temperatura', 'Sólidos Totales'	0.925	0.939	0.977	0.964	0.968	0.894	0.907	0.936	0.917	0.941	0.937 ± 0.026	0.259
'Coliformes Fecales', 'Nitratos', 'Turbiedad', 'Sólidos Totales'	0.929	0.945	0.975	0.959	0.971	0.871	0.909	0.949	0.926	0.936	0.937 ± 0.030	0.314

* Promedio = R^2

Fuente: Elaboración propia.

Entre los modelos entrenados, se destacan dos por su excelente rendimiento. El primero es el modelo entrenado con Coliformes fecales, Nitratos, Fosfatos y Temperatura, el cual logró una desviación estándar de 0.024 y un promedio de 0.946. El segundo es el modelo entrenado con Coliformes fecales, Nitratos, Fosfatos y Sólidos totales, el cual alcanzó una desviación estándar de 0.027 y un promedio de 0.946. Estos modelos demuestran una alta precisión y consistencia en la evaluación de la calidad del agua. Por otro lado, el modelo entrenado con Oxígeno disuelto, Coliformes fecales, pH y Nitratos mostró un rendimiento inferior en comparación. Este modelo obtuvo un promedio de 0.926 y una desviación estándar de 0.030. Estos valores indican que sus predicciones fueron menos precisas y más variables en comparación con los modelos de mejor rendimiento.

Los resultados de las pruebas de hipótesis revelaron que no hay evidencia de diferencias significativas entre todos los modelos entrenados con 4 parámetros y el modelo base, Bosque aleatorio-1. Esto se evidencia en los valores p obtenidos en estas pruebas, los cuales son mayores al valor crítico establecido de 0.05. Estos resultados indican que los modelos con 4 parámetros no presentan un rendimiento estadísticamente distinto al del modelo base.

Una vez más, estos resultados destacan la importancia de las variables seleccionadas para el entrenamiento de los modelos. Es relevante señalar que el modelo que incorpora el parámetro de Oxígeno disuelto mostró el valor p más bajo de 0.074, lo que lo acerca al valor crítico de 0.05. Aunque este valor p no alcanza el nivel de significancia establecido, resulta interesante observar que el modelo que considera el Oxígeno disuelto podría presentar características distintas en términos de rendimiento en comparación con el modelo base. Aunque no se puede afirmar con certeza debido a la falta de suficiente evidencia estadística, este hallazgo sugiere que el Oxígeno disuelto podría desempeñar un papel relevante en la evaluación de la calidad del agua.

4.7 Elección del modelo más apropiado con la capacidad similar al ICA NSF

Además de los indicadores estadísticos mencionados anteriormente que se utilizaron para la selección de los modelos, se consideró conveniente tener en cuenta la logística requerida por los análisis de laboratorio de cada parámetro. Cada parámetro, como físico, químico y biológico, requiere una metodología específica tanto en el campo como en el laboratorio. Por lo tanto, al evaluar esta condición, se realizó una calificación cualitativa para determinar la selección más adecuada (ver Tabla 11).

Tabla 11: Calificación cualitativa para los 9 parámetros requeridos por ICA NSF.

Parámetro	Facilidad de muestreo en campo	Facilidad de análisis en laboratorio	Requerimiento de equipos especializados
Oxígeno Disuelto	Moderada	Moderada	Medio
Coliformes Fecales	Moderada	Moderada	Medio
pH	Simple	No aplica	Bajo
DBO ₅	Moderada	Compleja	Alto
Nitratos	Simple	Simple	Medio
Fosfatos	Simple	Simple	Medio
Temperatura	Simple	No aplica	Bajo
Turbiedad	Simple	No aplica	Medio
Sólidos Totales	Simple	Moderada	Alto

Fuente: Elaboración propia.

4.7.1 Modelo de Bosque aleatorio entrenado con 3 parámetros

Después de realizar un exhaustivo análisis de los modelos creados con 3 parámetros y realizar una calificación cualitativa para cada parámetro requerido por el ICA NSF, hemos llegado a la conclusión de que el modelo de Bosque aleatorio por regresión entrenado con los parámetros de Coliformes Fecales, pH y Nitratos es la elección más adecuada. Este modelo tiene un promedio de 0.930 y una desviación estándar de 0.026. La razón principal detrás de nuestra elección radica en que los modelos que demuestran un rendimiento altamente preciso y consistente en la evaluación de la calidad del agua siempre incluyen los parámetros de Coliformes Fecales y Nitratos como indicadores clave de contaminación. En ausencia de estos dos parámetros, el rendimiento del modelo se ve considerablemente afectado.

Además, hemos considerado el pH como un parámetro crucial debido a su facilidad de muestreo en campo, la ausencia de la necesidad de un análisis de laboratorio y el bajo requerimiento de equipos especializados. Por lo tanto, hemos priorizado el pH sobre los Fosfatos. Es importante destacar que el modelo entrenado con Coliformes Fecales, Nitratos y Fosfatos se ha demostrado como el mejor entre todos los modelos entrenados con tres parámetros, con un promedio de 0.942 y una desviación estándar de 0.028. Sin embargo, debemos tener en cuenta que los Fosfatos sí

requieren un análisis de laboratorio y el requerimiento de equipos especializados es medio, lo cual aumenta la complejidad y los recursos necesarios para su inclusión en el modelo.

4.7.2 Modelo de Bosque aleatorio entrenado con 4 parámetros

Después de realizar un análisis exhaustivo tanto de los modelos creados con 4 parámetros como de la calificación cualitativa para cada parámetro requerido por el ICA NSF, hemos llegado a la conclusión de que el modelo de Bosque aleatorio por regresión entrenado con los parámetros de Coliformes Fecales, pH, Nitratos y Temperatura es la elección más apropiada. Este modelo tiene un promedio de 0.936 y una desviación estándar de 0.026. La razón principal detrás de nuestra elección es que, al igual que en los modelos creados con 3 parámetros, los modelos que demuestran un rendimiento altamente preciso y consistente en la evaluación de la calidad del agua siempre incluyen los parámetros de Coliformes Fecales y Nitratos como indicadores clave de contaminación. La ausencia de estos dos parámetros tiene un impacto significativo en el rendimiento del modelo.

Además, hemos considerado tanto el pH como la Temperatura como parámetros importantes debido a su facilidad de muestreo en campo, la falta de necesidad de un análisis de laboratorio y el bajo requerimiento de equipos especializados. Por lo tanto, hemos priorizado el pH y la Temperatura sobre los Fosfatos y los Sólidos Totales. Es importante destacar que el modelo entrenado con Coliformes Fecales, Nitratos, Fosfatos y Sólidos Totales ha demostrado ser el mejor entre todos los modelos entrenados con 4 parámetros, con un promedio de 0.946 y una desviación estándar de 0.027. Sin embargo, debemos tener en cuenta que los Fosfatos requieren un análisis de laboratorio, lo cual aumenta la complejidad y los recursos necesarios para su inclusión en el modelo. Del mismo modo, los Sólidos Totales tienen una facilidad de análisis de laboratorio moderada y el requerimiento de equipos especializados es alto.

5 Conclusiones

En conclusión, la recopilación exitosa de datos hidrológicos y meteorológicos, así como la disponibilidad de una base de datos completa de parámetros físicos, químicos y microbiológicos de la cuenca del Río Yanuncay proporcionada por ETAPA E. P., ha sido de suma importancia para el estudio y comprensión de los recursos hídricos en esta área. La utilización del software ArcMap ha sido esencial para analizar y visualizar estos datos, lo que ha proporcionado una sólida base para la toma de decisiones fundamentadas. Asimismo, el uso del lenguaje de programación Python ha sido fundamental en la obtención de modelos basados en datos, cumpliendo con los resultados esperados en este trabajo. Todo esto ha contribuido a un enfoque integral y efectivo en la evaluación de los recursos hídricos en la cuenca del Río Yanuncay.

El análisis realizado en cada estación de monitoreo revela una aparente estabilidad en las condiciones hídricas de los ríos y quebradas en la cuenca del Río Yanuncay durante el período examinado. Esta estabilidad es un indicador positivo de la calidad del agua en la cuenca, ya que los niveles de los parámetros medidos se encuentran dentro de rangos aceptables y no alcanzan las categorías más bajas de clasificación, como “Mala” o “Pésima”. No obstante, es fundamental tener en cuenta que se requiere un monitoreo continuo y a largo plazo en estaciones específicas, como la E-5 con 3 monitoreos en el año 2014 y la E-6 con 2 monitoreos en el año 2014 y uno en el año 2015, ubicadas en el Río Galgal, así como la estación E-9 con 2 monitoreos en el año 2014 y uno en el año 2015, ubicada en el Río Yanuncay, y finalmente la estación E-14 con un monitoreo en el año 2020 y otro en el año 2021 ubicada en el Río Soldados. Este seguimiento a largo plazo permitiría obtener una imagen más precisa y completa de la calidad del agua tanto en esas áreas puntuales como en toda la cuenca en general.

Los resultados de este estudio han demostrado la efectividad de los modelos basados en Árboles de regresión para predecir la calidad del agua en la cuenca del Río Yanuncay, a pesar de utilizar un número limitado de parámetros en comparación con el Índice de la Fundación Nacional de Saneamiento lo cual demuestra la hipótesis planteada. Estos hallazgos sugieren la posibilidad de optimizar los requisitos de monitoreo sin comprometer la precisión en la estimación del Índice de Calidad del Agua NSF. Entre los diversos modelos evaluados, el modelo de Bosque aleatorio por regresión, entrenado con los parámetros de Coliformes Fecales, pH y Nitratos, ha demostrado ser la elección más acertada para concluir este trabajo experimental. Este modelo en particular ha

obtenido un R cuadrado de 0.930 y una desviación estándar de 0.026, lo cual indica un rendimiento sobresaliente en la predicción de la calidad del agua en la cuenca del Río Yanuncay.

6 Recomendaciones

Recomendamos enfáticamente la utilización del modelo Bosque aleatorio por regresión, ya que ha demostrado una ventaja sobresaliente en términos de simplicidad y facilidad de uso de parámetros. Este enfoque no requiere ajustes adicionales ni configuraciones complicadas, lo que simplifica su implementación y aplicación. Esta característica resulta sumamente beneficiosa para aquellos interesados en obtener resultados precisos sin tener que invertir una cantidad considerable de tiempo y recursos en la configuración y optimización del modelo. Al aprovechar la simplicidad de este enfoque, los usuarios podrán obtener rápidamente estimaciones precisas de la calidad del agua en la cuenca del Río Yanuncay, lo que les permitirá tomar decisiones informadas y realizar acciones correctivas de manera más eficiente.

A pesar de los resultados alentadores obtenidos, es de vital importancia implementar medidas efectivas de control y monitoreo en las áreas de la cuenca del Río Yanuncay. Esto garantizará una evaluación constante de la calidad del agua y la detección oportuna de cualquier cambio o tendencia preocupante en los parámetros medidos. Además, se requiere una promoción activa de prácticas sostenibles y de tratamiento de aguas residuales en la zona urbana, con el fin de lograr una mejora continua en la calidad del agua y preservar los valiosos recursos hídricos para las generaciones futuras. Es esencial establecer una gestión integral del agua que aborde no solo la predicción y monitoreo, sino también la implementación de medidas preventivas y correctivas en toda la cuenca. Esto incluye la educación y participación de la comunidad local, así como la colaboración con entidades gubernamentales y organizaciones ambientales, con el objetivo de fomentar una gestión responsable y sostenible del agua en la cuenca del Río Yanuncay.

7 Bibliografía

Alvaro. (2023). *Árboles de decisión en Python*. Machine Learning Para Todos. <https://machinelearningparatodos.com/arboles-de-decision-en-python/>

Baird, R. B., Eaton, A. D., & Rice, E. W. (2012). Standard Methods for the Examination of Water and Wastewater Standard Methods for the Examination of Water and Wastewater. *Public Health*, 51(1), 940–940. <http://www.ajph.org/cgi/doi/10.2105/AJPH.51.6.940-a>

Bartram, J., & Ballance, R. (1996). Water Quality Monitoring- A Practical Guide to the Design and Implementation of Freshwater Quality Studies and Monitoring Programmes. In *United Nations Environment Programme; World Health Organization*. E & FN Spon. <https://doi.org/10.1002/ejoc.201200111>

Bear, J. (2007). *Hydraulics of Groundwater* (pp. 70–75). [https://books.google.es/books?hl=es&lr=&id=yQogKirZW-MC&oi=fnd&pg=PP1&dq=Introduction+to+Groundwater+Hydraulics&ots=ijdeaQ_ft2&sig=C3K0fe6HQuH5FC7muTjrLAWCa_E#v=onepage&q=Introduction to Groundwater Hydraulics&f=false](https://books.google.es/books?hl=es&lr=&id=yQogKirZW-MC&oi=fnd&pg=PP1&dq=Introduction+to+Groundwater+Hydraulics&ots=ijdeaQ_ft2&sig=C3K0fe6HQuH5FC7muTjrLAWCa_E#v=onepage&q=Introduction+to+Groundwater+Hydraulics&f=false)

Behmel, S., Damour, M., Ludwig, R., & Rodriguez, M. J. (2016). Water quality monitoring strategies — A review and future perspectives. *Science of the Total Environment*, 571, 1312–1329. <https://doi.org/10.1016/j.scitotenv.2016.06.235>

Bekele, E., Page, D., Vanderzalm, J., Kaksonen, A., & Gonzalez, D. (2018). Water recycling via aquifers for sustainable urban water quality management: Current status, challenges and opportunities. In *Water (Switzerland)* (Vol. 10, Issue 4, p. 457). Multidisciplinary Digital Publishing Institute. <https://doi.org/10.3390/w10040457>

Bergamasco, A., Nguyen, H. Q., Caruso, G., Xing, Q., & Carol, E. (2021). Advances in water quality monitoring and assessment in marine and coastal regions. In *Water (Switzerland)* (Vol. 13, Issue 14, p. 1926). Multidisciplinary Digital Publishing Institute. <https://doi.org/10.3390/w13141926>

Bhardwaj, J., Gupta, K. K., & Gupta, R. (2015, April 27). A review of emerging trends on water quality measurement sensors. *Proceedings - International Conference on Technologies for Sustainable Development, ICTSD 2015*. <https://doi.org/10.1109/ICTSD.2015.7095919>

Bilonick, R. A. (1991). An Introduction to Applied Geostatistics. *Technometrics*, 33(4), 483–485. <https://doi.org/10.1080/00401706.1991.10484886>

Bishop, C. M. (2006). Pattern Recognition and Machine Learning. *Springer*, 738. <https://link.springer.com/book/9780387310732>

Biswas, A. K. (1991). Water for sustainable development in the 21st century: a global perspective. *GeoJournal*, 24(4), 341–345. <https://doi.org/10.1007/BF00578255>

Boyd, C. (2019). *Water Quality: An Introduction*. Springer. <https://books.google.es/books?hl=es&lr=&id=h0mvDwAAQBAJ&oi=fnd&pg=PR5&dq=Water+Quality:+An+Introduction&ots=YIIFRz6BND&sig=K47xKJHBt-B4QxXTMZU5LvLdHo#v=onepage&q=Water+Quality%3A+An+Introduction&f=false>

Boyle, D. P., Gupta, H. V., & Sorooshian, S. (2000). Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods. *Water Resources Research*, 36(12), 3663–3674. <https://doi.org/10.1029/2000WR900207>

Brown, R. M., McClelland, N. I., Deininger, R. A., & Tozer, R. G. (1970). A water quality index-do we dare. *Water and sewage works*, 117(10).

Bui, D. T., Khosravi, K., Tiefenbacher, J., Nguyen, H., & Kazakis, N. (2020). Improving prediction of water quality indices using novel hybrid machine-learning algorithms. *Science of the Total Environment*, 721, 137612. <https://doi.org/10.1016/j.scitotenv.2020.137612>

Cabrera Mendieta, D. R. (2018). *Modelado de sistemas dinámicos con MachineLearning: aplicaciones al mantenimiento basado en la condición*.

Cañar Uyaguari, M. I. (2023). *Desarrollo de un sistema para determinar la calidad de agua en la cuenca del río Paute, basado en la identificación de macroinvertebrados acuáticos como indicadores de calidad de agua, utilizando técnicas de visión por computador*. <http://dspace.ups.edu.ec/handle/123456789/26620>

Descy, J.-P., & Coste, M. (1991). A test of methods for assessing water quality based on diatoms. *SIL Proceedings, 1922-2010*, 24(4), 2112–2116. <https://doi.org/10.1080/03680770.1989.11899905>

EPA, (Environmental Protection Agency). (2003). National Management Measures to Control Nonpoint Source Pollution from Agriculture. In *National Management Measures to Control Nonpoint Source Pollution from Agriculture*. <http://www.epa.gov/owow/nps/urbanmm/index.html>.

EPA, (Environmental Protection Agency). (2017). Water quality standards handbook. In *Water quality standards handbook: Chapter 3: Water Quality Criteria* (pp. 1–26). https://books.google.es/books?hl=es&lr=&id=BxSBa2FYt_wC&oi=fnd&pg=PA21&dq=Water+Quality+Standards+Handbook&ots=NxAW1hB5CV&sig=1bPOhoan27kFVKfD2JdQ2_B5Ijo#v=onepage&q=Water+Quality+Standards+Handbook&f=false

Espinosa Ramírez, A. J. (2018). El agua, un reto para la salud pública: la calidad del agua y las oportunidades para la vigilancia en salud ambiental. *Repositorio.Unal.Edu.Co*, 26, 11–18. <https://repositorio.unal.edu.co/handle/unal/63149>

ESRI, (Environmental Systems Research Institute). (2021). *Qué es ArcMap*. ArcGIS Desktop. <https://desktop.arcgis.com/es/arcmap/latest/map/main/what-is-arcmap-.htm>

Fernández, N., & Solano, F. (2007). Capítulo III: índices de calidad (ICAs) y de contaminación (ICOs) del agua de importancia mundial. *Índices De Calidad Y De Contaminación Del Agua*, 76. http://www.unipamplona.edu.co/unipamplona/portaIIG/home_10/recursos/general/pag_contenido/libros/06082010/icatest_capitulo3.pdf

Fewtrell, L., & Bartram, J. (2001). *Water quality: guidelines, standards, and health*. 431. https://books.google.com/books/about/Water_Quality.html?hl=es&id=Wl6UQUV6dF4C

Flores, A. (2021). Lista de las mejores librerías de Python para el 2022. *Desarrollo Web y Tecnología*, 1–22. <https://www.crehana.com/blog/transformacion-digital/librerias-python/>

Freeman, H. M. (1995). *Industrial pollution prevention handbook*. McGraw-Hill, Inc., New York, NY (United States).

Fu, B., Horsburgh, J. S., Jakeman, A. J., Gualtieri, C., Arnold, T., Marshall, L., Green, T. R., Quinn, N. W. T., Volk, M., Hunt, R. J., Vezzano, L., Croke, B. F. W., Jakeman, J. D., Snow, V., & Rashleigh, B. (2020). Modeling Water Quality in Watersheds: From Here to the Next Generation. In *Water Resources Research* (Vol. 56, Issue 11, p. e2020WR027721). John Wiley & Sons, Ltd. <https://doi.org/10.1029/2020WR027721>

Gao, D., Chen, T., Yang, K., Zhou, J., & Ao, T. (2021). Projecting the impacts of climate change on streamflow in the upper reaches of the Yangtze river basin. *Journal of Water and Climate Change*, 12(5), 1724–1743. <https://doi.org/10.2166/wcc.2020.082>

Golet, G. H., Roberts, M. D., Luster, R. A., Werner, G., Larsen, E. W., Unger, R., & White, G. G. (2006). Assessing societal impacts when planning restoration of large alluvial rivers: A case study of the Sacramento River Project, California. *Environmental Management*, 37(6), 862–879. <https://doi.org/10.1007/s00267-004-0167-x>

Goovaerts, P. (1997). Geostatistics for natural resources evaluation. *Applied Geostatistics*, 483.

Gribbin, J. E. (2013). *Introduction to Hydraulics & Hydrology: With Applications for Stormwater Management*. Cengage Learning.

Hastie, T., Friedman, J., & Tibshirani, R. (2001). *The Elements of Statistical Learning*. Springer. <https://doi.org/10.1007/978-0-387-21606-5>

Ibekwe, A. M., Murinda, S. E., & Graves, A. K. (2011). Microbiological evaluation of water quality from urban watersheds for domestic water supply improvement. *International Journal of Environmental Research and Public Health*, 8(12), 4460–4476. <https://doi.org/10.3390/ijerph8124460>

Inegi. (2012). *Modelos Digitales de Elevación (MDE) - Descripción*. <https://www.inegi.org.mx/contenidos/temas/mapas/relieve/continental/metadatos/mde.pdf>

Jiménez, M., & Vélez, M. (2006). Análisis comparativo de indicadores de la calidad de agua superficial. *Avance En Recursos Hidráulicos*, 53–69. <http://www.redalyc.org/articulo.oa?id=145020399004>

Khatib, J. M., Baydoun, S., & ElKordi, A. A. (2018). Water pollution and urbanisation trends in Lebanon: Litani river basin case study. In *Urban Pollution: Science and Management* (pp. 397–415). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781119260493.ch30>

Koranga, M., Pant, P., Kumar, T., Pant, D., Bhatt, A. K., & Pant, R. P. (2022). Efficient water quality prediction models based on machine learning algorithms for Nainital Lake, Uttarakhand. *Materials Today: Proceedings*, 57, 1706–1712. <https://doi.org/10.1016/j.matpr.2021.12.334>

Krige, D. G. (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of Southern African Institute of Mining and Metallurgy*, 52(6), 119–139. https://doi.org/10.10520/AJA0038223X_4792

Lazorchak, J., Hill, B., Averill, D., Peck, D., & Klemm, D. (1998). Environmental Monitoring and Assessment Program - Surface Waters : Field Operations and Methods for Measuring the Ecological Condition of Non-wadeable Rivers and Streams. *Epa/620/R-94/00F, September*, 1–208.

Li, M.-C. C., Liang, S.-X. X., Sun, Z. C., & Zhang, G. Y. (2012). Optimal dynamic temporal-spatial parameter inversion methods for the marine integrated element water quality model using a data-driven neural network. *Journal of Marine Science and Technology (Taiwan)*, 20(5), 575–583. <https://doi.org/10.6119/JMST-012-0109-1>

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, 2(3), 18–22. <http://www.stat.berkeley.edu/>

Lin, L., Yang, H., & Xu, X. (2022). Effects of Water Pollution on Human Health and Disease Heterogeneity: A Review. In *Frontiers in Environmental Science* (Vol. 10, p. 880246). Frontiers Media S.A. <https://doi.org/10.3389/fenvs.2022.880246>

Martínez-Dalmau, J., Gutiérrez-Martín, C., Kahil, T., & Berbel, J. (2023). Impact of alternative water policies for drought adaptation in the Guadalquivir Mediterranean river basin, southern Spain. *Journal of Hydrology: Regional Studies*, 47, 101444. <https://doi.org/10.1016/j.ejrh.2023.101444>

McKinney, W. (2010). Data Structures for Statistical Computing in Python. *Proceedings of the 9th Python in Science Conference*, 56–61. <https://doi.org/10.25080/majora-92bf1922-00a>

Merrington, G., Nfa, D. L. W., Parkinson, R., Redman, M., & Winder, L. (2003). *Agricultural Pollution: Environmental Problems and Practical Solutions*. Spon Press. https://books.google.com/books/about/Agricultural_Pollution.html?hl=es&id=FPFdxxEcxfAC

Montgomery, D. C. (2019). *Introduction to statistical quality control* (p. 5). https://books.google.com/books/about/Introduction_to_Statistical_Quality_Cont.html?hl=es&id=YWLNEAAAQBAJ

Palaniappan, M., Gleick, P. H., Allen, L., Cohen, M. J., Smith, J. C., & Smith, C. (2011). Water quality. *Encyclopedia of Earth Sciences Series, 14*, 1043–1046. https://doi.org/10.5822/978-1-59726-228-6_3/COVER

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research, 12*(85), 2825–2830. <http://scikit-learn.sourceforge.net>.

Pinto, U., & Maheshwari, B. (2014). A framework for assessing river health in peri-urban landscapes. *Ecohydrology and Hydrobiology, 14*(2), 121–131. <https://doi.org/10.1016/j.ecohyd.2014.04.001>

Puma, H. (2020). *Bosques Aleatorios*. Medium. <https://medium.com/@hpumah/bosques-aleatorios-482163ace92e>

Python Software Foundation. (2023). *Python 3.12.1 Documentation: Math Module*. Recuperado de <https://docs.python.org/3/library/math.html>

Rahat, S. H., Steissberg, T., Chang, W., Chen, X., Mandavya, G., Tracy, J., Wasti, A., Atreya, G., Saki, S., Bhuiyan, M. A. E., & Ray, P. (2023). Remote sensing-enabled machine learning for river water quality modeling under multidimensional uncertainty. *Science of the Total Environment, 898*, 165504. <https://doi.org/10.1016/j.scitotenv.2023.165504>

Safoniuk, M. (2004). Wastewater Engineering: Treatment and Reuse. *Chemical Engineering, 111*(7), 10–12. <https://go.gale.com/ps/i.do?p=AONE&sw=w&issn=00092460&v=2.1&it=r&id=GALE%7CA119904663&sid=googleScholar&linkaccess=fulltext>

SNET. (2007). *ÍNDICE DE CALIDAD DEL AGUA GENERAL "ICA"*.
<https://www.snet.gob.sv/Hidrologia/Documentos/calculoICA.pdf>

Sutton, R., & Barto, A. (2018). *Reinforcement Learning, second edition: An Introduction*. MIT Press.

https://books.google.es/books?hl=es&lr=&id=uWV0DwAAQBAJ&oi=fnd&pg=PR7&dq=Reinforcement+Learning:+An+Introduction&ots=mjnKr4W0j9&sig=oszC19qntJWe_KuSL8A6G8RI nWg#v=onepage&q=Reinforcement Learning%3A An Introduction&f=false

Svirčev, Z. B., Tokodi, N., Drobac, D., & Codd, G. A. (2014). Cyanobacteria in aquatic ecosystems in Serbia: Effects on water quality, human health and biodiversity. In *Systematics and Biodiversity* (Vol. 12, Issue 3, pp. 261–270). Taylor & Francis.
<https://doi.org/10.1080/14772000.2014.921254>

Truchy, A., Göthe, E., Angeler, D. G., Ecke, F., Sponseller, R. A., Bundschuh, M., Johnson, R. K., & McKie, B. G. (2019). Partitioning spatial, environmental, and community drivers of ecosystem functioning. *Landscape Ecology*, 34(10), 2371–2384. <https://doi.org/10.1007/s10980-019-00894-9>

TYC-GIS. (2015). *Delimitar una Cuenca Hidrográfica con la herramienta "Hydrology" de ArcGIS*. c. <https://www.cursosgis.com/delimitar-una-cuenca-hidrografica-con-la-herramienta-hydrology-de-arcgis/>

Valsami-Jones, E. (2004). *Phosphorus in Environmental Technology*. 272–273.
<https://books.google.es/books?hl=es&lr=&id=J7bwK8k3DOIC&oi=fnd&pg=PR13&dq=Phosphorus+in+Environmental+Technologies&ots=OEITPI8eOF&sig=UACgS8IxlkoXuQ4pKmb0BF NuULY#v=onepage&q=Phosphorus in Environmental Technologies&f=false>

Van Rossum, G. (2007). Python (programming language). In *USENIX Annual Technical Conference*, 1–14. http://kelas-karyawan-bali.kurikulum.org/IT/en/2420-2301/Python_3721_kelas-karyawan-bali-kurikulumngetesumum.html

Vienna, J., & Spearing, D. (2012). Environmental Issues and Waste Management Technologies in the Ceramic and Nuclear Industries -IX. *Materials and Manufacturing Processes*, 12(3), 563–564. <https://doi.org/10.1080/10426919708935167>

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., ... Vázquez-Baeza, Y. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3), 261–272. <https://doi.org/10.1038/s41592-019-0686-2>

Wani, S. P., & Garg, K. K. (2009). Watershed Management Concept and Principles. *150 Problemas de ECG, January*, 219–220.

Ward, R. C. (2001). Development and use of water quality criteria and standards in the United States. *Regional Environmental Change*, 2(2), 66–72. <https://doi.org/10.1007/s101130100028>

Warren, C. (1971). Biology and Water Pollution Control. *The Journal of Wildlife Management*, 38(2), 379. <https://doi.org/10.2307/3800757>

WHO. (2008). Guidelines for Drinking-water Quality. *Guidelines for Drinking-Water Quality*, 1.

Witten, D., & James, G. (2013). An introduction to statistical learning with applications in R. *Springer publication*.

WordPress. (n.d.). *Diagrama de Flujo*. TECNOLOGÍA E INFORMATICA. Retrieved January 21, 2024, from <https://tecnoinfoenelcolegio.wordpress.com/diagrama-de-flujo/>

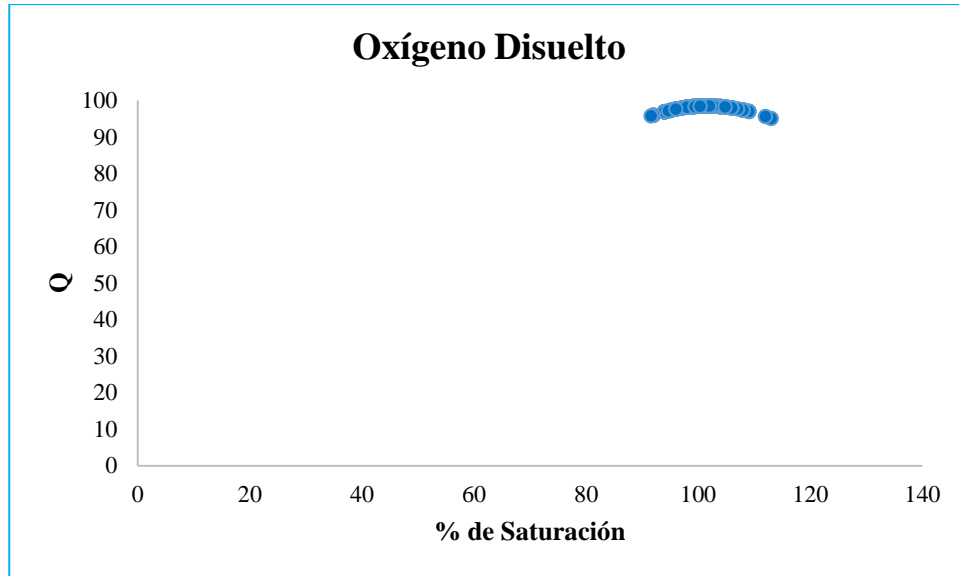
Yourdon, E., & Constantine, L. L. (1979). Structured design. Fundamentals of a discipline of computer program and systems design. *Englewood Cliffs: Yourdon Press*. <https://ui.adsabs.harvard.edu/abs/1979sdfd.book.....Y/abstract>

8 Anexos

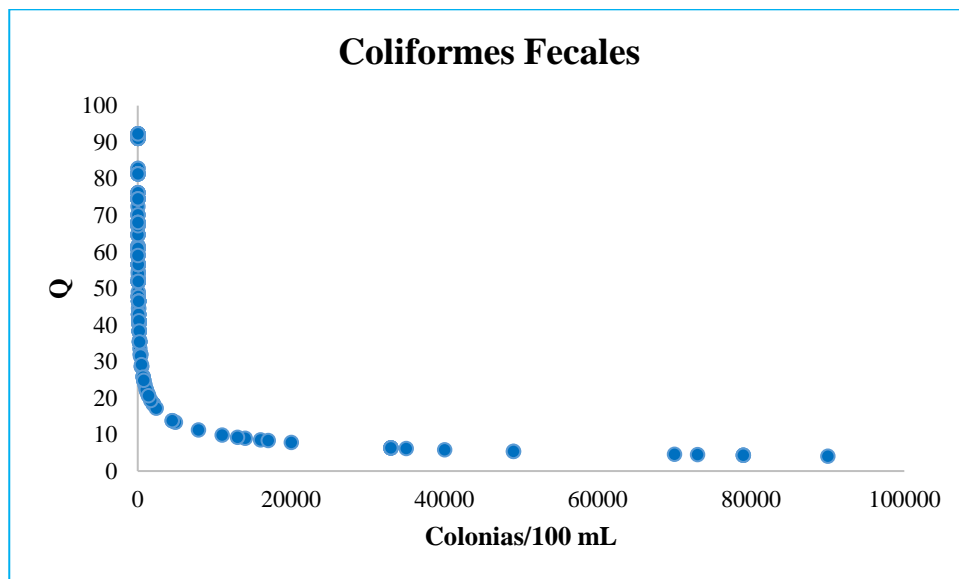
Anexo 1: Información general de las estaciones de monitoreo de la cuenca del Río Yanuncay.

Estación	Código	Coordenada X	Coordenada Y	Altitud
E-1	YAN-YH-CONTROL	689764	9668485	3575
E-2	YAN-SH-CONTROL	689772	9668446	3575
E-3	YAN-CAN-CONTROL	690289	9668664	3555
E-4	YAN-QUIN-040	690459	9668650	3577
E-5	YAN-GAL-020	690109	9677448	3942
E-6	YAN-GAL-005	692703	9672372	3655
E-7	YAN-GAL-CONTROL	693056	9670700	3515
E-8	YAN-BER-CONTROL	696022	9671334	3380
E-9	YAN-YAN-005	695835	9673164	3330
E-10	YAN-IZH-020	692859	9679312	3816
E-11	YAN-IZH-CONTROL	695700	9673389	3366
E-12	YAN-YAN-010	695909	9673285	3325
E-13	YAN-YAN-015	697240	9674930	3230
E-14	YAN-SOL-020	696185	9678325	3449
E-15	YAN-SOL-CONTROL	697698	9674882	3197
E-16	YAN-YAN-018	698967	9675686	3230
E-17	YAN-YAN-020	701631	9675583	3076
E-18	YAN-YAN-023	702398	9674806	3076
E-19	YAN-YAN-025	705296	9673697	2940
E-20	YAN-YAN-030	711293	9676264	2777
E-21	YAN-YAN-040	716785	9679214	2605
E-22	YAN-YAN-060	722025	9677449	2510
E-23	YAN-YAN-CONTROL	723897	9678232	2477

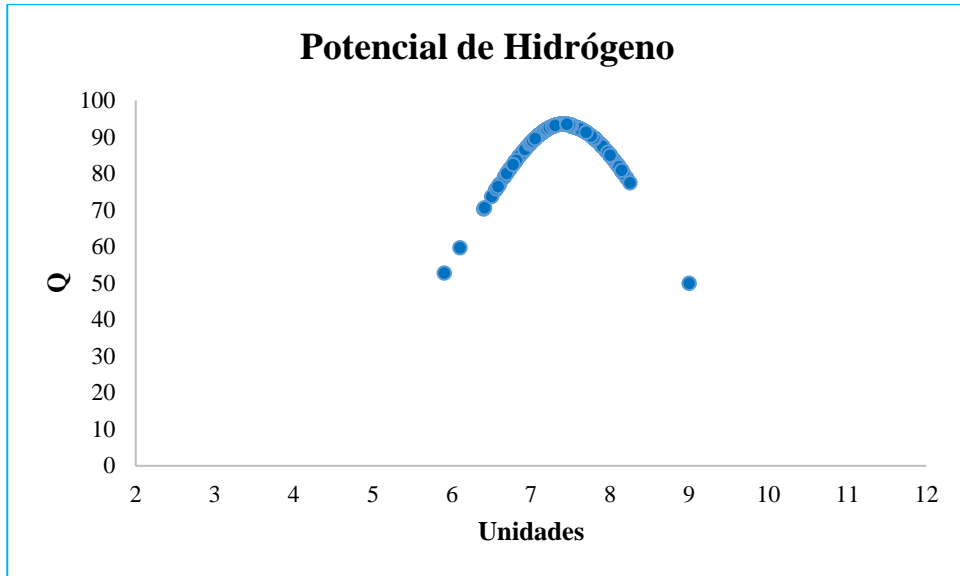
Anexo 2: Gráfico del factor de escala Q correspondiente al Oxígeno Disuelto.



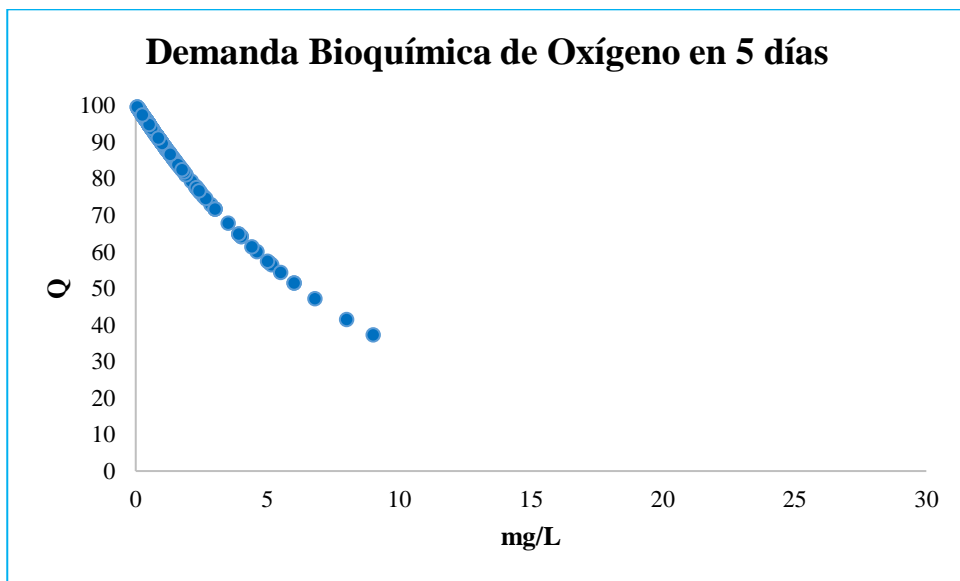
Anexo 3: Gráfico del factor de escala Q correspondiente a Coliformes Fecales.



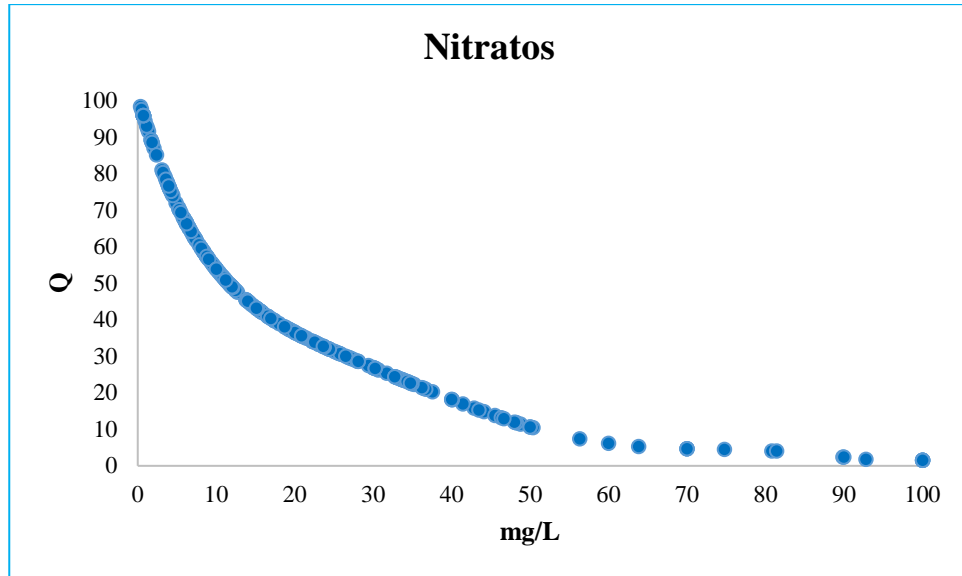
Anexo 4: Gráfico del factor de escala Q correspondiente al pH.



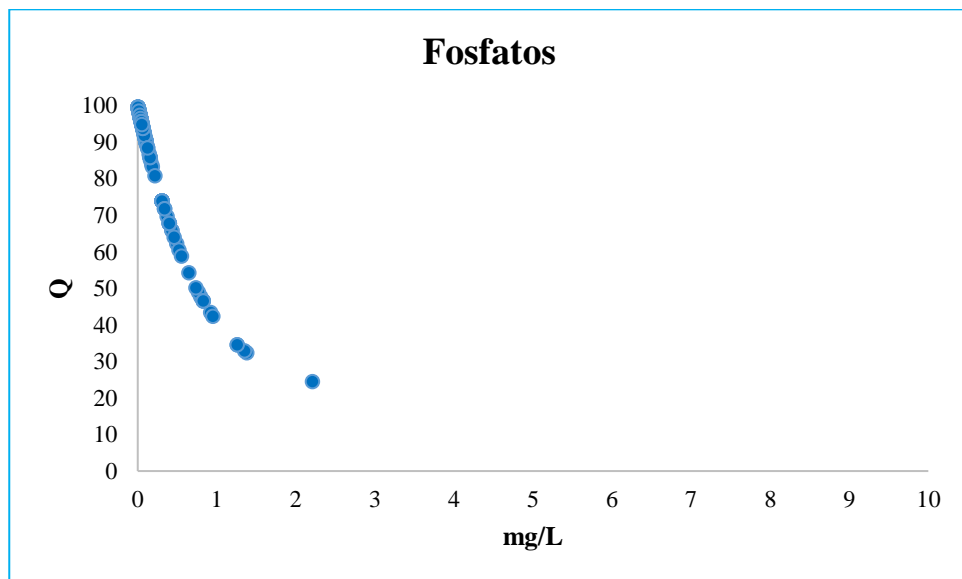
Anexo 5: Gráfico del factor de escala Q correspondiente a la DBO₅.



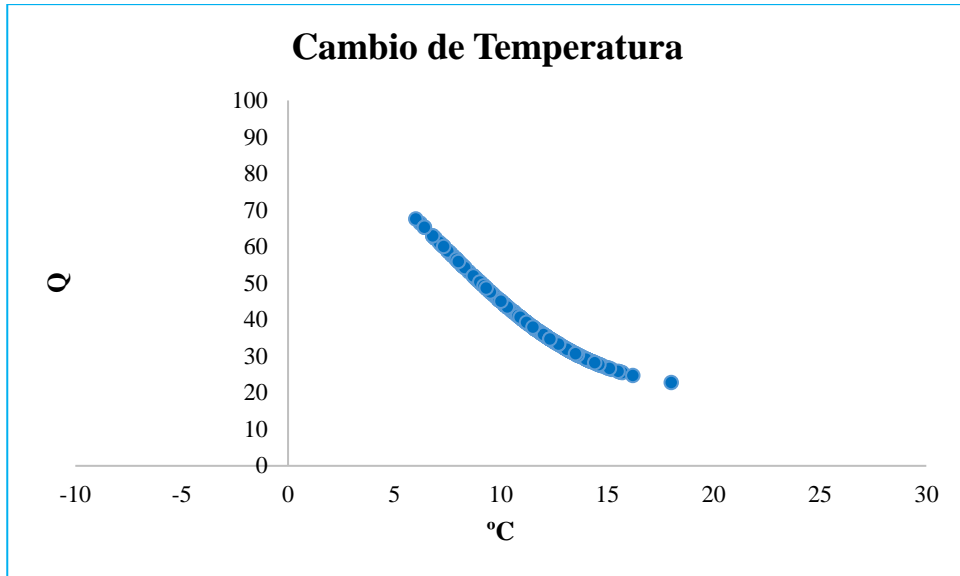
Anexo 6: Gráfico del factor de escala Q correspondiente a los Nitratos.



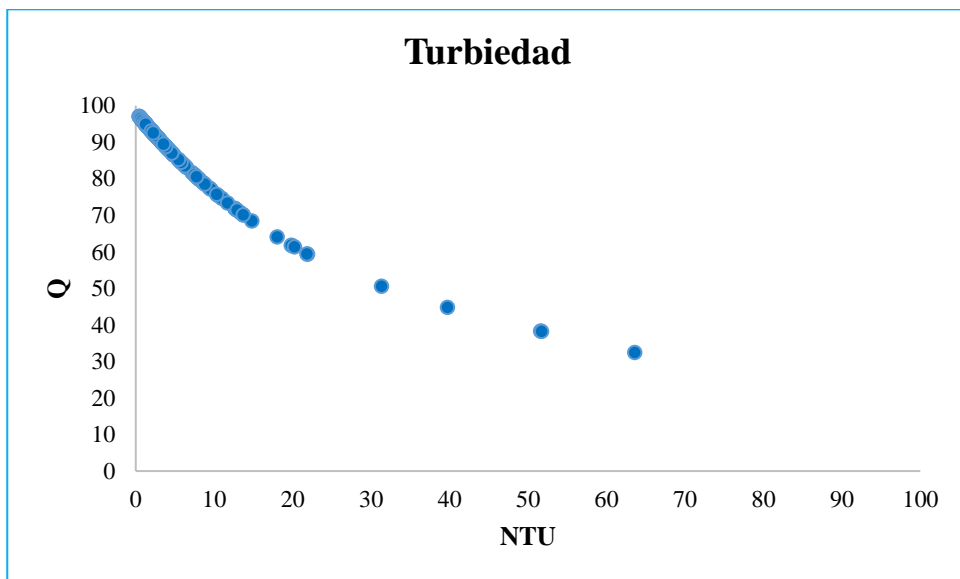
Anexo 7: Gráfico del factor de escala Q correspondiente a los Fosfatos.



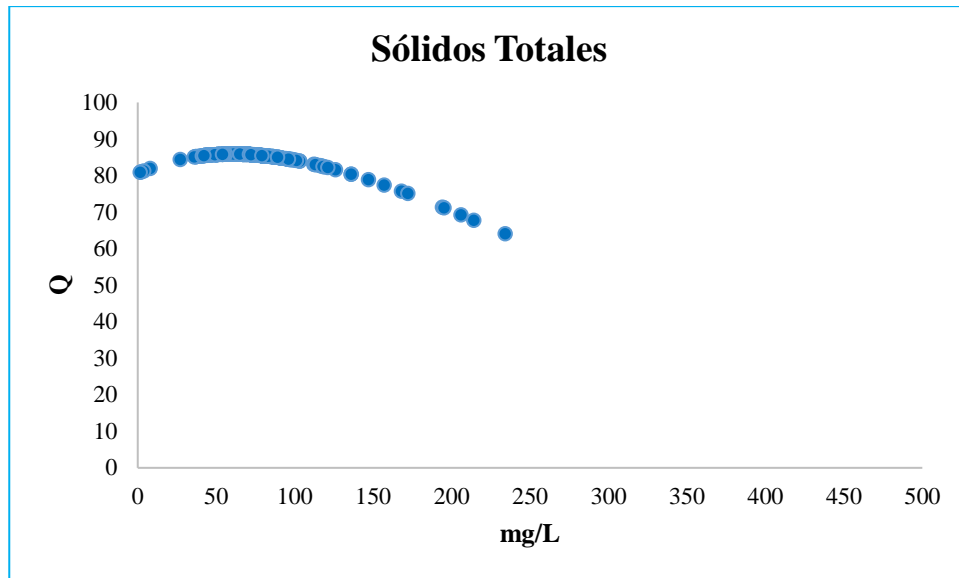
Anexo 8: Gráfico del factor de escala Q correspondiente a la Temperatura.



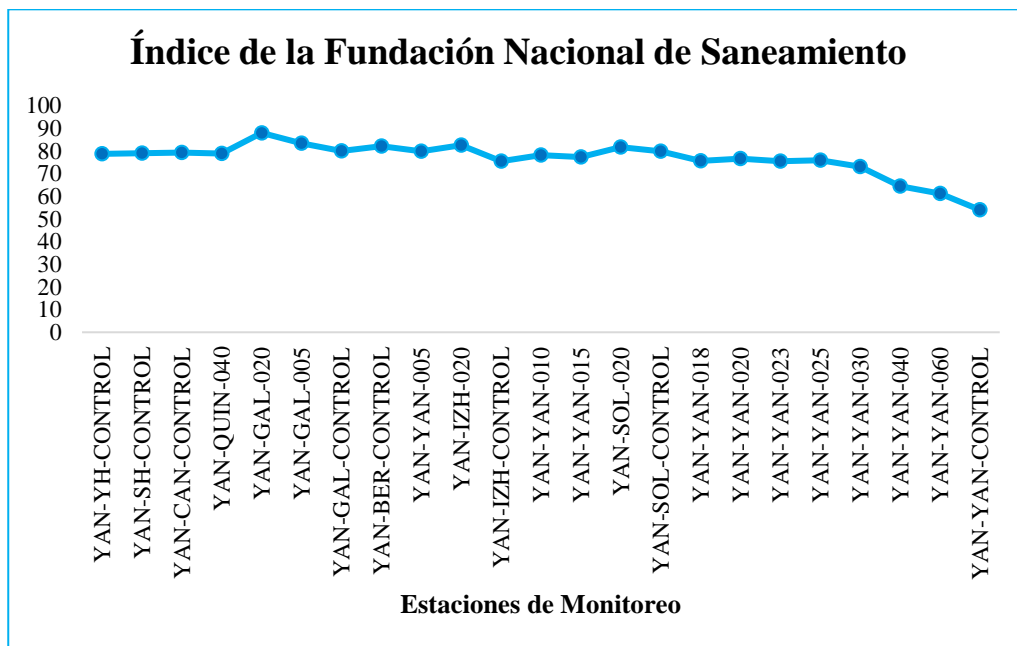
Anexo 9: Gráfico del factor de escala Q correspondiente a la Turbiedad.



Anexo 10: Gráfico del factor de escala Q correspondiente a los Sólidos Totales.



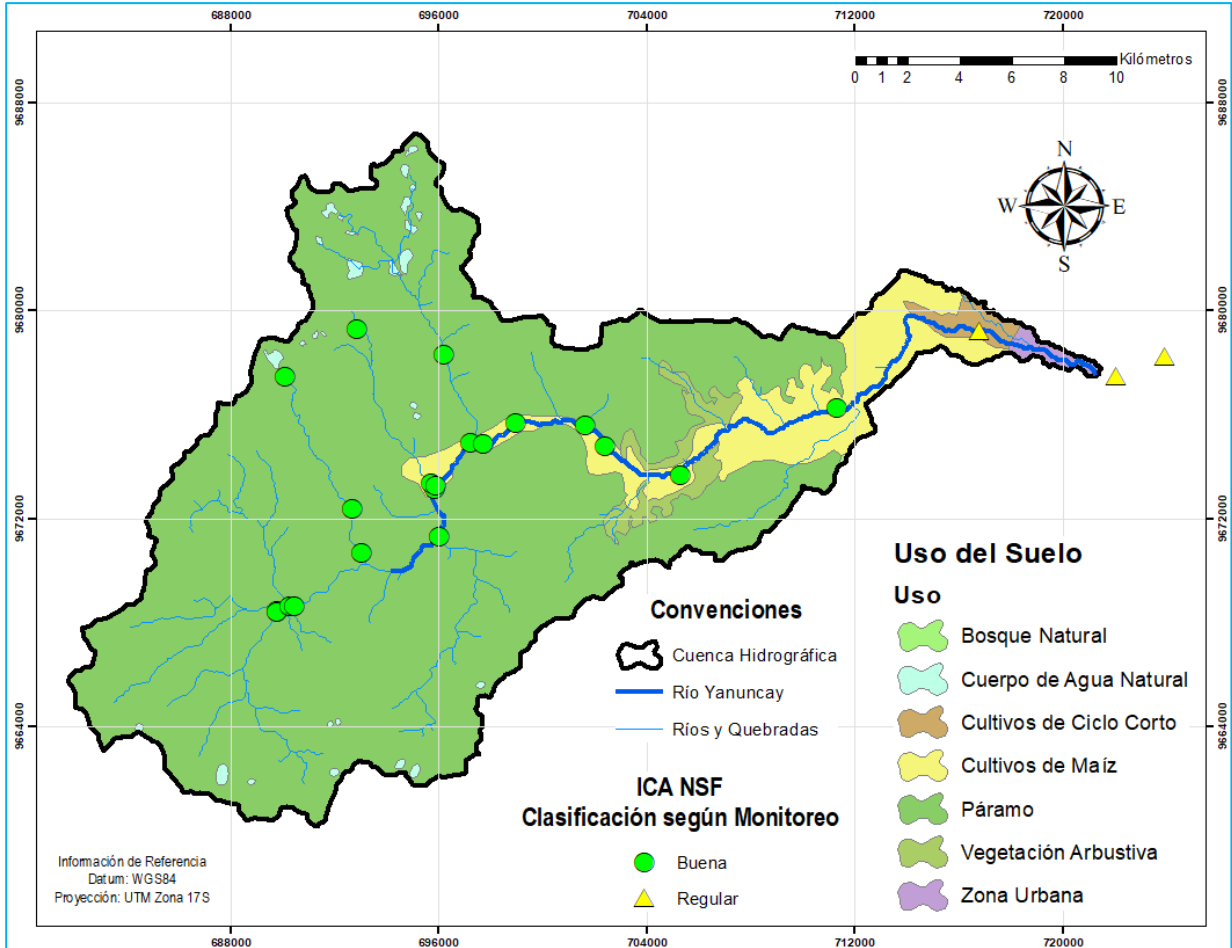
Anexo 11: Gráfico de la calidad del agua en cada estación de monitoreo.



Anexo 12: Resumen del promedio y desviación estándar de la calidad del agua en cada estación de monitoreo.

Estación	Código	Promedio ±
E-1	YAN-YH-CONTROL	78.70 ± 3.78
E-2	YAN-SH-CONTROL	79.01 ± 2.84
E-3	YAN-CAN-CONTROL	79.32 ± 4.11
E-4	YAN-QUIN-040	78.89 ± 4.20
E-5	YAN-GAL-020	87.97 ± 1.89
E-6	YAN-GAL-005	83.32 ± 1.30
E-7	YAN-GAL-CONTROL	80.03 ± 3.81
E-8	YAN-BER-CONTROL	82.13 ± 3.69
E-9	YAN-YAN-005	79.91 ± 1.58
E-10	YAN-IZH-020	82.57 ± 3.69
E-11	YAN-IZH-CONTROL	75.45 ± 4.29
E-12	YAN-YAN-010	78.20 ± 3.13
E-13	YAN-YAN-015	77.27 ± 3.58
E-14	YAN-SOL-020	81.65 ± 3.16
E-15	YAN-SOL-CONTROL	79.84 ± 4.37
E-16	YAN-YAN-018	75.68 ± 2.60
E-17	YAN-YAN-020	76.66 ± 2.90
E-18	YAN-YAN-023	75.44 ± 1.97
E-19	YAN-YAN-025	75.85 ± 2.98
E-20	YAN-YAN-030	73.12 ± 3.23
E-21	YAN-YAN-040	64.40 ± 3.48
E-22	YAN-YAN-060	61.14 ± 2.68
E-23	YAN-YAN-CONTROL	54.03 ± 4.54

Anexo 13: Calidad del agua en función del uso del suelo.



Anexo 14: Resultados de modelos con un rendimiento mayor a 0.90.

Modelos	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	Promedio*
('Coliformes Fecales', 'Nitratos')	0.888	0.937	0.965	0.945	0.947	0.879	0.890	0.931	0.901	0.947	0.923
('Oxígeno Disuelto', 'Coliformes Fecales', 'Nitratos')	0.903	0.940	0.976	0.947	0.950	0.870	0.881	0.906	0.900	0.945	0.922
('Coliformes Fecales', 'pH', 'Nitratos')	0.923	0.935	0.969	0.957	0.933	0.887	0.895	0.942	0.907	0.956	0.930
('Coliformes Fecales', 'DBO5', 'Nitratos')	0.918	0.962	0.975	0.952	0.948	0.845	0.922	0.938	0.951	0.915	0.932
('Coliformes Fecales', 'Nitratos', 'Fosfatos')	0.935	0.953	0.913	0.960	0.977	0.878	0.934	0.964	0.935	0.970	0.942
('Coliformes Fecales', 'Nitratos', 'Temperatura')	0.901	0.938	0.976	0.954	0.956	0.890	0.882	0.922	0.898	0.945	0.926
('Coliformes Fecales', 'Nitratos', 'Turbiedad')	0.923	0.937	0.980	0.954	0.959	0.881	0.908	0.933	0.912	0.952	0.934
('Coliformes Fecales', 'Nitratos', 'Sólidos Totales')	0.920	0.947	0.973	0.954	0.967	0.874	0.906	0.944	0.899	0.931	0.931
('Oxígeno Disuelto', 'Coliformes Fecales', 'pH', 'Nitratos')	0.915	0.938	0.974	0.954	0.935	0.894	0.872	0.929	0.896	0.955	0.926
('Oxígeno Disuelto', 'Coliformes Fecales', 'DBO5', 'Nitratos')	0.910	0.960	0.972	0.948	0.955	0.863	0.926	0.945	0.949	0.927	0.935
('Oxígeno Disuelto', 'Coliformes Fecales', 'Nitratos', 'Fosfatos')	0.904	0.959	0.909	0.954	0.973	0.879	0.930	0.960	0.940	0.964	0.937

('Oxígeno Disuelto', 'Coliformes Fecales', 'Nitratos', 'Temperatura')	0.920	0.943	0.981	0.955	0.954	0.900	0.852	0.922	0.917	0.949	0.929
('Oxígeno Disuelto', 'Coliformes Fecales', 'Nitratos', 'Turbiedad')	0.924	0.944	0.976	0.956	0.958	0.871	0.892	0.939	0.908	0.954	0.932
('Oxígeno Disuelto', 'Coliformes Fecales', 'Nitratos', 'Sólidos Totales')	0.918	0.950	0.975	0.957	0.965	0.879	0.902	0.949	0.906	0.935	0.933
('Coliformes Fecales', 'pH', 'DBO5', 'Nitratos')	0.926	0.956	0.972	0.961	0.951	0.873	0.920	0.937	0.953	0.928	0.938
('Coliformes Fecales', 'pH', 'Nitratos', 'Fosfatos')	0.934	0.957	0.920	0.965	0.969	0.876	0.932	0.964	0.946	0.973	0.944
('Coliformes Fecales', 'pH', 'Nitratos', 'Temperatura')	0.925	0.932	0.977	0.960	0.949	0.904	0.889	0.941	0.916	0.965	0.936
('Coliformes Fecales', 'pH', 'Nitratos', 'Turbiedad')	0.928	0.936	0.972	0.961	0.946	0.878	0.908	0.936	0.922	0.961	0.935
('Coliformes Fecales', 'pH', 'Nitratos', 'Sólidos Totales')	0.935	0.943	0.976	0.961	0.951	0.878	0.908	0.946	0.911	0.939	0.935
('Coliformes Fecales', 'DBO5', 'Nitratos', 'Fosfatos')	0.919	0.961	0.935	0.966	0.968	0.875	0.943	0.964	0.951	0.955	0.944
('Coliformes Fecales', 'DBO5', 'Nitratos', 'Temperatura')	0.931	0.959	0.980	0.957	0.962	0.894	0.926	0.942	0.956	0.930	0.944
('Coliformes Fecales', 'DBO5', 'Nitratos', 'Turbiedad')	0.931	0.950	0.975	0.957	0.966	0.858	0.924	0.951	0.946	0.945	0.940
('Coliformes Fecales', 'DBO5', 'Nitratos', 'Sólidos Totales')	0.931	0.951	0.978	0.969	0.958	0.888	0.928	0.948	0.946	0.945	0.944

('Coliformes Fecales', 'Nitratos', 'Fosfatos', 'Temperatura')	0.938	0.961	0.915	0.964	0.983	0.904	0.923	0.954	0.948	0.971	0.946
('Coliformes Fecales', 'Nitratos', 'Fosfatos', 'Turbiedad')	0.923	0.953	0.935	0.966	0.977	0.880	0.928	0.963	0.941	0.967	0.943
('Coliformes Fecales', 'Nitratos', 'Fosfatos', 'Sólidos Totales')	0.931	0.956	0.945	0.970	0.976	0.882	0.943	0.964	0.922	0.967	0.946
('Coliformes Fecales', 'Nitratos', 'Temperatura', 'Turbiedad')	0.930	0.937	0.982	0.958	0.966	0.894	0.903	0.933	0.935	0.960	0.940
('Coliformes Fecales', 'Nitratos', 'Temperatura', 'Sólidos Totales')	0.925	0.939	0.977	0.964	0.968	0.894	0.907	0.936	0.917	0.941	0.937
('Coliformes Fecales', 'Nitratos', 'Turbiedad', 'Sólidos Totales')	0.929	0.945	0.975	0.959	0.971	0.871	0.909	0.949	0.926	0.936	0.937
('Oxígeno Disuelto', 'Coliformes Fecales', 'pH', 'DBO5', 'Nitratos')	0.920	0.958	0.973	0.957	0.957	0.883	0.920	0.939	0.953	0.927	0.939
('Oxígeno Disuelto', 'Coliformes Fecales', 'pH', 'Nitratos', 'Fosfatos')	0.905	0.960	0.923	0.960	0.969	0.890	0.932	0.957	0.940	0.968	0.940
('Oxígeno Disuelto', 'Coliformes Fecales', 'pH', 'Nitratos', 'Temperatura')	0.920	0.935	0.979	0.955	0.947	0.909	0.862	0.928	0.908	0.960	0.930
('Oxígeno Disuelto', 'Coliformes Fecales', 'pH', 'Nitratos', 'Turbiedad')	0.928	0.940	0.976	0.963	0.948	0.879	0.891	0.942	0.935	0.959	0.936
('Oxígeno Disuelto', 'Coliformes Fecales', 'pH', 'Nitratos', 'Temperatura')	0.929	0.945	0.975	0.958	0.957	0.885	0.900	0.950	0.897	0.940	0.934

'Nitratos', 'Sólidos Totales')											
('Oxígeno Disuelto', 'Coliformes Fecales', 'DBO5', 'Nitratos', 'Fosfatos')	0.932	0.960	0.940	0.963	0.970	0.878	0.944	0.961	0.946	0.960	0.946
('Oxígeno Disuelto', 'Coliformes Fecales', 'DBO5', 'Nitratos', 'Temperatura')	0.916	0.961	0.978	0.958	0.960	0.891	0.923	0.939	0.950	0.928	0.940
('Oxígeno Disuelto', 'Coliformes Fecales', 'DBO5', 'Nitratos', 'Turbiedad')	0.933	0.958	0.979	0.958	0.968	0.863	0.920	0.950	0.950	0.939	0.942
('Oxígeno Disuelto', 'Coliformes Fecales', 'DBO5', 'Nitratos', 'Sólidos Totales')	0.934	0.957	0.972	0.968	0.962	0.877	0.921	0.954	0.945	0.942	0.943
('Oxígeno Disuelto', 'Coliformes Fecales', 'Nitratos', 'Fosfatos', 'Temperatura')	0.911	0.961	0.916	0.963	0.975	0.903	0.930	0.958	0.938	0.966	0.942
('Oxígeno Disuelto', 'Coliformes Fecales', 'Nitratos', 'Fosfatos', 'Turbiedad')	0.925	0.956	0.919	0.965	0.974	0.887	0.929	0.969	0.939	0.965	0.943
('Oxígeno Disuelto', 'Coliformes Fecales', 'Nitratos', 'Fosfatos', 'Sólidos Totales')	0.927	0.956	0.947	0.963	0.978	0.883	0.933	0.967	0.914	0.963	0.943
('Oxígeno Disuelto', 'Coliformes Fecales',	0.927	0.943	0.977	0.965	0.965	0.885	0.884	0.940	0.933	0.962	0.938

'Nitratos', 'Temperatura', 'Turbiedad')												
('Oxígeno Disuelto', 'Coliformes Fecales', 'Nitratos', 'Temperatura', 'Sólidos Totales')	0.933	0.949	0.979	0.962	0.965	0.893	0.905	0.941	0.915	0.938	0.938	
('Oxígeno Disuelto', 'Coliformes Fecales', 'Nitratos', 'Turbiedad', 'Sólidos Totales')	0.929	0.948	0.976	0.956	0.966	0.874	0.904	0.951	0.910	0.945	0.936	
('Coliformes Fecales', 'pH', 'DBO5', 'Nitratos', 'Fosfatos')	0.919	0.961	0.935	0.968	0.966	0.881	0.940	0.962	0.950	0.964	0.945	
('Coliformes Fecales', 'pH', 'DBO5', 'Nitratos', 'Temperatura')	0.929	0.954	0.976	0.963	0.957	0.892	0.929	0.943	0.948	0.932	0.942	
('Coliformes Fecales', 'pH', 'DBO5', 'Nitratos', 'Turbiedad')	0.931	0.953	0.973	0.958	0.957	0.876	0.922	0.957	0.943	0.946	0.942	
('Coliformes Fecales', 'pH', 'DBO5', 'Nitratos', 'Sólidos Totales')	0.943	0.956	0.975	0.968	0.959	0.889	0.917	0.947	0.946	0.955	0.946	
('Coliformes Fecales', 'pH', 'Nitratos', 'Fosfatos', 'Temperatura')	0.927	0.958	0.921	0.970	0.977	0.909	0.931	0.959	0.950	0.977	0.948	
('Coliformes Fecales', 'pH', 'Nitratos', 'Fosfatos', 'Turbiedad')	0.927	0.950	0.914	0.971	0.974	0.894	0.932	0.969	0.937	0.971	0.944	
('Coliformes Fecales', 'pH', 'Nitratos', 'Fosfatos', 'Sólidos Totales')	0.910	0.963	0.939	0.971	0.977	0.886	0.938	0.969	0.931	0.966	0.945	

('Coliformes Fecales', 'pH', 'Nitratos', 'Temperatura', 'Turbiedad')	0.940	0.931	0.976	0.963	0.959	0.901	0.898	0.934	0.937	0.963	0.940
('Coliformes Fecales', 'pH', 'Nitratos', 'Temperatura', 'Sólidos Totales')	0.938	0.942	0.976	0.961	0.962	0.900	0.904	0.946	0.911	0.947	0.939
('Coliformes Fecales', 'pH', 'Nitratos', 'Turbiedad', 'Sólidos Totales')	0.940	0.943	0.974	0.962	0.963	0.888	0.902	0.942	0.918	0.952	0.938
('Coliformes Fecales', 'DBO5', 'Nitratos', 'Fosfatos', 'Temperatura')	0.911	0.965	0.942	0.970	0.974	0.901	0.958	0.955	0.962	0.963	0.950
('Coliformes Fecales', 'DBO5', 'Nitratos', 'Fosfatos', 'Turbiedad')	0.919	0.957	0.943	0.968	0.976	0.881	0.946	0.963	0.953	0.959	0.946
('Coliformes Fecales', 'DBO5', 'Nitratos', 'Fosfatos', 'Sólidos Totales')	0.918	0.961	0.943	0.970	0.976	0.895	0.947	0.970	0.944	0.960	0.949
('Coliformes Fecales', 'DBO5', 'Nitratos', 'Temperatura', 'Turbiedad')	0.936	0.952	0.982	0.965	0.967	0.905	0.928	0.952	0.946	0.947	0.948
('Coliformes Fecales', 'DBO5', 'Nitratos', 'Temperatura', 'Sólidos Totales')	0.941	0.961	0.977	0.970	0.968	0.906	0.932	0.949	0.945	0.955	0.950
('Coliformes Fecales', 'DBO5', 'Nitratos', 'Turbiedad', 'Sólidos Totales')	0.938	0.951	0.974	0.965	0.965	0.885	0.922	0.955	0.939	0.951	0.944
('Coliformes Fecales', 'Nitratos', 'Fosfatos', 'Temperatura', 'Turbiedad')	0.927	0.953	0.934	0.974	0.980	0.897	0.932	0.964	0.939	0.972	0.947

('Coliformes Fecales', 'Nitratos', 'Fosfatos', 'Temperatura', 'Sólidos Totales')	0.928	0.959	0.948	0.969	0.982	0.887	0.939	0.964	0.934	0.967	0.948
('Coliformes Fecales', 'Nitratos', 'Fosfatos', 'Turbiedad', 'Sólidos Totales')	0.918	0.956	0.938	0.968	0.980	0.882	0.936	0.971	0.926	0.965	0.944
('Coliformes Fecales', 'Nitratos', 'Temperatura', 'Turbiedad', 'Sólidos Totales')	0.931	0.941	0.976	0.966	0.973	0.894	0.913	0.939	0.922	0.947	0.940
('Oxígeno Disuelto', 'Coliformes Fecales', 'pH', 'DBO5', 'Nitratos', 'Fosfatos')	0.916	0.962	0.937	0.966	0.966	0.898	0.939	0.958	0.954	0.957	0.945
('Oxígeno Disuelto', 'Coliformes Fecales', 'pH', 'DBO5', 'Nitratos', 'Temperatura')	0.924	0.952	0.978	0.957	0.958	0.899	0.924	0.936	0.949	0.932	0.941
('Oxígeno Disuelto', 'Coliformes Fecales', 'pH', 'DBO5', 'Nitratos', 'Turbiedad')	0.931	0.959	0.976	0.960	0.966	0.894	0.906	0.956	0.942	0.940	0.943
('Oxígeno Disuelto', 'Coliformes Fecales', 'pH', 'DBO5', 'Nitratos', 'Sólidos Totales')	0.938	0.957	0.970	0.969	0.960	0.893	0.915	0.952	0.940	0.947	0.944
('Oxígeno Disuelto', 'Coliformes Fecales', 'pH', 'Nitratos', 'Fosfatos', 'Temperatura')	0.922	0.959	0.914	0.966	0.977	0.891	0.931	0.949	0.950	0.976	0.944

('Oxígeno Disuelto', 'Coliformes Fecales', 'pH', 'Nitratos', 'Fosfatos', 'Turbiedad')	0.918	0.953	0.915	0.966	0.969	0.888	0.926	0.968	0.948	0.969	0.942
('Oxígeno Disuelto', 'Coliformes Fecales', 'pH', 'Nitratos', 'Fosfatos', 'Sólidos Totales')	0.911	0.959	0.948	0.971	0.974	0.886	0.937	0.970	0.932	0.961	0.945
('Oxígeno Disuelto', 'Coliformes Fecales', 'pH', 'Nitratos', 'Temperatura', 'Turbiedad')	0.929	0.938	0.975	0.964	0.956	0.902	0.880	0.936	0.939	0.963	0.938
('Oxígeno Disuelto', 'Coliformes Fecales', 'pH', 'Nitratos', 'Temperatura', 'Sólidos Totales')	0.933	0.944	0.978	0.966	0.965	0.899	0.893	0.947	0.918	0.945	0.939
('Oxígeno Disuelto', 'Coliformes Fecales', 'pH', 'Nitratos', 'Turbiedad', 'Sólidos Totales')	0.930	0.946	0.973	0.965	0.961	0.886	0.902	0.950	0.917	0.950	0.938
('Oxígeno Disuelto', 'Coliformes Fecales', 'DBO5', 'Nitratos', 'Fosfatos', 'Temperatura')	0.919	0.962	0.942	0.968	0.976	0.901	0.956	0.954	0.955	0.960	0.949
('Oxígeno Disuelto', 'Coliformes Fecales', 'DBO5', 'Nitratos', 'Fosfatos', 'Turbiedad')	0.918	0.958	0.945	0.967	0.977	0.883	0.934	0.965	0.949	0.958	0.946
('Oxígeno Disuelto', 'Coliformes Fecales', 'DBO5', 'Nitratos', 'Fosfatos', 'Sólidos Totales')	0.897	0.961	0.958	0.970	0.976	0.882	0.950	0.969	0.947	0.960	0.947

('Oxígeno Disuelto', 'Coliformes Fecales', 'DBO5', 'Nitratos', 'Temperatura', 'Turbiedad')	0.933	0.950	0.980	0.964	0.968	0.897	0.924	0.953	0.944	0.947	0.946
('Oxígeno Disuelto', 'Coliformes Fecales', 'DBO5', 'Nitratos', 'Temperatura', 'Sólidos Totales')	0.942	0.960	0.975	0.970	0.966	0.900	0.933	0.951	0.947	0.943	0.949
('Oxígeno Disuelto', 'Coliformes Fecales', 'DBO5', 'Nitratos', 'Turbiedad', 'Sólidos Totales')	0.938	0.954	0.976	0.964	0.973	0.893	0.921	0.957	0.926	0.950	0.945
('Oxígeno Disuelto', 'Coliformes Fecales', 'Nitratos', 'Fosfatos', 'Temperatura', 'Turbiedad')	0.923	0.956	0.916	0.967	0.980	0.904	0.927	0.967	0.950	0.972	0.946
('Oxígeno Disuelto', 'Coliformes Fecales', 'Nitratos', 'Fosfatos', 'Temperatura', 'Sólidos Totales')	0.906	0.961	0.938	0.971	0.980	0.898	0.935	0.967	0.940	0.965	0.946
('Oxígeno Disuelto', 'Coliformes Fecales', 'Nitratos', 'Fosfatos', 'Turbiedad', 'Sólidos Totales')	0.926	0.957	0.939	0.967	0.980	0.890	0.931	0.970	0.926	0.965	0.945
('Oxígeno Disuelto', 'Coliformes Fecales', 'Nitratos', 'Temperatura', 'Turbiedad', 'Sólidos Totales')	0.939	0.949	0.977	0.965	0.972	0.906	0.903	0.944	0.924	0.953	0.943

('Coliformes Fecales', 'pH', 'DBO5', 'Nitratos', 'Fosfatos', 'Temperatura')	0.932	0.964	0.937	0.972	0.971	0.909	0.950	0.959	0.962	0.965	0.952
('Coliformes Fecales', 'pH', 'DBO5', 'Nitratos', 'Fosfatos', 'Turbiedad')	0.917	0.959	0.939	0.973	0.970	0.896	0.946	0.964	0.955	0.960	0.948
('Coliformes Fecales', 'pH', 'DBO5', 'Nitratos', 'Fosfatos', 'Sólidos Totales')	0.919	0.965	0.951	0.973	0.967	0.889	0.948	0.972	0.946	0.965	0.950
('Coliformes Fecales', 'pH', 'DBO5', 'Nitratos', 'Temperatura', 'Turbiedad')	0.932	0.954	0.976	0.963	0.968	0.907	0.921	0.946	0.952	0.949	0.947
('Coliformes Fecales', 'pH', 'DBO5', 'Nitratos', 'Temperatura', 'Sólidos Totales')	0.941	0.955	0.972	0.970	0.967	0.910	0.932	0.949	0.948	0.951	0.949
('Coliformes Fecales', 'pH', 'DBO5', 'Nitratos', 'Turbiedad', 'Sólidos Totales')	0.943	0.952	0.970	0.966	0.966	0.897	0.922	0.951	0.938	0.952	0.946
('Coliformes Fecales', 'pH', 'Nitratos', 'Fosfatos', 'Temperatura', 'Turbiedad')	0.930	0.956	0.925	0.971	0.976	0.900	0.928	0.963	0.946	0.978	0.947
('Coliformes Fecales', 'pH', 'Nitratos', 'Fosfatos', 'Temperatura', 'Sólidos Totales')	0.939	0.962	0.939	0.973	0.984	0.905	0.933	0.965	0.932	0.974	0.950
('Coliformes Fecales', 'pH', 'Nitratos', 'Fosfatos', 'Turbiedad', 'Sólidos Totales')	0.920	0.956	0.942	0.971	0.973	0.891	0.928	0.971	0.924	0.968	0.944

('Coliformes Fecales', 'pH', 'Nitratos', 'Temperatura', 'Turbiedad', 'Sólidos Totales')	0.944	0.943	0.978	0.967	0.965	0.890	0.912	0.941	0.926	0.946	0.941
('Coliformes Fecales', 'DBO5', 'Nitratos', 'Fosfatos', 'Temperatura', 'Turbiedad')	0.933	0.954	0.942	0.972	0.975	0.903	0.951	0.961	0.956	0.967	0.951
('Coliformes Fecales', 'DBO5', 'Nitratos', 'Fosfatos', 'Temperatura', 'Sólidos Totales')	0.920	0.959	0.962	0.974	0.979	0.911	0.954	0.964	0.950	0.965	0.954
('Coliformes Fecales', 'DBO5', 'Nitratos', 'Fosfatos', 'Turbiedad', 'Sólidos Totales')	0.925	0.958	0.956	0.972	0.976	0.885	0.953	0.966	0.937	0.961	0.949
('Coliformes Fecales', 'DBO5', 'Nitratos', 'Temperatura', 'Turbiedad', 'Sólidos Totales')	0.943	0.952	0.977	0.971	0.971	0.898	0.936	0.956	0.935	0.952	0.949
('Coliformes Fecales', 'Nitratos', 'Fosfatos', 'Temperatura', 'Turbiedad', 'Sólidos Totales')	0.917	0.952	0.946	0.975	0.984	0.896	0.944	0.967	0.935	0.968	0.948
('Oxígeno Disuelto', 'Coliformes Fecales', 'pH', 'DBO5', 'Nitratos', 'Fosfatos', 'Temperatura')	0.918	0.963	0.938	0.967	0.974	0.912	0.949	0.957	0.960	0.964	0.950
('Oxígeno Disuelto', 'Coliformes Fecales', 'pH', 'DBO5', 'Nitratos', 'Fosfatos', 'Turbiedad')	0.907	0.960	0.932	0.965	0.969	0.898	0.937	0.964	0.950	0.960	0.944

('Oxígeno Disuelto', 'Coliformes Fecales', 'pH', 'DBO5', 'Nitratos', 'Fosfatos', 'Sólidos Totales')	0.915	0.964	0.951	0.971	0.969	0.902	0.947	0.967	0.944	0.962	0.949
('Oxígeno Disuelto', 'Coliformes Fecales', 'pH', 'DBO5', 'Nitratos', 'Temperatura', 'Turbiedad')	0.932	0.955	0.981	0.963	0.969	0.903	0.910	0.951	0.936	0.947	0.945
('Oxígeno Disuelto', 'Coliformes Fecales', 'pH', 'DBO5', 'Nitratos', 'Temperatura', 'Sólidos Totales')	0.936	0.953	0.976	0.969	0.971	0.902	0.926	0.950	0.944	0.952	0.948
('Oxígeno Disuelto', 'Coliformes Fecales', 'pH', 'DBO5', 'Nitratos', 'Turbiedad', 'Sólidos Totales')	0.935	0.958	0.973	0.967	0.967	0.886	0.907	0.959	0.932	0.950	0.943
('Oxígeno Disuelto', 'Coliformes Fecales', 'pH', 'Nitratos', 'Fosfatos', 'Temperatura', 'Turbiedad')	0.934	0.957	0.926	0.967	0.976	0.907	0.934	0.969	0.945	0.976	0.949
('Oxígeno Disuelto', 'Coliformes Fecales', 'pH', 'Nitratos', 'Fosfatos', 'Temperatura', 'Sólidos Totales')	0.885	0.959	0.944	0.970	0.980	0.903	0.939	0.965	0.931	0.972	0.945
('Oxígeno Disuelto', 'Coliformes Fecales', 'pH', 'Nitratos', 'Fosfatos', 'Turbiedad', 'Sólidos Totales')	0.923	0.958	0.942	0.972	0.976	0.897	0.926	0.975	0.931	0.966	0.947

('Oxígeno Disuelto', 'Coliformes Fecales', 'pH', 'Nitratos', 'Temperatura', 'Turbiedad', 'Sólidos Totales')	0.938	0.939	0.977	0.965	0.965	0.886	0.895	0.947	0.923	0.953	0.939
('Oxígeno Disuelto', 'Coliformes Fecales', 'DBO5', 'Nitratos', 'Fosfatos', 'Temperatura', 'Turbiedad')	0.914	0.956	0.936	0.970	0.976	0.917	0.947	0.961	0.953	0.963	0.949
('Oxígeno Disuelto', 'Coliformes Fecales', 'DBO5', 'Nitratos', 'Fosfatos', 'Temperatura', 'Sólidos Totales')	0.901	0.964	0.960	0.973	0.978	0.911	0.957	0.963	0.954	0.965	0.953
('Oxígeno Disuelto', 'Coliformes Fecales', 'DBO5', 'Nitratos', 'Fosfatos', 'Turbiedad', 'Sólidos Totales')	0.926	0.959	0.951	0.968	0.972	0.885	0.944	0.968	0.939	0.962	0.947
('Oxígeno Disuelto', 'Coliformes Fecales', 'DBO5', 'Nitratos', 'Temperatura', 'Turbiedad', 'Sólidos Totales')	0.939	0.953	0.976	0.967	0.973	0.902	0.924	0.955	0.936	0.951	0.947
('Oxígeno Disuelto', 'Coliformes Fecales', 'Nitratos', 'Fosfatos', 'Temperatura', 'Turbiedad', 'Sólidos Totales')	0.909	0.954	0.938	0.974	0.981	0.897	0.946	0.965	0.933	0.967	0.946
('Coliformes Fecales', 'pH', 'DBO5', 'Nitratos',	0.912	0.958	0.940	0.974	0.975	0.907	0.949	0.959	0.958	0.972	0.950

'Fosfatos', 'Temperatura', 'Turbiedad')											
('Coliformes Fecales', 'pH', 'DBO5', 'Nitratos', 'Fosfatos', 'Temperatura', 'Sólidos Totales')	0.924	0.963	0.948	0.975	0.979	0.907	0.955	0.964	0.949	0.965	0.953
('Coliformes Fecales', 'pH', 'DBO5', 'Nitratos', 'Fosfatos', 'Turbiedad', 'Sólidos Totales')	0.933	0.956	0.949	0.975	0.972	0.893	0.945	0.969	0.945	0.962	0.950
('Coliformes Fecales', 'pH', 'DBO5', 'Nitratos', 'Temperatura', 'Turbiedad', 'Sólidos Totales')	0.933	0.952	0.976	0.968	0.967	0.916	0.926	0.955	0.936	0.954	0.948
('Coliformes Fecales', 'pH', 'Nitratos', 'Fosfatos', 'Temperatura', 'Turbiedad', 'Sólidos Totales')	0.933	0.960	0.937	0.975	0.980	0.895	0.938	0.969	0.925	0.970	0.948
('Coliformes Fecales', 'DBO5', 'Nitratos', 'Fosfatos', 'Temperatura', 'Turbiedad', 'Sólidos Totales')	0.931	0.959	0.953	0.976	0.980	0.908	0.958	0.959	0.952	0.968	0.955
('Oxígeno Disuelto', 'Coliformes Fecales', 'pH', 'DBO5', 'Nitratos', 'Fosfatos', 'Temperatura', 'Turbiedad')	0.896	0.956	0.952	0.973	0.976	0.903	0.936	0.966	0.957	0.964	0.948
('Oxígeno Disuelto', 'Coliformes Fecales', 'pH', 'DBO5', 'Nitratos', 'Fosfatos', 'Temperatura', 'Sólidos Totales')	0.927	0.962	0.949	0.973	0.976	0.910	0.956	0.959	0.949	0.969	0.953

('Oxígeno Disuelto', 'Coliformes Fecales', 'pH', 'DBO5', 'Nitratos', 'Fosfatos', 'Turbiedad', 'Sólidos Totales')	0.900	0.960	0.939	0.972	0.971	0.892	0.942	0.969	0.939	0.962	0.945
('Oxígeno Disuelto', 'Coliformes Fecales', 'pH', 'DBO5', 'Nitratos', 'Temperatura', 'Turbiedad', 'Sólidos Totales')	0.941	0.955	0.975	0.967	0.967	0.895	0.927	0.955	0.937	0.951	0.947
('Oxígeno Disuelto', 'Coliformes Fecales', 'pH', 'Nitratos', 'Fosfatos', 'Temperatura', 'Turbiedad', 'Sólidos Totales')	0.932	0.958	0.948	0.974	0.979	0.907	0.942	0.968	0.937	0.971	0.952
('Oxígeno Disuelto', 'Coliformes Fecales', 'DBO5', 'Nitratos', 'Fosfatos', 'Temperatura', 'Turbiedad', 'Sólidos Totales')	0.910	0.956	0.956	0.972	0.976	0.904	0.952	0.966	0.949	0.965	0.951
('Coliformes Fecales', 'pH', 'DBO5', 'Nitratos', 'Fosfatos', 'Temperatura', 'Turbiedad', 'Sólidos Totales')	0.919	0.957	0.951	0.974	0.976	0.899	0.947	0.966	0.954	0.972	0.951
('Oxígeno Disuelto', 'Coliformes Fecales', 'pH', 'DBO5', 'Nitratos', 'Fosfatos', 'Temperatura', 'Turbiedad', 'Sólidos Totales')	0.906	0.956	0.947	0.974	0.972	0.917	0.950	0.964	0.946	0.966	0.950*

* Promedio = R^2

* *Modelo base.*