



Acta de Correcciones al Proyecto de Grado Ingeniería Electrónica

Fecha: Mayo 11 2022

Autores: Iván Camilo Dallos Bedoya y Victor Daniel Córdoba Dorado

Nombre del Proyecto de Grado: Predicción del transporte de sólidos en el valle alto del río Cauca a la altura de la planta Puerto Mallarino por medio de análisis de imágenes satelitales y procesamiento digital de señales

Director: Hernan Dario Benitez Restrepo

Como indica el artículo 2.27 de las Directrices de Trabajo de Grado, he verificado que los estudiantes indicados arriba han implementado todas las correcciones que los Jurados del Proyecto de Grado definieron que se efectuaran, como consta en el Acta de Calificación correspondiente.

Hernán Benítez

Firma de Director(a) del Proyecto de Grado

Nota de Aceptación

Aprobado por el Comité de Trabajo de Grado en cumplimiento de los requisitos exigidos por la Pontificia Universidad Javeriana para optar el título de Ingeniero Electrónico.

Camilo Rocha

Dr. Hernán Camilo Rocha Niño
Decano de la Facultad de Ingeniería



ING. LUIS EDUARDO TOBON LLANO
Director Carrera Ingeniería Electronica.

Hernán Benítez

ING. Hernán Darío Benítez Restrepo
Director(a) Trabajo



ING. Luis Eduardo Tobón Llano
Jurado 1



ING. Hernán Darío Vargas Cardona
Jurado 2

Pontificia Universidad Javeriana Cali



Tesis para el grado de ingeniería electrónica

Predicción del transporte de sólidos en el valle alto del río Cauca a la altura de la planta Puerto Mallarino por medio de análisis de imágenes satelitales y procesamiento digital de señales

Iván Camilo Dallos Bedoya

Víctor Daniel Córdoba Dorado

Facultad de Ingeniería y Ciencias

2022

Santiago de Cali, 17 de Enero de 2022

Dr. Luis Eduardo Tobon
Director
Carrera de Ingeniería Electrónica
Pontificia Universidad Javeriana-Cali

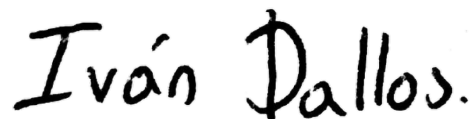
Cordial Saludo

Nos permitimos presentar a su consideración el trabajo de grado titulado “Predicción del transporte de sólidos en el valle alto del río Cauca a la altura de la planta Puerto Mallarino por medio de análisis de imágenes satelitales y procesamiento digital de señales” con el fin de cumplir con los requisitos exigidos por la Universidad para optar al título de Ingeniero Electrónico. Al firmar aquí, damos fe que entendemos y conocemos las directrices para la presentación de trabajos de grado de la Facultad de Ingeniería y Ciencias aprobadas el 26 de noviembre de 2009, donde se establecen los plazos y normas para el desarrollo del del trabajo de grado.

Atentamente



Víctor Daniel Córdoba
Id: 8936096



Ivan Camilo Dallos
Id: 8935432

Santiago de Cali, 17 de Enero de 2022

Dr. Luis Eduardo Tobon
Director
Carrera de Ingeniería Electrónica
Pontificia Universidad Javeriana-Cali

Cordial Saludo

Por medio de la presente me permito informarle que los estudiantes de Ingeniería Electrónica Víctor D. Córdoba (código: 8936096) e Iván C. Dallos (código: 8935432), trabajan bajo mi dirección en el proyecto de grado titulado “Predicción del transporte de sólidos en el valle alto del río Cauca a la altura de la planta Puerto Mallarino por medio de análisis de imágenes satelitales y procesamiento digital de señales”.

Atentamente:

Hernán Benítez

Dr. Hernan D. Benitez R

Dedicatoria

Este trabajo de grado está dedicado a todas las personas que alguna vez creyeron en nosotros. A nuestros familiares, profesores y amigos por apoyarnos cuando más lo necesitábamos. A las personas que en verdad estuvieron a nuestro lado animándonos a salir adelante.

Resumen

En este documento, se propone como problemática de estudio la turbidez como factor causante de la falta de agua potable en la población caleña a través de su afectación directa en la planta Puerto Mallarino en la ciudad de Cali. Se analiza la problemática utilizando las suspensiones no programadas, no avisadas reportadas por la SUI (Sistema Único de Información de Servicios Públicos Domiciliarios), las cuales muestran un gran número de horas y suscriptores afectados comparado con otro tipo de suspensiones y se desarrolla con los datos de turbidez del año 2015 a 2020 proporcionados por EMCALI en la bocatoma de la planta de Puerto Mallarino sumado a los datos de reflectancia de las bandas del satélite Sentinel-2 en este mismo periodo de tiempo.

Con estos recursos, desde la ingeniería electrónica y más específicamente desde el Procesamiento Digital de Señales, se propone una solución con el procesamiento de imágenes satelitales provenientes del satélite Sentinel-2 aprovechando el fenómeno de la reflectancia presente en los cuerpos de agua sedimentados y las diferentes magnitudes de ondas electromagnéticas reflejadas que se pueden medir haciendo uso de las bandas espectrales de este mismo satélite y análisis de datos medidos in situ.

Como finalidad, se pretende determinar un modelo lógico-matemático (algoritmo) que permita con el mayor grado de exactitud posible entregar una aproximación a los valores reales de turbidez según la información espectral disponible a través de las imágenes satelitales.

Palabras Claves: Agua, planta, bocatoma, río Cauca, Puerto Mallarino, turbidez, sedimentación, reflectancia, NTU, Sentinel-2, bandas espectrales.

In this document, turbidity is proposed as a study problem as a factor causing the lack of drinking water in the population of Cali through its direct impact on the Puerto Mallarino plant in the city of Cali. The problem is analyzed using unscheduled, unannounced suspensions reported by the SUI (Single Information System for Home Public Services), which show a large number of hours and subscribers affected compared to other types of suspensions and is developed with the data of turbidity from 2015 to 2020 provided by EMCALI at the intake of the Puerto Mallarino plant added to the reflectance data of the bands of the Sentinel-2 satellite in this same period of time.

With these resources, from electronic engineering and more specifically from Digital Signal Processing, a solution is proposed with the processing of satellite images from the Sentinel-2 satellite, taking advantage of the phenomenon of reflectance present in sedimented bodies of water and the different magnitudes of reflected electromagnetic waves that can be measured using the spectral bands of this same satellite and analysis of data measured in situ.

As a purpose, it is intended to determine a logical-mathematical model (algorithm) that allows, with the highest possible degree of accuracy, to deliver an approximation to the real turbidity values according to the spectral information available through satellite images.

Keywords: Water, plant, intake, Cauca River, Puerto Mallarino, turbidity, sedimentation, reflectance, NTU, Sentinel-2, spectral bands.

Índice general

1. Introducción	1
2. Descripción del problema	2
2.1. Planteamiento del problema	2
2.1.1. Formulación	3
2.1.2. Sistematización	3
2.2. Objetivos	3
2.2.1. Objetivo general	3
2.2.2. Objetivos específicos	3
2.3. Justificación	3
2.4. Delimitaciones y alcance	4
2.4.1. Entregables	5
3. Desarrollo	6
3.1. Estructura	6
3.1.1. Áreas Temáticas	6
3.1.2. Marco teórico	6
3.1.3. Trabajos relacionados	10
3.2. Metodología	10
3.2.1. Tipo de estudio	10
3.2.2. Ocupaciones	10
3.3. Selección de información	11
3.3.1. Selección de datos por imágenes satelitales	11
3.3.2. Selección de datos por EMCALI	12
4. Resultados	15
4.1. Construcción del modelo	15
4.1.1. Modelo Único	15
4.2. Modelo Doble	18
4.3. Validación del modelo	20
4.4. Mejora del Modelo	21
4.4.1. Recolección de datos	21
4.4.2. Uso de imágenes de satélites de revisita diaria	21
4.4.3. Uso de cámara multiespectral	21
5. Conclusiones	23

Índice de figuras

3.1. Comparación de agua sedimentada y agua clara con posición en bandas espectrales	7
3.2. Puntos de referencia en el río Cauca.	11
3.3. Área de interés río Cauca	12
3.4. Casos de imagen satelital en área de interés PTAP Puerto Mallarino.	13
3.5. Comparación turbidez en Puerto Mallarino y punto Milán.	14
4.1. Coeficientes de correlación de las bandas del Sentinel-2	16
4.2. Relación de banda vs Datos de turbidez reales	17
4.3. Fecha de datos de turbidez vs Turbidez	18
4.4. Digrama de flujo cálculo de turbidez con el modelo doble	19
4.5. Comparación Turbidez real vs predicción modelo único y modelo doble	19
4.6. Comparación predicción modelo doble y dato real de turbidez	20
4.7. Validación del modelo doble con los datos reales de turbidez.	21

Índice de tablas

2.1. Significado de los códigos de suspensión del servicio de acueducto	2
3.1. Bandas espectrales del satélite Sentinel-2	8
3.2. Bandas espectrales del satélite LandSat-8	8
4.1. Índices de correlación resultado de comparación con modelo.	17
4.2. Datos de error del modelo único	18
4.3. Índices de correlación en modelo doble.	18
4.4. Comparación datos de error: Modelo único y modelo doble	19
4.5. Datos de error para la validación del modelo.	20

Capítulo 1

Introducción

Santiago de Cali es una ciudad en la que habitan 2'497.562 personas para el 2020 según estimaciones del DANE, a las cuales se les debe garantizar acceso al servicio de agua potable para esto empresas municipales de Cali o EMCALI es la que maneja el servicio de acueducto de los caleños. Esta tiene a su disposición diferentes plantas de tratamiento de agua como lo son: La Ribera, Río Cali, Río Cauca, La Reforma y Puerto Mallarino, siendo esta última la más importante debido a que suministra el 60 % del agua potable a los caleños.

La planta de Puerto Mallarino toma sus aguas del río Cauca al oriente de la ciudad de Cali, al cual le llegan los sedimentos de sus ríos tributarios volviendo turbia el agua. Adicionalmente, en el sector de Navarro, al río Cauca se suman las aguas servidas del canal sur contaminándolo de los desechos de la ciudad. También fallas en la prestación del servicio con poca duración son habituales debido a las reparaciones de la gran infraestructura que posee la ciudad para la prestación del servicio. Cuando el río tiene demasiada turbidez o tiene un nivel bajo oxígeno disuelto, mayor a 3000 NTU (Unidad de turbidez nefelométrica) y menor a 2.5 mg/L. respectivamente, se tiene que cerrar obligatoriamente la bocatoma para no generar daños en los filtros de la planta y garantizar un agua en la salida apta para el consumo humano. Para estas situaciones, la planta de Puerto Mallarino activa sus embalses dándole autonomía por 7 horas, siendo en determinadas ocasiones insuficiente para la duración de sedimentación o contaminantes fluyendo en el río Cauca. Esto a su vez, genera una suspensión sin aviso previo al servicio de acueducto de la ciudad, lo cual según los reportes de la SUI son los que presentan mayor duración y afectan a más suscriptores comparándolos con las suspensiones alertadas. Un avance en las tecnologías de detección en sedimentación para estos casos es lo ideal para que estas afecciones no sean graves para la planta y no se generen daños a la infraestructura de la misma. Por ello, en las siguientes páginas se muestra la descripción completa del problema y como se puede llegar a la solución de este sistema.

Capítulo 2

Descripción del problema

2.1. Planteamiento del problema

Uno de los objetivos propuestos por la organización mundial de la salud en el ámbito del agua es garantizar que las personas tengan acceso al agua potable [1]. Por ello, en una ciudad como Cali, existe una clara dependencia por parte de la comunidad hacia las empresas de servicios públicos, por lo que es imperativo que estas empresas garanticen un servicio continuo y confiable.

Los reportes de suspensión del servicio de acueducto disponibles en el SUI hablan de las veces que se paraliza el servicio de agua potable en Cali. Estos informes no especifican una causa específica por la cual los suscriptores dejaron de brindar el servicio, sino que se generalizan mediante códigos que van del 1 al 5.

Código	Tipo
1	Reparaciones técnicas y mantenimiento periódico, con aviso oportuno a los usuarios.
2	Racionamiento por fuerza mayor, con aviso a los usuarios.
3	Previsiones contra la inestabilidad inmobiliaria.
4	Suspensiones no programadas ni anunciadas.
5	Suspensiones por no ofrecer el servicio y no relacionadas con ninguno de los tipos de suspensiones anteriores.

Tabla 2.1: Significado de los códigos de suspensión del servicio de acueducto [2]

El informe, además de la cantidad de suscriptores afectados, también indica el número de horas que se suspendió el servicio junto con la fecha exacta del día en que ocurrió la suspensión [2].

Ahora, realizando un análisis de datos, se encuentra que para las suspensiones que involucran el código 4 que son suspensiones no programadas, no anunciadas, es cuando se presenta la mayor cantidad de abonados afectados con una cantidad de horas de afección mayor en comparación con el promedio, lo que se puede pensar que la planta de Puerto Mallarino ha dejado de brindar su servicio porque abastece a la mayor parte de Cali.

“El 60% del agua potable en Cali es provista por la planta de Puerto Mallarino” [3] ubicada al noroeste de la ciudad por el sector Juanchito, que toma el agua del río Cauca para potabilizarla.

Según la SUI, de 2018 a 2025 la planta Puerto Mallarino tiene la posibilidad de tomar 8000 L/s del caudal del río Cauca, lo que casi cuadruplica las concesiones de las otras tres plantas que abastecen a Cali [4].

En este mismo documento se menciona que la planta de Puerto Mallarino cuenta con dos embalses, uno de 100.000 m^3 y otro de 80.000 m^3 que permiten que la planta funcione de forma autónoma durante 7 horas, que se utilizan cuando la toma del río se ha detenido por turbidez o por baja cantidad de oxígeno disuelto.

Teniendo en cuenta que los códigos 1 y 2 son eventos anticipados mientras que los códigos 3 y 4 son eventos aleatorios que resultan en fallas en la prestación del servicio sin previo aviso está el problema de la turbidez. La turbidez es una causa directa de sedimentación. La planta cuenta con 4 procesos en los que se filtran sedimentos gruesos y delgados. En primer lugar, una criba inicial que se limpia automáticamente,

luego se continúa con una etapa de desarenado donde se retiran arena, grava y lodos y, finalmente, decantación y filtración donde se depositan los lodos y se eliminan devolviendo los sedimentos, a través de tuberías al río. Para la filtración es necesario cambiar los filtros cada 24 horas o menos si la turbidez del agua aumenta [4].

Si se miden a tiempo los niveles de turbidez del río Cauca para la toma de la planta de Puerto Mallarino, en los momentos donde la turbidez pueda afectar la planta o el proceso, se podría cerrar la toma de agua cruda para no afectar los filtros y poner en funcionamiento oportuno los reservorios.

Para ello, mediante la reflectancia del agua captada por el satélite Sentinel-2 en el valle alto del río Cauca, se pretende obtener el nivel de turbidez mediante el procesamiento digital de imágenes satelitales utilizando sus respectivas bandas espectrales y un combinación de estas, tomando como referencia los trabajos presentados en el apartado 3.1.3 de este texto, buscando calcular la turbidez presente en el río Cauca a la altura de la planta de tratamiento de Puerto Mallarino y realizar un proceso de calibración con las bases de datos provistas por EMCALI.

2.1.1. Formulación

¿Cómo respaldar el sistema actual de medición de turbidez del río Cauca en la planta de Puerto Mallarino utilizando imágenes satelitales y procesamiento de bandas espectrales?

2.1.2. Sistematización

- ¿Cuál es la periodicidad con la que el satélite Sentinel-2 frecuenta el suroeste de Colombia?
- ¿Qué nivel máximo de turbidez puede soportar la planta antes de cambiar el filtro o decidir cerrarlo?
- ¿Cuál es el nivel máximo de nubosidad con el que se podrían trabajar las imágenes de satélite?

2.2. Objetivos

2.2.1. Objetivo general

Predecir los niveles de turbidez del río Cauca que pueden llegar a la toma de la planta de Puerto Mallarino mediante el análisis de imágenes satelitales proporcionadas por el satélite Sentinel-2.

2.2.2. Objetivos específicos

- Construir una base de datos que contenga índices de reflectancia obtenidos a partir de imágenes de satélite y datos de turbidez medidos en estaciones de medición.
- Aplicar algoritmos de regresión en la base de datos construida para predecir los niveles de turbidez en el río Cauca.
- Evaluar los algoritmos de regresión aplicados en términos de métricas tales como el error cuadrático medio (RMSE) y el R^2 .

2.3. Justificación

El ingeniero Juan Carlos Escobar, jefe de operaciones de la planta de PM¹ explica que los factores que se toman en cuenta para cerrar la planta son oxígeno disuelto y turbidez del agua. El oxígeno disuelto tiene una relación inversamente proporcional a la cantidad de material biológico en el agua, es decir, desechos orgánicos. La turbidez es una medida de la transparencia del agua, proporcionalmente relacionada con la cantidad de partículas inorgánicas en el agua (sedimento).

En el caso de la planta de Puerto Mallarino, existen 2 estaciones de medición, la primera ubicada río arriba llamada Milán en el sector de Navarro y la segunda ubicada en la toma de la planta de tratamiento (Con un desfase de 3 horas aproximadamente en la medición). En estas dos estaciones miden de forma permanente parámetros como la turbidez del agua y en el caso de que se mantenga un nivel superior a 3000 NTU, por un período de una hora cesará la captación del afluente. [5]

¹PM: Puerto Mallarino.

La turbidez por sí misma es un motivo para cerrar la captación de agua, en primer lugar, al ser material inorgánico, la ingestión de este puede resultar nocivo y es innecesario para el organismo. Según la OMS, no es saludable consumir agua con una medida de turbidez superior a 2 NTU, y plantas de tratamiento como la de PM deben asegurar un resultado de 0,6 NTU en la salida de agua potable [6]. La sedimentación en altos niveles ayudan a que las bacterias proliferen y requiera más cloro para la desinfección. Junto a esto, existen riesgos de obstruir un proceso en la planta (que no requiere cambio de filtros) y generar costos de mantenimiento imprevistos y mayor dificultad en proporcionar el servicio.

Cuando se cierra la captación, se cambia el suministro hacia los embalses auxiliares con aproximadamente 7 horas de autonomía, hasta que ya no se pueda prestar el servicio y se considere un fallo en la prestación del servicio.

Las fallas en la prestación del servicio de acueducto afectan gravemente a la población de Cali; para el año 2019, hubo un total de 48717 (suma total de horas individuales de cada suspensión en la ciudad) horas en las que no se brindó el servicio de acueducto debido a las diferentes situaciones que enumeran los códigos [7]. Solo el código 4, para el año 2019 representa solo el 18,73% de los casos registrados. Asimismo, este porcentaje equivale al 72,25% del total de horas en las que se detuvo el servicio de acueducto, lo que sugiere que un caso aleatorio afecta el servicio mucho más que un caso anticipado [7]. Un caso similar ocurre desde enero hasta junio de 2020 donde las suspensiones del código 4 representan el 17,56% de los fallos, pero equivale al 70,31% de los abonados afectados [7].

Las mediciones in situ de estudios granulométricos pueden proporcionar información sobre la cantidad de sedimento transportado por un río. Aunque útiles, ofrecen un resultado temporal como ya había ocurrido en el informe CVC de estudios granulométricos [8], al igual que las mediciones regulares en los puntos de Puerto Mallarino y Milán, que miden permanentemente la turbidez y el oxígeno disuelto. Estos puntos tienen costos de mantenimiento regulares y la información proporcionada se limita al nivel de la variable medida y al tiempo que se tarda en tomar la muestra. El análisis de la información proporcionada por EMCALI muestra que estos sensores son sensibles a pequeñas variaciones de muestra y fácilmente pueden dar información incorrecta, que hace necesario un sistema auxiliar que corrobore la información y no se base en la misma tecnología.

Se espera que el proyecto tenga un impacto ambiental y económico mínimo, ya que se desarrollará en gran parte en plataformas de software y una vez implementado se podrá evitar la construcción de nuevas estaciones de monitoreo a lo largo del río y además proporcionar una propuesta plausible de un sistema auxiliar.

La viabilidad de este proyecto se refleja en la disponibilidad de programas como Google Earth Engine² (Open Access), que tienen la posibilidad de crear índices de turbidez basados en la combinación de bandas espectrales. Asimismo, estos programas cuentan con un historial de imágenes satelitales desde 1993 hasta la actualidad, con lo cual, sumado a las bases de datos proporcionadas por EMCALI que muestran las medidas específicas de turbidez cada minuto desde 2013, se pueden contrastar los datos inferidos con los reales.

En el campo del conocimiento; la disponibilidad de imágenes satelitales abre las puertas al uso de procesamiento de imágenes digitales, no solo en ríos sino también en áreas forestales, linderos urbanos, crecimiento controlado de plantas, etc. En la sección 2.1.2 se describen las especialidades de cada banda satelital y sus posibles usos. El desarrollo obtenido en este proyecto es un punto de partida para el uso de estas aplicaciones en el Valle del Cauca.

2.4. Delimitaciones y alcance

El estudio se aplicará para las fechas en las que se disponga de imágenes satelitales de los satélites antes mencionados y el análisis será temporal, con datos de 2017 a 2020.

El estudio se limitará al del río Cauca en las cercanías de la ciudad de Cali.

Los siguientes fenómenos como sequías, deslizamientos y causas extraordinarias que afectan la cantidad de sedimento transportado por un río no serán relevantes en el estudio ya que, al no ser tan comunes, pueden o no (según la investigación) ser tomados en cuenta.

El proyecto busca solucionar la falta de un sistema de información basado en imágenes de satélite inexistente para el momento.

²Sitio oficial: <https://earthengine.google.com>

2.4.1. Entregables

- Modelo de turbidez según relaciones de bandas satelitales.
- Base de datos de reflectancia y turbidez del año 2015 a 2020.
- Algoritmo para la adquisición de la relación de bandas satelitales.
- Algoritmo para la adquisición de datos de reflectancia en Google Earth Engine.

Capítulo 3

Desarrollo

3.1. Estructura

3.1.1. Áreas Temáticas

- Computación y procesamiento de información - informática - Matlab.
- Computación y procesamiento de información - Programación.
- Computación y procesamiento de información - Procesamiento de datos - Análisis de datos.
- Computación y procesamiento de información - Procesamiento de imágenes - Análisis de color de imágenes.
- Descomposición de imágenes. Secuencias de imágenes.
- Geociencias y teledetección - Técnicas de medición geofísica - Procesamiento de imágenes geofísicas.
- Geociencias y teledetección - Geociencias - Ríos. Sedimentos.
- Geociencia y teledetección - Teledetección - Imagen hiperespectral.
- Imágenes: imágenes multiespectrales.
- Instrumentación y medición - Monitorización - Monitorización remota.
- Procesamiento de señales: procesamiento de señales geofísicas.
- Ingeniería y teoría de sistemas - Análisis y diseño de sistemas - Modelado de sistemas.

3.1.2. Marco teórico

Turbiedad:

La turbiedad se refiere al grado de transparencia de un líquido incoloro debido a las partículas suspendidas. Cuanto mayor sea el número de partículas en suspensión, mayor será su nivel de turbiedad.

En Colombia, según el Ministerio de Salud, la turbiedad se mide en NTU (Unidad Nefelométrica de Turbiedad) utilizando un nefelómetro o turbidímetro que según BMG LabTech¹ desarrolladores de nefelómetros y microplacas. Las partículas suspendidas se cuantifican utilizando un haz de luz que es dispersado por las partículas y recogido por un detector colocado en un lado de 90° de la fuente de luz.

En la PTAP² de PM³ y en la estación de monitoreo en Navarro llamada Milán se han adquirido los datos de turbiedad y oxígeno disuelto del río Cauca cada minuto desde 2013 y como se indica en la justificación, la turbiedad no debe ser superior a 3000 NTU y el oxígeno disuelto no debe ser inferior a 2,5. mg / L para que la planta cierre su bocatoma. [5]

La turbiedad es una causa directa de sedimentación y la relación entre estos dos términos se obtiene estadísticamente con estudios gravimétricos ya que es relativa al lugar donde se obtienen las muestras.

¹Laboratorio de diseño de microplacas <https://www.bmglabtech.com/es/>

²Planta de Tratamiento de Agua Potable.

³Puerto Mallarino

Reflectancia del agua:

La reflectividad es la fracción de energía incidente reflejada por una cubierta terrestre, que cambia en función de la longitud de onda. La distribución de la reflectividad en función de la longitud de onda se denomina firma espectral, que es diferente para los diferentes tipos de cobertura y caracteriza a los objetos en la superficie terrestre, por lo que es posible lograr, en cierta medida, su discriminación. [9]

Una característica de los cuerpos de agua terrestres es la reflectividad casi nula en el infrarrojo medio. Además, la reflectividad del agua depende de la profundidad del cuerpo y del sustrato del lecho. [9]

Los cuerpos de agua clara presentan una mayor reflectancia en la banda verde y azul (560 nm) a comparación de las bandas de infrarrojos (739 nm). Las aguas con presencia de turbidez por sedimentación aumentan su reflectividad hacia la banda roja (664 nm). En la figura 3.1 muestra cómo se comporta la reflectancia detectada por las bandas del satélite Sentinel-2 ya sea por agua clara o agua turbia debido a la sedimentación. [10]

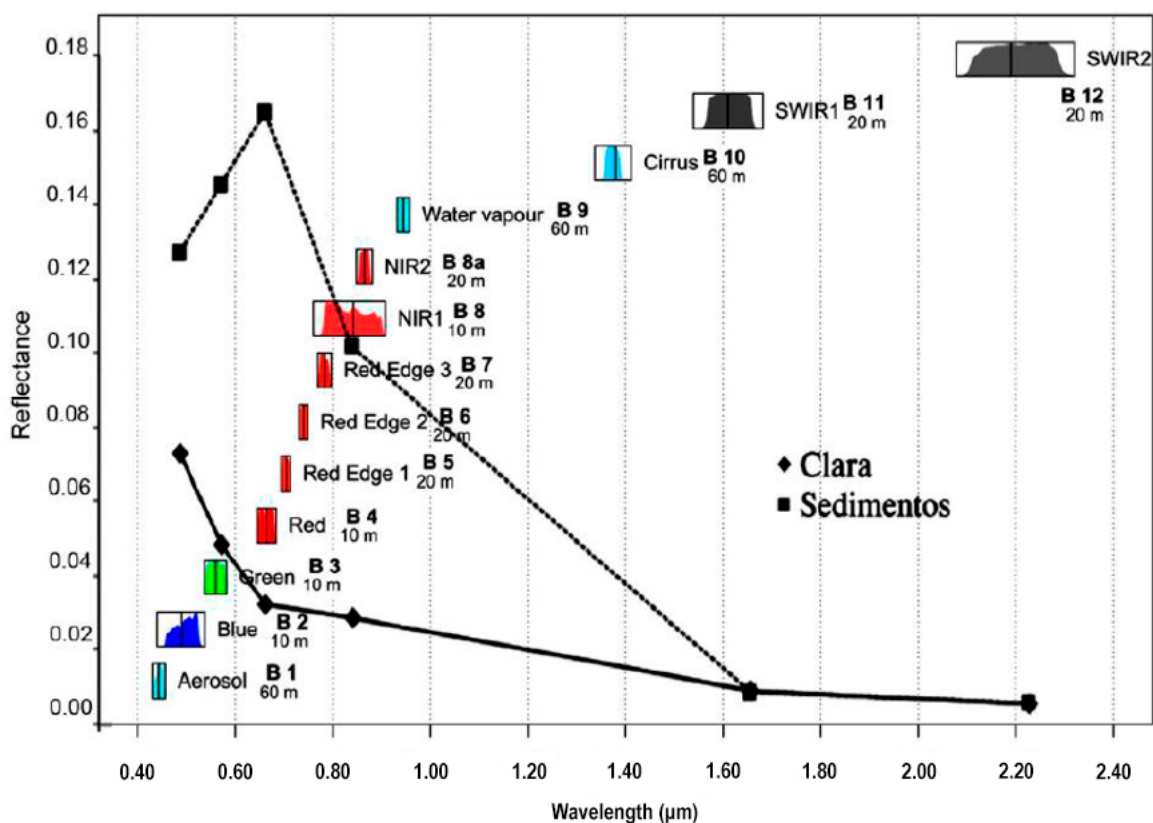


Figura 3.1: Comparación de agua sedimentada y agua clara con posición en bandas espectrales [9] y [11]

Satélites de observación de la Tierra:

Existen varios satélites de observación de la Tierra como LandSat-7 [12], LandSat-8 [13], Sentinel-2 [14], Superview [15], Pleiades 1 [16] y KompSat [17] que tienen diferentes bandas espectrales con diferentes períodos de revisita. Los satélites LandSat-7, LandSat-8 y Sentinel 2 son de libre acceso y por tanto representan una gran ventaja para la realización de este proyecto. Seguido con su amplio espectro de bandas que se muestran en las tablas 3.1 y 3.2.

En la tabla 4.5 las siglas S2A y S2B hacen referencia al satélite Sentinel-2A y Sentinel-2B respectivamente.

El factor decisivo para la elección del satélite Sentinel-2 de LandSat-8 y LandSat-7 como fuente de imágenes satelitales para el proyecto se debe a su período de revisita y a la calidad de imágenes; esto debido a que mientras el LandSat-8 tiene un período de revisión de 16 días, el satélite Sentinel-2 revisita cada 5 días ofreciendo una mayor cantidad de información por mes, además el satélite LandSat-7 presenta líneas blancas en las imágenes (errores de reflexión electromagnética). [18]

Sensor	N° de banda	Nombre	λ central S2A (μm)	$B\omega$ S2A (μm)	λ central S2B (μm)	$B\omega$ S2B (μm)	Resolución (metros)
MSI	1	Aerosol costero	0.44	0.02	0.44	0.02	60
MSI	2	Azul	0.50	0.07	0.49	0.07	10
MSI	3	Verde	0.56	0.04	0.56	0.04	10
MSI	4	rojo	0.66	0.03	0.67	0.03	10
MSI	5	Vegetación borde rojo	0.70	0.02	0.70	0.02	20
MSI	6	Vegetación borde rojo	0.74	0.02	0.74	0.02	20
MSI	7	Vegetación borde rojo	0.78	0.02	0.78	0.02	20
MSI	8	NIR	0.84	0.12	0.83	0.12	10
MSI	8b	NIR estrecha	0.86	0.02	0.86	0.02	20
MSI	9	Vapor de agua	0.95	0.02	0.94	0.02	60
MSI	10	SWIR Cirrus	1.37	0.03	1.38	0.03	60
MSI	11	SWIR	1.61	0.09	1.61	0.09	20
MSI	12	SWIR	2.20	0.18	2.19	0.18	20

Tabla 3.1: Bandas espectrales del satélite Sentinel-2 [18]

LandSat-8					
Sensor	N° de banda	Nombre	λ (μm)	Resolución (m)	Aplicaciones
OLI	1	Costero	0.43 0.45	a 30	Estudios costeros y de aerosoles
OLI	2	Azul	0.45 0.51	a 30	Cartografía batimétrica, que distingue el suelo de la vegetación y la vegetación caducifolia de la vegetación de coníferas
OLI	3	Verde	0.53 0.59	a 30	Destaca los picos de máxima vegetación, que son útiles para evaluar el vigor de las plantas
OLI	4	Rojo	0.63 0.67	a 30	Distingue las laderas de vegetación
OLI	5	NIR	0.85 0.88	a 30	Destaca el contenido de biomasa y las costas
OLI	6	SWIR-1	1.57 1.65	a 30	Distingue la humedad del suelo y de la vegetación; penetra a través de nubes finas
OLI	7	SWIR-2	2.11 2.29	a 30	Mejora de la lectura de la humedad del suelo y la vegetación y la penetración a través de nubes finas
OLI	8	Pancromática (Pan)	0.50 0.68	a 15	Resolución de 15 metros, definición de imagen más nítida
OLI	9	Cirrus	1.36 1.38	a 30	Mejor detección de la contaminación en cirros
TIRS	10	Sensor Térmico Infrarrojo 1 (TIRS 1)	10.60 11.19	a 30	Resolución de 100 metros, mapeo térmico y humedad estimada del suelo
TIRS	11	Sensor Térmico Infrarrojo 2 (TIRS 2)	11.50 12.51	a 30	Resolución de 100 metros, mapeo térmico y humedad estimada del suelo

Tabla 3.2: Bandas espectrales del satélite LandSat-8[18]

Bandas espectrales:

Una banda espectral se refiere a una parte del espectro electromagnético (EM) que es captada por un sensor, luego el espectro electromagnético se divide en varias bandas que se diferencian entre sí por la frecuencia de la señal de luz reflejada. Cada objeto recibe todo el espectro electromagnético, pero refleja o absorbe solo ciertas bandas de este. De esta forma, es posible diferenciar qué tipo de objeto se captura en una imagen analizando las fotos tomadas por cada sensor óptico en la banda respectiva para la que fue diseñado. Actualmente existen 13 bandas espectrales para el satélite LandSat-8 las cuales se han relacionado con diferentes tipos de materiales u objetos que los reflejan con mayor facilidad: [19]

1. **Aerosol costero (0,434 - 0,454 μm):** particularmente útil en estudios costeros, batimétricos y de aerosoles. La banda costera refleja azules y púrpuras y tiene la capacidad de penetrar en el agua (hasta 20-30 m en agua clara). Es menos absorbido por el agua y es más sensible a las nubes, el humo y la bruma, lo que permite observar el color del océano y las aguas costeras. Útil para firmas espectrales para vegetación marina como pastos marinos y otros hábitats bentónicos submarinos, también se utiliza para filtrar nubes en el procesamiento de imágenes.
2. **Azul (0,464 - 0,529 μm):** la luz visible como la banda azul puede alcanzar profundidades de hasta 20-30 metros. ideal para obtener imágenes de aguas profundas, reconocer: humo y nubes.
3. **Verde (0,543 - 0,578 μm):** la banda verde forma parte del espectro visible que cubre el pico de reflectancia de las superficies de las hojas de la planta (de ahí el color verde que vemos). Esto también significa que la reflectancia espectral es baja en las regiones azul y roja del espectro debido a la absorción por la clorofila durante la fotosíntesis. Esta banda es útil para evaluar el vigor de plantas y vegetación en mares, bosques y zonas urbanas.
4. **Roja (0,650 - 0,680 μm):** los suelos tropicales, el entorno construido y las características geológicas generalmente reflejan en gran medida esta banda. El rojo se usa en fórmulas como el índice de vegetación de diferencia normalizada (NDVI) debido a la cantidad de luz roja (y azul) que absorben las plantas. algunas aplicaciones pueden ser el análisis del tipo de suelo y área construida.
5. **Borde rojo (0,704 - 0,783 μm):** la banda del borde rojo está entre la banda roja y la del infrarrojo cercano. Se puede utilizar para analizar la salud de las plantas, ya que las plantas sanas absorben la luz roja, pero reflejan con mucha fuerza el infrarrojo de esta banda. Hay tres sub-bandas en esta banda y cuanto mayor es la longitud de onda, menos se ve afectada por las condiciones atmosféricas (bandas 5, 6 y 7). (bandas 5, 6 y 7).
6. **Infrarrojo cercano 1 - NIR (0,778 - 0,893 μm):** la reflectancia del infrarrojo cercano es una de las formas más poderosas de clasificar la vegetación saludable. Separar clases como el agua y la vegetación siempre es más fácil en la región del infrarrojo cercano. Esto se debe a que las plantas sanas reflejan más energía en el espectro del infrarrojo cercano, mientras que el agua la absorbe. Esta banda se emplea para evaluar cambios sutiles en las plantas de un área. Narrow NIR es una banda más estrecha encapsulada en este mismo rango (bandas 8 y 8b).
7. **Vapor de agua (0,935 - 0,955 μm):** existe una similitud en las características entre el infrarrojo cercano y las bandas de vapor de agua. La principal diferencia es que esta banda se ve menos afectada por las condiciones atmosféricas.
8. **Cirro (1,359 - 1,389 μm)** la banda de cirros se ganó su nombre porque es un especialista en la detección de nubes de cirros. La atmósfera absorbe casi toda la banda de cirros, ya que el suelo es apenas visible. Las nubes de gran altitud que no son visibles en otras bandas se detectan y seleccionan.
9. **Infrarrojos de onda corta 1 - SWIR-1 (1,569 - 1,659 μm):** el infrarrojo de onda corta (SWIR) puede ayudar a discriminar entre suelos secos y húmedos. También se utiliza en la firma espectral para geología y clasificación de suelos. SWIR también es conocido por su capacidad para penetrar nubes delgadas e incluso humo y neblina mejor que las bandas visibles.
10. **Infrarrojos de onda corta 2 - SWIR-2 (2,112 - 2,292 μm):** el infrarrojo-2 de onda corta tiene sus similitudes con el SWIR-1. SWIR-2 se utiliza principalmente para obtener imágenes de tipos de suelo, características geológicas y minerales como cobre y sulfatos. También es sensible a la vegetación y a las variaciones de humedad del suelo. La nieve, el hielo y las nubes característicos aparecen en tonos más oscuros.

Las bandas espectrales tienen la capacidad de operar entre ellas para obtener nuevos resultados, así como de diferenciar por paleta de colores indicando el nivel de reflectancia alcanzado por píxel. Herramientas como Google Earth Engine [20] o LandViewer [21] ofrecen la posibilidad de obtener y tratar imágenes de satélite operando sus bandas, logrando los resultados requeridos.

3.1.3. Trabajos relacionados

- Un estudio realizado en la Universidad Nacional Autónoma de México por Leonardo Arias bajo la dirección de Rodrigo Takashi (Magíster en Ingeniería) en agosto de 2014 habla sobre la estimación de turbidez mediante teledetección en embalses del estado de Aguascalientes en México, específicamente en el Represas Niágara y presidente Calles donde utilizan métodos de teledetección a través de imágenes satelitales para obtener un mapa de turbidez y estado trófico. Adicionalmente, el autor toma muestras de agua de 250 ml en la presa presidente Calles para realizar estudios de turbidez utilizando un turbidímetro para corroborar los datos obtenidos en la teledetección. [22]
- La Universidad Central de Venezuela realizó un estudio similar por parte de Jacinto Artigas y José Luis López en noviembre de 2016 titulado “Estimación de la concentración de sedimentos en el río Orinoco mediante sensores remotos”. En este estudio, los autores mencionan que para el estudio se tiene en cuenta la profundidad del río y se realizan correcciones atmosféricas a la imagen satelital. Como resultado, comparan las concentraciones de sedimentación calculadas con las medidas en diferentes tramos del río Orinoco con las obtenidas a través de imágenes satelitales, concluyendo que la relación obtenida entre reflectancia y concentración de sedimentos fue insatisfactoria debido a que la reflectancia se ve afectada por otras variables que deben ser identificadas y ponderadas. [23]
- Un artículo de la revista Remote Sensing publicado el 20 de noviembre de 2018 por Q. V. Pham, N. T. T. Ha, N. Pahlevan, L. T. Oanh, T. B. Nguyen y N. T. Nguyen menciona que hacen uso de las imágenes del satélite LandSat-8 para cuantificar la concentración de sedimentos en el Río Rojo en Vietnam ya que existen altos niveles de turbidez debido a la actividad campesina y la alta pluviosidad en el sector. Los autores mencionan que las correcciones atmosféricas se realizan primero a las imágenes de satélite utilizando el método LaSRC. Luego encuentran que las bandas espectrales 2, 3, 4 y 5 responden a la reflectancia proveniente de las aguas del río, especialmente las bandas 3 y 4, por lo que a través de diferentes configuraciones entre bandas espectrales encuentran que la operación $(B4 + B5) / B3$ presenta el error cuadrático medio más pequeño (0.305) en comparación con la entrega del resultado de sedimentación medido en los 40 puntos de estudio ubicados a lo largo del río. [24]
- Estudio publicado en la Revista Internacional de Ciencia y Tecnología Ambiental, realizado por Torres-Bejarano, Arteaga-Hernández, Rodríguez-Ibarra, Mejía-Ávila¹ y González-Márquez, profesionales de la Universidad de Córdoba (Colombia) y la Universidad Autónoma de Occidente (México). En este estudio se utiliza la teledetección a través de imágenes satelitales para monitorear la calidad del agua en un complejo de humedales en el norte de Colombia, en el bajo Sinú, se utiliza: imágenes del satélite Sentinel 2, sistemas de información geográfica (GIS) y modelos empíricos para estimar la temperatura, conductividad eléctrica, la turbidez y el pH del agua del humedal. Los resultados indican una correlación entre los modelos utilizados y las variables medidas físicamente con un valor mínimo de R^2 de 0,6 (para temperatura) y en el caso de sedimentación, de 0.88 logrando evaluar los niveles de estas variables de estudio con considerable confianza. [25]

3.2. Metodología

3.2.1. Tipo de estudio

Es una investigación aplicada como tesis de pregrado.

3.2.2. Ocupaciones

- Objetivo: Construir una base de datos que contenga índices de reflectancia obtenidos de imágenes de satélite y datos de sedimentación medidos en las estaciones de muestreo.

Actividad 1: Solicitar los datos de turbidez tomados en la estación de Milán y en la captación de PM a los encargados de estas dos estaciones de medición.

Actividad 2: Descargar los datos de reflectancia de cada banda del satélite Sentinel-2 y adjuntar a la base de datos que contiene los datos de turbidez.

Actividad 3: Ajustar la ventana de tiempo para poder juntar estos 2 tipos de datos.

Actividad 4: Descartar datos de reflectancia que según por su imagen satelital, no se puedan utilizar.

Entregable: base de datos en Excel que muestra el nivel de turbidez según la fecha requerida.

- Objetivo: Aplicar algoritmos de regresión en la base de datos construida para predecir los niveles de sedimentación en el río Cauca.

Actividad 1: Encontrar la relación de bandas adecuada que tenga la correlación más alta.

Actividad 2: Probar diferentes algoritmos de regresión en la base de datos que permitan predecir niveles de turbidez.

Entregable: Información gráfica y/o numérica sobre la turbidez del río Cauca en la imagen tomada y resultados de los métodos de regresión utilizados para predecir el valor de turbidez como tal.

- Objetivo: Evaluar los algoritmos de regresión aplicados en términos de métricas como el error cuadrático medio (RMSE) y R^2

Actividad 1: Compare las mediciones de turbidez puntual de Milán y los resultados de los modelos de combinación lineal de bandas espectrales.

Entregable: valores RMSE y R^2 para los algoritmos de regresión aplicados.

3.3. Selección de información

3.3.1. Selección de datos por imágenes satelitales

Para obtener la base de datos de las imágenes, se utiliza Google Earth Engine[20] debido a que es Open Access. En esta herramienta, se colocan un total de 50 puntos a lo largo del río en las cercanías de la PTAP de Puerto Mallarino como se ve en la figura 3.2. Es de aclarar que esta figura es representativa, pues son 50 puntos los que en realidad se utilizaron.



Figura 3.2: Puntos de referencia en el río Cauca.

Luego para condensar los datos de todos estos puntos se obtiene la media ya que es menos propensa a errores de valores extremos a comparación del promedio.

El script que se programó retorna un .csv con los datos de reflectancia de cada banda con una nubosidad del 20 % para no obtener demasiadas imágenes afectadas por este problema. Además, se descarga la imagen como tal para realizar un proceso de descarte manual dado el caso de sombra en distintos puntos alrededor del río. El recuadro verde en la figura 3.3 indica el área de interés.



Figura 3.3: Área de interés río Cauca

Al visualizar las imágenes se observa que el filtro de nubosidad utilizado en la plataforma de Google Earth Engine toma el 20 % de nubosidad del total de la imagen satelital y está de por sí es más grande que el área de interés, por lo que algunas imágenes están en su totalidad cubiertas por nubes o sombras. Esto se ejemplifica mejor en la figura 3.4.

Por ello se decide remover este filtro y se realiza una depuración manual con todas las imágenes registradas desde el 2015 (lanzamiento del satélite) hasta la actualidad de los datos de EMCALI obtenidos (2020). De 260 imágenes son útiles sólo 36 de tipo “Ideal” llegando a 45 imágenes sumando las de tipo “Posible”.

Para realizar el modelo matemático se toman en cuenta solo los casos ideales, ya que se puede asegurar que la representación los datos entregados son fiables.

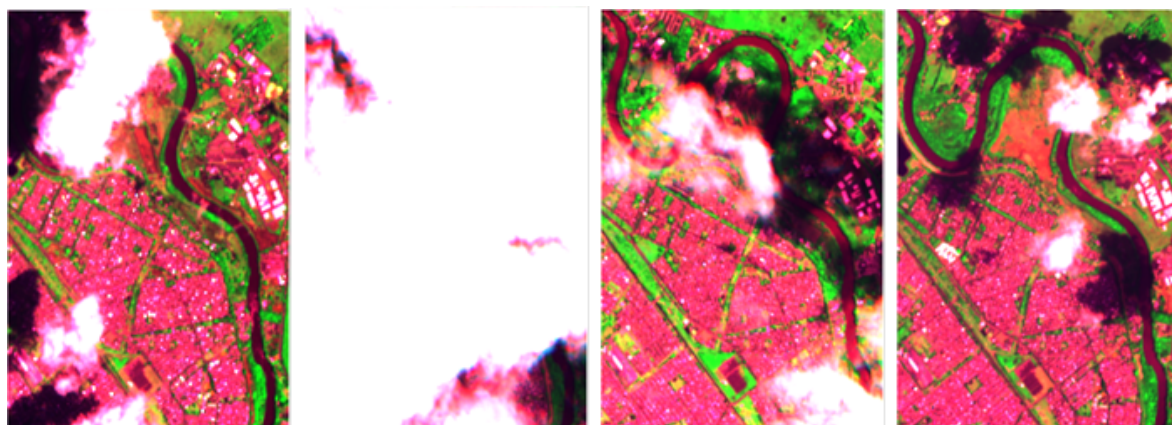
Los datos relacionados a la reflectancia se guardan junto a los de turbidez de la planta de Puerto Mallarino y punto Milán en la misma hoja de datos. Se adjunta el dato real de turbidez de cada fecha de la imagen satelital tomada para proceder luego en el software Matlab a realizar el modelo matemático.

3.3.2. Selección de datos por EMCALI

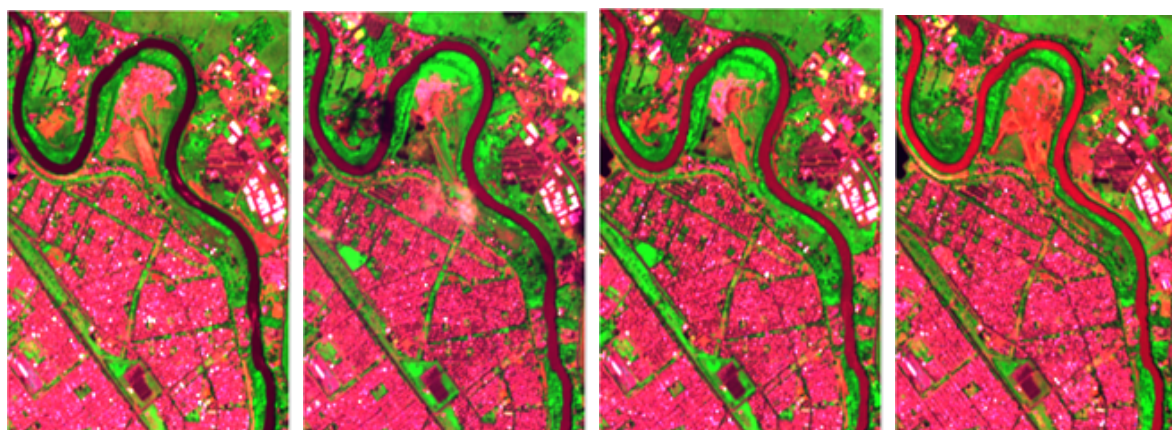
El ingeniero Juan Carlos Escobar compartió datos de turbidez desde el año 2013 hasta noviembre del 2020 de la PTAP de Puerto Mallarino y monitoreo Milán que en palabras de él “Los datos solo tienen un desfase de 3 horas aproximadamente entre la PTAP de Puerto Mallarino y punto Milán”.

El dato de turbidez se toma cada minuto las 24 horas del día los 7 días a la semana, por lo que se toma el promedio de un día para así ajustar la ventana de tiempo a los datos retornados por las imágenes satelitales.

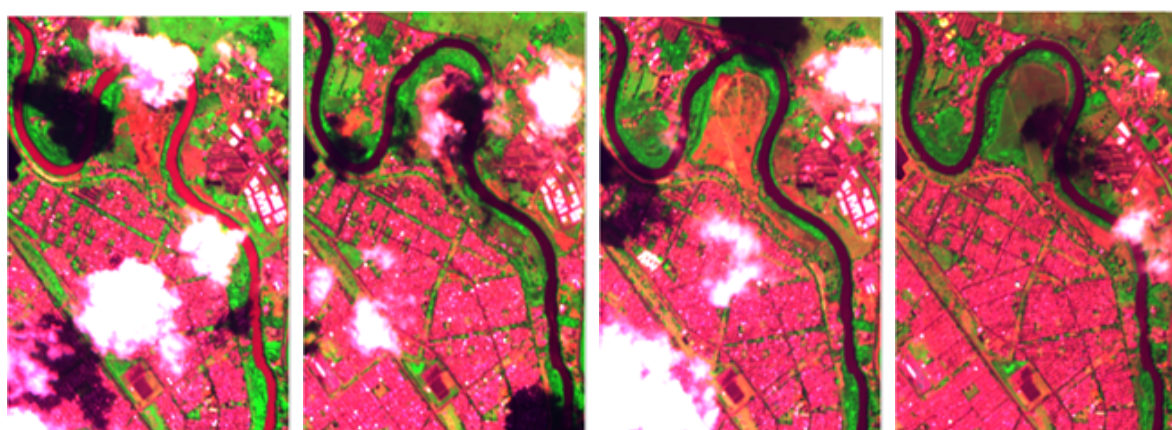
Los datos de turbidez de punto de muestreo Milán presentan inconsistencias en un mismo día e incluso en minutos lo cual da pie a errores o a conjeturas erradas. Por ello, se prefiere excluir todos los datos de este sitio debido a la alta variabilidad presentada. Esto se ve representado en la Figura 3.5.



(a) Nube tapando gran parte del río. No se toma en cuenta. (b) Gran nubosidad. No se toma en cuenta. (c) Baja nubosidad genera sombras no deseadas sobre el río. No se toma en cuenta. (d) Varios puntos de sombra sobre el río. No se toma en cuenta.



(e) Caso ideal. Baja turbidez por observación. (f) Caso ideal. Media turbidez por observación. (g) Caso ideal. Media turbidez por observación. (h) Caso ideal. Alta turbidez por observación.



(i) Caso posible. Puntos de nubosidad sobre el río. (j) Caso posible. Puntos de sombra sobre el río. (k) Caso posible. Sombra única sobre sector del río. (l) Caso posible. Punto de sombra y nube sobre el río.

Figura 3.4: Casos de imagen satelital en área de interés PTAP Puerto Mallarino.

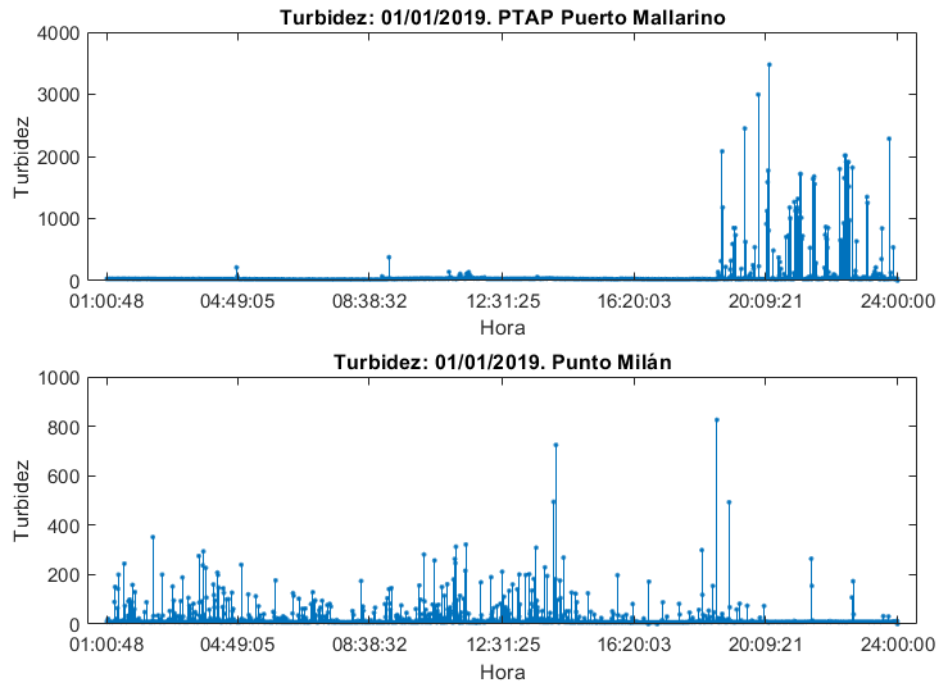


Figura 3.5: Comparación turbidez en Puerto Mallarino y punto Milán.

En la Figura 3.5 se observa que para la turbidez de 1 día en la bocatoma de Puerto Mallarino se tiene una tendencia estable con alguna variabilidad. Mientras que para el punto Milán está variabilidad se presenta durante todo el día.

Esta situación en el punto de muestreo Milán puede ser debido a las aguas residuales del canal sur que son depositas en esta sección del río.

Capítulo 4

Resultados

4.1. Construcción del modelo

Para lograr la predicción de turbidez con bajos índices de error, se plantean 2 modelos: modelo único y modelo doble.

4.1.1. Modelo Único

En los trabajos consultados en el capítulo 3.1.3 no se indican como se obtienen los modelos de predicción basados en bandas de reflectancia; Solo indican que se basaron en la correlación de las bandas satelitales y los datos reales de turbidez (medición in situ). Por lo que, para iniciar el modelo matemático de este trabajo, se usan los coeficientes de correlación como punto de referencia.

Los coeficientes de correlación utilizados son el coeficiente de Kendall, el coeficiente de Spearman y el coeficiente de Pearson de cada banda con respecto al dato real de turbidez los cuales indican una medida basada en rangos para los dos primeros y una medida lineal respectivamente [26]. Las fórmulas correspondientes se indican en las ecuaciones (4.1 - 4.2 - 4.3).

$$Rho_{Pearson}(x, y) = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad (4.1)$$

$$Rho_{Spearman}(x, y) = 1 - \frac{6 \sum d_i^2}{n^3 - n} \quad (4.2)$$

$$Rho_{Kendall}(x, y) = \frac{2K}{n(n-1)} \quad (4.3)$$

Donde en orden de aparición:

- σ_{xy} : Covarianza de x,y.
- σ_x, σ_y : Desviación típica de x y y respectivamente.
- d_i : Vector de diferencias entre x y y.
- n : Longitud de datos.
- K : Diferencia entre la suma del conteo de todos los mayores a $Y(i)$ para cada $Y(j > i)$ y la suma del conteo de todos los menores a $Y(i)$ para cada $Y(j > i)$.

Para reducir la complejidad de estas ecuaciones, se utiliza la función “corr” de MATLAB [26] la cual retorna 2 valores: el Rho que es el valor de la correlación como tal y el Pval que es una medida de que tan confiable es este Rho. Un Pval cercano a 0 indica que el Rho entregado es un valor confiable mientras que un Pval cercano a 1 indica todo lo contrario.

Por facilidad de nomenclatura:

- RhoP significa Rho-Pearson. PvalP significa Pval – Pearson.
- RhoS significa Rho-Spearman. PvalS significa Pval – Spearman.

- RhoK significa Rho-Kendall. PvalK significa Pval – Kendall.

En la Figura 4.1 se muestra los valores de las correlaciones obtenidas de cada banda individual con el valor real de turbidez:

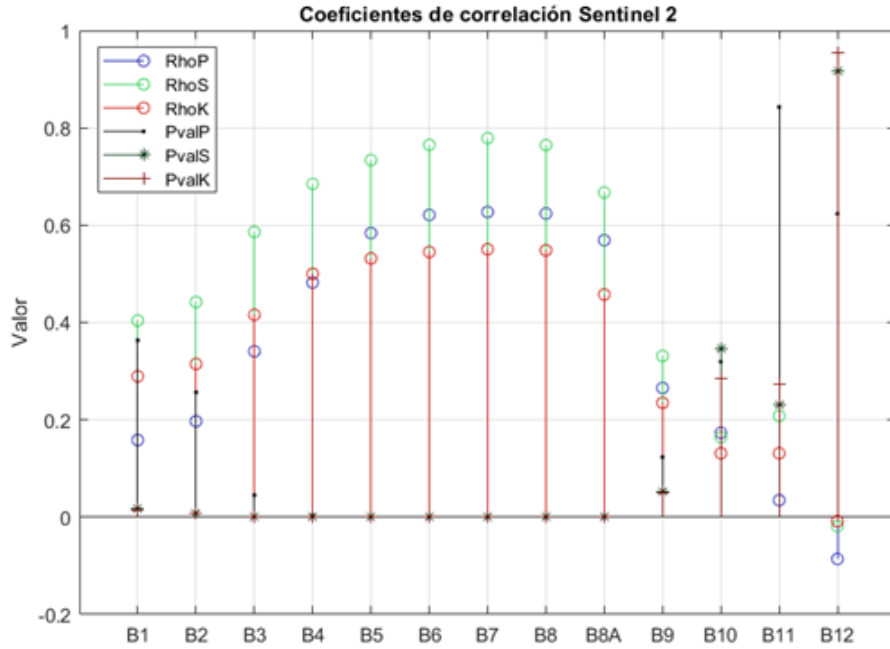


Figura 4.1: Coeficientes de correlación de las bandas del Sentinel-2

Como se puede observar las bandas B7, B6 y B8 presentan la más alta correlación de Spearman, por lo que tiene sentido que el modelo final este relacionado con estas bandas.

Para confirmar el enunciado anterior, se desarrolla un script llamado “RelacionBanda”[27] el cual toma las bandas desde B2 hasta a la B12 con el fin de obtener un modelo acertado.

El script consiste en una serie de ciclos anidados con el objetivo de combinar todas las bandas posibles quedándose con los valores más altos de correlación. Por observaciones empíricas se denota que los datos tienden a tener un coeficiente de Pearson más alto que el coeficiente Kendall por lo que la condición de triunfo será que la correlación tenga un valor más alto que 0.8.

El script tiene 3 salidas como se muestra en la ecuación 4.4:

$$\begin{aligned}
 & 1. \frac{B_{N1}}{B_{N2}} \\
 & 2. \frac{B_{N1}}{B_{N2}} + \frac{B_{N3}}{B_{N4}} \\
 & 3. \frac{B_{N1}}{B_{N2}} + \frac{B_{N3}}{B_{N4}} + \frac{B_{N5}}{B_{N6}}
 \end{aligned} \tag{4.4}$$

Cada B_{Ni} es un ciclo, por lo que entre más términos tenga la relación final más tiempo de procesamiento tendrá el código.

Para solucionar este imprevisto la relación de 3 términos de salida se ingresa a otro script llamado “MasRelacionBanda”[28] que toma como base esta relación y añade dos términos más sin necesidad de aumentar significativamente el tiempo de procesamiento. Por lo que al final se puede obtener una relación de bandas de infinitos términos.

Además, para obtener las constantes se construye otro script, “CTERelacion”[29] similar a los dos anteriores scripts, solo que toma cada término y lo multiplica con un vector de constantes aleatorias determinados que busca aumentar aún más el valor de correlación. La relación final obtenida es la que se muestra en la ecuación 4.5.

$$Relacion_{Banda} = \frac{1,75 * B_2}{B_5} + \frac{1,75 * B_6}{B_{8A}} + \frac{1,5 * B_{12}}{B_5} \tag{4.5}$$

Tipo	Resultado
Rho - Pearson	0.91528
Pval - Pearson	1.3927e-14
Rho - Spearman	0.91317
Pval - Spearman	0
Rho - Kendall	0.73445
Pval - Kendall	2.1801e-12

Tabla 4.1: Índices de correlación resultado de comparación con modelo.

Los resultados de los índices de correlación se indican en la tabla 4.1.

Con la relación obtenida, se hace un diagrama de dispersión en Excel con los datos de correlación y los datos de EMCALI de turbidez Figura 4.2. Luego se agrega la ecuación de tendencia lineal para obtener el modelo final que predice la turbidez según los datos de reflectancia dados por cada banda del satélite Sentinel-2.

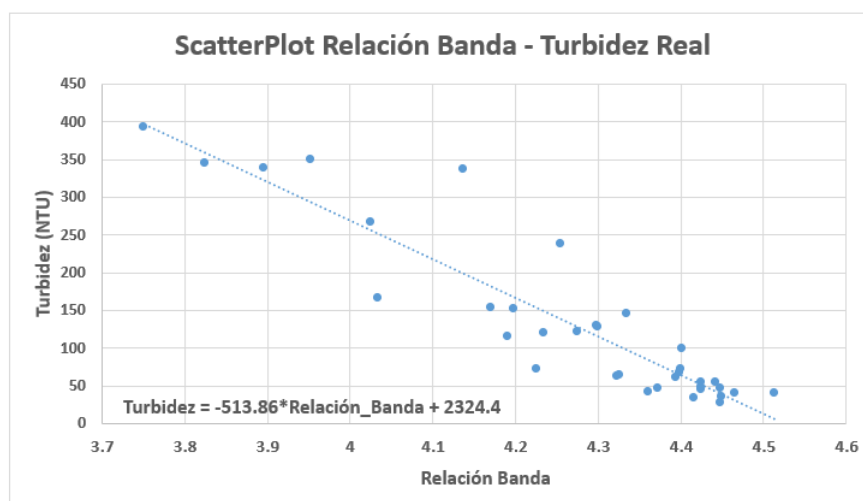


Figura 4.2: Relación de banda vs Datos de turbidez reales

La ecuación 4.6 con la que se obtiene los valores correspondientes a la turbidez es la siguiente:

$$Turbidez = -513,86 * Relacion_{Banda} + 2324,4 \quad (4.6)$$

Con la ecuación obtenida (4.6) se tiene la gráfica perteneciente a la Figura 4.3 que compara los datos reales de turbidez y los resultados de turbidez calculados.

Basado en el análisis de la Figura 4.3, los datos relevantes para evaluar el correcto funcionamiento del modelo se consignan en la tabla 4.2. En esta tabla se observa que el error porcentual promedio (promedio de cada error porcentual de los datos individuales) es del 30.17% y en llamada con el Ing. Juan Carlos Escobar jefe de operación de la PTAP de Puerto Mallarino indica que el error máximo que puede tener un dato si se predice debe ser del 10%.

Como se puede apreciar en la Figura 4.2, el rango de datos de turbidez para la construcción se encuentra entre los 26.88 NTU y los 392.42 NTU. Esto es debido a la poca cantidad de datos con los que se trabaja, pues no se tienen valores de turbidez cercanos a los 3000 NTU con su contraparte en imágenes satelitales para trabajar.

Para reducir este porcentaje de error, se observa que en la gráfica 4.2 a partir de 150 NTU se comienza a tener una alta dispersión a comparación de niveles bajos de turbidez. Por lo que se decide dividir este modelo en dos partes: una para niveles altos turbidez y otra para niveles bajos.

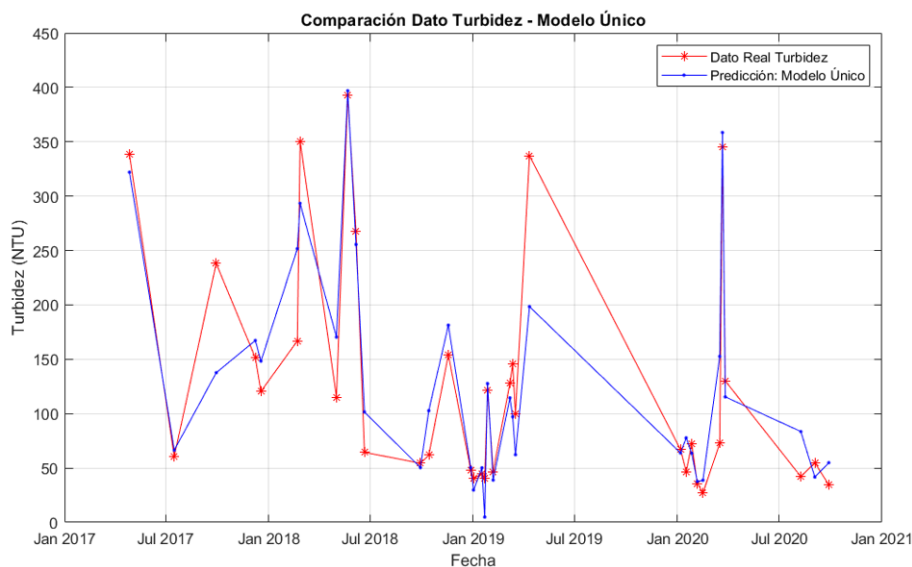


Figura 4.3: Fecha de datos de turbidez vs Turbidez

Métrica	Resultado
RMSE	42.717
R^2 Ordinary	0.84062
R^2 Adjusted	0.83579
Error promedio porcentual	30.67%

Tabla 4.2: Datos de error del modelo único

4.2. Modelo Doble

El modelo doble se crea a partir de la idea de reducir el error promedio de cada dato teniendo un modelo específico para datos de turbidez mayores de 150 NTU y otro para menores o iguales a 150 NTU. Este punto de cambio específico se da al analizar el scatter plot de los datos.

Para tener el modelo doble, se realiza el mismo procedimiento de obtención del modelo único, solo que primero se toman los datos de baja turbidez (menor a 150 NTU) y luego los de mayor turbidez (mayor o igual a 150 NTU) para obtener la relación con mayor correlación en cada caso (4.7).

$$Relacion_{Bajo} = \frac{2,5 * B_2}{B_5} + \frac{3 * B_{12}}{B_6} + \frac{2 * B_{11}}{B_5} + \frac{1,5 * B_5}{B_7} + \frac{2 * B_3}{B_4} \quad (4.7)$$

$$Relacion_{Alto} = \frac{0,75 * B_2}{B_5} + \frac{0,75 * B_9}{B_2} + \frac{3 * B_4}{B_5}$$

Los resultados de los índices de correlación se indican en la Tabla 4.3.

Tipo	Relación Bajo	Relación Alto
Rho - Pearson	0.7658	0.88021
Pval - Pearson	0.006001	0.00034997
Rho - Spearman	0.85455	0.9
Pval - Spearman	0.0016492	0.00022532
Rho - Kendall	0.67273	0.74545
Pval - Kendall	0.0031063	0.00075953

Tabla 4.3: Índices de correlación en modelo doble.

Con la relación de banda de turbidez baja y alta, se consigue la ecuación (4.8) para el cálculo de turbidez baja y la ecuación (4.9) para el cálculo de turbidez alta del mismo modo que el modelo único:

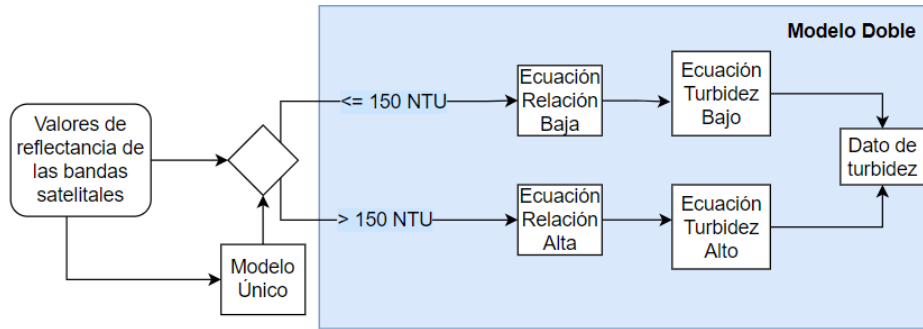


Figura 4.4: Digrama de flujo cálculo de turbidez con el modelo doble

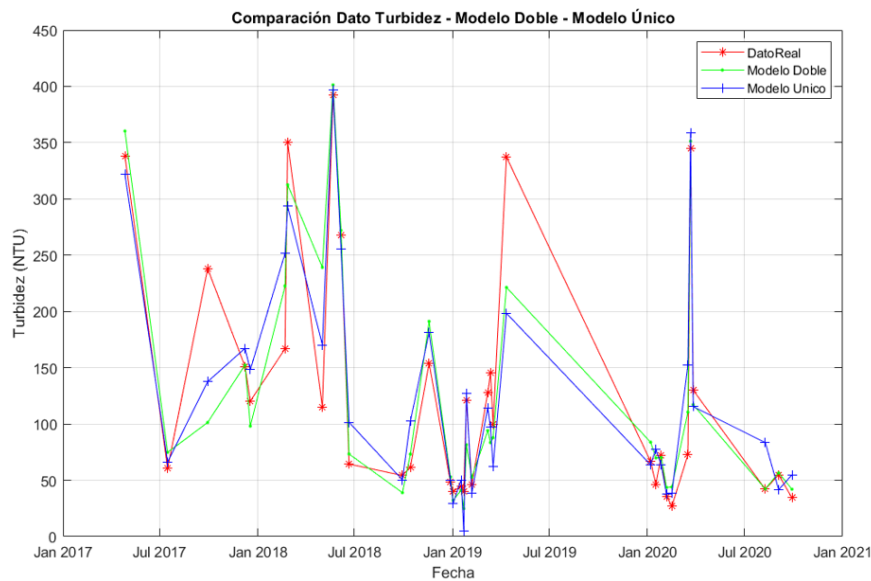


Figura 4.5: Comparación Turbidez real vs predicción modelo único y modelo doble

$$Turbidez_{Baja} = -93,329 * Relacion_{Bajo} + 955,12 \tag{4.8}$$

$$Turbidez_{Alta} = -686,74 * Relacion_{Alto} + 2880,9 \tag{4.9}$$

Para alternar entre modelos de turbidez alta y baja, el criterio de cambio de modelo lo tiene el modelo único ya que a pesar de que presenta fallas de predicción individual, es preciso en decir si la turbidez supera el límite dicho. Para entender esto un poco mejor, se muestra el diagrama de flujo del modelo doble en la Figura 4.4

Con los modelos obtenidos, se realiza la comparación grafica (Figura 4.5) de los datos reales de turbidez, turbidez calculada del modelo doble y turbidez calculada del modelo único. Luego los datos relevantes para evaluar el modelo doble frente al modelo unico y el dato real se consignan en la tabla 4.4.

Métrica	Modelo Único	Modelo Doble
RMSE	42.717	43.1432
R^2 Ordinary	0. 84062	0.84065
R^2 Adjusted	0. 83579	0.83582
Error promedio porcentual	30.67 %	23.98 %

Tabla 4.4: Comparación datos de error: Modelo único y modelo doble

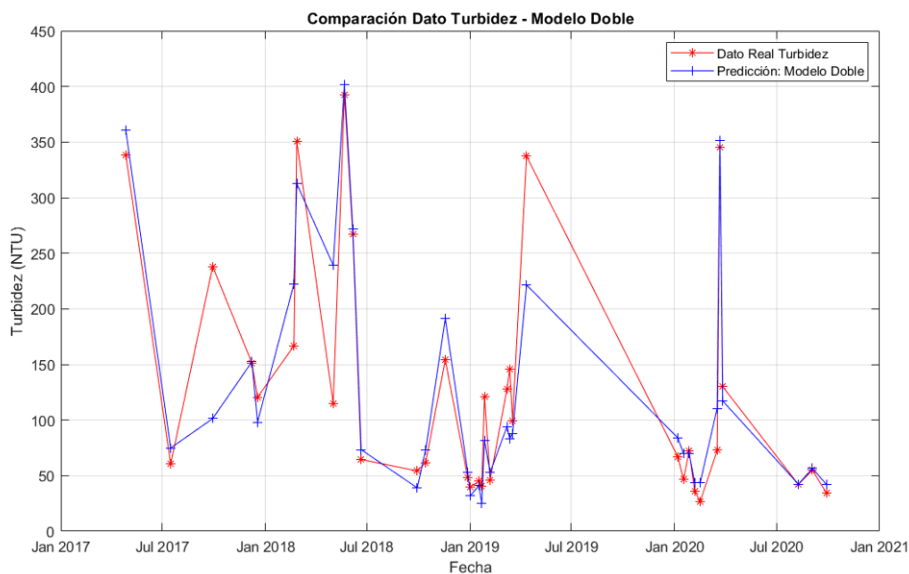


Figura 4.6: Comparación predicción modelo doble y dato real de turbidez

Métrica	Modelo Único	Modelo Doble
RMSE	107.59	102.27
R^2 Ordinary	0.562	0.624
R^2 Adjusted	0.555	0.618
Error promedio porcentual	37.10 %	27.99 %

Tabla 4.5: Datos de error para la validación del modelo.

La gráfica 4.6 muestra los datos reales de turbidez con la turbidez calculada con el modelo doble indicando una similitud mayor y una mejor adaptación a los picos junto a una mejora en el error porcentual promedio. El modelo doble mejora aspectos del modelo único, por lo que este es el modelo final de este trabajo.

4.3. Validación del modelo

No se ha tenido acceso a los datos turbidez en la bocatoma de la PTAP de Puerto Mallarino a partir de noviembre del 2020 pues estos no han sido compartidos por EMCALI. Por lo que no se puede hacer una validación detallada y actualizada.

Sin embargo, se tienen los datos de las imágenes con etiqueta “posible” que aunque no fueron usadas para la construcción del modelo porque presentan pequeñas porciones de nubes o sombras a lo largo del río, no son lo suficientemente invasivas como para descartarlas por completo.

La gráfica que se muestra en la Figura 4.7 indica el resultado del cálculo de turbidez utilizando el modelo doble para estas imágenes catalogadas como posibles sumadas a las imágenes utilizadas para la construcción del modelo y su comparación con los datos reales proporcionados por EMCALI hasta el 2019.

Para la gráfica de la Figura 4.7 se tiene un dato de 1086.2 NTU el cual se encuentra fuera del rango de construcción del modelo añadiendo error a las métricas utilizadas. Para este tipo de datos no se puede realizar un reajuste del modelo porque es solo un dato. Si se tuvieran más datos cercanos a los 3000 NTU o que ampliaran el rango de construcción se podría pensar en un reajuste.

Los datos de error para ambos modelos son consignados en la Tabla 4.5. En la cual se observa que el modelo doble sigue siendo mejor en el sentido de las métricas utilizadas para la predicción de turbidez.

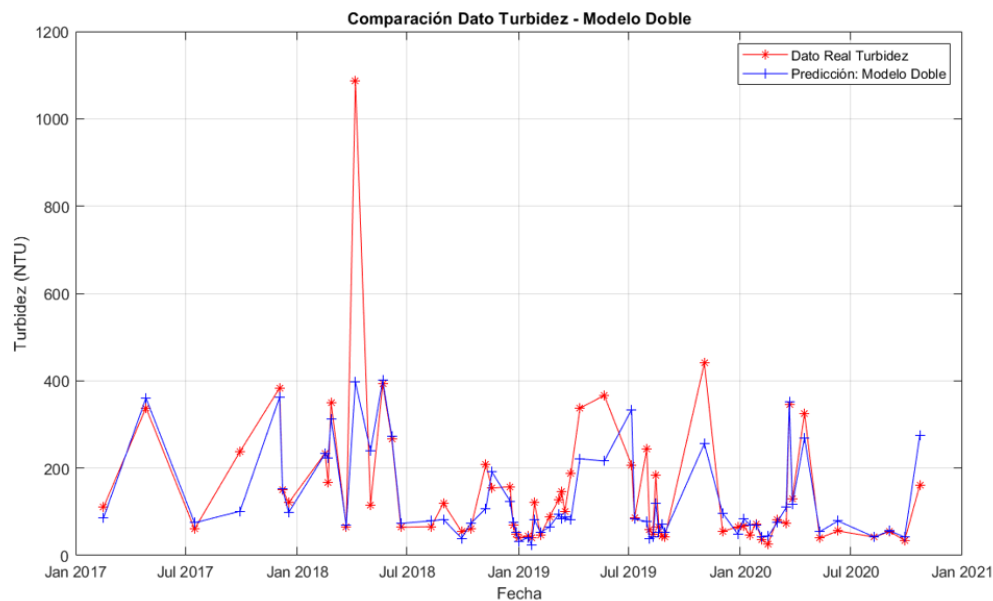


Figura 4.7: Validación del modelo doble con los datos reales de turbidez.

4.4. Mejora del Modelo

4.4.1. Recolección de datos

El modelo final de este trabajo se creó utilizando las imágenes satelitales del satélite Sentinel-2 del 2017 a noviembre del 2020 debido a que es en este periodo de tiempo que se tienen los datos reales de turbidez de la PTAP de Puerto Mallarino.

- De 262 imágenes satelitales consultadas, se eliminaron 50 por no tener contraparte con datos reales de turbidez de EMCALI.
- De 212 imágenes satelitales con dato real de turbidez, solo se utilizaron 66 para la validación del modelo, ya que tenían baja nubosidad.
- De 66 imágenes satelitales con dato real de turbidez y baja nubosidad, solo se usaron 35 imágenes satelitales para la construcción del modelo ya que presentaban la nubosidad más baja.

Para mejorar el modelo en este aspecto, se pueden agregar a la base de datos las imágenes y los datos de turbidez del año 2021 ya que no fueron utilizados para el modelo debido a que no se compartieron por parte de EMCALI. Así mismo es de resaltar que la alta nubosidad que presenta la región dificultó un poco la recolección adecuada de datos.

4.4.2. Uso de imágenes de satélites de revisita diaria

El satélite Sentinel-2 tiene un periodo de revisita cada 5 días, lo que indica que al año habrá 73 imágenes aproximadamente. Existen otros satélites que tienen periodos de revisita más pequeños como el conjunto de satélites SuperView1 que tiene un periodo de 97 minutos [15] o el satélite Kompsat3/3A con periodo de revisita de 1.3 días [17].

Estos satélites a pesar de que no cuentan con las mismas bandas que tiene el Sentinel-2, se puede realizar la recalibración del modelo con las bandas existentes, ya que la cantidad de datos aumentaría considerablemente.

4.4.3. Uso de cámara multispectral

El uso de una cámara multispectral en un principio se saldría del tema de este trabajo de grado, ya que omitiría el uso de bandas satelitales.

Esta cámara bien puede ser instalada en un sector planificado del río, descargar los datos recolectados y realizar la calibración del modelo según las bandas espectrales que pueda leer. Esto eliminaría la problemática de la nubosidad y la frecuencia de los datos.

Por otro lado, se incluiría un gasto de energía eléctrica para el mantenimiento de este dispositivo, más la protección necesaria para que sobreviva a la intemperie.

Capítulo 5

Conclusiones

- Los índices de reflectancia consignados en la base de datos además de ser útiles para este proyecto, con el tiempo y con mayor acumulación puede convertirse en una base robusta para numerosos estudios relacionados al río Cauca en la zona de la planta Puerto Mallarino. Aplicaciones como la medición de oxígeno disuelto que es otra de las causantes de cierre en la bocatoma pueden ser predecidas del mismo modo que la turbidez usando la reflectancia de las imagenes satelitales.
- Con una base de datos de 36 items, se logra construir dos modelos de predicción, uno para turbidez baja y otro para turbidez alta que si bien en un principio se pensaba utilizar un modelo único para la predicción de turbidez, esta idea fue descartada debido a la alta dispersión que se presenta para valores de turbidez altos debido a la poca cantidad de datos siendo esta la causante de la división del modelo. Sin embargo, el modelo único a pesar de fallar en predecir el valor específico de turbidez, es el que indica si el dato es mayor o menor a 150 NTU que es cuando se produce el cambio en el modelo doble siendo menores a 150 NTU para el uso del modelo de turbidez baja y mayores a 150 NTU para el uso del modelo de turbidez alta. Los scripts utilizados para la elaboración de este trabajo junto a las bases de datos se encuentran en el enlace a GitHub: <https://github.com/Victordanco/TrabajodeGrado>
- Para llegar a los modelos de las ecuaciones 4.8 y 4.9 se pasaron por métricas como el RMSE, R^2 y el error porcentual promedio. Los resultados finales se pueden ver consignados en la tabla 4.4. Según el ingeniero Juan Carlos Escobar el error porcentual no puede superar el 10 % sin embargo el logrado a lo largo de este documento en la validación de datos alcanza el 27.99 % que a pesar de no estar dentro de lo valido, es suficiente para la baja cantidad de datos obtenidos en la construcción del modelo.

Bibliografía

- [1] UNESCO, *AGUA Y CAMBIO CLIMÁTICO*. París: Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura, 2020.
- [2] Superintendencia de Servicios Públicos Residenciales, *ANEXO RESOLUCIÓN No. SSPD - 20101300048765 DEL 14-12-2010*, Bogotá, 2010.
- [3] EMCALI, “Planta de tratamiento de agua potable Puerto Mallarino,” Disponible en <https://www.youtube.com/watch?v=j3kWjYNF1XQ&featu> (Consultado: 2020/08/26), Cali - Colombia, 2009.
- [4] Dirección Técnica de Gestión de Acueductos, Alcantarillado y Saneamiento, “Evaluación integral de proveedores empresas municipales de CALI EICEESP - EMCALI EICEESP,” pp. 34–104, 2018.
- [5] Contraloría General de la República de Santiago de Cali, “Informe final de auditoría gubernamental con enfoque especial integral articulado al abastecimiento de agua potable y su medición - vigencia 2013-2017,” 2018.
- [6] World Health Organization, “Guidelines for drinking-water quality: fourth edition incorporating the first addendum,” 2017.
- [7] V. Córdoba and I. Dallos, “Análisisreportes,” Disponible en: <https://github.com/Victordanco/TrabajodeGrado/blob/main/AnálisisReportes.m>, (Código Matlab).
- [8] Escuela de Ingeniería de Recursos Naturales, “Caracterización de los ríos tributarios del río Cauca,” 2003.
- [9] Palacios, L. Pellat, F. Oropeza-Mota, J. Figueroa, Benjamín y Ménez, Mario y Ortiz, Carlos y Exebio-García, Adolfo, “Clasificador genérico de objetos en imágenes ETM +,” *Agrociencia*, vol. 40, no. 5, pp. 613–626, 2006.
- [10] Universidad de Murcia, “Apuntes de clase para fundamentos físicos de la teledetección,” *Licenciatura en Geografía*, 2005 - 2006.
- [11] M. Immitzer, F. Vuolo, C. Atzberger, “First experience with sentinel-2 data for crop and tree species classifications in central europe»,” *Remote Sensing*, 2016.
- [12] Earth Observing System(EOS), “Landsat 7,” Disponible en: <https://eos.com/find-satellite/landsat-7/>(Consultado: 2020/10/28), (Pagina web).
- [13] Earth Observing System (EOS), “Landsat 8,” Disponible en: <https://eos.com/find-satellite/landsat-8/>(Consultado: 2020/10/28), (Pagina web).
- [14] Earth Observing System(EOS), “Sentinel-2,” Disponible en: <https://eos.com/find-satellite/sentinel-2/>(Consultado: 2020/10/28), (Pagina web).
- [15] Earth Observing System (EOS), “Superview1,” Disponible en: <https://eos.com/find-satellite/superview-1/>(Consultado: 2020/10/28), (Pagina web).
- [16] Earth Observing System(EOS), “Pléiades,” Disponible en: <https://eos.com/find-satellite/pleiades-1/>(Consultado: 2020/10/28), (Pagina web).
- [17] Earth Observing System (EOS), “Kompsat 3/3a,” Disponible en: <https://eos.com/find-satellite/kompsat-3-3a/>(Consultado: 2020/10/28), (Pagina web).

- [18] Earth Observing System(EOS), “Sistema de observación de la tierra,” Disponible en: <https://eos.com/es/>(Consultado: 2020/10/18), (En línea).
- [19] GISGeography, “Spectral signature cheatsheet - spectral bands in remote sensin,” Disponible en: <https://gisgeography.com/spectral-signature/>(Consultado: 2020/10/18), (En línea).
- [20] Google Earth Engine, “Platform,” Disponible en: <https://earthengine.google.com/platform/>(Consultado: 2020/10/30), (En línea).
- [21] Land Viewer, “My landviewer,” Disponible en: https://eos.com/products/landviewer/user-guide/my_landviewer/(Consultado: 2020/10/29), (En línea).
- [22] L. Arias, “Estimación de turbidez mediante percepción remota en depósitos del estado de Aguascalientes,” *México: Universidad Nacional Autónoma de México*, 2014, biblioteca central: Tesis digitales.
- [23] J. Artigas, JL López, “Estimación de la concentración de sedimentos en el río Orinoco,” *Caracas: Universidad Central de Venezuela*, conferencia, Jornadas de Investigación Facultad de Ingeniería JIFI 2016 UCV.
- [24] Q. V. Pham, N. T. T. Ha, N. Pahlevan, L. T. Oanh, T. B. Nguyen, N. T. Nguyen, “Using Landsat-8 images for quantifying suspended sediment concentration in Red River,” *Remote Sens*, vol. 10, no. 11, p. 1841, 2018.
- [25] F. Torres-Bejarano, F. Arteaga-Hernández, D. Rodríguez-Ibarra, D. Mejía-Ávila¹, L. C. González-Márquez, “Water quality assessment in a wetland complex using Sentinel 2,” *International Journal of Environmental Science and Technology*, vol. 18, p. 2345–2356, 2020.
- [26] Matlab R2021b, “Corr: Linear or rank correlation references,” Disponible en: https://la.mathworks.com/help/stats/corr.html#mw_1b19e0d5-7906-4577-a0a5-b20311da7faf(Consultado: 2021/03/15).
- [27] V. Córdoba, I. Dallos., “«relacionbanda» script de los autores,” Disponible en: <https://github.com/Victordanco/TrabajodeGrado/blob/main/RelacionBanda.m>, (Código Matlab).
- [28] V. Córdoba, I. Dallos, “«masrelacionbanda» script de los autores,” Disponible en: <https://github.com/Victordanco/TrabajodeGrado/blob/main/MasRelacionBanda.m>, (Código Matlab).
- [29] V. Córdoba, I. Dallos., “«cterelacionbanda» script de los autores,” Disponible en: <https://github.com/Victordanco/TrabajodeGrado/blob/main/CTERelacionBanda.m>, (Código Matlab).