

**CUANTIFICACIÓN DE CONCENTRACIONES DE DETERMINANTES DE CALIDAD
DEL AGUA A PARTIR DE INFORMACIÓN TELEDETECTADA, EN EL RÍO
MAGDALENA.**



CRISTIAN CAMILO ORDÓÑEZ ALDANA

**UNIVERSIDAD SANTO TOMÁS
FACULTAD DE INGENIERÍA AMBIENTAL
DIVISIÓN DE INGENIERIAS
BOGOTÁ D.C.**

2021

CUANTIFICACIÓN DE CONCENTRACIONES DE DETERMINANTES DE CALIDAD
DEL AGUA A PARTIR DE INFORMACIÓN TELEDETECTADA, EN EL RÍO
MAGDALENA.

CRISTIAN CAMILO ORDÓÑEZ ALDANA

Trabajo de grado presentado como requisito para optar al título de
Ingeniero Ambiental

Director:

DAVID ANDRÉS ZAMORA ÁVILA

Ingeniero Civil

UNIVERSIDAD SANTO TOMÁS
FACULTAD DE INGENIERÍA AMBIENTAL
DIVISIÓN DE INGENIERIAS
BOGOTÁ D.C.

2021

TABLA DE CONTENIDO

RESUMEN	8
ABSTRACT	9
INTRODUCCIÓN	10
OBJETIVOS.....	12
OBJETIVO GENERAL.....	12
OBJETIVOS ESPECÍFICOS	12
1. MARCO TEÓRICO	13
1.1 Área de estudio.....	13
1.2 Determinantes e índices de la calidad del agua	16
1.3 Conceptos generales en la teledetección	18
1.4 Algoritmos de teledetección para índices de calidad del agua	21
1.5 Modelos de regresión	23
1.6 Normatividad del proyecto	30
2. METODOLOGÍA.....	31
2.1 FASE 1. Identificación de puntos de monitoreo en el río Magdalena...31	
2.2 FASE 2. Cálculo de índices Clorofila-a, Turbiedad y Ficocianina	31
2.3 FASE 3. Evaluación de modelos para cuantificar determinantes.....	35
3. ANÁLISIS Y DISCUSIÓN DE RESULTADOS	38
3.1 INFORMACIÓN IN-SITU Y SATELITAL RECOPIADA	38
3.2 ÍNDICES CALCULADOS	40
3.3 APLICACIÓN DE MODELOS	47
4. IMPACTO SOCIAL Y HUMANÍSTICO DEL PROYECTO	57
5. CONCLUSIONES	58
6. RECOMENDACIONES.....	60
7. REFERENCIAS	61
8. ANEXOS.....	68

LISTA DE TABLAS

Tabla 1. Marco normativo aplicable a la investigación.....	30
Tabla 2. Punto de monitoreo para ejemplo de extracción de reflectancia.....	32
Tabla 3. Algoritmos evaluados para la detección de los índices.....	34
Tabla 4. Estructura de los modelos utilizados.....	35
Tabla 5. Puntos de monitoreo seleccionados, con rango de fecha y determinantes	39
Tabla 6. Desempeño de predicción para cada modelo utilizado.....	56

LISTA DE FIGURAS

Figura 1. Área de estudio río Magdalena, Colombia.....	14
Figura 2. Puntos de monitoreo seleccionados para los años 2006, 2013 y 2016 en el río Magdalena	15
Figura 3. Visualización de los índices Clorofila-a, Ficocianina y Turbiedad en el agua.....	17
Figura 4. Espectro electromagnético	18
Figura 5. Ilustración de sensores satelitales pasivos y activos	19
Figura 6. Gráfico de energía restrodispersada de limo en suspensión	22
Figura 7. Diagrama de caja y bigotes o Boxplot.....	24
Figura 8. Gráficos de diagnóstico para la evaluación de modelos lineales.....	27
Figura 9. Gráficas de residuos con dispersión constante y patrón simétrico	28
Figura 10. Vista general del servidor EarthData de la NASA	32
Figura 11. Localización de punto de monitoreo Regidor para el año 2016	33
Figura 12. Delimitación y extracción de reflectancia en estación Regidor	33
Figura 13. Esquema de la metodología del proyecto por fases	37
Figura 14. Evolución espacial y temporal de valores mínimos Turbiedad respecto a Sólidos Suspendidos Totales.....	40
Figura 15. Evolución espacial y temporal de valores medios Turbiedad respecto a Sólidos Suspendidos Totales.....	41
Figura 16. Evolución espacial y temporal de valores mínimos Clorofila-a respecto a Nitratos.....	42
Figura 17. Evolución espacial y temporal de los valores medios Clorofila-a respecto a Nitratos.....	43
Figura 18. Evolución espacial y temporal de valores mínimos Ficocianina respecto a Fósforo Total	44
Figura 19. Evolución espacial y temporal de valores medios Ficocianina respecto a Fósforo Total.....	44
Figura 20. Representación temporal y espacial de Sólidos Suspendidos Totales año 2006, 2013 y 2016	45
Figura 21. Representación temporal y espacial de Nitratos año 2006, 2013 y 2016.....	46
Figura 22. Representación temporal y espacial de Fósforo Total año 2013 y 2016	46
Figura 23. Distribución y correlación de los SST respecto a los valores de mínimos y medias del algoritmo TurbMoore80Red.....	47
Figura 24. Relación Nitratos, mínimos y medias algoritmo 2BDA.....	48
Figura 25. Relación Fósforo Total, mínimos y medias índice NDVI	48
Figura 26. Identificación de datos atípicos mediante Boxplot para los índices evaluados.....	49
Figura 27. Identificación de datos atípicos mediante Boxplot para los determinantes evaluados.....	50

Figura 28. Relación mejorada de mínimos algoritmo TurbMoore80Red respecto a Sólidos Suspendidos Totales	50
Figura 29. Relación mejorada de mínimos algoritmo 2BDA respecto a Nitratos	51
Figura 30. Relación de Nitratos calibrados con los estimados por el modelo lineal de transformación logarítmica.....	52
Figura 31. Diagnóstico estadístico del modelo lineal de transformación logarítmica para Nitratos.....	52
Figura 32. Gráfico de residuos de modelo lineal de transformación logarítmica para Nitratos	53
Figura 33. Relación de Nitratos calibrados con los estimados por el modelo lineal	53
Figura 34. Diagnóstico estadístico del modelo lineal para Nitratos.....	54
Figura 35. Gráficos de diagnóstico del modelo lineal para Nitratos	54

LISTA DE ECUACIONES

Ecuación 1. Algoritmo TurbMoore80Red para cálculo del índice Turbiedad	21
Ecuación 2. Algoritmo 2BDA para el cálculo del índice Clorofila-a.	22
Ecuación 3. Índice NDVI para el cálculo de la Ficocianina	23
Ecuación 4. Función utilizada por el modelo de regresión lineal simple.	24
Ecuación 5. Función utilizada por el modelo lineal de transformación logarítmica.	25
Ecuación 6. Función utilizada por el modelo lineal con múltiples variables.	25
Ecuación 7. Error Cuadrático Medio MSE.....	29
Ecuación 8. Error Absoluto Medio MAE	29
Ecuación 9. Coeficiente de determinación R^2	29

RESUMEN

Los sistemas de teledetección se han vuelto cada vez más importantes en la actualidad al proporcionar información confiable y de fácil adquisición, subsanando la escasez de datos que impiden el seguimiento y evaluación de la dinámica de la calidad del agua a lo largo de un río. La gran mayoría de los ríos en Colombia tiene limitada información que permita caracterizar la calidad de sus aguas en el tiempo y el espacio. Por lo tanto, el objetivo del presente estudio es utilizar información teledetectada para cuantificar concentraciones de los determinantes de calidad del agua Nitratos, Fósforo Total y Sólidos Suspendidos Totales, en puntos de monitoreo del río Magdalena, Colombia. Para su desarrollo se utilizó el producto MOD13Q1 que proviene de la información registrada por el satélite MODIS Terra y media el uso de reflectancia de las bandas 645 nm y 858 nm para calcular los índices Turbiedad, Clorofila-a y Ficocianina mediante algoritmos de teledetección. También fueron usados las concentraciones de los determinantes de la calidad del agua mencionados que fueron proporcionadas por las autoridades ambientales CORMAGDALENA e IDEAM. A partir de esta información, se construyeron modelos de regresión para cuantificar los determinantes en función de los índices calculados y se evaluó su desempeño con el coeficiente de determinación.

Para los determinantes Sólidos Suspendidos Totales y Fósforo Total no se presentaron buenos resultados, a diferencia de los Nitratos que, mediante el modelo de regresión lineal de transformación logarítmica utilizado para su cuantificación en función de los valores mínimos del índice Clorofila-a, se obtuvo un $R^2 = 0.5625$, lo que lleva a concluir que existe un ajuste favorable entre los datos y el potencial uso de la información teledetectada para caracterizar algunos determinantes de calidad del agua en lugares con escasa información o difícil acceso para ejecutar actividad de monitoreo.

Palabras clave: algoritmos de teledetección, calidad del agua, modelos de regresión, MODIS.

ABSTRACT

Remote sensing systems have become increasingly important today by providing reliable and readily available information, filling the data gaps that prevent the monitoring and evaluation of water quality dynamics along a river. The vast majority of rivers in Colombia have limited information to characterize the quality of their waters in time and space. Therefore, the objective of this study is to use remote-detected information to quantify concentrations of the water quality determinants Nitrates, Total Phosphorus and Total Suspended Solids, at monitoring points of the Magdalena River, Colombia. For its development the product MOD13Q1 was used that comes from the information recorded by the satellite MODIS Terra and means the use of reflectance of the bands 645 nm and 858 nm to calculate the indices Turbidity, Chlorophyll-a and Phycocyanin by teledetection algorithms. The concentrations of the aforementioned water quality determinants that were provided by the environmental authorities CORMAGDALENA and IDEAM were also used. Based on this information, regression models were constructed to quantify the determinants according to the calculated indices and their performance was evaluated with the coefficient of determination.

For the determinants Total Suspended Solids and Total Phosphorus did not present good results, unlike the Nitrates that, by the logarithmic transformation linear regression model used for its quantification as a function of the minimum values of the Chlorophyll-a index, an $R^2 = 0.5625$ was obtained, This leads to the conclusion that there is a favorable adjustment between the data and the potential use of remote sensing information to characterize some determinants of water quality in places with scarce information or difficult access to execute monitoring activity.

Keywords: satellite technologies, remote sensing algorithms, water quality, regression models, MODIS

INTRODUCCIÓN

En la actualidad, el manejo, seguimiento y control de la calidad del agua de fuentes superficiales sigue siendo un problema, ya que el realizar un monitoreo constante requiere tiempo y costos elevados, además de la dificultad de acceder a ciertos puntos debido a la naturaleza y problemas sociales del lugar. Por lo cual el uso de las tecnologías satelitales enfocadas en temas meteorológicos y de control del ambiente tienen una mayor importancia y están siendo evaluadas en diferentes campos.

El trabajo de los entes de control del ambiente y la salud pública en todo el mundo ha incrementado, a causa de que los problemas sociales por enfermedades son mayores [1]. Frente a esta situación, las entidades encargadas del recurso hídrico en Colombia también presentan problemáticas a causa del crecimiento que está teniendo al querer ser un país desarrollado, aumentando los proyectos con alta demanda de agua, lo que genera una carga de contaminantes cada vez mayor al ambiente. Por los objetivos de desarrollo sostenible, más específicamente, el referido a la salud y bienestar de las personas, el de agua y saneamiento, y el de la vida submarina [2]. De igual manera, por los temas sociales de conflicto armado mantenidos a causa de este recurso [3] que, si bien han disminuido, siguen presentes en las zonas rurales del país, haciendo que sea más difícil el manejo de las aguas, tanto en su cantidad, como en su calidad.

Adicionalmente, existen limitaciones en la capacidad operativa que tienen las Corporaciones o Entidades encargadas del seguimiento y evaluación de este recurso, impidiendo su correcta caracterización. También, el presupuesto con el que cuentan es limitado para llevar a cabo la actividad de monitoreo, ya que se debe seleccionar una frecuencia y periodicidad de muestreos, junto con un indicador de otros parámetros que son más costosos y complejos medir, como es el caso de la conductividad eléctrica que aplica como indicador de los sólidos disueltos totales [4].

Países como Brasil y México que tienen similitudes en cuanto a los problemas de monitoreo y conflicto armado, han recurrido al uso de tecnologías de teledetección [5] que permitan realizar de manera más eficiente la obtención de información, tratamientos de datos y análisis de estos, sin la necesidad de hacer medición in-situ [6] [7] [8] [9]. De la misma forma en Colombia se han realizado estudios en ríos y embalses que implementan este medio de adquisición de información para analizar las condiciones del agua en algunas zonas del país [10] [11] [12] [13] [14].

La presente investigación busca analizar cómo la información teledetectada recogida por el instrumento MODIS a bordo del satélite Terra, puede ser utilizada para cuantificar los determinantes de calidad del agua Sólidos Suspendidos Totales, Nitratos y Fósforo Total medidos in-situ, en un caso de estudio en diferentes puntos de monitoreo de calidad del agua del río Magdalena, Colombia.

Para su desarrollo, se solicitaron los datos de calidad del agua a lo largo del río Magdalena a las autoridades ambientales IDEAM y CORMAGDALENA, encargadas de su monitoreo, para luego obtener la información teledetectada del instrumento MODIS mediante el servidor EarthData, coincidente con los datos in-situ suministrados y así procesar las imágenes satelitales en RStudio y ArcMap 10.5. Posteriormente se seleccionaron algoritmos de teledetección para extraer índices que permitieran construir modelos y cuantificar los determinantes en función de estos, evaluando su efectividad mediante la métrica de desempeño R^2 , siendo el determinante Nitrato el único en presentar resultados favorables.

Este trabajo inicia contextualizando el área de estudio y la importancia que tiene para Colombia tanto a nivel ambiental como económico, junto con la importancia de implementar la información teledetectada. Continúa explicando toda la teoría fundamental sobre conceptos de teledetección y los modelos necesarios para entender el proyecto, junto con la normativa nacional y los objetivos de desarrollo sostenible aplicables al estudio. Luego se muestra la metodología utilizada y los resultados encontrados al terminar la investigación. Finaliza mencionando la importancia social y humanística del trabajo.

OBJETIVOS

OBJETIVO GENERAL

- Cuantificar concentraciones de Nitrógeno Total, Fósforo Total y Sólidos Suspendidos Totales en puntos de monitoreo del río Magdalena, Colombia, con base en información teledetectada.

OBJETIVOS ESPECÍFICOS

- Identificar los puntos de monitoreo del río Magdalena con información disponible de calidad del agua e imágenes satelitales que coincidan temporalmente.
- Calcular los índices Clorofila-a, Turbiedad y Ficocianina a través de algoritmo derivados de información teledetectada, en los puntos de monitoreo seleccionados.
- Evaluar modelos regresivos para cuantificar Nitrógeno Total, Fósforo Total y Sólidos Suspendidos Totales en función de los índices obtenidos de las imágenes satelitales.

1. MARCO TEÓRICO

El desarrollo de las tecnologías de teledetección de cambios meteorológicos en las últimas dos décadas ha dado un impulso notable a nuevas investigaciones y estudios enfocados en analizar concentraciones de diferentes determinantes e índices presentes en cuerpos de agua.

Por otra parte, debido a que el monitoreo fuentes hídricas sigue siendo un problema a causa de los costos que trae consigo, del desplazamiento de personal, uso de equipos y dificultad de acceso a algunas zonas; el uso de imágenes satelitales ha incrementado en todos los países en desarrollo [15]. Esta implementación favorece la obtención de datos de lagos, ríos y embalses, disminuye costos, y muestra resultados favorables para la gestión del recurso hídrico.

1.1 Área de estudio

El río Magdalena es considerado el cuerpo de agua más importante en Colombia, teniendo un recorrido de 1540 kilómetros y atravesando once departamentos desde su nacimiento en el Huila [16], más específicamente en el páramo de las Papas, zona de la Laguna de La Magdalena localizada al interior de un cráter gigantesco de un extinto volcán en la cima de la Cordillera Central de los Andes a 3685 metros sobre el nivel del mar, en el Macizo Colombiano [17]; sigue su recorrido por el Tolima, Cundinamarca, Boyacá, Caldas, Antioquia, Santander, Cesar, Magdalena, Bolívar y Atlántico, hasta su desembocadura en Bocas de Ceniza en el municipio de Barranquilla, mar Caribe. En el lugar conocido como el estrecho del río Magdalena, ubicado a unos kilómetros del municipio de San Agustín, Huila; el río cuenta con un ancho aproximado de 2.20 metros, siendo su parte más angosta; lo contrario a su anchura en el municipio de Plato, Magdalena, donde posee una extensión de 1073 metros de ancho.

Este río se encuentra ubicado dentro de la cuenca río Magdalena-Cauca, contando con una división en 4 sectores específicos, cuenca alta, media, baja y el río Cauca. El área de la cuenca cubre un total de 725 municipios equivalente al 65 % de los municipios de todo el país. De estos, 129 son ribereños y administrativamente reúne 18 departamentos. En lo que respecta a la importancia socioeconómica del río como corriente natural para Colombia, se resalta que cerca del 80 % de la población se encuentra en la cuenca, genera el 70 % de producción hidroeléctrica, también genera el 95% de la producción termoeléctrica, representa el 75 % de la producción agropecuaria y con los demás sectores productivos asentados en la cuenca contribuyen en un 85% del Producto Interno Bruto [18].

En la Figura 1 se muestra espacialmente el río Magdalena a lo largo de Colombia, con sus divisiones departamentales.

Figura 1. Área de estudio río Magdalena, Colombia



Fuente: autor.

1.1.1 Caracterización de datos disponibles.

Las corporaciones ambientales CORMAGDALENA e IDEAM, encargadas del monitoreo de la calidad del agua a lo largo del río Magdalena, suministraron información de las campañas realizadas en los años 2006, 2007, 2011, 2013, 2015 y 2016; de las cuales fueron seleccionadas para el estudio las correspondientes a los años 2006, 2013 y 2016 [19] [20] [21], al ser las campañas que contaban con mayor cantidad de datos de los tres determinantes de calidad del agua a estudiar, con su fecha correspondiente de monitoreo y la georreferenciación de la estaciones donde se realizaron los muestreos.

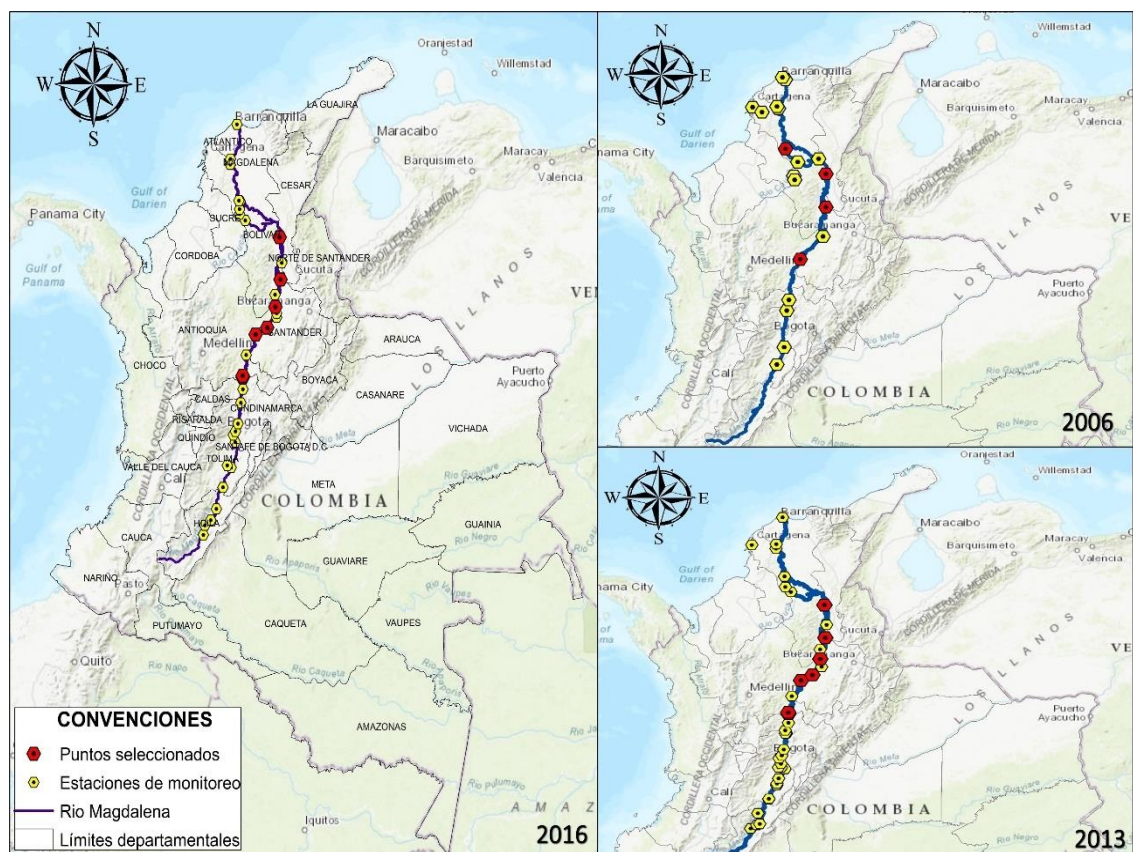
Los puntos de monitoreo de calidad del agua seleccionados para estos tres años a estudiar fueron determinados con base en un ancho del río igual o superior a 250 metros debido a la resolución espacial de MODIS, y descartando zonas con

pixeles sin valor. Al realizar esto, la zona hidrográfica donde se enfocaron los puntos fue el Medio Magdalena.

Para el año 2006 hay 18 estaciones de monitoreo con dos fechas de muestreos cada una, de las cuales se utilizaron 4, de acuerdo con lo mencionado anteriormente; para un total de 16 datos de los determinantes, debido a que no posee datos de Fósforo Total. En el año 2013, hay 20 puntos de monitoreo, 8 seleccionados, dando un total de 24 datos de muestreos. Por último, para el año 2016, se tienen 23 estaciones, siendo utilizadas 8 y un total de 24 datos. En el **ANEXO A** se mencionan los puntos seleccionados para cada año, con el número de muestras utilizadas para cada determinante.

En la Figura 2 se muestran espacialmente las estaciones de monitoreo de calidad del agua para cada año, junto con los puntos seleccionados en la zona hidrográfica del Medio Magdalena, con sus límites departamentales.

Figura 2. Puntos de monitoreo seleccionados para los años 2006, 2013 y 2016 en el río Magdalena



Fuente: autor.

1.2 Determinantes e índices de la calidad del agua

Los índices son utilizados para determinar la presencia de algunos determinantes presentes en el agua sin necesidad de realizar una medición de este. Para la investigación se utilizaron los índices Turbiedad para Sólidos Suspendedos Totales, Clorofila-a para Nitrógeno Total y Ficocianina para Fósforo Total.

1.2.1 Índices.

1.2.1.1 Clorofila-a.

Es un pigmento verde presente en algas y plantas que permite la absorción de la luz solar para que realicen el proceso de fotosíntesis. Generalmente es usada como indicador de productividad de algas en ecosistemas acuáticos, es decir que, mediante la concentración de Clorofila-a presente en el agua, se determina el aumento de biomasa de algas. Por lo general altas concentraciones indican mala calidad del agua. Su aumento se ve influenciado por la presencia de materia orgánica proveniente de las plantas de tratamiento de agua residual, de altos niveles de nutrientes de fertilizantes y escorrentía urbana [22].

1.2.1.2 Turbiedad.

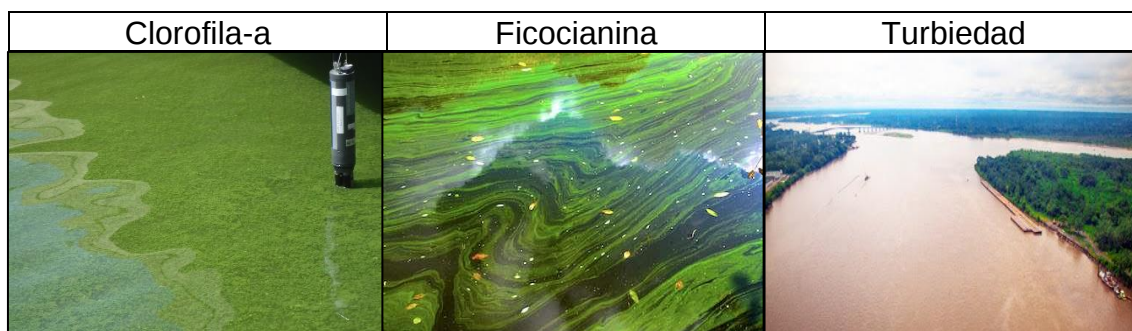
Es una medida de la claridad del agua, refiriéndose a lo turbia o clara que puede estar. Refleja el contenido de sustancias orgánicas, coloidales y minerales en el agua [23]. Una concentración alta es perjudicial para cuerpos de agua al reducir la cantidad de luz necesaria para la fotosíntesis de las plantas, ocasionando menor producción de oxígeno y haciendo que los niveles de Oxígeno Disuelto sean bajos [24]. Su aumento se ve influenciado por los sedimentos procedentes de la erosión de tierras cercanas.

1.2.1.3 Ficocianina.

Es un pigmento verde producido por las cianobacterias, el cual ayuda a estimar la concentración de éstas en el agua, que en altas cantidades pueden generar toxicidad. También es utilizada como indicador de floraciones de algas nocivas y vigilancia de procesos de eutrofización [25].

En la Figura 3 se muestra cómo se ven los índices mencionados anteriormente, en un cuerpo de agua.

Figura 3. Visualización de los índices Clorofila-a, Ficocianina y Turbiedad en el agua



Fuente: [26] [27] [28]

1.2.2 Determinantes.

1.2.2.1 Sólidos Suspendedos Totales (SST).

Son la cuantificación de sólidos en partículas, sedimentables y en suspensión, sin incluir sólidos solubles, tanto inorgánicos como orgánicos [29], que están presentes en corrientes de agua superficial y se consideran como la cantidad de residuos que se retienen en un filtro de fibra de vidrio de poro nominal de 0.45 micras. Este material particulado incluye limos, arcillas, desechos orgánicos finos, algas, etc., que en alta presencia impiden la entrada de luz, afectando el desarrollo del ecosistema acuático. La fuente de SST incluye aguas residuales, erosión del suelo, fertilizantes, vertimientos industriales, escorrentía natural y es medida en miligramos por litro (mg/L) [30].

1.2.2.2 Nitrógeno Total.

Es una medida que totaliza la concentración de las diferentes formas del nitrógeno presentes en un cuerpo de agua. El nitrógeno se encuentra presente en forma de nutriente, necesario para el crecimiento de animales y plantas acuáticas. Es abundante en el medio ambiente, pero también se introduce por fertilizantes, aguas residuales y escorrentía. Su presencia en exceso causa una eutrofización en el agua, generando un crecimiento fuera de lo normal de algas y plantas presentes en la fuente hídrica, disminuyendo así los organismos y microorganismos presentes en ella [31].

1.2.2.3 Fósforo Total.

Constituye todas las formas del fósforo, siendo este componente esencial para el desarrollo de las plantas. Sin embargo, cuando tiene concentraciones altas ocasiona una aceleración en la eutrofización, generando la reducción del Oxígeno Disuelto en el agua por el incremento de nutrientes [32]. Es común que su aumento en cuerpos de agua superficiales se vea influenciado por desechos orgánicos, efluentes industriales, fertilizantes orgánicos y por escorrentía [33].

1.3 Conceptos generales en la teledetección

Para realizar la teledetección de cualquier fenómeno natural u objeto en la tierra, se debe tener presente tres factores principales; una fuente energética de radiación electromagnética (sol o generada), interacción del objeto de estudio con la radiación (agua, suelo, etc.) y el sistema que recibe la radiación reflejada (sensores). Al reunir estas tres condiciones, la información captada mediante el sensor a bordo de un satélite es representada visualmente en una imagen satelital [34]. A continuación, se describen los principales conceptos de teledetección.

1.3.1 Resolución.

Es la capacidad de registro que tiene un sensor. La radiación tanto emitida como reflejada de la tierra se divide en 4 dimensiones (espacial, temporal, espectral y radiométrica) [35].

1.3.1.1 Espacial.

Es la encargada de definir el tamaño del píxel de una imagen satelital, cuanto menor sea, mayor es la claridad de la imagen.

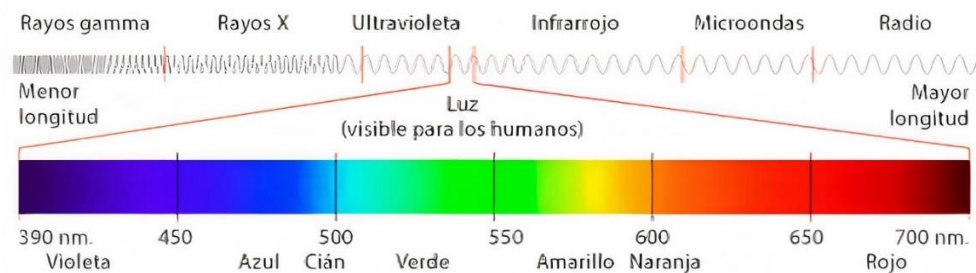
1.3.1.2 Temporal.

Describe el tiempo que tarda un satélite para orbitar y volver a visitar un área específica, y se define como el tiempo que transcurre entre dos tomas de imágenes.

1.3.1.3 Espectral.

Es el número de bandas de radiación en el espectro electromagnético que un satélite puede muestrear [36]. En la Figura 4 se muestra el espectro electromagnético con su longitud de onda en nanómetros y la radiación de cada uno.

Figura 4. Espectro electromagnético



Fuente: [37]

1.3.1.4 Radiométrica.

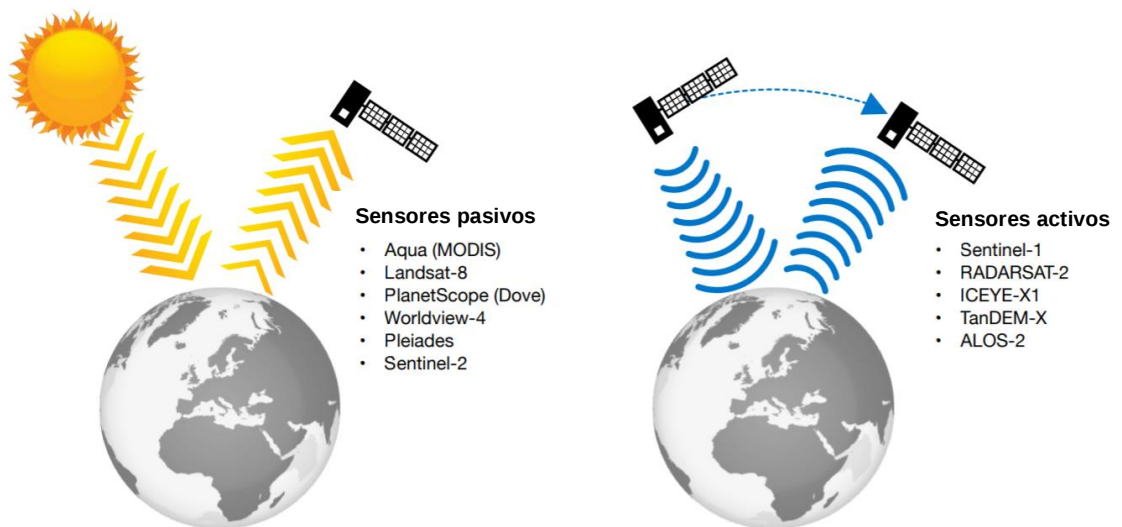
Se refiere al número de intervalos de intensidad que puede captar un sensor, expresado generalmente en bits necesarios para un píxel. Entre mayor sea la precisión, la imagen tendrá mejor nivel de detalle [38].

1.3.2 Misiones satelitales.

Las misiones satelitales que tienen por objeto evaluar el comportamiento de la superficie o la atmósfera de la tierra, por lo general, están instrumentadas con sensores activos o pasivos, necesarios para realizar la captación, almacenamiento y transmisión de imágenes satelitales a distancia.

Los sensores activos, son capaces de generar su propia radiación y recibirla rebotada; los pasivos miden la radiación electromagnética reflejada por la superficie terrestre y usan energía solar [39]. En la Figura 5 se muestra lo mencionado anteriormente y se listan algunos sensores de acuerdo con su clasificación entre pasivo y activo.

Figura 5. Ilustración de sensores satelitales pasivos y activos



Fuente: [35]

Un aspecto importante para determinar el sensor que debe utilizarse en un estudio, es el cuerpo hídrico que se piensa analizar, ya que el agua refleja de maneras diferentes las ondas electromagnéticas, dependiendo de su contenido. Por ejemplo, el agua con alta turbiedad refleja de manera distinta al agua clara o la que tiene contenido de vegetación [40]. También se debe tener en cuenta las bandas utilizadas por el sensor, ya que son las encargadas de captar la luz reflejada por el agua.

En la presente investigación fue utilizado el sensor pasivo MODIS del satélite Terra. Los sensores pasivos se pueden clasificar de la siguiente manera:

1.3.2.1 Radiómetros de barrido y empuje.

Son sistemas óptico-electrónicos que descomponen en varias longitudes de onda la radiación, luego cada longitud es enviada a un grupo de detectores que la transforman en señal eléctrica y la convierten en un valor numérico llamado Nivel Digital (ND). Los de barrido son los más comunes; como lo explica [38], cuentan con un espejo móvil que oscila perpendicularmente al curso de su trayectoria, lo que permite explorar una franja de terreno más amplia. Por otra parte, los radiómetros de empuje no cuentan con el espejo oscilante debido a que poseen varios detectores que cubren todo el campo visual del sensor, permitiendo aumentar la resolución y reduciendo errores.

1.3.2.2 Radiómetros de microondas.

Están compuestos por una antena que recibe y aumenta la señal de microondas emitidas por el objeto que refleja la radiación. La longitud de onda es directamente proporcional al diámetro de la antena, mientras que la resolución espacial es inversamente proporcional, por lo cual es necesario un tamaño de píxel bastante grande. Al poseer resolución espacial pequeña, solo funcionan para estudios globales [41].

1.3.3 Espectrorradiómetro de imágenes de media resolución (MODIS).

El espectrorradiómetro de imágenes de media resolución MODIS, fue seleccionado debido sus características, ya que rastrea más signos de la Tierra que cualquier otro sensor Terra, teniendo un tamaño de imagen cerca de 2330 km², una resolución espacial de 250 a 1000 metros, resolución temporal de 1 a 2 días, una resolución Espectral de 36 bandas disponibles, de las cuales se usaron la 1 Roja Visible (longitud de onda 620-670 nanómetros) y 2 Infrarrojo cercano (longitud de onda 841-876nm), y una resolución radiométrica (ND) 1024 [39]. Las características principales de las bandas MODIS se muestran en el **ANEXO B**.

1.3.3.1 Producto terrestre MOD13Q1.

El producto fue seleccionado ya que MODIS posee mayor temporalidad, es decir, abarca todos los años de estudio y permite obtener más imágenes que otros satélites, que si bien poseen mejores resoluciones espaciales como WorldView-2 y 3, Sentinel-2 y Landsat-8, no contaban con imágenes para las fechas de los muestreos.

De igual manera, MOD13Q1 cuenta con resolución espacial de 250 metros, la cual puede ser utilizada respecto al ancho del río Magdalena. Posee dos capas que muestran el estado general de la vegetación de la zona; la primera es el Índice de Vegetación de Diferencia Normalizada NDVI, derivada de un radiómetro avanzado de alta resolución perteneciente a la Administración Nacional Oceánica y Atmosférica (NOAA-AVHRR), la segunda es el índice de vegetación mejorado (EVI), para regiones de alta biomasa. También incluye dos capas de calidad de píxeles, las bandas de reflectancia MODIS 1 (rojo), 2 (infrarrojo cercano), 3 (azul) y 7 (infrarrojo medio) y cuatro de observación.

Este producto cuenta con un algoritmo que escoge el valor de píxel con mejor comportamiento durante un periodo de adquisición de 16 días, bajo un criterio de poca presencia de nubes, ángulo de visión bajo que se refiere al área de cobertura del sensor por cada tramo de la imagen y el valor más alto de la relación NDVI/EVI, y genera una sola imagen de mejor calidad, utilizando los datos diarios durante 16 días [42]. Otra característica importante es que ya trae consigo un preprocesamiento que incluye la corrección atmosférica y geométrica [43] [44], debido a los criterios que maneja para la recolección de información.

1.4 Algoritmos de teledetección para índices de calidad del agua

Son los encargados de calcular los índices mediante su ecuación correspondiente. Básicamente, extraen la reflectancia de cada píxel en las imágenes satelitales, por cada banda del instrumento seleccionado y arrojan un valor numérico [45]. Los algoritmos utilizados en el trabajo correspondieron a TurbMoore80Red para calcular Turbiedad, 2BDA (Two Bands Algorithm) para Clorofila-a y NDVI (Normalized Difference Vegetation Index) para Ficocianina.

1.4.1.1 TurbMoore80Red.

Hace parte de los 45 algoritmos del paquete de R [46] waterquality, el cual tiene como función principal convertir fácil y rápidamente imágenes satelitales generadas por WorldView-2 y 3, Sentinel-2, Landsat-8, MODIS, MERIS y OLCI, en uno o varios índices diseñados para detección de Turbiedad, Clorofila-a y cianobacterias (Ficocianina). TurbMoore80Red es uno de los tres algoritmos que detectan Turbiedad, aplicables a MODIS en el entorno RStudio [47]. Utiliza solamente la banda 1 (roja visible) del instrumento, y selecciona un valor de reflectancia a 658 nanómetros como se muestra en la Ecuación 1.

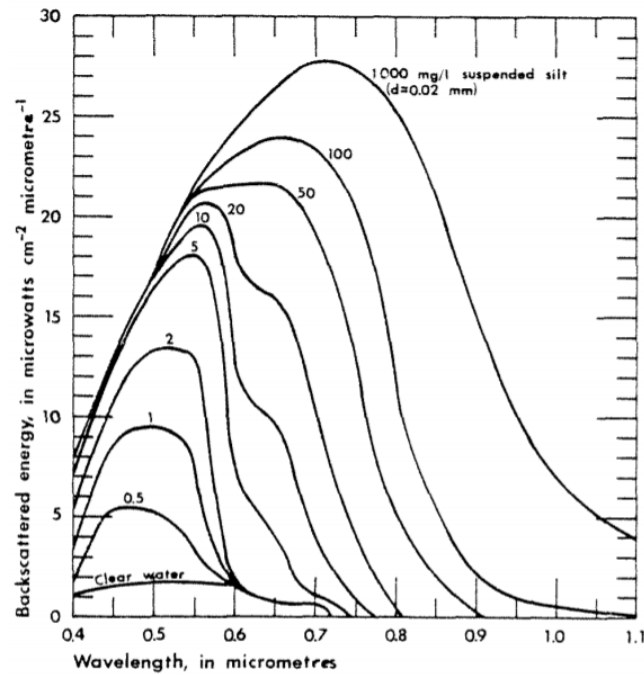
Ecuación 1. Algoritmo TurbMoore80Red para cálculo del índice Turbiedad

$$TurbMoore80Red = function(w658, MODIS)$$

Fuente: [48]

Su desarrollo está fundamentado en el estudio [48], donde utilizan un gráfico de flujo de energía retrodispersada (energía que se refleja en el objeto), frente a la longitud de onda para varias concentraciones de limo suspendido por litro como se muestra en la Figura 6.

Figura 6. Gráfico de energía restrodispersada de limo en suspensión



Fuente: [48]

Muestra básicamente que el flujo retrodispersado aumenta con las concentraciones de limo en suspensión (calculado para partículas diámetro de 0,02 mm) [48].

1.4.1.2 Two Band Algorithm (2BDA).

El algoritmo utiliza las bandas 1 (rojo) y 2 (infrarrojo) del instrumento MODIS, realizando una relación de bandas [49], como se muestra en la Ecuación 2, para calcular el índice Clorofila-a, teniendo como referencia los valores de reflectancia de cada píxel de la imagen en la zona de estudio.

Ecuación 2. Algoritmo 2BDA para el cálculo del índice Clorofila-a.

$$2BDA = \frac{MOD13Q1 [b2(float (857nm))]}{MOD13Q1 [b1(float (646nm))]}$$

Fuente: [49]

1.4.1.3 Normalized Difference Vegetation Index (NDVI).

Este índice es utilizado para dar una estimación de cantidad, calidad y desarrollo de la vegetación, dependiendo su uso; midiendo la diferencia entre la luz absorbida (rojo visible), con la luz reflejada (infrarrojo cercano), dividida en la suma de estas dos intensidades [50] como se muestra en la Ecuación 3. En este caso de estudio se utilizó para estimar la Ficocianina presente en las algas verdiazules (cianobacterias).

Ecuación 3. Índice NDVI para el cálculo de la Ficocianina

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)}$$

Fuente: [50].

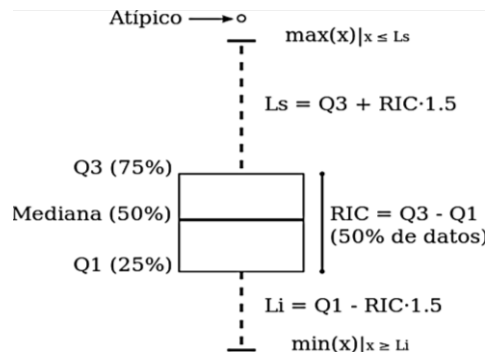
1.5 Modelos de regresión

Estos modelos se basan en un análisis de regresión el cual permite estudiar la relación o dependencia entre dos o más variables [51]. En esta investigación, la intensidad de dicha relación es cuantificada por el coeficiente de correlación de Pearson, el cual asigna un valor en un rango de -1 a 1, indicando que no existe correlación con un valor 0, que la relación es positiva total cuando es 1 y negativa total cuando es -1. Es decir que, si existe una correlación positiva alta, las dos variables estudiadas tienen un comportamiento lineal y aumentan al mismo tiempo, pero si la correlación es cercana a 0, indica que cuando una variable aumenta la otra disminuye [52].

La función principal de un modelo es utilizar un conjunto de información y proporcionar un resumen concreto y de baja dimensión para explicar la realidad por medio de supuestos o predicciones [53].

Antes de realizar la implementación de un modelo es necesario identificar datos atípicos en cada variable, en este caso se utiliza la metodología Boxplot o diagrama de caja y bigotes, los cuales muestran gráficamente un resumen de los datos, identifican los valores atípicos y permiten intuir la simetría de los datos [54]. Poseen cinco medidas descriptivas como se muestra la Figura 7.

Figura 7. Diagrama de caja y bigotes o Boxplot.



Fuente: [54]

La primera es la mediana o cuartil 2, que divide en dos partes el diagrama, dejando la misma cantidad de valores a un lado que a otro. La segunda es el cuartil 1 (Q1), que muestra el 25% de los datos menores o iguales a este valor. Luego se encuentra el tercer cuartil (Q3), mostrando el 75% de los valores menores o iguales a este valor. También se encuentra el Rango Intercuartílico (RIC), que es la diferencia entre el tercer y primer cuartil. Por último, están el límite inferior (Li) que muestra el valor mínimo de los datos y límite superior (Ls) que muestra el valor máximo de los datos; los valores por encima o debajo de estos se denominan atípicos. En lo que respecta a las líneas o bigotes a los extremos de la caja, su tamaño depende de que tan dispersos o distantes están los datos hacia esos valores, más no que exista más datos a un lado que al otro [54].

1.5.1 Modelo lineal simple.

Los modelos de regresión lineal simple (simple = una variable independiente) [55], permiten estimar como cambia la variable dependiente o respuesta a medida que lo hace una variable independiente o explicativa, ajustando una línea recta a los datos observados, mientras que los modelos de regresión no lineal y regresión logística usan una línea curva [56]. En otras palabras, el modelo lineal usa una función matemática que permite expresar una variable a partir de otra, mediante una regresión lineal, como se muestra en la Ecuación 4.

Ecuación 4. Función utilizada por el modelo de regresión lineal simple.

$$y = ax + b$$

Fuente: [55]

y = variable dependiente o respuesta
x = variable independiente o explicativa
a = pendiente o coeficiente a ser determinado
b = intersección, cuando $x=0$, $y=b$

1.5.1.1 Modelo lineal de transformación logarítmica

La razón para realizar este tipo de transformación se debe a la no linealidad de las variables dependiente e independiente, con la cual se busca que exista varianza constante u homocedasticidad, y una relación más simétrica entre los datos [57]. Es importante tener en cuenta que los valores utilizados sean > 0 para poder aplicar el modelo. Esta función permite determinar las estimaciones no lineales (ponderadas) de mínimos cuadrados de los parámetros de un modelo [58] y su función general se muestra en la Ecuación 5.

Ecuación 5. Función utilizada por el modelo lineal de transformación logarítmica.

$$y = a * \log(x) + b$$

Fuente: [58]

y = variable dependiente o respuesta

x = logaritmo natural de la variable independiente o explicativa

a = pendiente o coeficiente a ser determinado

b= intersección, cuando $x=0$, $y=b$

1.5.1.2 Modelo lineal con múltiples variables

Este modelo utiliza la regresión lineal múltiple como base, la cual trata de ajustar un modelo lineal donde la variable respuesta (y), es determinada por un conjunto de variables predictoras ($x_1, x_2, x_3 \dots$) [59]. Pueden utilizarse para predecir un valor de la variable respuesta o para analizar la influencia de los predictores sobre ésta. Los modelos lineales múltiples siguen la Ecuación 6, donde $+$ indica la inclusión de una variable predictora, no adición.

Ecuación 6. Función utilizada por el modelo lineal con múltiples variables.

$$y = a_1x_1 + a_2x_2 + a_nx_n + b$$

Fuente: [59]

y = variable dependiente o respuesta

x_n = variables independientes o explicativas

a_n = coeficientes a ser determinados

b= intersección, cuando $x=0$, $y=b$

Una característica para tener en cuenta es que los predictores deben ser independientes entre ellos, es decir, no tener colinealidad. Si esta relación existe, no se puede evaluar el efecto que tiene cada uno sobre la variable dependiente y provocar grandes cambios en las estimaciones [60].

1.5.1.3 Diagnóstico de regresión

Luego de aplicar un modelo, es necesario hacer un diagnóstico de regresión para verificar si es correcto usarlo o no. Este puede presentarse mediante un gráfico de diagnóstico como se muestra en la Figura 8, el cual permite analizar mediante gráficas si el modelo presenta linealidad de los datos, normalidad de los residuos, homogeneidad de la varianza de los residuos e independencia de los términos de error de los residuos [61]. De igual manera, mediante la estadística del modelo es posible realizar un mejor diagnóstico.

▪ Estadística del modelo.

Se analiza con el fin de determinar si el modelo puede ser aplicado. La estadística del modelo se analiza para determinar si la mediana es cercana a cero, los valores mínimos y máximos de los residuos son cercanos entre sí en valor absoluto; el p-value para establecer el grado de significancia de las variables predictoras respecto a la variable respuesta, mediante los coeficientes de regresión, y el R^2 para medir que tanto explica el modelo los datos. En un modelo no lineal o de transformación se revisa el error residual estándar que entre más cercano esté a 0, mejor es el modelo [62].

Al tener estas características, se puede rechazar la hipótesis nula y aceptar la hipótesis alternativa, indicando así una asociación aceptable entre el predictor y la variable resultado, permitiendo aplicar el modelo [63].

▪ Gráficos de diagnóstico.

La interpretación de las cuatro gráficas mostradas en la Figura 8 para diagnóstico de modelos se entiende de la siguiente manera:

Residuos vs Ajustados. Es utilizado para comprobar supuestos de relación lineal. Cuando la línea roja es relativamente horizontal y no presenta variaciones drásticas, indica linealidad de los datos. En este ejemplo, no hay un patrón de los residuos, lo que indica una relación lineal entre las variables.

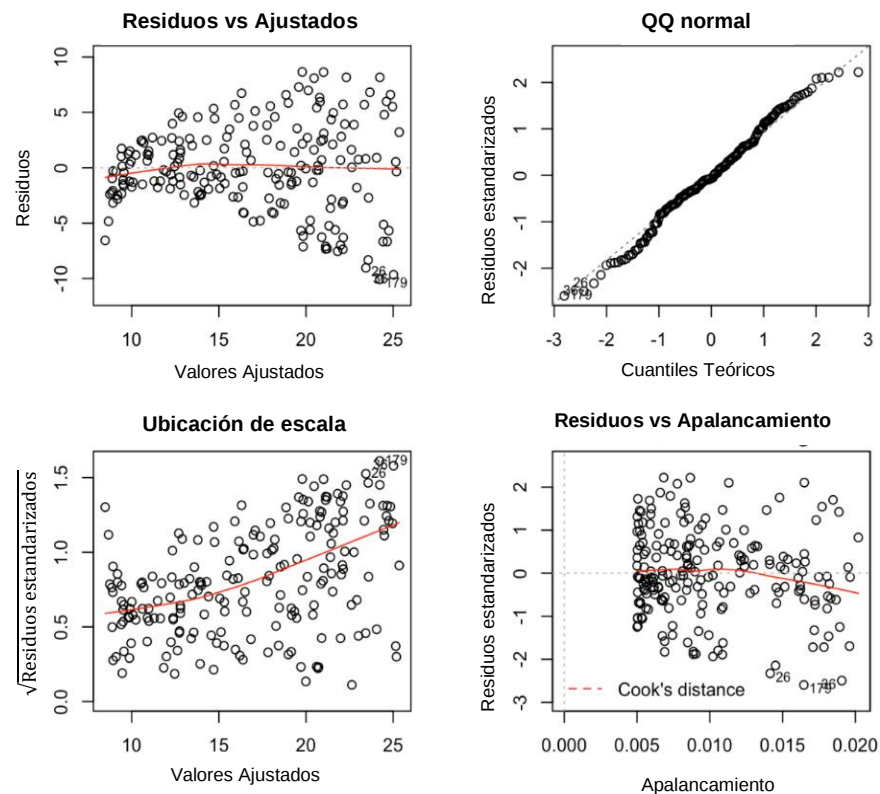
QQ normal. Muestra la normalidad de los residuos. Cuando los puntos residuales siguen la tendencia de la línea recta, se supone que los errores residuales se distribuyen normalmente como es el caso presentado.

Ubicación de escala (o ubicación de extensión). Con este gráfico se comprueba si los residuos tienen una varianza constante (homocedasticidad); si los puntos están distribuidos de igual manera alrededor de la línea horizontal, existe homocedasticidad; en el ejemplo mostrado hay heterocedasticidad.

Residuos vs apalancamiento. Se utiliza para saber si los datos tienen valores que puedan influir en el resultado de la regresión al incluirlos o eliminarlos del análisis. Estos valores influyentes pueden ser datos atípicos en la variable

resultado (y), o puntos de alto apalancamiento (valores extremos), en la variable predictora (x). En la gráfica de ejemplo se observan tres puntos más extremos (#26, #36, #179), con residuos estandarizados debajo de -2. Sin embargo, es bueno porque no existen datos atípicos que estén por encima de las 3 desviaciones estándar, es decir que no existen puntos de alto apalancamiento

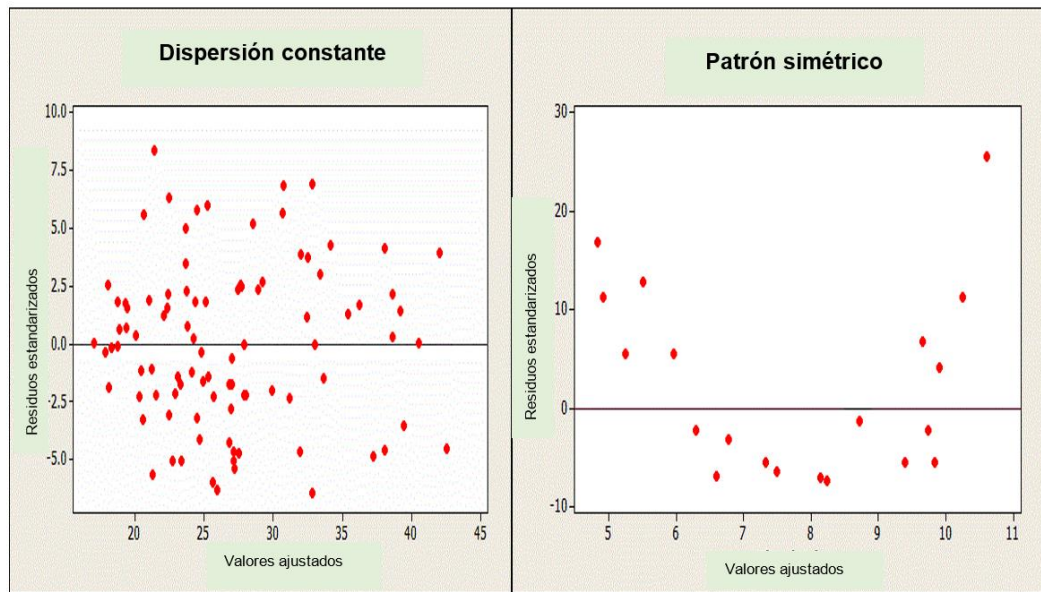
Figura 8. Gráficos de diagnóstico para la evaluación de modelos lineales.



Fuente: [61]

Para realizar el diagnóstico de un modelo no lineal o con alguna transformación como la logarítmica, se deben revisar los residuos estandarizados respecto a los valores ajustados por el modelo. Es decir que los residuos deben tener una distribución aleatoria o dispersa y no tener un patrón como se muestra en la Figura 9.

Figura 9. Gráficas de residuos con dispersión constante y patrón simétrico



Fuente: [62]

1.5.1.4 Métricas de desempeño.

Un criterio de desempeño es básicamente relaciones que señalan la capacidad que tiene un modelo para representar la realidad. Pueden ser usadas en un entorno de calibración, validación, optimización y evaluación de incertidumbre [64].

Generalmente, la creación de un modelo está basada en un principio de retroalimentación. Se construye el modelo, se mide su desempeño con las métricas y se realizan mejoras hasta conseguir una precisión aceptable, ya que no solo se trata de su construcción si no de que ofrezca una alta precisión con el uso de datos de muestra [65].

La métrica se elige de acuerdo con el tipo de modelo y la forma de implementación. Las métricas más comunes para evaluar predicciones de regresión son RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), R^2 (Coeficiente de determinación) [66].

- **Error Cuadrático Medio (MSE)**

Es una de las métricas más comunes utilizadas para evaluar regresiones y se represente como lo muestra la Ecuación 7.

Ecuación 7. Error Cuadrático Medio MSE

$$MSE = \frac{1}{n} \sum_{j=1}^n (y_i - \hat{y}_i)^2$$

Fuente: [66].

Mide básicamente la diferencia cuadrada entre los datos observados (y_i) y los datos estimados ($\hat{y}_i$), para luego promediar esos valores. Entre más grande sea, peor es el modelo y no puede arrojar valores negativos. En un modelo perfecto su valor sería 0 [66].

▪ **Error Absoluto Medio (MAE)**

Este error es la media de la diferencia absoluta entre los datos observados (y_i) y los datos estimados ($\hat{y}_i$), mostrado en la Ecuación 8. Es decir, que todas las diferencias individuales se ponderan por igual en el promedio. Entre más lejano este de cero, peor es la predicción.

Ecuación 8. Error Absoluto Medio MAE

$$MAE = \frac{1}{n} \sum_{j=1}^n |y_i - \hat{y}_i|$$

Fuente: [66].

▪ **Coefficiente de determinación R^2**

Esta métrica ayuda a evaluar el rendimiento porcentual de un modelo de regresión y se interpreta como la variación total de la variable respuesta, explicada por la variación de las variables predictoras, es decir, muestra el ajuste de las variables independientes a la variable dependiente. Se puede ver su sintaxis básica en la Ecuación 9, en el denominador la sumatoria de los errores residuales y en el denominador la sumatoria de las desviaciones. Una ventaja es que no posee escala y se mueve en un rango de $0 \leq R^2 \leq 1$, donde valores cercanos a 0 muestran que el modelo no explica la variable resultado, y los cercanos a 1 indican que el ajuste del modelo a la variable que se quiere explicar es favorable [65].

Ecuación 9. Coeficiente de determinación R^2

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_i)^2}$$

Fuente: [65].

1.6 Normatividad del proyecto

Se considera un marco normativo de acuerdo con la salud y saneamiento ambiental, su control e importancia para el bienestar de las personas, con el fin de mostrar la normativa aplicable a la investigación.

Tabla 1. Marco normativo aplicable a la investigación

NORMA		FECHA	CONTENIDO
Art. 49 Constitución Política de Colombia		Julio 4,1991	Establece la atención de la salud y el saneamiento ambiental, y ordena al Estado encargarse de organizar, dirigir y reglamentar estos.
Art. 79 Constitución Política de Colombia		Julio 4,1991	Consagra el derecho de todas las personas residentes en el país de gozar de un ambiente sano.
Art. 80 Constitución Política de Colombia		Julio 4,1991	Establece como deber del Estado la planificación del manejo y aprovechamiento de los recursos naturales, para garantizar su desarrollo sostenible, su conservación, restauración o sustitución.
Ley 99		Dic 22,1993	Por la cual se crea el Ministerio del Medio Ambiente, se reordena el Sector Público encargado de la gestión y conservación del medio ambiente y los recursos naturales renovables, se organiza el Sistema Nacional Ambiental, SINA, y se dictan otras disposiciones.
Ley 9		Enero 24,1979	Código sanitario nacional
			Art. 51 a 54: Control y prevención de las aguas para consumo humano.
			Art. 55 aguas superficiales.
			Art. 69 a 79: potabilización de agua
Decreto 1575		Mayo 9, 2007	Por el cual se establece el Sistema para la Protección y Control de la Calidad del Agua para Consumo Humano
Resolución 2115		Junio 22, 2007	Por medio de la cual se señalan características, instrumentos básicos y frecuencias del sistema de control y vigilancia para la calidad del agua para consumo humano
Decreto 3930		Oct 25, 2010	Por el cual se reglamenta parcialmente el Título I de la Ley 9 de 1979, así como el Capítulo 11 del Título VI-Parte 11I- Libro 11 del Decreto - Ley 2811 de 1974 en cuanto a usos del agua y residuos líquidos y se dictan otras disposiciones.
Decreto 1076		Mayo 26, 2015	Por medio del cual se expide el Decreto Único Reglamentario del Sector Ambiente y Desarrollo Sostenible
ISO 26000	Objetivo 3	2018	Conjunto de guías que establecen líneas en materia de Responsabilidad Social establecidas por la Organización Internacional de Normalización, para dar cumplimiento a los Objetivos de Desarrollo Sostenible. Objetivo 3 Salud y Bienestar. Objetivo 6 Agua limpia y Saneamiento.
	Objetivo 6		

Fuente: autor.

2. METODOLOGÍA

Para realizar la presente investigación se desarrolló una metodología de acuerdo con los objetivos propuestos, comprendiendo tres fases principales como se muestra en la Figura 13.

2.1 FASE 1. Identificación de puntos de monitoreo en el río Magdalena

Esta fase inició solicitando a las autoridades ambientales IDEAM y CORMAGDALENA, encargadas de realizar el monitoreo en la zona, toda la información in-situ disponible de los determinantes de calidad del agua a lo largo del río Magdalena. Seguidamente se escogieron las campañas que contaban con mayor cantidad de datos, para luego seleccionar los puntos de monitoreo en el río que poseían datos de Sólidos Suspendidos Totales, Nitrógeno Total y Fósforo Total, con la fecha respectiva de su monitoreo y las estaciones georreferenciadas. Se aclara que en lo que respecta al Nitrógeno, fueron incluidos los Nitratos, debido a que no había información suficiente de Nitrógeno Total.

Posteriormente, se revisaron los **MODIS Land Products** [67], para seleccionar el que mejor se ajustara al propósito del trabajo; teniendo como referencia que el ancho mínimo del río es de 220 metros y la parte más grande es de 1000 metros, para así determinar qué imágenes contaban con resoluciones espaciales aptas. Una característica de estos productos es que ya tienen integrada la corrección atmosférica y geométrica.

Luego, mediante el servidor Earthdata de la NASA, se buscaron los datos disponibles del producto escogido anteriormente en la zona de estudio, teniendo en cuenta que las fechas de muestreos de las campañas coincidiera con la resolución temporal de las imágenes satelitales, para así filtrar la información y seleccionar las mismas estaciones para los años de estudio.

2.2 FASE 2. Cálculo de índices Clorofila-a, Turbiedad y Ficocianina

En esta fase se procesan en RStudio y Arcgis10.5, las imágenes obtenidas del instrumento. Posteriormente, se seleccionan los algoritmos aplicables para la detección de las métricas a trabajar y se calculan los índices Clorofila-a, Turbiedad y Ficocianina mediante códigos en el entorno RStudio [47]. Por último, al tener los valores de reflectancia por cada estación, arrojados por los algoritmos, se hace una estadística para extraer los valores mínimos, máximos y la media por cada punto de monitoreo para representar temporal y espacialmente su comportamiento en el tiempo por medio de diagramas de dispersión.

2.2.1 Extracción de reflectancia para puntos de monitoreo

Inicialmente se convirtieron las imágenes de formato HDF a TIFF, extrayendo únicamente las bandas de reflectancia 1 (rojo.tif), 2 (infrarrojo cercano.tif) y 3 (azul.tif), todas con resolución espacial de 250m. Luego, se realizó un stack para crear una imagen que contuviera solo las tres bandas y poder así trabajar en la extracción de reflectancia de cada estación. A continuación, se muestra un ejemplo con un punto de monitoreo llamado **Regidor**.

- I. Seleccionar punto para extracción de reflectancia, como se muestra en la Tabla 2.

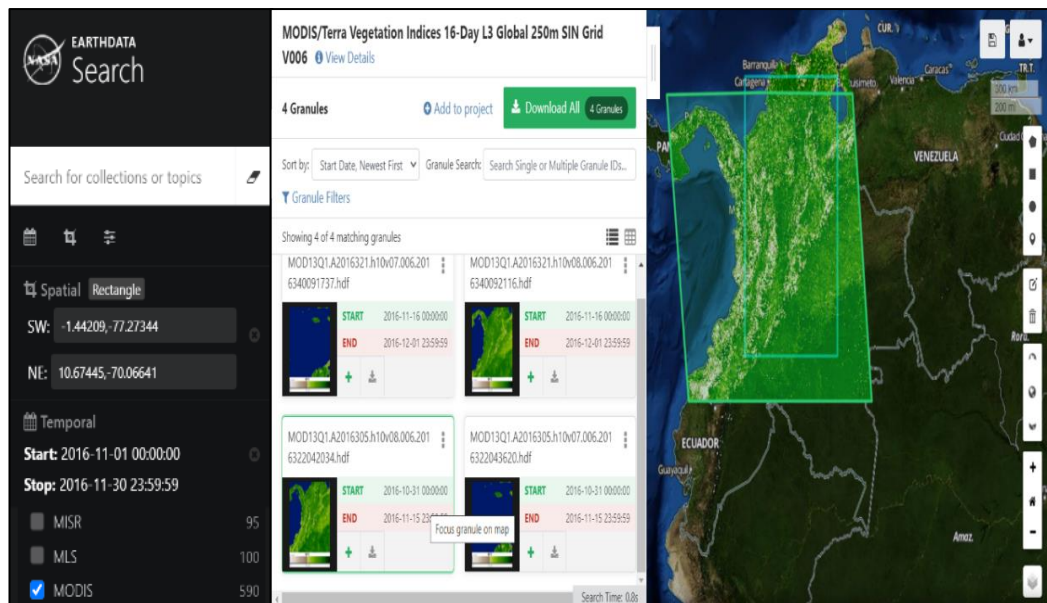
Tabla 2. Punto de monitoreo para ejemplo de extracción de reflectancia.

FECHA DE MUESTREO	CODIGO DE LABORATORIO	NOMBRE ESTACIÓN	LATITUD	LONGITUD	ELEVACION (msnm)	CORRIENTE	MUNICIPIO	DEPARTAMENTO
03/11/2016	25351	Puerto Salgar	05°28'92"	74°39'43"	172	MAGDALENA	PUERTO SALGAR	CUNDINAMARCA
06/11/2016	25363	Desembocadura Río Sogamoso	07°11'55"	73°54'35"	68	SOGAMOSO	PUERTO WILCHES	SANTANDER
06/11/2016	25364	Río Cimitarra Desembocadura	07°27'34"	73°56'10"	70	CIMITARRA	SAN PABLO	BOLIVAR
07/11/2016	25365	Sitio Nuevo R11	07°28'30"	73°55'26"	70	MAGDALENA	PUERTO WILCHES	SANTANDER
07/11/2016	25366	Río Lebrija Desembocadura	08°07'32,2"	73°46'23"	45	LEBRIJA	AGUACHICA	CESAR
08/11/2016	25367	Regidor	08°39'59"	73°49'15"	35	MAGDALENA	REGIDOR	BOLIVAR

Fuente: [21]

- II. Obtener imagen satelital del producto MODIS que corresponda a la fecha de muestreo realizada en la estación de monitoreo en el servidor Earthdata mostrado en la Figura 10.

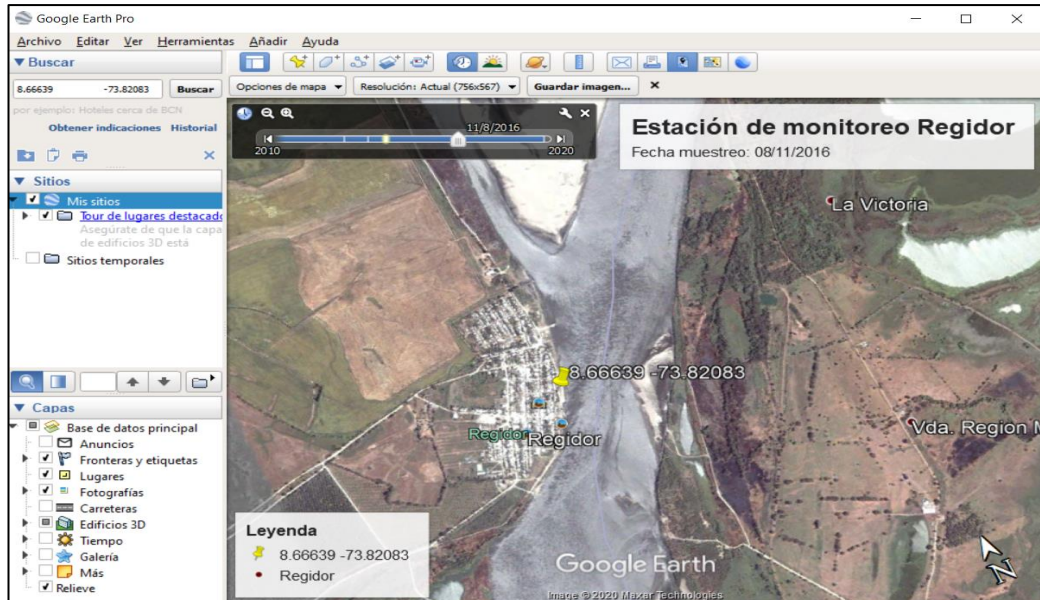
Figura 10. Vista general del servidor EarthData de la NASA



Fuente: [68]

- III. Con la ayuda de Google Earth se observó el punto de monitoreo en la fecha de muestreo, con el fin de ver su localización y cómo se encontraba el cauce del río, como se ve en la Figura 11.

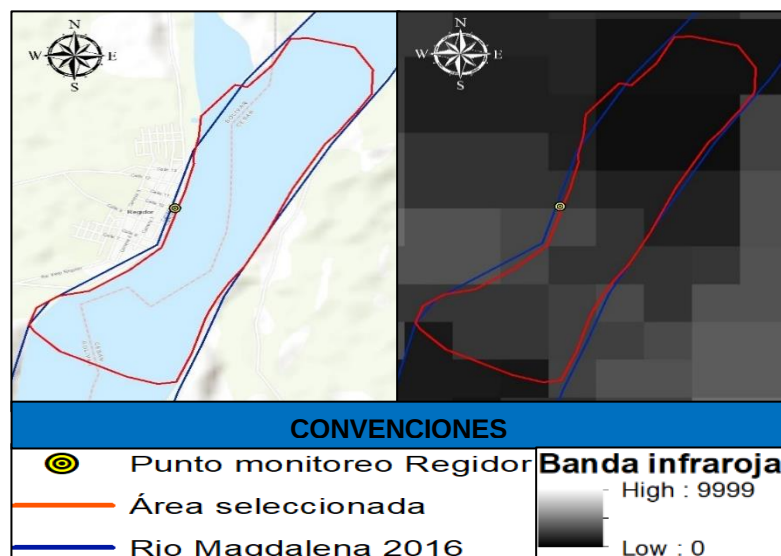
Figura 11. Localización de punto de monitoreo Regidor para el año 2016



Fuente: autor mediante Google Earth Pro.

- IV. Una vez realizado el procesamiento de la imagen, se delimita un área de acuerdo con la orilla del río y una extensión de 1000 metros aguas arriba y aguas abajo del punto de muestreo mostrado en la Figura 12, para obtener una cantidad significativa de píxeles y así poder realizar el cálculo de los índices.

Figura 12. Delimitación y extracción de reflectancia en estación Regidor



Fuente: autor.

2.2.2 Selección de algoritmos de teledetección

Para seleccionar los algoritmos aplicables a las imágenes MODIS se realizó una revisión bibliográfica, junto con los disponibles en el paquete *waterquality* [46] de R, el cual cuenta con 45 algoritmos de detección de la calidad del agua derivados por satélites (Worldview 2, Sentinel 2, Landsat 8, MODIS y MERIS), utilizados para determinar los índices de Clorofila-a, Ficocianina y Turbiedad. Al explorar el paquete se encontraron 3 algoritmos aplicables a las imágenes MODIS, solamente para la detección de la Turbiedad; siendo TurbDox02NIRoverRed, que hace una relación de la banda 2 a una reflectancia de 857 nanómetros sobre la banda 1 a una reflectancia de 658 nanómetros, TurbHarr92NIR que selecciona una reflectancia de 857 nanómetros de la banda 2 y TurbMoore80Red que utiliza una reflectancia de 658 nanómetros de la banda 1 [46].

En lo que respecta a la Clorofila-a, fueron escogidos tres algoritmos diferentes. El primero fue log10 (Chlorophyll-a concentration), desarrollado por A. Germán et, en el artículo *Análisis de una serie temporal de clorofila-a a partir de imágenes MODIS de un embalse eutrófico* [69], siendo solicitado a los investigadores para ser evaluado con los datos del presente trabajo, el cual no pudo aplicarse debido a que fue ajustado con datos de campo de un embalse en particular (San Roque, Córdoba, Argentina). El segundo fue 2BDA (Two Band Algorithm) [47], que utiliza la relación de bandas entre el infrarrojo cercano NIR y el rojo visible RED. OSAVI Index (Optimized Soil- Adjusted Vegetation Index) [70], fue el tercero en ser evaluado, sin presentar valores que se ajustaran a los datos evaluados y NDVI (Normalized Difference Vegetation Index) [50], utilizado para Ficocianina. En la Tabla 3 se muestran los algoritmos utilizados con el índice que calculan, su ecuación y las bandas que usan.

Tabla 3. Algoritmos evaluados para la detección de los índices

ALGORITMOS			
Nombre	Índice	Ecuación	Bandas
TurbDox02NIRoverRed	Turbiedad	function (w857, w658) {(w857/w658)}	1,2
TurbHarr92NIR		function (w857,MODIS) { (w857)}	2
TurbMoore80Red		function (w658,MODIS) { (w658)}	1
log10 (Chlorophyll-a concentration)	Clorofila-a	$\text{Log } 10 (\text{Chl} - a) = 2.58 - 0.896 * \text{RED} / \text{NIR}$	1,2
2BDA (two-band-algorithm)		NIR/RED	1,2
OSAVI Index		$((\text{NIR}-\text{RED})/(\text{NIR}+\text{RED}+0.16))$	1,2
NDVI	Ficocianina	$((\text{NIR}-\text{RED})/(\text{NIR}+\text{RED}))$	1,2

Fuente: autor.

2.3 FASE 3. Evaluación de modelos para cuantificar determinantes

Primero, se graficaron diagramas de dispersión de varias variables en R para seleccionar los datos que tuvieran una mejor correlación. Estos permiten relacionar los valores de los determinantes de calidad del agua estudiados con los valores mínimos y medios de su índice correspondiente, mediante una regresión lineal y arrojan el valor de correlación de Pearson.

Luego, mediante metodología Boxplot se identificaron y eliminaron datos atípicos tanto en los determinantes de calidad estudiados, como en los índices calculados.

Una vez definidas las mejores correlaciones, se construye el modelo lineal simple, el modelo lineal de transformación logarítmica y el modelo lineal de múltiples variables para los tres determinantes mostrado en la Tabla 4, utilizando las funciones *lm* (*linear model*), encargada de ajustar modelos lineales y también realizar regresiones, análisis de covarianza y de varianza de un solo estrato [71]; y *nls* (*Nonlinear Least Squares*) que permite determinar las estimaciones no lineales (ponderadas) de mínimos cuadrados de los parámetros de un modelo [56] de R en el entorno RStudio.

Tabla 4. Estructura de los modelos utilizados

MODELOS	
Modelo lineal	lm (determinante ~ valores algoritmo, data= datos)
Modelo lineal de transformación logarítmica	nls (determinante ~ A * log (valores algoritmo) + B, data= datos)
Modelo lineal de múltiples variables	lm (determinante ~ algoritmo 1+ algoritmo 2+ algoritmo 3, data= datos)

Fuente: autor.

Después se realizaron las gráficas de diagnóstico incorporados en R, comprobando si existe o no linealidad, si hay una distribución normal de los residuos, su heterocedasticidad u homocedasticidad y los puntos de apalancamiento. Así mismo, se utiliza la estadística del modelo para comprobar que la mediana sea cercana a cero, los valores mínimos y máximos sean cercanos en valor absoluto y el p-value para determinar si la variable predictora y la intersección en los modelos son significativas, con el fin de rechazar la hipótesis nula y aceptar la hipótesis alternativa, indicando así una asociación aceptable entre el predictor y la variable resultado. Finalizado lo anterior, se seleccionan los modelos que pueden ser aplicados y se continua con la calibración y predicción.

- I. Primero se tomó una muestra con el comando *sample* de R, de acuerdo con la dimensión de la base de datos a trabajar, seleccionando 60% para calibrar y 40% para validar.

$$ii = \text{sample}(1 : \text{dim}(\text{datos})[1], \text{round}(\text{dim}(\text{datos})[1] * 0.6, 0), \text{replace} = \text{FALSE})$$

- II. Luego se crearon variables que tuvieran los datos de calibración y validación por separado, de los determinantes de calidad del agua y los algoritmos de teledetección, de acuerdo con las muestras seleccionadas anteriormente.

$$\begin{aligned} \text{Determinante}_{cal} &<- \text{datos}[ii, \text{columna de valores a trabajar}] \\ \text{Determinante}_{val} &<- \text{datos}[-ii, \text{columna de valores a trabajar}] \\ \text{Algoritmo}_{cal} &<- \text{datos}[-ii, \text{columna de valores a trabajar}] \\ \text{Algoritmo}_{val} &<- \text{datos}[ii, \text{columna de valores a trabajar}] \end{aligned}$$

- III. Después se construyó el modelo con las variables de calibración y validación creadas en el paso II, para determinar los coeficientes del modelo.

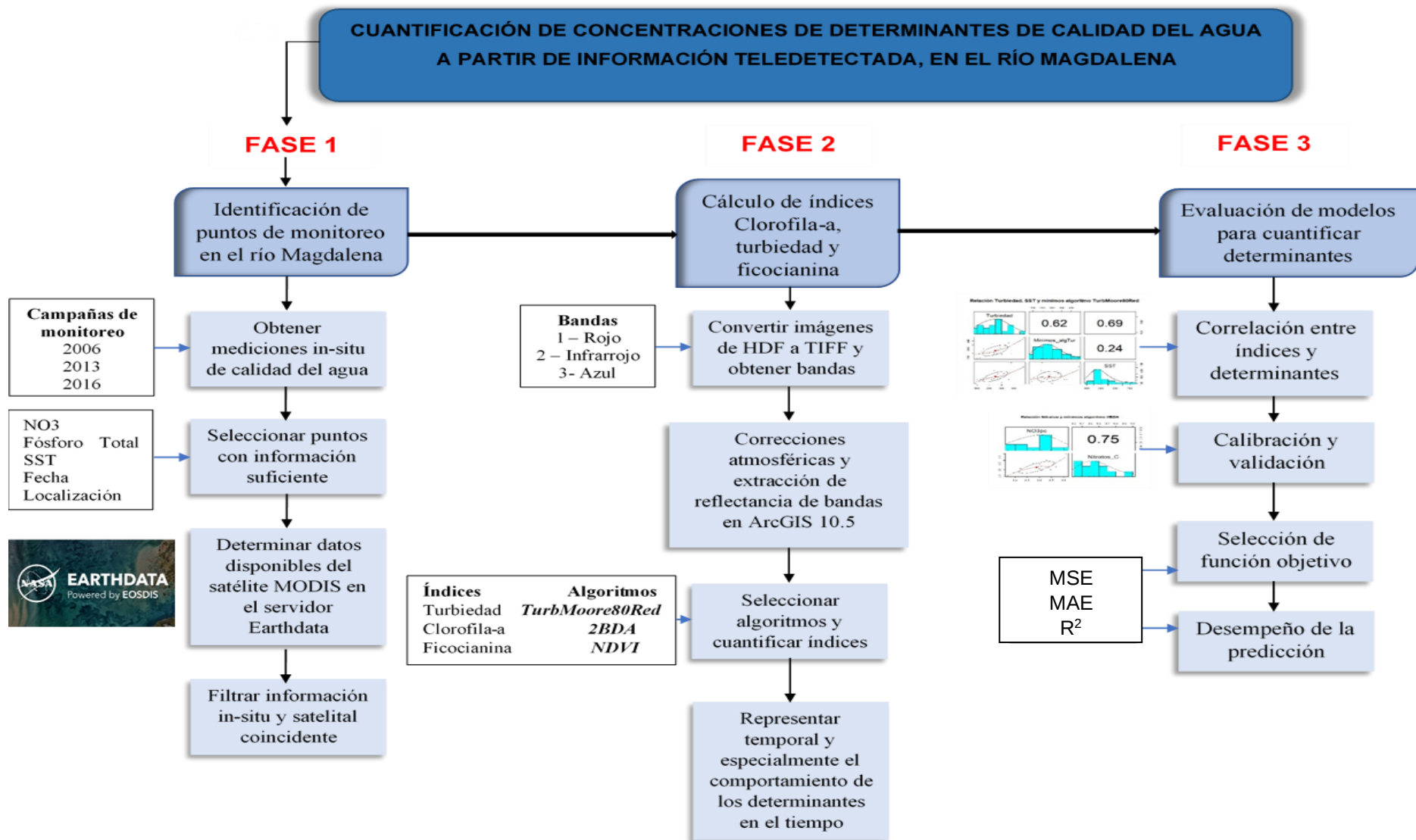
$$\text{modelo} <- \text{lm}(\text{Determinante}_{cal} \sim \text{Algoritmo}_{cal}, \text{data} = \text{datos})$$

- IV. Por último, se realizaron predicciones con la función *predict* de R, y posteriormente se validó el modelo con los valores de los coeficientes calculados.

$$\begin{aligned} \text{Det}_{pred} &<- \text{predict}(\text{modelo}, \text{data.frame}(\text{Algoritmo}_{cal})) \\ \text{Det}_{val} &<- (\text{coef}(\text{summary}(\text{modelo}))[2] * \text{Algoritmo}_{val}) \\ &\quad + (\text{coef}(\text{summary}(\text{modelo}))[1]) \end{aligned}$$

Por último, en esta fase se selecciona una función objetivo entre MSE, MAE y R^2 para evaluar el desempeño de la predicción y así seleccionar los modelos que tienen mejor ajuste de los datos.

Figura 13. Esquema de la metodología del proyecto por fases



Fuente: autor.

3. ANÁLISIS Y DISCUSIÓN DE RESULTADOS

En este apartado se muestran los resultados obtenidos al realizar la adquisición de datos tanto in-situ como teledetectados. La evolución temporal y espacial de los índices y determinantes en los puntos de monitoreo del río Magdalena. Finalizando con la evaluación de los modelos que pueden ser aplicados para la predicción y cuantificación de los determinantes de calidad del agua.

3.1 INFORMACIÓN IN-SITU Y SATELITAL RECOPIADA

3.1.1 Campañas y puntos seleccionados

Las autoridades ambientales IDEAM y CORMAGDALENA suministraron documentos técnicos finales de campañas de monitoreo de calidad y cantidad del agua realizadas en los años 2006, 2007, 2011, 2013, 2015 y 2016. De los cuales fueron seleccionados los correspondientes a los años 2006, 2013 y 2016, debido a que eran los únicos que contaban con las estaciones georreferenciadas, las fechas claras de cada muestreo y los determinantes necesarios para realizar la investigación.

Las campañas de los años 2013 y 2016 contaban con las mismas estaciones monitoreadas y un solo periodo de muestreo, en octubre para el 2013 y noviembre para el 2016. Para la campaña del año 2006, solo hay un punto monitoreo coincidente con los años anteriores llamado *Sitio Nuevo R-11* y tiene dos periodos de muestreos, el primero realizado en julio y el segundo en octubre. En la Tabla 5, se muestran los puntos de monitoreo seleccionados, junto con el rango de fecha de muestreo y determinantes. Para el Nitrógeno Total no se pudo realizar una evaluación debido a la escasez de datos, por tal razón fue utilizado el determinante Nitrato, presentando resultados favorables.

Tabla 5. Puntos de monitoreo seleccionados, con rango de fecha y determinantes

FECHA	ESTACIONES	CONTENIDO DE SÓLIDOS	NITRÓGENO		FÓSFORO
		SÓLIDOS SUSPENDIDOS TOTALES (mg SST/L)	NITRÓGENO TOTAL (mg N/L)	NITRATO (mg N-NO ₃ /L)	FÓSFORO TOTAL (mg P/L)
11/07/2006 ----- 14/10/2006	Magangué	130-440	NA	0.5-1	NA
	La Gloria				
	Sitio Nuevo R-11				
	Puerto Berrio				
	Magangué				
	La Gloria				
11/10/2013 ----- 20/10/2013	Sitio Nuevo R-11	128-630	0.76 -2.5	0.31-0.89	0.2-1.4
	Puerto Berrio				
	Magangué				
	La Gloria				
	Sitio Nuevo R-11				
	Puerto Berrio				
03/11/2016 ----- 08/11/2016	Desembocadura Río Negro	28-760	NA	0.17-0.59	0.12-0.72
	Desembocadura Río La Miel				
	Desembocadura Río Alicante				
	Desembocadura Río Carare				
	Desembocadura Río Sogamoso				
	Sitio Nuevo R-11				
	Regidor				
	Desembocadura Río Cesar				
	Desembocadura Río Negro				
	Desembocadura Río La Miel				
	Desembocadura Río Alicante				
	Desembocadura Río Carare				
	Desembocadura Río Sogamoso				
	Sitio Nuevo R-11				
	Regidor				
	Desembocadura Río Cesar				
	Desembocadura Río Negro				
	Desembocadura Río La Miel				

Fuente:[19][20][21]

Debido a que la mayoría de los informes suministrados por las autoridades no contaban con las estaciones georreferenciadas o las fechas en que se realizaba el muestreo, hubo una escasez de datos para poder realizar un análisis de mayor peso a lo largo del río Magdalena.

3.1.2 Datos satelitales del instrumento MODIS

De los productos MODIS encontrados se seleccionó el producto o imagen satelital Vegetation Indices 16-Day L3 Global, con ID MOD13Q1 del satélite Terra; ya que posee una resolución espacial de 250m, contiene la corrección atmosférica y geométrica incluida y las bandas necesarias para extraer los valores de reflectancia suficientes para realizar los cálculos.

En el **ANEXO C** se muestran los puntos de monitoreo de calidad del agua seleccionados, junto con sus fechas de monitoreos y se mencionan las imágenes satelitales disponibles encontradas del producto MOD13Q1, en formato HDF, descargadas del servidor EarthData [68], del área de estudio.

Se adquirieron tres productos satelitales para el año 2006, dos para el 2013 y uno para todo el periodo del 2016; los cuales tuvieron la resolución temporal coincidente con las fechas de muestreo de los ocho puntos de monitoreo de calidad del agua seleccionados.

Para el primer periodo de muestreo del año 2006, se utilizó una imagen satelital con la estación Magangué y otra imagen para las estaciones La Gloria, Sitio Nuevo R-11 y Puerto Berrio. Así mismo, para el segundo periodo del año fue utilizada una sola imagen para todas las estaciones.

Por otra parte, para el año 2013 hubo dos imágenes satelitales. La primera abarcó las mediciones realizadas el 11,13 y 15 de octubre, mientras que la otra las mediciones del 17, 19 y 20 del mismo mes.

3.2 ÍNDICES CALCULADOS

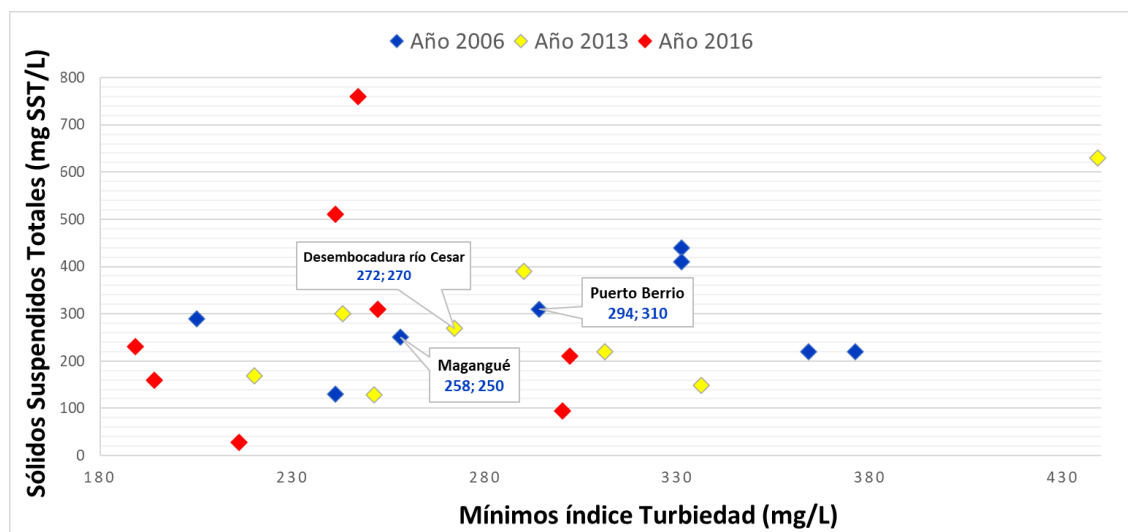
Los algoritmos seleccionados para calcular cada índice fueron, TurbMoore80Red para Turbiedad, 2BDA para Clorofila-a y NDVI para Ficocianina; descritos en el marco teórico. Los valores que arrojaron fueron utilizados para analizar la relación que tienen con los determinantes de calidad del agua objeto de estudio (Sólidos Suspendidos Totales, Nitratos, Fósforo Total).

Es este apartado se muestran diagramas de dispersión para representar la relación espacial y temporal de los índices calculados respecto a su determinante de calidad del agua en los años de estudio, utilizando los valores mínimos y medios de cada índice, evidenciando que los datos que tienen mayor relación con los determinantes de calidad del agua son los valores mínimos.

3.2.1 Evolución espacial y temporal de la Turbiedad en puntos de monitoreo del río Magdalena.

Se representaron los valores mínimos y medios de la Turbiedad calculada por el algoritmo TurbMoore80Red respecto al determinante Sólidos Suspendidos Totales para los años 2006, 2013 y 2016. En la Figura 14 se muestra el diagrama para los valores mínimos y en la Figura 15 para los valores medios.

Figura 14. Evolución espacial y temporal de valores mínimos Turbiedad respecto a Sólidos Suspendidos Totales

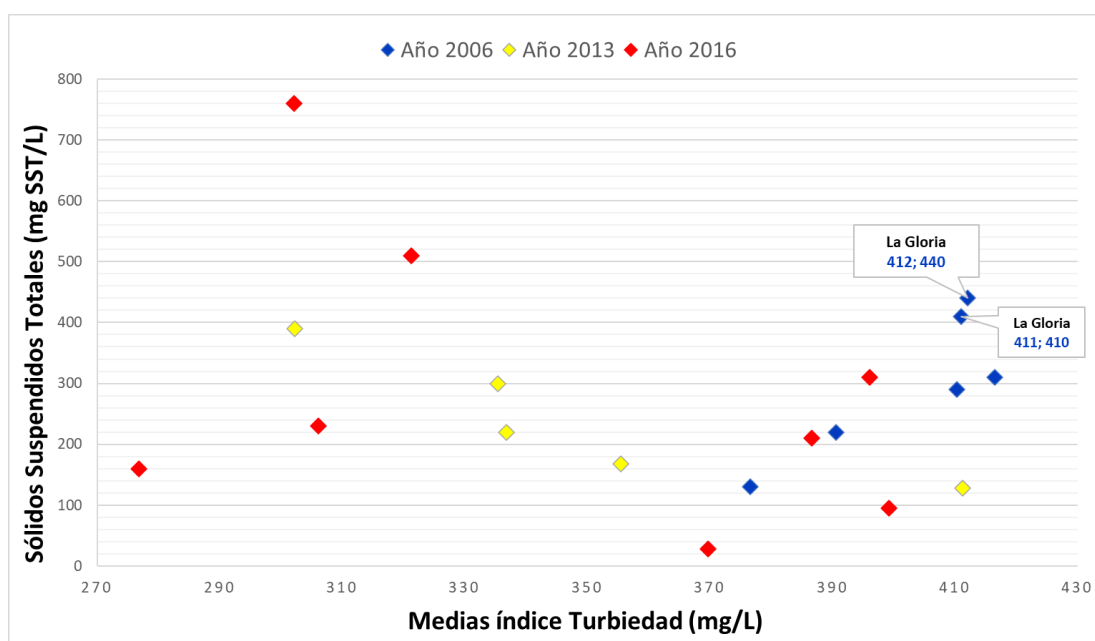


Fuente: autor.

Se observa que para el año 2006, el punto de monitoreo Magangué presenta valores cercanos de 250 (mg/L SST), respecto a 258 (mg/L) del valor mínimo del índice de Turbiedad. De igual manera, en el punto Puerto Berrio, el valor de Sólidos Suspendedos Totales corresponde a 310 (mg/L SST) y el del índice a 294 (mg/L)

En cuanto al año 2013, solamente el punto Desembocadura río Cesar tiene valores cercanos de 270 (mg/L SST) y 272 (mg/L) del índice Turbiedad. Para el 2016 no hay valores similares de Sólidos respecto a los mínimos del algoritmo.

Figura 15. Evolución espacial y temporal de valores medios Turbiedad respecto a Sólidos Suspendedos Totales



Fuente: autor.

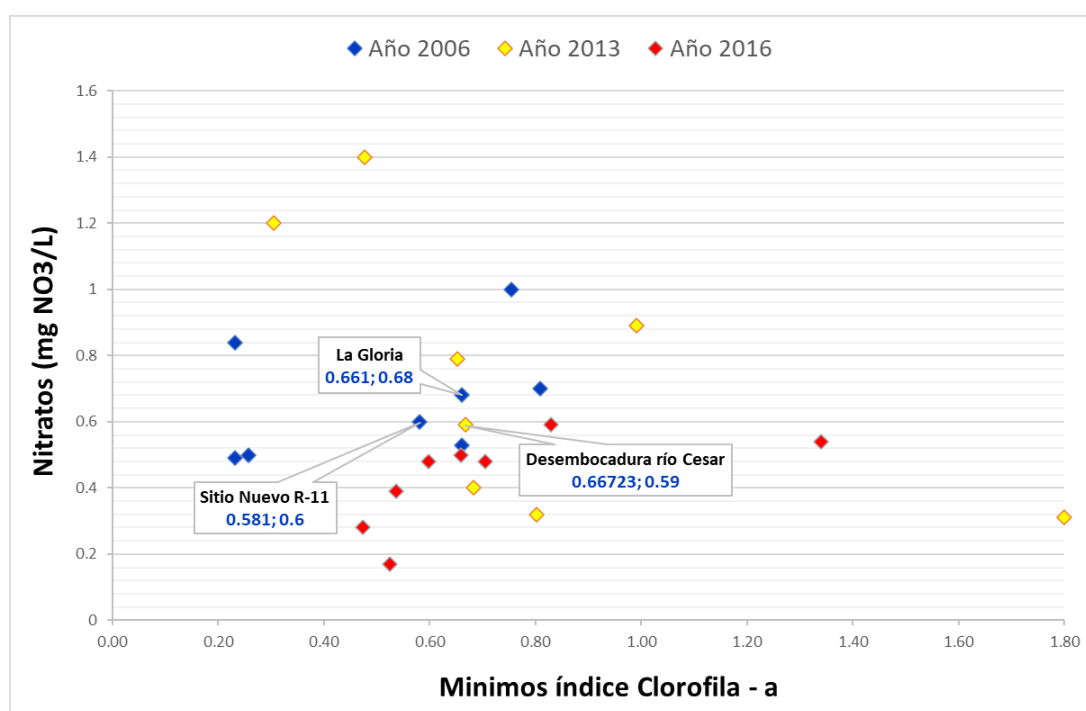
Para los años 2013 y 2016 no existen datos cercanos entre los valores medios calculados por el algoritmo TurbMoore80Red y los Sólidos Suspendedos Totales otorgados por las autoridades ambientales. Sin embargo, para el año 2006 se evidencia que, para los dos periodos de mediciones, el punto de monitoreo La Gloria tiene cercanía entre sus valores. Para el primer periodo correspondiente a julio, se presenta un dato de 440 (mg/L SST) respecto a 412 (mg/L) del valor medio de Turbiedad; mientras que, para el segundo periodo de muestreo en octubre de 2006, el dato de Sólidos Suspendedos Totales es de 410 (mg/L SST), con 411 (mg/L) del valor medio medido por el índice.

3.2.2 Evolución espacial y temporal de la Clorofila-a en puntos de monitoreo del río Magdalena.

El valor del algoritmo 2BDA fue calculado sin procesar, es decir sin ajustarlo a unas unidades específicas, esto es comúnmente utilizado para este algoritmo con el fin de dar simplicidad y facilitar su aplicación [72] [73].

Los valores mínimos y medios de la Clorofila calculada por el algoritmo 2BDA (Two Band Algorithm) respecto al determinante Nitrato para los años 2006, 2013 y 2016 se representaron en diagramas de dispersión. En la Figura 16 se muestra el diagrama para los valores mínimos del algoritmo 2BDA respecto a los valores del Nitrato en (mg/L NO₃).

Figura 16. Evolución espacial y temporal de valores mínimos Clorofila-a respecto a Nitratos



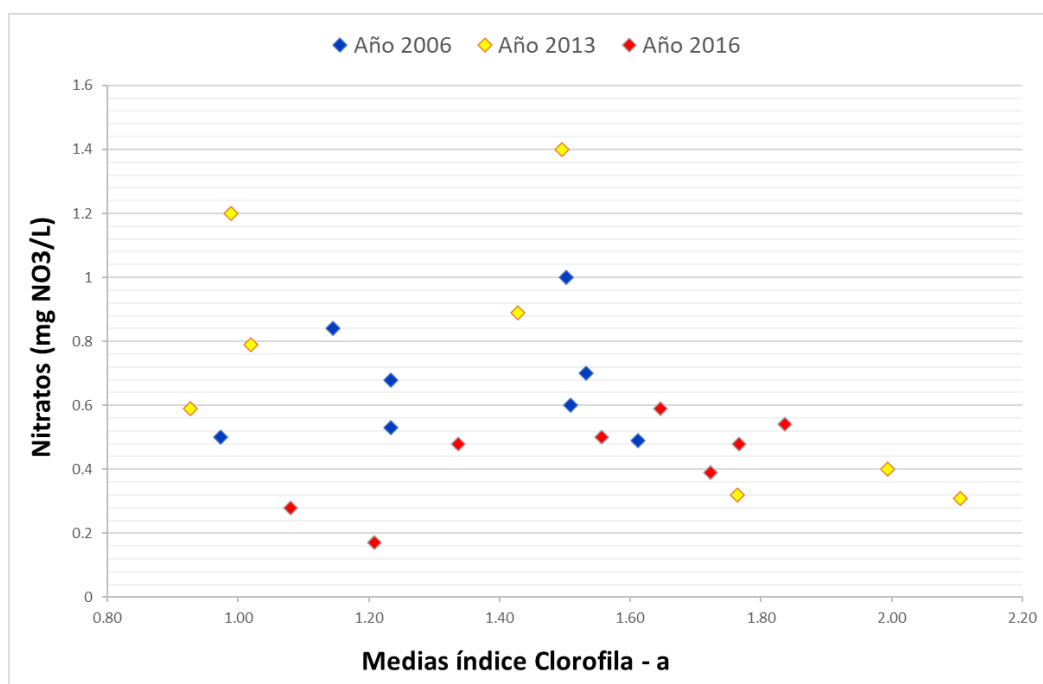
Fuente: autor.

El punto de monitoreo La Gloria en el año 2006 presentó un valor medio del índice de 0.66 (mg/L NO₃) respecto a un valor de 0.68 (mg/L NO₃) del determinante Nitrato. De igual manera, el punto de monitoreo Sitio Nuevo R-11 tuvo un valor de 0.58 (mg/L NO₃) para el índice Clorofila y 0.6 (mg/L NO₃) de Nitrato medido por la autoridad en el año 2006.

En el año 2013 solo el punto Desembocadura río Cesar tuvo valores cercanos de 0.66 (mg/L NO₃) para el índice Clorofila-a y 0.6 (mg/L NO₃) de Nitrato. En lo que respecta al año 2016 no se presentaron valores significativos.

En la Figura 17 se muestra el diagrama de dispersión para los valores medios del índice respecto a los valores de Nitrato proporcionados por la autoridad. Para estos valores no se presentaron datos similares en ninguno de los tres años, sin embargo, se representa para ver su comportamiento temporal.

Figura 17. Evolución espacial y temporal de los valores medios Clorofila-a respecto a Nitratos



Fuente: autor.

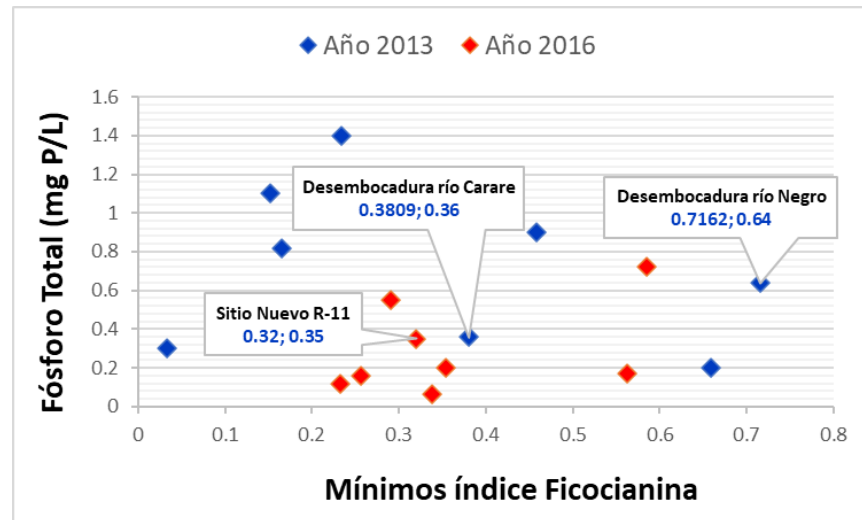
3.2.3 Evolución espacial y temporal de la Ficocianina en puntos de monitoreo del río Magdalena.

El cálculo de este índice fue realizado para los años 2013 y 2016, debido a que no había datos de Fósforo Total para el año 2006 en los informes entregados por las autoridades ambientales.

En la Figura 18 se representan los valores mínimos de Ficocianina respecto a los valores de Fósforo Total proporcionados por las autoridades. Se observa que para el año 2013 el punto Desembocadura río Carare presenta una similitud entre los datos calculados, siendo 0.38 (mg P/L) el arrojado por el índice y 0.36 (mg P/L) de Fósforo Total. Igualmente, para el punto de monitoreo Desembocadura río Negro se evidencia un valor de 0.716 (mg P/L) calculado por el índice y 0.64 (mg P/L) del determinante de calidad del agua.

Por otra parte, para el año 2016, el único punto que presentó valores similares fue Sitio Nuevo R-11, con un valor de 0.32 (mg P/L) del índice y 0.35 (mg P/L) de Fósforo Total.

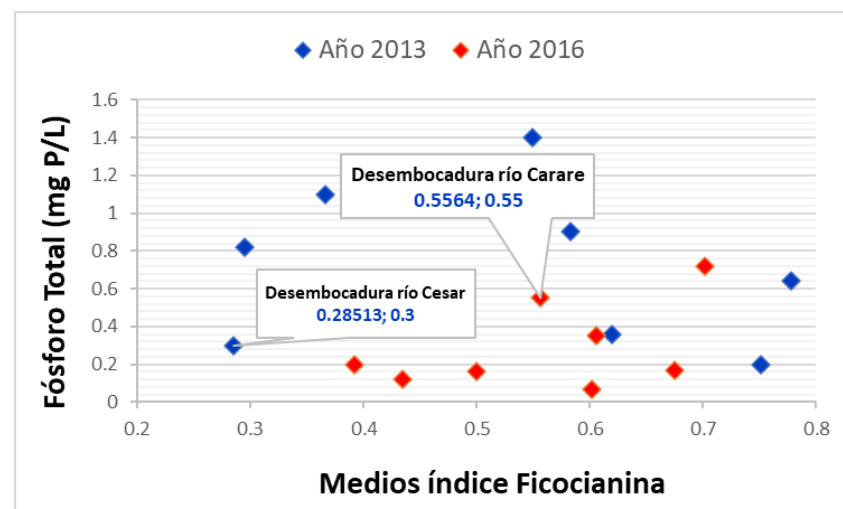
Figura 18. Evolución espacial y temporal de valores mínimos Ficocianina respecto a Fósforo Total



Fuente: autor

Para los valores medios del índice Ficocianina mostrados en la Figura 19, se tuvo para el punto de monitoreo Desembocadura río Carare 0.556 (mg P/L) medido por el índice con 0.55 (mg P/L) de Fósforo Total, esto en el año 2016. El otro punto significativo fue Desembocadura río Cesar, dando un valor del índice de 0.285 (mg P/L) respecto a 0.3 (mg P/L) del determinante para el 2013.

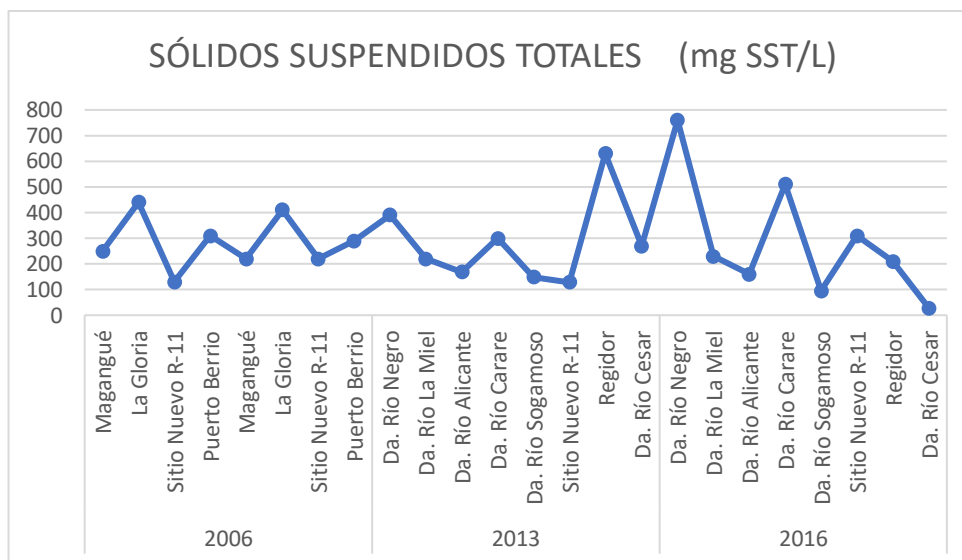
Figura 19. Evolución espacial y temporal de valores medios Ficocianina respecto a Fósforo Total



Fuente: autor

Se muestra comportamiento de los tres determinantes Sólidos Suspendidos Totales en la Figura 20, Nitratos en la Figura 21 y Fósforo Total en la Figura 22, durante los años 2006, 2013 y 2016.

Figura 20. Representación temporal y espacial de Sólidos Suspendidos Totales año 2006, 2013 y 2016



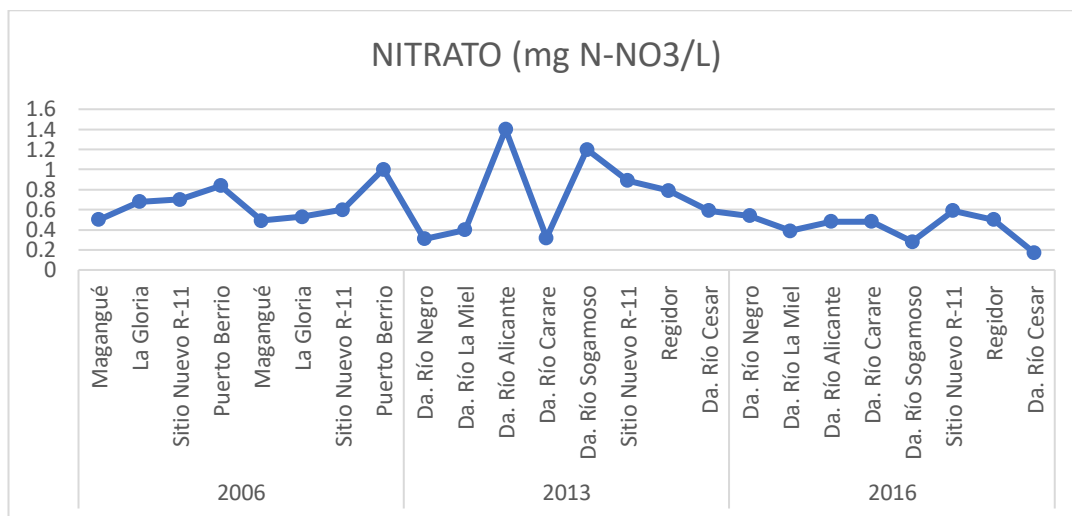
Fuente: autor.

Puede observarse que los niveles de sólidos en el año 2006 presentan una carga significativa superior a 200 (mg SST/L), lo que indica aguas contaminadas o con gran cantidad de partículas de material suspendido [74]. Solamente la estación Sitio Nuevo R-11 presenta condiciones aceptables teniendo un valor de 130 (mg SST/L).

Para el año 2013, hubo un valor extremadamente alto de 630 (mg SST/L), en la estación Regidor, indicando aguas fuertemente contaminadas posiblemente por descargas de aguas residuales al río, lo cual puede ser perjudicial para la economía basada en la piscicultura y agricultura, al no ser apta para su uso [74]. Nuevamente el punto Sitio Nuevo R-11 es el que presenta menor cantidad de sólidos.

La estación Desembocadura río Negro para el año 2016 fue la que mayor presencia de sólidos mostró durante los tres años evaluados, con una cantidad de 760 (mg SST/L), teniendo un nivel significativo de partículas aportadas al río Magdalena.

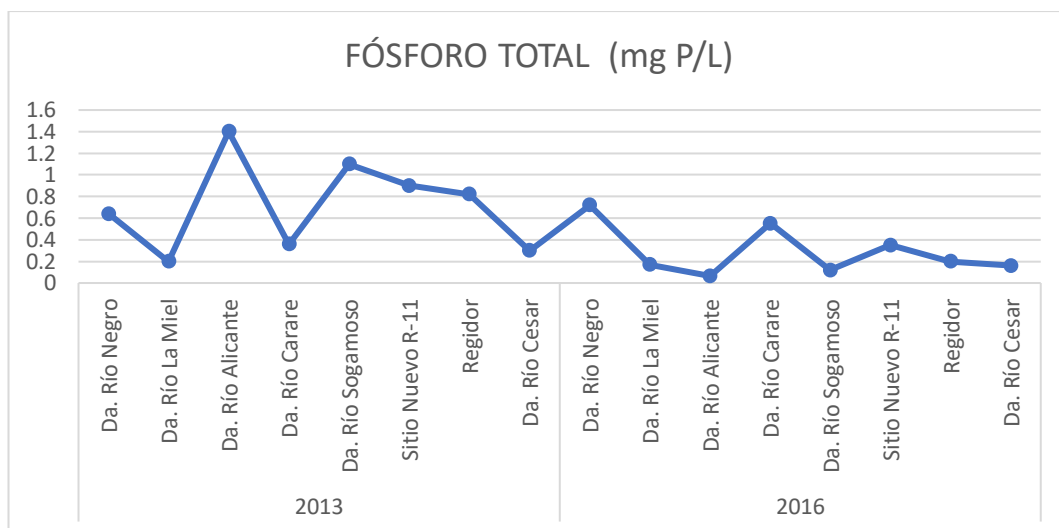
Figura 21. Representación temporal y espacial de Nitratos año 2006, 2013 y 2016



Fuente: autor.

El valor más representativo es 1.4 (mg N-NO₃/L) en la estación Desembocadura río Alicante para el año 2013; sin embargo, no es significativo para afectar la calidad del agua del río.

Figura 22. Representación temporal y espacial de Fósforo Total año 2013 y 2016



Fuente: autor.

Los niveles de Fósforo Total son relativamente bajos para los años 2013 y 2016, siendo Desembocadura río Alicante el más alto con 1.4 (mg P/L), por lo que puede decirse que en esta estación existió una actividad económica que aportó nutrientes, ya que también se ve presente en los Nitratos para el mismo año.

3.3 APLICACIÓN DE MODELOS

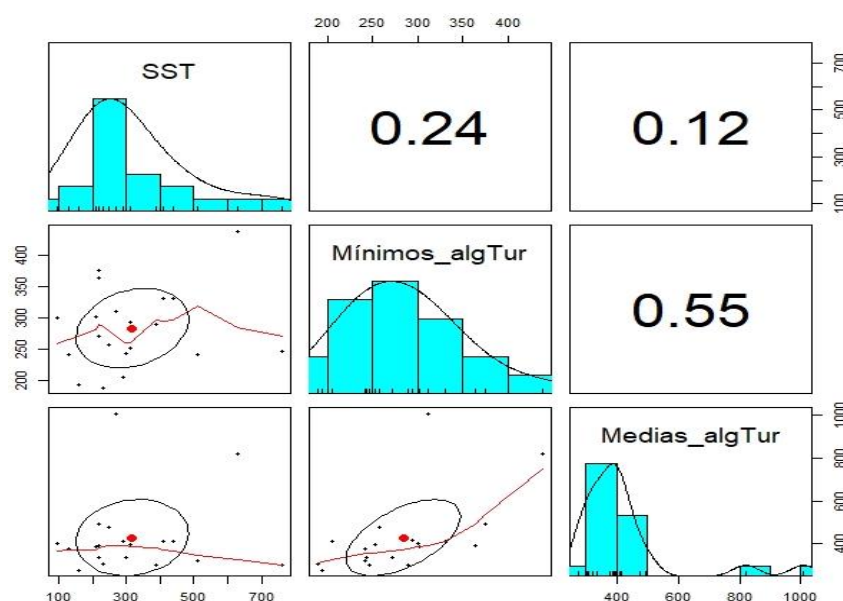
Se muestran las correlaciones realizadas entre los determinantes Sólidos Suspendidos Totales respecto al algoritmo TurbMoore80Red; Nitratos con el algoritmo 2BDA y el Fósforo Total con el índice NDVI. Luego, los datos atípicos encontrados para cada variable identificados mediante metodología Boxplot y finalmente los modelos que pudieron aplicarse. Para el Nitrógeno Total no se pudo realizar una evaluación debido a la escasez de datos, por tal razón fue utilizado el determinante Nitrato.

3.3.1 Correlaciones de índices respecto a determinantes

Las gráficas de varias variables mostradas en este apartado utilizan el coeficiente de correlación de Pearson para evaluar que tan relacionados linealmente se encuentran los datos medidos por la autoridad con los calculados por los algoritmos de teledetección, sin eliminación de atípicos. En la Figura 23 se muestra la distribución de los Sólidos Suspendidos Totales con relación a los valores mínimos y medios del índice Turbiedad.

Los datos de Sólidos Suspendidos Totales (SST) muestran una correlación baja de 0.24 respecto a los valores mínimos y 0.12 a los medios calculados por el algoritmo TurbMoore80Red, sin la eliminación de datos atípicos para ninguna de las variables. Este resultado dio un indicio de que la aplicación de un modelo lineal para estos datos no tendría un comportamiento favorable.

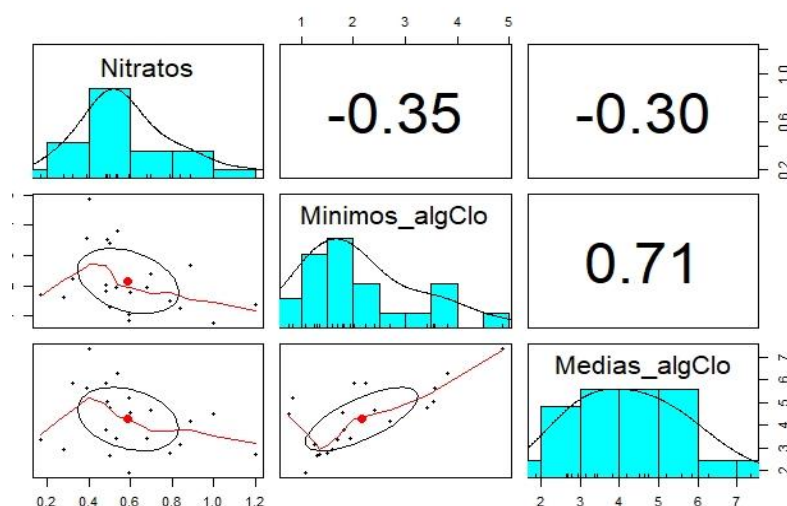
Figura 23. Distribución y correlación de los SST respecto a los valores de mínimos y medias del algoritmo TurbMoore80Red



Fuente: autor.

A diferencia del algoritmo para Turbiedad, el comportamiento de los valores de Nitratos respecto a los valores mínimos del algoritmo 2BDA, muestra una mayor correlación de Pearson negativa de -0.35 como se ve en la Figura 24, sin embargo, sigue siendo muy baja para considerarse aceptable. Por otro lado, se observa en el gráfico de varias variables de dispersión que el Nitrato no presenta una relación lineal con los mínimos del algoritmo, pero que si existe una relación entre estas dos variables.

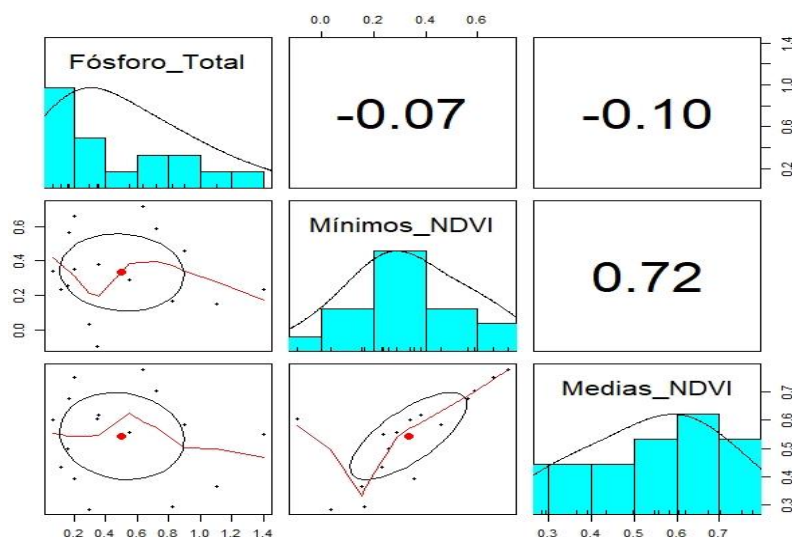
Figura 24. Relación Nitratos, mínimos y medias algoritmo 2BDA



Fuente: autor.

El índice NDVI respecto al Fósforo Total no tiene correlación para los valores mínimos y medios calculados como se muestra en la Figura 25.

Figura 25. Relación Fósforo Total, mínimos y medias índice NDVI



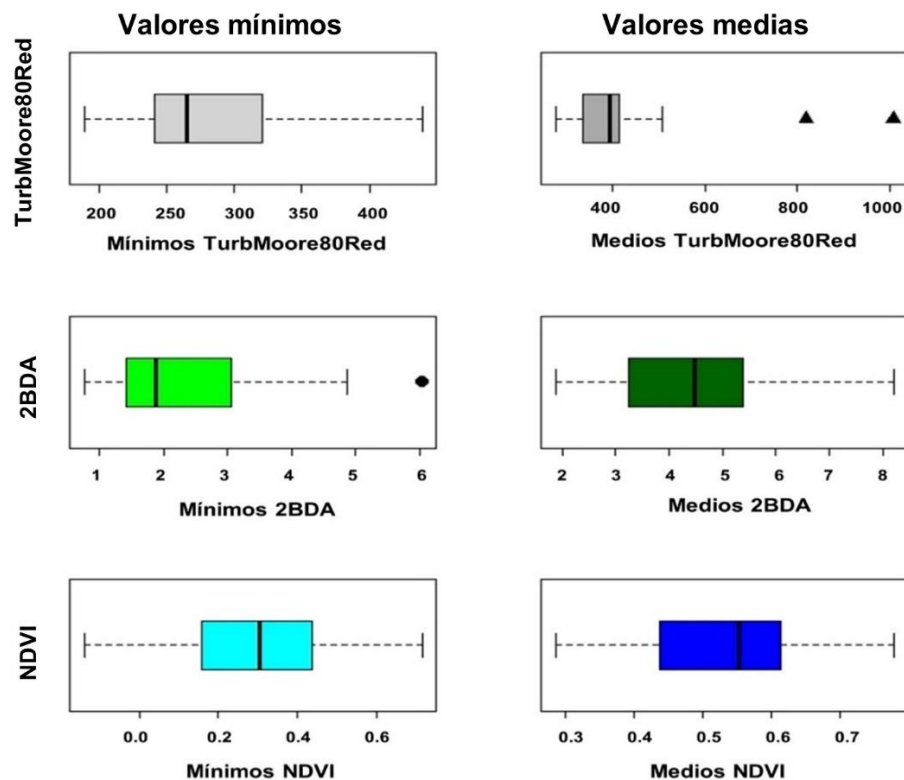
Fuente: autor.

Por otra parte, la mayoría de las relaciones no fueron buenas debido a que la resolución espacial de las imágenes MODIS a pesar de ser de 250 m, sigue siendo muy grande para analizar el comportamiento de la calidad en el río Magdalena. De igual manera, debido a que las estaciones de monitoreo se encuentran distribuidas espacialmente a lo largo del río, la información por cada punto no es suficiente, a diferencia de otros datos que son recolectados de un solo sitio.

3.3.2 Determinación de datos atípicos y correlaciones mejoradas.

Se eliminaron los dos valores atípicos para las medias del algoritmo TurbMoore80Red y el dato atípico de los mínimos 2BDA, identificados mediante los Boxplot como se ve en la Figura 26, para poder realizar correlaciones con los determinantes y ver la mejora que presentan.

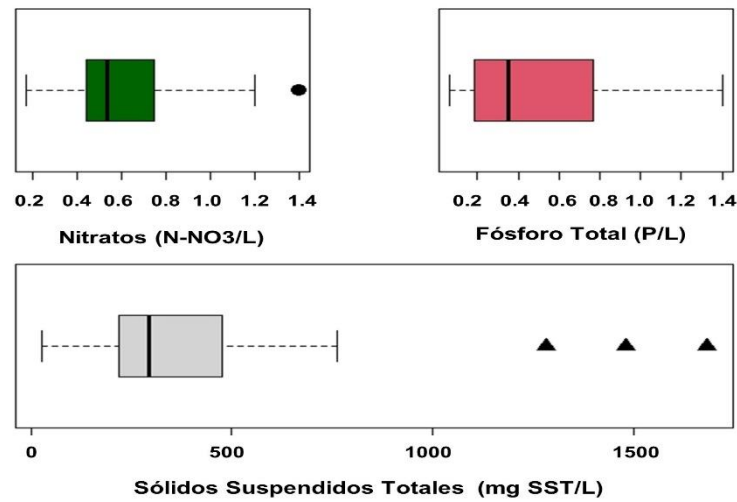
Figura 26. Identificación de datos atípicos mediante Boxplot para los índices evaluados



Fuente: autor.

Los Sólidos Suspendidos Totales muestran tres datos atípicos y el Nitrato solamente uno, para los valores de Fósforo Total no existen datos atípicos, como se ve en la Figura 27.

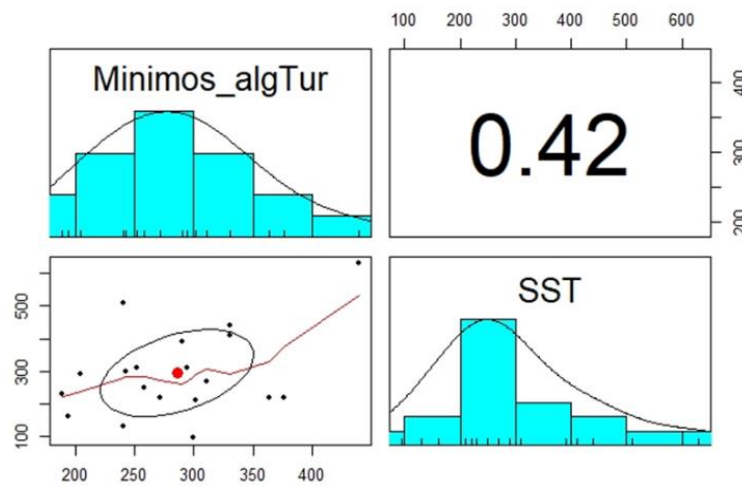
Figura 27. Identificación de datos atípicos mediante Boxplot para los determinantes evaluados



Fuente: autor.

Una vez eliminados los datos atípicos encontrados en los valores calculados por los algoritmos y los presentes en los determinantes, se encontró que la correlación de Pearson mejoró de 0.24 a 0.42 relación positiva moderada [75] como se ve en la Figura 28, entre las variables Sólidos Suspendidos Totales con mínimos del algoritmo de Turbiedad.

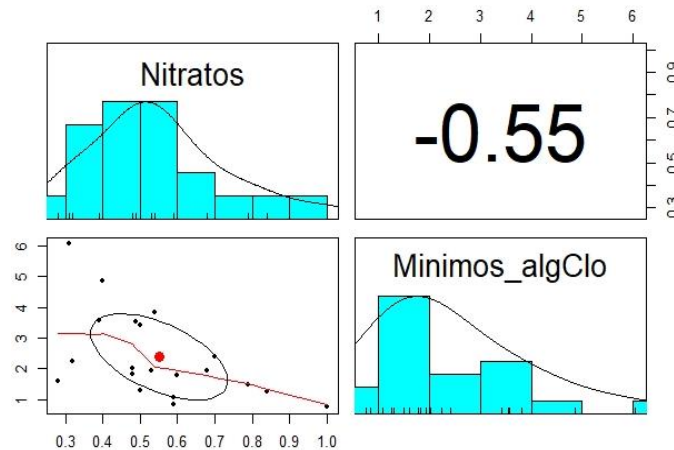
Figura 28. Relación mejorada de mínimos algoritmo TurbMoore80Red respecto a Sólidos Suspendidos Totales



Fuente: autor.

Entre los Nitratos respecto al algoritmo de identificación de Clorofila-a también hubo una mejora de -0.35 a -0.55 del coeficiente de correlación de Pearson mostrado en la Figura 29, considerándose una relación negativa modera [75].

Figura 29. Relación mejorada de mínimos algoritmo 2BDA respecto a Nitratos



Fuente: autor.

Las correlaciones mejorarían al tener más datos en un mismo punto para aplicar modelos de mayor peso que representen la dinámica espacial de los determinantes a lo largo del río Magdalena.

3.3.3 Evaluación de los modelos de regresión

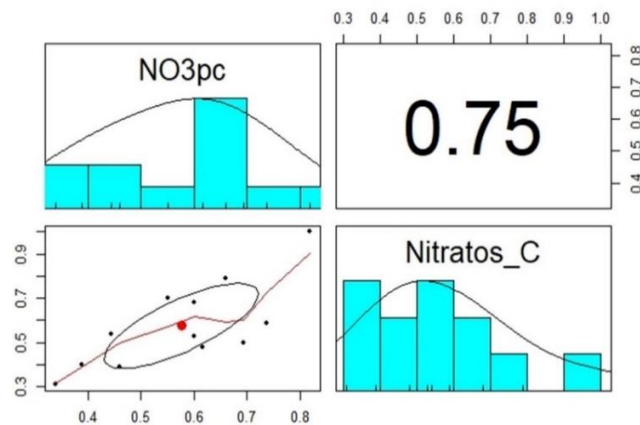
Se debe tener en cuenta que los algoritmos fueron utilizados para calcular los índices, que cumplen la función de ser las variables independientes, necesarias para estimar los determinantes de calidad del agua (variable dependiente), mediante la construcción de los modelos de regresión. Una vez construidos los modelos, se evaluó si pueden ser utilizados para cuantificar concentraciones de los determinantes de calidad del agua.

Los datos empleados fueron los valores mínimos de los índices, al ser los que presentaron mejor correlación con los determinantes de calidad del agua. Hay que resaltar que estos valores arrojados por los algoritmos no pudieron ser verificados, ya que no era el objeto de la investigación y no se tenía información de laboratorio de estos, en los datos proporcionados por las autoridades ambientales.

Al calibrar y aplicar los modelos lineales de regresión simple, lineales de transformación logarítmica y lineales de múltiples variables, se encontró que el modelo lineal de transformación logarítmica aplicado a los Nitratos en función de los valores mínimos del índice Clorofila-a parametrizados con logaritmo natural, mostró mejores resultados. Se obtuvo una correlación positiva de 0.75 entre los valores calibrados de Nitratos proporcionados por la autoridad (Nitratos_C) y los valores estimados de Nitratos (NO3pc) por el modelo, como se muestra en la Figura 30.

Si existiera una mayor cantidad de monitoreos en un mismo punto del río y la distancia de las estaciones no fuera tan grande, la calibración podría mejorar. También es importante mencionar que no fue posible realizar la validación ya que, al eliminar valores atípicos, la base de datos resultante fue muy pequeña

Figura 30. Relación de Nitratos calibrados con los estimados por el modelo lineal de transformación logarítmica



Fuente: autor.

Así mismo, al realizar su diagnóstico se verificó que los coeficientes A y B de las variables utilizadas en el modelo son significativos de acuerdo con los códigos de significancia y que el error residual estándar es cercano a cero como se muestra en la Figura 31, pudiendo rechazar la hipótesis nula y aceptar la hipótesis alternativa, indicando así una asociación no perfecta, pero si aceptable entre el predictor y la variable resultado, permitiendo aplicar el modelo.

Figura 31. Diagnóstico estadístico del modelo lineal de transformación logarítmica para Nitratos

```
Formula: Nitratos ~ A * log(Minimos_algClo, base = exp(1)) + B

Parameters:
  Estimate Std. Error t value Pr(>|t|)
A -0.19369   0.06116  -3.167  0.00534 **
B  0.68936   0.05535  12.455 2.77e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

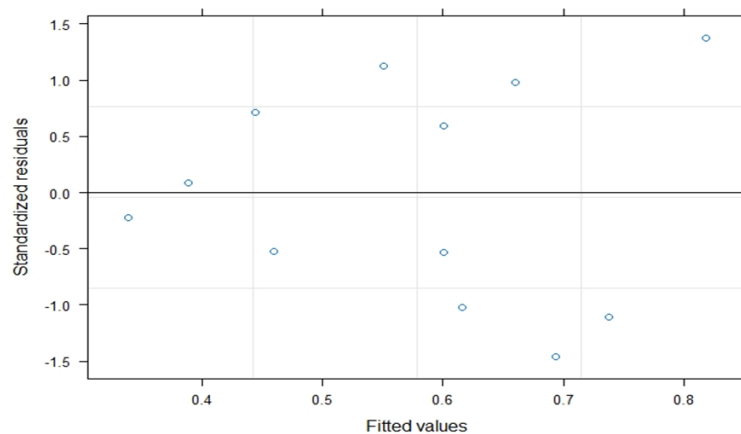
Residual standard error: 0.1511 on 18 degrees of freedom

Number of iterations to convergence: 1
Achieved convergence tolerance: 1.367e-07
```

Fuente: autor.

De igual manera, se realizó la gráfica de residuos, la cual presentó un comportamiento disperso, indicando que los residuos se distribuyen con normalidad mostrado en la Figura 32.

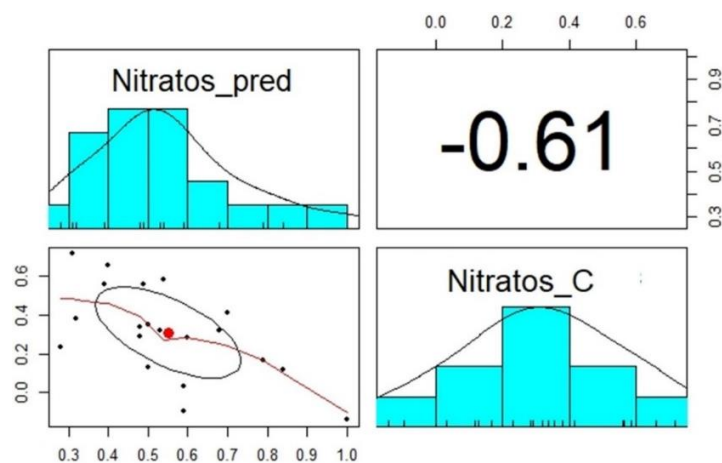
Figura 32. Gráfico de residuos de modelo lineal de transformación logarítmica para Nitratos



Fuente: autor.

Por otra parte, al ser los Nitratos los datos que tuvieron mejor correlación con los valores calculados por los algoritmos, se realizó un modelo lineal para determinar la correlación de los nitratos calibrados (Nitratos_C), en función de los valores del índice Ficocianina estimados por el modelo (Nitratos_pred), evidenciando una correlación negativa moderada de -0.61 entre los valores calibrados y estimados mostrado en la Figura 33.

Figura 33. Relación de Nitratos calibrados con los estimados por el modelo lineal



Fuente: autor.

Se realizó el diagnóstico estadístico, teniendo como resultado que los valores mínimos y máximos son cercanos en valor absoluto, que la media es cercana a cero y que los coeficientes del modelo son significativos de acuerdo con los códigos de significancia. De igual manera se evidencia un $R^2 = 0.3794$ con un p-value < a 0.05 como se muestra en la Figura 34.

Figura 34. Diagnóstico estadístico del modelo lineal para Nitratos

```
Call:
lm(formula = Nitratos_C ~ Minimos_algFic_C, data = datosatipT)

Residuals:
    Min       1Q   Median       3Q      Max
-0.247090 -0.021593 -0.002831  0.031292  0.264225

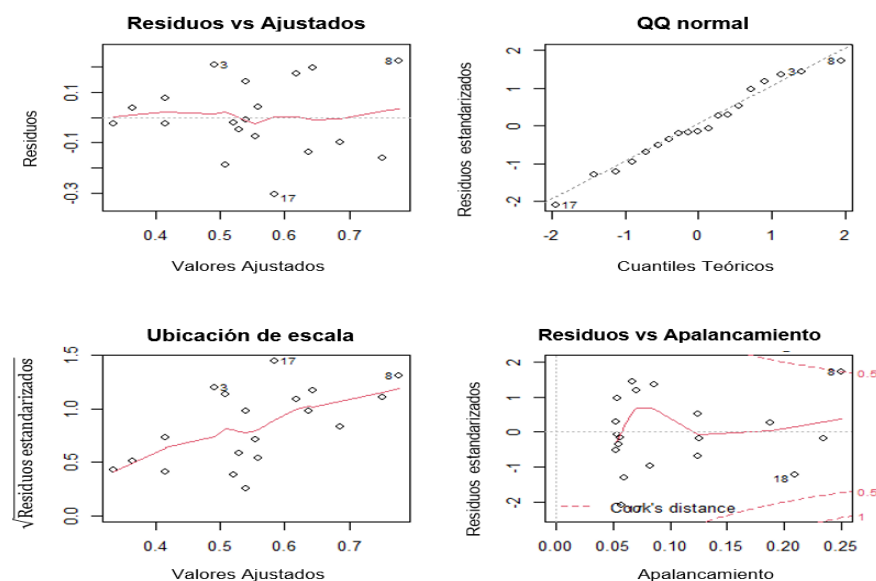
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.6241     0.0715   8.729 1.1e-05 ***
Minimos_algFic_C -0.4175     0.1780  -2.346  0.0436 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1255 on 9 degrees of freedom
Multiple R-squared:  0.3794,    Adjusted R-squared:  0.3104
F-statistic: 5.502 on 1 and 9 DF,  p-value: 0.04362
```

Fuente: autor.

De las gráficas de diagnóstico presentadas en la Figura 35 se puede decir que existe linealidad de los datos, que los errores residuales se distribuyen normalmente y que presenta un punto de alto apalancamiento. También, que presenta heterocedasticidad.

Figura 35. Gráficos de diagnóstico del modelo lineal para Nitratos



Fuente: autor.

A pesar de que el modelo presenta coeficientes significativos y linealidad de los datos, no puede ser aplicado debido a que su p-value es muy cercano a 0.05 y un R^2 muy pequeño, lo que puede arrojar errores de gran magnitud al realizar predicciones con dicho modelo. De igual manera, de acuerdo con el gráfico de diagnóstico, los residuos no tienen una varianza constante, es decir presentan heterocedasticidad.

3.3.3.1 Métrica de desempeño R^2

La selección de la métrica coeficiente de determinación R^2 se debió a que es una medida adimensional fácil de calcular e interpretar al mostrar la bondad de ajuste o cómo se ajustan los datos observados con los estimados, expresado mediante un porcentaje [76]. Para que su interpretación fuera la correcta se revisaron los residuos en el diagnóstico, para evitar incurrir en afirmaciones falsas [77]. Las métricas MSE y MAE no fueron usadas debido a que presentan limitaciones por la escasez de datos utilizados, lo que no permite captar el comportamiento real del error para evaluar los modelos de regresión [66]. Así también, es preferible utilizar las métricas MSE y MAE para comparar el rendimiento entre diferentes modelos de regresión, para que el valor de estas métricas tenga una explicación adecuada [78] [79].

Se pudieron observar valores cercanos a cero en el cálculo de los coeficientes de determinación para todos los modelos aplicados a los determinantes Sólidos Suspendidos Totales y Fósforo Total. Estos indican que no existe un ajuste de los datos teledetectados con los observados, o que existe un 0% de explicación del modelo a los datos en torno a su media [80], es decir que los modelos no siguen la tendencia de los datos [81].

El modelo de transformación logarítmica aplicado para la cuantificación de nitratos a partir de los valores mínimos del índice Clorofila-a, tuvo un $R^2 = 0.5625$, indicando que es el modelo con mejor ajuste de los datos estimados respecto a los observados, en otras palabras, el modelo explica la variable respuesta Nitrato un 56%.

Por otra parte, el modelo de regresión lineal simple aplicado para la cuantificación de nitrato en función de los valores mínimos del índice Ficocianina, arrojó un $R^2 = 0.379$, lo que indica que el modelo explica la variable nitrato en un 37%. A pesar de que fue el segundo mejor modelo, no es apto para aplicar debido a que presenta un coeficiente de determinación muy bajo y por sus características estadísticas mencionadas en su diagnóstico.

En la Tabla 6 se muestran los modelos aplicados para cada determinante, con el algoritmo utilizado e índice calculado por este, junto con la métrica de desempeño *coeficiente de determinación* R^2 ; resaltando los modelos que tuvieron un mejor comportamiento para su aplicación en la detección del determinante Nitrato.

Tabla 6. Desempeño de predicción para cada modelo utilizado

DETERMINATE	ALGORITMO / ÍNDICE	MODELOS	METRICA DE EVALUACIÓN
			R ²
Sólidos Suspendidos Totales SST	TurbMoore80Red / Turbiedad	modelo1 <- lm (SST~Minimos_algTur)	0.1764
		modelo1nlslog <- nls (SST~A*log(Minimos_algTur)+B, star=c(A=0.003,B=0.0001),control=nls.control(warnOnly = TRUE))	El modelo no puede aplicarse
		modSSTvv <- lm(SST ~ Minimos_algTur + Minimos_algClo + Minimos_algFic)	0.1313
Nitrato NO3-	2BDA / Clorofila-a	modNO3lm <- lm(Nitratos~Minimos_algClo)	0.3141
		modNO3log <- nls(Nitratos~A*log(Minimos_algClo, base = exp(1))+B, star=c(A=0.003,B=0.0001),control=nls.control(warnOnly = TRUE))	0.5625
		modNO3vv <- lm(Nitratos ~ Minimos_algFic+Minimos_algClo+Minimos_algTur)	0.05343
	NDVI / Ficocianina	modNO3Fic <- lm(Nitratos~Minimos_algFic, data = datosatipT)	0.3794
Fosforo P	NDVI / Ficocianina	modPlm <- lm(P_Total ~ Min_algFic)	0.06585
		modPnls <- nls(P_Total ~ A*log(Min_algFic, base = exp(1))+B, start = c(A=0.003,B=0.0001), control = nls.control(warnOnly = TRUE))	El modelo no puede aplicarse
		modPvv <- lm(P_Total ~ Min_algFic + Min_algClo+Min_algTur)	0.1055

Fuente: autor.

4. IMPACTO SOCIAL Y HUMANÍSTICO DEL PROYECTO

El trabajo tiene un aporte a la sociedad ya que busca mejorar el seguimiento y control del recurso hídrico superficial de Colombia debido a que la información disponible es nula o escasa, en lugares que por el conflicto armado [3], difícil acceso, baja capacidad operativa de entidades ambientales o costos, no es posible caracterizar las dinámicas de la calidad del agua. De esta manera, el uso de información teledetectada permitirá estimar la presencia y magnitud de los determinantes en el agua, orientando la toma de decisiones en la gestión del recurso hídrico, disminuyendo costos y a su vez, contribuyendo en el avance y cumplimiento del tercer Objetivo de Desarrollo Sostenible que busca la salud y bienestar de las personas, junto con el sexto que se enfoca en la disponibilidad de agua para todos, saneamiento y gestión sostenible de este recurso; ya que el no cumplir con estos influye tanto en la seguridad alimentaria, como en la educación para familias más necesitadas y la economía del país.

5. CONCLUSIONES

- De las seis campañas de monitoreo de calidad del agua lo largo del río Magdalena, suministradas por las autoridades ambientales CORMAGDALENA e IDEAM, solo fue posible utilizar las correspondientes a los años 2006, 2013 y 2016, al ser las que contaban con datos y fechas de muestreos, y con sus puntos de monitoreo georreferenciados. Las campañas realizadas en los años 2007, 2011 y 2015 eran informes técnicos sin los datos recolectados en los monitoreos, sin las fechas de muestreos y sin los puntos georreferenciados, por lo cual no pudieron ser utilizados para esta investigación.
- En el año 2006, de los 18 puntos de monitoreo de calidad del agua, se seleccionaron 4, cada uno con dos periodos de muestreos, para un total de 8 datos de concentraciones de Nitrato y 8 de Sólidos Suspendidos Totales, debido a que no hubo datos de Fósforo Total. Para el año 2013, de los 20 puntos de monitoreo fueron seleccionados 8, dando un total de 24 datos, correspondientes a 8 muestreas de Sólidos Suspendidos Totales, 8 de Nitratos y 8 de Fósforo Total. Por último, se utilizaron 8 puntos de monitoreo de los 23 que había en el año 2016, para un total de 24 datos. Al realizar esta selección, la zona hidrográfica donde se enfocaron los puntos fue el medio magdalena.
- Debido a la escasez de datos suministrados por las corporaciones ambientales, no fue posible realizar un análisis de mayor peso y modelos con mejores resultados, ya que los puntos de monitoreo se encuentran muy distribuidos espacialmente a lo largo del río Magdalena y no hay los suficientes datos de muestreos en un punto específico. Así mismo, no fue posible realizar la validación ya que, al eliminar valores atípicos, la base de datos resultante fue muy pequeña.
- Se determinó que la información de teledetección proporcionada por MODIS es una herramienta de adquisición de datos más sencilla y fue seleccionada al ser la que contaba con mayor cantidad de información disponible y por ser gratuita.
- La recolección de datos por satélites es limitada en ocasiones a nivel óptico por las nubes y por píxeles sin valor; a pesar de esto, se evidenció que pueden existir más imágenes que monitoreos in situ, por lo cual se podría analizar un momento en un año con el uso de los índices calculados y los modelos regresivos, permitiendo caracterizar la dinámica de la calidad del agua en lugares donde nunca se ha tenido acceso a información o no se ha podido realizar monitoreo.
- Los satélites WorldView-2 y 3, Sentinel-2 y Landsat-8, incluidos en los algoritmos del paquete waterquality cuentan con una resolución espacial más fina (es decir, tamaño de pixel más pequeño) que la de MODIS. Sin embargo, no contaban con imágenes satelitales para las fechas de muestreos realizados por las autoridades ambientales, por lo cual no pudieron utilizarse y algunas de

ellas debían ser pagadas para su adquisición como las de WorldView y otras de Sentinel.

- Los sistemas de teledetección traen beneficios en aspectos como: reducción de costos, al adquirir la información necesaria para extraer los parámetros de calidad en el agua directamente de las imágenes satelitales; también en tiempos de monitoreo, debido a que proporciona una mayor periodicidad y frecuencia de datos, lo que permite que haya un control más eficaz por parte de los organismos encargados de realizar la actividad de seguimiento a los cuerpos de agua, pudiendo así manejar y mitigar los impactos que se generan en el recurso hídrico.
- La Clorofila-a fue el índice que obtuvo mejores resultados con sus valores mínimos, al presentar una correlación de Pearson moderada de -0.55 respecto al determinante de calidad del agua Nitrato. Para los valores mínimos del índice Turbiedad, se obtuvo una correlación de Pearson baja de 0.42 respecto Sólidos Suspendidos Totales. En lo que respecta a los valores mínimos de la Ficocianina, existió una correlación nula de -0.07 con el Fósforo Total.
- El modelo lineal de transformación logarítmica para la cuantificación del determinante Nitrato fue el modelo que tuvo mejor rendimiento de acuerdo con el coeficiente de determinación, explicando la variable respuesta Nitrato en un 56%. Por otra parte, los modelos de regresión lineal simple, de transformación logarítmica y de múltiples variables, evaluados para cuantificar Sólidos Suspendidos Totales y Fósforo Total, presentaron valores < 0.2 del coeficiente de determinación, concluyendo que no pueden ser aplicados para la cuantificación de los determinantes mencionados.

6. RECOMENDACIONES

- En futuras investigaciones es recomendable utilizar imágenes de satélites que cuenten con mayor resolución espacial para obtener un valor de reflectancia con más exactitud y así mejores resultados.
- También es pertinente medir más variables como Redox, pH, conductividad para ampliar el esquema de posibilidades de combinar mediciones.
- El realizar el estudio en cuerpos de agua con mayor tamaño tales como un embalse, lagos o lagunas son favorables porque se obtienen mayor cantidad de valores de reflectancia, ya que las imágenes van a tener un área con más píxeles. Esto también permite utilizar bandas de menor resolución espacial (500 y 1000 metros) y por ende más algoritmos de teledetección.
- Existen muchos más algoritmos aplicables a los índices calculados, los cuales pueden ser evaluados en futuras investigaciones. De igual manera, hay diferentes índices que pueden ser aplicados para la detección de estos determinantes.
- Las autoridades ambientales encargadas de la toma de datos de calidad del agua pueden interesarse aún más en la implementación de las tecnologías satelitales, y realizar investigaciones de mejor calidad al contar con mayores fondos y personal especializado.

7. REFERENCIAS

- [1] S. B. Schroeder et al, "Passive remote sensing technology for mapping bull kelp (*Nereocystis luetkeana*): A review of techniques and regional case study," *Global Ecology and Conservation*, vol. 19, pp. e00683, 2019. Available: <https://www.sciencedirect.com/science/article/pii/S2351989419301076>. DOI: 10.1016/j.
- [2] United Nations, "About the Sustainable Development Goals," 2015. Available: <https://www.un.org/sustainabledevelopment/sustainable-development-goals/>.
- [3] D. Quevedo and N. Walteros, *El Recurso Hídrico En El Conflicto Armado En Colombia, Y Su Relevancia En El Proceso De Posconflicto Y Construcción De Paz*. (Pachón, Maria ed.) Bogotá: Universidad Externado de Colombia, 2019.
- [4] UICN (2018). *Guía de Monitoreo Participativo de la Calidad de Agua*. Quito – Ecuador: UICN.
- [5] M. Labrador, J. Évora and M. Pérez, *Satélites De Teledetección Para La Gestión Del Territorio*; 2012.
- [6] Novo, Evlyn Márcia Leão de Moraes et al, "Proposal for a remote sensing trophic state index based upon Thematic Mapper/Landsat images," *Revista Ambiente & Água*, vol. 8, (3), pp. 65-82, 2013. Available: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1980-993X2013000300006&lng=en&tlng=en. DOI: 10.4136/ambi-agua.1229.
- [7] R. Aguirre Gómez, "Análisis espectral del Lago de Guadalupe, mediante imágenes de satélite y datos in situ," *Investigaciones Geográficas, Boletín Del Instituto De Geografía*, vol. 2015, (86), pp. 15-24, 2015. Available: <https://www.sciencedirect.com/science/article/pii/S0188461115300030>. DOI: 10.14350/rig.42338.
- [8] T. W. P. Rodrigues et al, "DELINEAMENTO AMOSTRAL EM RESERVATÓRIOS UTILIZANDO IMAGENS LANDSAT-8/OLI: UM ESTUDO DE CASO NO RESERVATÓRIO DE NOVA AVANHANDAVA (ESTADO DE SÃO PAULO, BRASIL)," *Boletim De Ciências Geodésicas*, vol. 22, (2), pp. 303-323, 2016. Available: https://www.openaire.eu/search/publication?articleId=dedup_wf_001::434fef22a269f33da6d9aee7e0bbf8f. DOI: 10.1590/S1982-21702016000200017.
- [9] C. M. Rudorff, Novo, Evlyn M. L. M and L. S. Galvão, "Análise derivativa de dados hiperespectrais medidos em nível de campo e orbital para caracterizar a composição de águas opticamente complexas na Amazônia Derivative analysis of hyperspectral data measured at field and orbital level to characterize the composition of optically complex waters in the amazon," *Acta Amazonica*, vol. 37, (2), pp. 269-280, 2007. Available: <https://www.openaire.eu/search/publication?articleId=doajarticles::a91edbf97c1f0f347b64fd7c9e3e6a7>. DOI: 10.1590/S0044-59672007000200014.

- [10] Alexander Perdomo Vanegas, "Predicción de parámetros físico químicos de calidad del agua mediante el uso de sensores remotos: caso de estudio embalse del Neusa," 2015.
- [11] Gómez Díaz, J.L., & Delance-Martinic, J.S. (2014). Determinación del parámetro sólidos sus-pendidos totales (sst) mediante imágenes de sensores ópticos en un tramo de la cuenca media del río Bogotá (Colombia). *UD y la Geomática*, (9), pp 19-27.
- [12] I. Carrillo and J. Medina, "Suspended Solids in the Gulf of Urabá Colombia – Annual Average Estimation Using MODIS MYD09Q1 Images," *Trends and Advances in Information Systems and Technologies*, pp. 88-97, 2018. Available: https://link-springer-com.crai-ustadigital.usantotomas.edu.co/chapter/10.1007/978-3-319-77712-2_9. DOI: 10.1007/978-3-319-77712-2_9.
- [13] L. C. González-Márquez et al, "Use of LANDSAT 8 images for depth and water quality assessment of El Guájaró reservoir, Colombia," *Journal of South American Earth Sciences*, vol. 82, pp. 231-238, 2018. Available: <https://www.sciencedirect.com/science/article/pii/S0895981117303802>. DOI: 10.1016/j.jsames.2018.01.004.
- [14] F. Torres-Bejarano et al, "La modelación hidrodinámica para la gestión hídrica del embalse del Guájaró, Colombia," *Revista Internacional De Métodos Numéricos Para Cálculo Y Diseño En Ingeniería*, vol. 32, (3), pp. 163-172, 2016. Available: <https://www.sciencedirect.com/science/article/pii/S0213131515000310>. DOI: 10.1016/j.rimni.2015.04.001.
- [15] F. Sacristán Romero, "La teledetección satelital y los sistemas de protección ambiental," *AquaTIC: Revista Electrónica De Acuicultura*, (24), pp. 13-41, 2006. Available: <http://dialnet.unirioja.es/servlet/oaiart?codigo=2041195>.
- [16] Convenio especial de cooperación en ciencia y tecnología No., 1-0016-2016., "Aunar esfuerzos técnicos para la realización de una campaña de monitoreo de cantidad y calidad de agua en el área hidrográfica del río magdalena-cauca y sus tributarios, en el marco de la política nacional para la gestión integral del recurso hídrico (GIRH) y el plan de acción trianual de CORMAGDALENA." Feb, 2017.
- [17] ORGANIZACIÓN COLPARQUES , 2020. Lagunas de La Magdalena y Santiago. Available: <http://www.colparques.net>.
- [18] C. Garay et al, "Campaña de monitoreo de calidad de aguas de la cuenca Magdalena-Cauca," 2016.
- [19] IDEAM - CORMAGDALENA, "Informe final - convenio nº 024/2005," Bogotá, 2006.
- [20] IDEAM - CORMAGDALENA, "Monitoreo de calidad y cantidad del agua superficial en el área hidrográfica de los ríos magdalena cauca y sus principales

afluentes en jurisdicción de la corporación autónoma regional del río grande de la magdalena," Noviembre. 2014.

[21] Convenio especial de cooperación en ciencia y tecnología No., 1-0016-2016., "Aunar esfuerzos técnicos para la realización de una campaña de monitoreo de cantidad y calidad de agua en el área hidrográfica del río magdalena-cauca y sus tributarios, en el marco de la política nacional para la gestión integral del recurso hídrico (GIRH) y el plan de acción trianual de CORMAGDALENA," Feb, 2017.

[22] United States Environmental Protection Agency and O. W. US EPA, "Indicators: Chlorophyll a," Us Epa, -11-21T09:24:54-05:00, 2013.

[23] M. Martínez-Orjuela et al, "Evaluación de la turbiedad como parámetro indicador del tratamiento en una planta potabilizadora municipal," Revista UIS Ingenierías, vol. 19, (1), pp. 15-24, 2020. Available: <https://revistas.uis.edu.co/index.php/revistauisingenierias/article/view/9929>. DOI: 10.18273/revuin.v18n4-2020001.

[24] Innovación y Cualificación, S. L. and S. L. Target asesores, Experto En Gestión Medioambiental. (2nd ed.) ic editorial, 2016.

[25] G. J. Mchau et al, "Phycocyanin as a proxy for algal blooms in surface waters: case study of Ukerewe Island, Tanzania," Water Practice and Technology, vol. 14, (1), pp. 229-239, 2019. Available: <https://iwaponline.com/wpt/article/14/1/229/65408/Phycocyanin-as-a-proxy-for-algal-blooms-in-surface>. DOI: 10.2166/wpt.2019.005.

[26] YSI Inc. and Xylem Inc, "The basics of chlorophyll measurement in surface water explained," 2021. Available: <https://www.ysi.com/Accessory/id-6025/6025-Chlorophyll-Sensor>.

[27] ClickSonic, "Cyanobacteria, blue-green algae in water and their treatment," 2019. Available: <https://www.clicksonic.com/en/frequently-asked-questions/cyanobacteria-blue-green-algae-in-water-and-their-treatment.html>.

[28] CORMAGDALENA, "Río Magdalena," 2016. Available: <http://dc02eja.cormagdalena.com.co/index.php?idcategoria=372>.

[29] P. Foladori, G. Andreottola and G. Zigliio, Sludge Reduction Technologies in Wastewater Treatment Plants. London: 2010.

[30] COMUNIDAD ANDINA. Manual de Estadísticas Ambientales. CAN: Santa Cruz de la Sierra, 2005. p. 31-45.

[31] United States Geological Survey, "Nitrogen and Water," May 17, 2020.

[32] United States Geological Survey, "Phosphorus and Water," May 21, 2020.

[33] N. Chang, S. Imen and B. Vannah, "Remote Sensing for Monitoring Surface Water Quality Status and Ecosystem State in Relation to the Nutrient Cycle: A 40-Year Perspective," Critical Reviews in Environmental Science and Technology, vol. 45, (2), pp. 101-166, 2015. Available:

<http://www.tandfonline.com/doi/abs/10.1080/10643389.2013.829981>. DOI: 10.1080/10643389.2013.829981.

[34] S. Milrad, "11 - satellite imagery," in Synoptic Analysis and Forecasting, S. Milrad, Ed. 2018, Available: <http://www.sciencedirect.com/science/article/pii/B9780128092477000119>.

[35] Radiant Earth Foundation et al, "The Science of Collecting Information about our Radiant Earth," 2016. Available: <https://www.radiant.earth/infographic/the-view-from-above/>.

[36] Infraestructura de Datos Espaciales de Bogotá, 2019. Resolución espectral. Available: <https://ideca.gov.co> .

[37] G. Sánchez, "Espectro electromagnético ," Enero, 2021.

[38] Universidad de Murcia, "Plataformas, sensores y canales," SF. Available: <https://www.um.es/geograf/sigmur/teledet/tema03.pdf>.

[39] A. García Varela, 2018 "Landsat" <https://es.scribd.com/document/52385174/LANDSAT>.

[40] Francis Gohin, "Twenty years of satellite and in situ observations of surface chlorophyll-a from the northern Bay of Biscay to the eastern English Channel. Is the water quality improving?" Remote Sensing of Environment, 2019.

[41] GeoService Perú, "INFORMACION TECNICA RESUMIDA –LANDSAT 7 ETM, 2019" .

[42] K. Didan, "MOD13Q1 v006. MODIS/Terra Vegetation Indices 16-Day L3 Global 250 m SIN Grid," NASA LP DAAC at the USGS EROS Center, Febrero 18, 2000.

[43] The United Nations Office for Outer Space Affairs, (UNOOSA). (Mar 7,). In Detail: Drought monitoring using the Standard Vegetation Index (SVI). Available: <https://un-spider.org/advisory-support/recommended-practices/recommended-practice-agricultural-drought-monitoring-svi/in-detail>.

[44] M. E. D. Chaves, E. Ferreira and A. A. A. Dantas, "Threshold's definition in MOD13Q1 and VGT-S10 time series for coffee crop area estimation in Triângulo Mineiro/Alto Paranaíba," Theoretical and Applied Engineering, vol. 3, (2), pp. 1-10, 2019. Available: <http://www.taaeufla.deg.ufla.br/index.php/TAAE/article/view/15>. DOI:10.31422/taae.v3i2.15.

[45] Vargas-Sanabria, D; Campos-Vargas, C. Sistema multialgoritmo para la clasificación de coberturas de la tierra en el bosque seco tropical del Área de Conservación Guanacaste, Costa Rica. Tecnología en Marcha. Vol. 31-1. Enero-marzo 2018. Pág 58-69. DOI: 10.18845/tm.v31i1.3497.

[46] Johansen R., Reif M., Emery E., Nowosad J., Beck R., Xu M., Liu H., waterquality: An Open-Source R Package para la detección y cuantificación de floraciones de algas nocivas por cianobacterias y calidad del agua. USACE ERDC / EL TR-19-20; DOI: 10.21079 / 11681/35053).

- [47] R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- [48] GERALD K. MOORE (1980) Satellite remote sensing of water turbidity / Sonde de télémessure par satellite de la turbidité de l'eau, Hydrological Sciences Journal, 25:4, 407-421, DOI: 10.1080/02626668009491950 .
- [49] Beck, R., S. Zhan, H. Liu, S. Tong, B. Yang, M. Xu, Z. Ye, Y. Huang, S. Shu, Q. Wu, S. Wang, K. Berling, A. Murray, E. Emery, M. Reif, J. Harwood, J. Young, C. Nietch, D. Macke, M. Martin, G. Stillings, R. Stumpf, AND H. Su. Comparison of satellite reflectance algorithms for estimating chlorophyll-a in a temperate reservoir using coincident hyperspectral aircraft imagery and dense coincident surface observations. REMOTE SENSING OF ENVIRONMENT. Elsevier Science Ltd, New York, NY, 178:15-30, (2016).
- [50] Eearth Observing System. (-10-17T14:11:08Z). NDVI: Normalized Difference Vegetation Index for Agriculture. Available: <https://eos.com/ndvi/>.
- [51] L. González, "Regresión Lineal Simple - Teoría," -11-30T11:34:31+00:00, 2018.
- [52] K. Yeager. (Jan 6,2021). LibGuides: SPSS Tutorials: Pearson Correlation. Available: <https://libguides.library.kent.edu/SPSS/PearsonCorr>.
- [53] G. Golemund and H. Wickham, 23 Modelos: Conceptos Básicos | R Para Ciencia De Datos. SFAvailable: <http://es.r4ds.hadley.nz/>.
- [54] M. Rodríguez, "Interpretación de los Gráficos de Caja en el Análisis Descriptivo e Inferencial." 2015. Available: <https://www.qvision.es/blogs/manuel-rodriguez/2015/03/30/interpretacion-de-los-graficos-de-caja-en-el-analisis-descriptivo-e-inferencial/>.
- [55] M. Paladino, "Modelos lineales con R," 2017. Available: https://www.institutomora.edu.mx/testU/SitePages/martinpaladino/modelos_lineales_con_R.html#introduccion.
- [56] J. Marín, "El procedimiento regresión lineal," in Análisis De Regresión Lineal, España: Universidad Carlos III de Madrid, 2019, pp. Capítulo 18.
- [57] C. Ford, "Interpreting Log Transformations in a Linear Model," University of Virginia Library, 2018. Available: <https://data.library.virginia.edu/interpreting-log-transformations-in-a-linear-model/>.
- [58] R core, "Nonlinear Least Squares - nls," SF. Available: <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/nls>.
- [59] Montero Granados. R (2016): Modelos de regresión lineal múltiple. Documentos de Trabajo en Economía Aplicada. Universidad de Granada. España.
- [60] J. Amat Rodrigo, "Introducción a la regresión lineal múltiple," Madrid, España, Julio, 2016.

- [61] A. Kassambara, Machine Learning Essentials: Practical Guide in R. (1st ed.) STHDA - Statistical tools for high-throughput data analysis, 2018.
- [62] Mangiafico, SS 2016. Resumen y análisis de la evaluación del programa de extensión en R, versión 1.18.6. rcompanion.org/handbook/ . (Versión en pdf: rcompanion.org/documents/RHandbookProgramEvaluation.pdf .)
- [63] F. Hernández and M. Mazo, 7 Pruebas De Hipótesis Parte I | Modelos De Regresión Con R. (1st ed.) 2020 Available: https://fhernanb.github.io/libro_regresion/ph1.html.
- [64] SOTO-BRAVO, Freddy and GONZALEZ-LUTZ, María Isabel. Análisis de métodos estadísticos para evaluar el desempeño de modelos de simulación en cultivos hortícolas. Agron. Mesoam [online]. 2019, vol.30, n.2 [cited 2021-01-21], pp.517-534. Available from: http://www.scielo.sa.cr/scielo.php?script=sci_arttext&pid=S1659-13212019000200517&lng=en&nrm=iso. ISSN 2215-3608. <http://dx.doi.org/10.15517/am.v30i2.33839>.
- [65] T. Srivastava, "11 Important Model Evaluation Metrics for Machine Learning Everyone should know," Analytics Vidhya, 2019. Available: <https://www.analyticsvidhya.com/blog/2019/08/11-important-model-evaluation-error-metrics/>.
- [66] J. Brownlee, "Metrics to Evaluate Machine Learning Algorithms in Python," Agosto 21, 2020. Available: <https://machinelearningmastery.com/metrics-evaluate-machine-learning-algorithms-python/>.
- [67] B. Maccherone and S. Frazier, "Data Products - MODIS," Modis.Gsfc.Nasa.Gov, Available: <https://modis.gsfc.nasa.gov/data/dataproduct/>.
- [68] B. S. NASA Official, "EARTHDATA OPEN ACCESS FOR OPEN SCIENCE," SF. Available: <https://earthdata.nasa.gov/>.
- [69] A. Germán et al, "Análisis de una serie temporal de clorofila-a a partir de imágenes MODIS de un embalse eutrófico," 2016. Available: https://www.researchgate.net/publication/309338823_Analisis_de_una_serie_temporal_de_clorofila-a_a_partir_de_imagenes_MODIS_de_un_embalse_eutrofico.
- [70] M. D. Steven, "The Sensitivity of the OSAVI Vegetation Index to Observational Parameters," Remote Sensing of Environment, vol. 63, (1), pp. 49-60, 1998. Available: <http://www.sciencedirect.com/science/article/pii/S0034425797001144>. DOI: 10.1016/S0034-4257(97)00114-4.
- [71] R-Core, "lm - Fitting Linear Models," SF. Available: <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/lm>.
- [72] S. Mishra and D. R. Mishra, "Normalized difference chlorophyll index: A novel model for remote estimation of chlorophyll-a concentration in turbid productive waters," Remote Sens. Environ., vol. 117, pp. 394-406, 2012. Available: <https://www.sciencedirect-com.crai->

ustadigital.usantotomas.edu.co/science/article/pii/S0034425711003737. DOI: <https://doi-org.crai-ustadigital.usantotomas.edu.co/10.1016/j.rse.2011.10.016>.

[73] R. Beck et al, "Comparison of satellite reflectance algorithms for estimating chlorophyll-a in a temperate reservoir using coincident hyperspectral aircraft imagery and dense coincident surface observations," *Remote Sens. Environ.*, vol. 178, pp. 15-30, 2016. Available: <https://www-sciencedirect-com.crai-ustadigital.usantotomas.edu.co/science/article/pii/S0034425716300943>. DOI: <https://doi-org.crai-ustadigital.usantotomas.edu.co/10.1016/j.rse.2016.03.002>.

[74] R. M. C. Castañeda, "Ingeniería ambiental II - contaminación hídrica - sólidos suspendidos totales," Nov 26, 2014.

[75] P. Vinuesa, "Tema 8 - correlación: Teoría y práctica," Centro de Ciencias Genómicas UNAM, México, octubre 14, 2016.

[76] E. Rodríguez, *Errores Frecuentes En La Interpretación Del Coeficiente De Determinación Lineal*. San Lorenzo del Escorial: Real Centro Universitario, pp 315-332, ISSN: 1133-3677, 2005.

[77] Minitab, "Análisis de Regresión: ¿Cómo Puedo Interpretar el R-cuadrado y Evaluar la Bondad de Ajuste?" *Análisis De Datos, Estadística*, Abril 18, 2019.

[78] A. Guanga. (01-02T15:43:49.664Z). *Understand Regression Performance Metrics*, 2019 Available: <https://becominghuman.ai/understand-regression-performance-metrics-bdb0e7fcc1b3>

[79] S. Wu. (May 23, 2020). 3 best metrics to evaluate Regression Model - R Square, Adjusted R Square, MSE, RMSE, MAE. Available: <https://towardsdatascience.com/what-are-the-best-metrics-to-evaluate-your-regression-model-418ca481755b>.

[80] B. Bose, (Ago 17, 2020). "Performance Metrics: Regression Model," Medium, 2020. Available: <https://medium.com/analytics-vidhya/performance-metrics-regression-model-69f68a18504f>.

[81] D. Mishra. (Dic 6, 2019). *Regression: An Explanation of Regression Metrics and What Can Go Wrong*. Available: <https://towardsdatascience.com/regression-an-explanation-of-regression-metrics-and-what-can-go-wrong-a39a9793d914>.

8. ANEXOS

ANEXO A. Puntos de monitoreo seleccionados con su número de muestras

ZONA HIDROGRÁFICA	AÑO	PUNTO DE MONITOREO	MUESTRAS		
			Sólidos Suspendidos Totales	Nitratos	Fósforo Total
Medio magdalena	2006	Magangué	8	8	0
		La Gloria	8	8	0
		Sitio Nuevo R-11	8	8	0
		Puerto Berrio	8	8	0
	2013	Desembocadura Río Negro	8	8	8
		Desembocadura Río La Miel	8	8	8
		Desembocadura Río Alicante	8	8	8
		Desembocadura Río Carare	8	8	8
		Desembocadura Río Sogamoso	8	8	8
		Sitio Nuevo R-11	8	8	8
		Regidor	8	8	8
		Desembocadura Río Cesar	8	8	8
	2016	Desembocadura Río Negro	8	8	8
		Desembocadura Río La Miel	8	8	8
		Desembocadura Río Alicante	8	8	8
		Desembocadura Río Carare	8	8	8
		Desembocadura Río Sogamoso	8	8	8
		Sitio Nuevo R-11	8	8	8
		Regidor	8	8	8
		Desembocadura Río Cesar	8	8	8

Fuente: autor.

ANEXO B. Características principales del instrumento MODIS- Terra

Banda	Longitud de onda (nm)	Resolución (m)	Uso primario
1	620-670	250	Límites de tierra / nube / aerosoles
2	841-876	250	
3	459-479	500	Terreno / nube / aerosoles propiedades
4	545-565	500	
5	1230-1250	500	
6	1628-1652	500	
7	2105-2155	500	
8	405-420	1000	Color del océano / fitoplancton / biogeoquímica
9	438-448	1000	
10	483-493	1000	
11	526-536	1000	
12	546-556	1000	
13	662-672	1000	
14	673-683	1000	
15	743-753	1000	
16	862-877	1000	Atmosférica vapor de agua
17	890-920	1000	
18	931-941	1000	
19	915-965	1000	
	um		
20	3.660-3.840	1000	Temperatura de superficie / nube
21	3.929-3.989	1000	
22	3.929-3.989	1000	
23	4.020-4.080	1000	
24	4.433-4.498	1000	Temperatura atmosférica
25	4.482-4.549	1000	
26	1.360-1.390	1000	Cirros vapor de agua
27	6.535-6.895	1000	
28	7.175-7.475	1000	
29	8.400-8.700	1000	Propiedades de la nube
30	9.580-9.880	1000	Ozono
31	10.780-11.280	1000	Temperatura de superficie / nube
32	11.770-12.270	1000	
33	13.185-13.485	1000	Altitud superior de la nube
34	13.485-13.785	1000	
35	13.785-14.085	1000	
36	14.085-14.385	1000	

Fuente: [39].

ANEXO C. Imágenes satelitales encontradas para cada año de estudio

IN-SITU		INFORMACIÓN	
Fecha muestreo	Estación	Nombre de la imagen	Resolución temporal
11/07/2006	Magangué	MOD13Q1.A2006177.h10v08.006.2015161173038	START 2006-06-26 00:00:00 END 2006-07-11 23:59:59
15/07/2006	La Gloria	MOD13Q1.A2006193.h10v08.006.2015161174421	START 2006-07-12 00:00:00
16/07/2006	Sitio Nuevo R-11		END 2006-07-27 23:59:59
18/07/2006	Puerto Berrio		
04/10/2006	Magangué	MOD13Q1.A2006273.h10v08.006.2015161203334	START 2006-09-30 00:00:00
09/10/2006	La Gloria		END 2006-10-15 23:59:59
11/10/2006	Sitio Nuevo R-11		
14/10/2006	Puerto Berrio		
11/10/2013	Desembocadura Río Negro	MOD13Q1.A2013273.h10v08.006.2015269190105	START 2013-09-30 00:00:00
11/10/2013	Desembocadura Río La Miel		END 2013-10-15 23:59:59
13/10/2013	Desembocadura Río Alicante		
13/10/2013	Desembocadura Río Carare		
15/10/2013	Desembocadura Río Sogamoso		
17/10/2013	Sitio Nuevo R-11	MOD13Q1.A2013289.h10v08.006.2015269222228	START 2013-10-16 00:00:00
19/10/2013	Regidor		END 2013-10-31 23:59:59
20/10/2013	Desembocadura Río Cesar		
03/11/2016	Desembocadura Río Negro	MOD13Q1.A2016305.h10v08.006.2016322042034	START 2016-10-31 00:00:00
04/11/2016	Desembocadura Río La Miel		END 2016-11-15 23:59:59
05/11/2016	Desembocadura Río Alicante		
05/11/2016	Desembocadura Río Carare		
06/11/2016	Desembocadura Río Sogamoso		
07/11/2016	Sitio Nuevo R-11		
08/11/2016	Regidor		
08/11/2016	Desembocadura Río Cesar		

Fuente: autor.