

UNIVERSIDAD
NACIONAL
DE COLOMBIA

**Un nuevo modelo para la estimación de
la biomasa viva y el carbono
almacenado en los bosques de manglar
usando sensoramiento remoto y
aprendizaje de máquina: caso de
estudio Tumaco-Nariño**

Laura Lozano Arias

Universidad Nacional de Colombia
Facultad de Ingeniería y Administración, Departamento de Ingeniería
Palmira, Colombia

2023

Un nuevo modelo para la estimación de la biomasa viva y el carbono almacenado en los bosques de manglar usando sensoramiento remoto y aprendizaje de máquina: caso de estudio Tumaco-Nariño

Laura Lozano Arias

Tesis presentada como requisito para optar al título de:
Magister en Ingeniería Ambiental

Director:

PhD. John Josephraj Selvaraj

Codirector:

Mgtr. Bryan Ernesto Gallego Pérez

Línea de Investigación:

Monitoreo, modelación y gestión de recursos naturales

Grupo de Investigación:

Recursos Hidrobiológicos

Universidad Nacional de Colombia

Facultad de Ingeniería y Administración, Departamento de Ingeniería

Palmira, Colombia

2023

*Honro a mis padres con este trabajo,
Nelson y Luz Ángela gracias por darme
siempre lo mejor que podían dar.*

*“Porque todas las cosas proceden de él, y
existen por él y para él. ¡A él sea la gloria por
siempre!”*

Romanos 11:36 (NVI)

Declaración de obra original

Yo declaro lo siguiente:

He leído el Acuerdo 035 de 2003 del Consejo Académico de la Universidad Nacional. «Reglamento sobre propiedad intelectual» y la Normatividad Nacional relacionada al respeto de los derechos de autor. Esta disertación representa mi trabajo original, excepto donde he reconocido las ideas, las palabras, o materiales de otros autores.

Cuando se han presentado ideas o palabras de otros autores en esta disertación, he realizado su respectivo reconocimiento aplicando correctamente los esquemas de citas y referencias bibliográficas en el estilo requerido.

He obtenido el permiso del autor o editor para incluir cualquier material con derechos de autor (por ejemplo, tablas, figuras, instrumentos de encuesta o grandes porciones de texto).

Por último, he sometido esta disertación a la herramienta de integridad académica, definida por la universidad.



Laura Lozano Arias

Fecha 07/11/2023

Agradecimientos

En primer lugar, agradezco al proyecto "Investigación de servicios ecosistémicos derivados de bosques de manglar en el Pacífico Colombiano: Valle del Cauca, Nariño, Cauca y Chocó" BPIN2020000100054, financiado con recursos del Sistema General de Regalías, por permitir el desarrollo de esta tesis. Especialmente, quiero reconocer y agradecer a la profesora Ángela Inés Guzmán por brindar sus mejores capacidades en la dirección del proyecto y facilitar las condiciones necesarias para que este trabajo se llevara a cabo con éxito. También quiero expresar mi gratitud al Consejo Comunitario Bajo Mira y Frontera y a su comunidad por colaborar conmigo en la recopilación de datos, sin su valiosa ayuda, esto no habría sido posible.

En segundo lugar, quiero expresar mi agradecimiento al profesor John Selvaraj por confiar en mí y supervisar el desarrollo de esta investigación. Su apoyo y la oportunidad que me brindó para participar en este proyecto son invaluable. Asimismo, agradezco a mi codirector, Bryan Gallego, por compartir generosamente sus conocimientos y por guiarme a lo largo de este proceso.

Además, deseo destacar el apoyo inquebrantable de mi esposo, Juan David, quien hizo mis sueños los suyos y dedicó tiempo y esfuerzo adicional para que pudiera completar este trabajo. También agradezco a mi hermano, Juan Daniel, por su disposición para compartir recursos y conocimientos. Mi agradecimiento se extiende a mis padres y amigos por sus oraciones y apoyo constantes.

Por último, pero no menos importante, agradezco a Dios por su constante compañía a lo largo de este proceso. Le agradezco por la sabiduría y los dones que me ha otorgado, por ser mi fuente de tranquilidad en momentos de caos, y por su dirección, protección y provisión.

Resumen

Un nuevo modelo para la estimación de la biomasa viva y el carbono almacenado en los bosques de manglar usando sensoramiento remoto y aprendizaje de máquina: caso de estudio Tumaco-Nariño

Los manglares desempeñan un papel crucial en la mitigación del cambio climático al absorber y retener hasta cinco veces más carbono que otros bosques. Es importante determinar la biomasa viva y el carbono almacenado en estos ecosistemas para proporcionar una base sólida en la planificación y gestión gubernamental. Este estudio presenta un enfoque innovador ya que utiliza herramientas de teledetección junto con datos recolectados en campo, imágenes Worldview 2 y evalúa dos algoritmos de aprendizaje automático, Random Forest y Support Vector Regression. El caso de estudio en el manglar de Tumaco, Nariño, incluyó el cálculo de la superficie del bosque por medio de una clasificación supervisada, la estimación de la biomasa viva (aboveground y belowground) y el carbono almacenado, y la evaluación de los modelos. Los resultados revelaron una precisión global del 87% en la clasificación de coberturas, con valores promedio de 192.50 ± 102.78 para la biomasa aérea, 79.95 ± 56.85 para la biomasa subterránea y 127.43 ± 73.49 para el carbono almacenado. El modelo basado en Random Forest destacó con un rendimiento sobresaliente, mostrando un RMSE de 140.68 ± 98.76 y un R^2 de 0.78 ± 0.28 , superando a modelos globales. Adicionalmente, se evidenció que los índices espectrales fortalecen la capacidad del modelo para explicar y predecir la biomasa aérea. Se sugiere explorar el uso de imágenes Lidar y datos SAR para mejorar la precisión en estudios locales con mayor resolución espacial.

Palabras clave: Biomasa, Reservas de carbono, Aprendizaje automático, Manglar, Teledetección, Worldview-2.

Abstract

A new approach for estimating living biomass and stored carbon in mangrove forests using remote sensing and machine learning: Tumaco-Nariño case study.

Mangroves play a crucial role in climate change mitigation by absorbing and sequestering up to five times more carbon than other forests. It is important to determine the living biomass and carbon stored in these ecosystems to provide a sound basis for government planning and management. This study presents an innovative approach using remote sensing tools together with field collected data, using Worldview 2 imagery and evaluating two machine learning algorithms, Random Forest and Support Vector Regression. The case study in the mangrove forest of Tumaco, Nariño, included the calculation of forest area by supervised classification, estimation of live biomass (aboveground and belowground) and carbon stock, and evaluation of the models. The results revealed an overall accuracy of 87% in cover classification, with average values of 192.50 ± 102.78 for aboveground biomass, 79.95 ± 56.85 for belowground biomass and 127.43 ± 73.49 for carbon stock. The Random Forest based model stood out with an outstanding performance, showing an RMSE of 140.68 ± 98.76 and an R^2 of 0.78 ± 0.28 , outperforming global models. Additionally, it was evidenced that the spectral indices strengthen the model's ability to explain and predict aerial biomass. It is suggested to explore the use of Lidar images and SAR data to improve accuracy in local studies with higher spatial resolution.

Keywords: Biomass, Carbon stocks, Machine learning, Mangrove, Remote sensing, Worldview-2.

Contenido

	Pág.
Resumen	IX
Lista de figuras	XIII
Lista de tablas	XIV
Introducción	1
1. Capítulo 1: Contexto de la investigación	3
1.1 Planteamiento del problema.....	3
1.2 Justificación.....	5
1.3 Pregunta problema.....	7
1.4 Objetivos	7
1.4.1 Objetivo general.....	8
1.4.2 Objetivos específicos	8
2. Capítulo 2: Marco Teórico y Estado del arte	9
2.1. Marco teórico	9
2.1.1. Bosques de Manglar	9
2.1.2. Almacenamiento de carbono en bosques de manglar	10
2.1.3. Sensores remotos.....	12
2.1.4. Técnicas para mapeo de manglares con sensores remotos	14
2.1.4.1. Clasificación no supervisada	14
2.1.4.2. Clasificación supervisada	14
2.1.5. Evaluación de la exactitud del mapeo de manglares con sensores remotos..	15
2.1.5.1. Matriz de error	16
2.1.5.2. Exactitud global, del productor y del usuario.....	17
2.1.5.3. Coeficiente de Kappa	18
2.1.6. Técnicas para la estimación de biomasa con sensores remotos.....	19
2.1.7. Evaluación de modelos de estimación de biomasa y carbono con sensores remotos	21
2.1.7.1. Validación cruzada	21
2.1.7.2. Métricas de evaluación.....	22
2.1.8. Aprendizaje de máquina aplicado en investigaciones en bosques de manglar	23
2.1.8.1. Máquinas de Vectores de Soporte.....	24
2.1.8.2. Bosques Aleatorios.....	25
2.2. Estado del arte	26

2.2.1. Estimación del carbono almacenado en bosques de manglar con sensores remotos.....	26
3. Capítulo 3: Materiales y métodos	31
3.1. Área de estudio.....	32
3.2. Adquisición y preprocesamiento de datos de campo	33
3.2.1. Recolección de datos en campo.....	33
3.2.2. Cálculo de biomasa aérea (AGB)	34
3.2.3. Validación de los datos de AGB	34
3.3. Preprocesamiento de Imágenes satelitales.....	35
3.3.1. Características de la imagen satelital utilizada.	35
3.3.2. Preprocesamiento de la imagen satelital	36
3.3.3. Cálculo de índices espectrales	37
3.4. Mapeo de la cobertura del bosque de manglar	39
3.4.1. Datos de entrada para la clasificación de cobertura de manglar.....	39
3.4.2. Esquema de clasificación	40
3.4.3. Post procesamiento de la imagen.....	41
3.4.4. Evaluación de la exactitud del mapa de cobertura.....	41
3.5. Estimación de AGB en el bosque de manglar.	42
3.5.1. Optimización de variables explicativas para modelos de AGB.....	42
3.5.2. Modelos para la estimación de AGB.....	43
3.5.3. Ajuste de hiperparámetros	44
3.5.4. Evaluación de los modelos de AGB.....	44
3.6. Estimación de BGB y el carbono almacenado en el bosque de manglar.....	45
4. Capítulo 4: Resultados y Discusión	46
4.1. Características de los datos de AGB recolectados en campo.	46
4.2. Selección de variables explicativas.....	48
4.3. Caracterización y cuantificación de la cobertura del bosque de manglar	51
4.4. Rendimiento de los modelos de AGB.....	53
4.4.1. Importancia de las variables al modelo de AGB con el algoritmo Random Forest ..	55
4.4. Cuantificación de AGB, BGB y Carbono almacenado en el bosque de manglar.	57
4.5. Posibles incertidumbres asociadas al modelo.....	65
5. Conclusiones y recomendaciones	67
5.1. Conclusiones	67
5.2. Recomendaciones	68
Bibliografía	69

Lista de figuras

	Pág.
Figura 1. Ejemplo matemático de la matriz de error.	17
Figura 2. Esquema metodológico general del flujo de trabajo.....	31
Figura 3. Ubicación de la zona de estudio en el municipio de Tumaco-Nariño.	32
Figura 4. Gráfico de probabilidad normal para las parcelas de AGB (t/ha) levantadas en campo.	47
Figura 5. Contribuciones de variables al componente: (a) Componente 1, (b) Componente 2.....	50
Figura 6. Mapa clasificación supervisada para el área de estudio.	51
Figura 7. Importancia de las variables explicativas para el modelo Random Forest: (a) Primera corrida con todas las variables, (b) Segunda corrida con variables del primer filtro, (c) Tercera corrida con variables del segundo filtro.	56
Figura 8. Mapa de distribución de la biomasa aérea [t/ha] de los bosques de manglar ubicados en el área de estudio.....	61
Figura 9. Mapa de distribución de la biomasa subterránea [t/ha] de los bosques de manglar ubicados en el área de estudio.....	62
Figura 10. Mapa de distribución del carbono almacenado en la Biomasa aérea y subterránea [tC/ha] de los bosques de manglar ubicados en el área de estudio.	63

Lista de tablas

	Pág.
Tabla 1. Resumen de los datos provenientes de sensores remotos utilizados para estudios en ecosistemas de manglar.	13
Tabla 2. Estudios de la estimación de carbono almacenado en ecosistemas de manglar con el uso de sensores remotos.	29
Tabla 3. Densidad de la madera utilizada para calcular AGB.....	34
Tabla 4. Resolución espectral de las bandas Worldview 2.	35
Tabla 5. Característica de la imagen satelital Worldview2.....	36
Tabla 6. Índices espectrales derivados de la imagen multiespectral Worldview-2.	37
Tabla 7. Agrupación de variables explicativas para entrenamiento de los modelos de AGB.	43
Tabla 8. Estadística descriptiva de las parcelas de AGB (t/ha) levantadas en campo.	47
Tabla 9. Above Ground Biomass estimado en campo en diferentes zonas del Pacífico Colombiano.....	48
Tabla 10. Precisión del productor y del usuario de la clasificación de la cobertura de manglar para el área de estudio ubicada en el Municipio de San Andrés de Tumaco.	52
Tabla 11. Métricas de evaluación para modelos de AGB (t/ha).....	54
Tabla 12. Cuantificación de AGB, BGB y Carbono almacenado en los bosques de manglar del área de estudio.	58
Tabla 13. Estudios de Biomasa aérea (AGB), Biomasa subterránea (BGB) y Carbono en Colombia y el Pacífico Colombiano con sensoramiento remoto.	60

Introducción

Colombia figura entre los 20 países que albergan bosques de manglar en su territorio, y aproximadamente el 70% de estos ecosistemas se ubican a lo largo de la costa Pacífica (Mejía-Rentería et al., 2018). Estos ecosistemas brindan una amplia gama de servicios ecosistémicos al país, siendo uno de los más destacados la captura y almacenamiento de carbono (Millennium Ecosystem Assessment, 2005). Los bosques de manglar son particularmente efectivos en esta función, con capacidad para capturar y almacenar carbono de manera significativa (Donato et al., 2011). La conservación de estos ecosistemas desempeña un papel crucial en la lucha contra el calentamiento global.

A pesar de su importancia, los bosques de manglar han enfrentado impactos negativos por actividades humanas como la acuicultura, el desarrollo costero y la urbanización, lo que ha resultado en una pérdida global de aproximadamente 6075 km² en cobertura de manglar entre los años 1996 y 2016 (Spalding & Leal, 2021). Además, la tala y destrucción de estos bosques interrumpe su capacidad de absorber el CO₂, lo que puede llevar a un aumento significativo en las emisiones de gases de efecto invernadero.

En un esfuerzo por abordar la pérdida de los bosques de manglar y contrarrestar el calentamiento global, los países están adoptando activamente políticas de carbono neutral (Zhu & Yan, 2022). Colombia se ha unido a algunos de estos acuerdos, como la Agenda para el Desarrollo Sostenible 2030 y el Programa REDD+. Para cumplir con estos acuerdos y programas, es esencial que Colombia realice estimaciones de biomasa y carbono almacenado en los ecosistemas de manglar. Estos estudios proporcionan herramientas esenciales para la toma de decisiones informadas y la planificación de acciones con el fin de abordar el cambio climático (Zhu & Yan, 2022).

Los enfoques basados en sensramiento remoto resultan adecuados para el mapeo y seguimiento de los ecosistemas de manglar, permitiendo la identificación de atributos como especies, biomasa y reservas de carbono. Además, esta herramienta ofrece ventajas en

términos de costos, precisión y cobertura para estudios a gran escala (Pham, Xia, et al., 2019).

A nivel global, los estudios que utilizan sensores remotos en bosques de manglar han estado en desarrollo desde 1959, inicialmente enfocados en el mapeo de estos ecosistemas; fue solo a partir de 2011 que se comenzó a estimar el carbono almacenado en el manglar (Aldrich et al., 1959; Wicaksono et al., 2011). A pesar de las ventajas de esta tecnología, su aplicación en Colombia con estos propósitos aún es escasa (Bolívar et al., 2018; Castellanos-Galindo et al., 2021).

Motivado por esta situación, el presente estudio pretende generar información sobre la biomasa viva y el carbono almacenado en los ecosistemas de manglar en la región del Pacífico, específicamente en el municipio de Tumaco, Nariño. Para lograrlo, se ha desarrollado un nuevo modelo que integra diversos elementos, tales como imágenes satelitales de alta resolución, índices de vegetación, datos de biomasa aérea recopilados en el terreno y la aplicación de algoritmos de aprendizaje automático para la modelación de los datos. Vale la pena destacar que la novedad de este modelo radica en el método empleado, el cual constituye una combinación única de factores que hasta ahora no han sido utilizados.

1. Capítulo 1: Contexto de la investigación

En este capítulo se expone el problema central de la investigación, la pregunta problema, la justificación por la cual es relevante llevar a cabo el presente estudio y los objetivos.

1.1 Planteamiento del problema

Los bosques de manglar están conformados por grandes árboles de mangle adaptados para sobrevivir en suelos con alta salinidad, pantanosos y constantemente inundados por las mareas (Spalding & Leal, 2021). Estos ecosistemas prestan un amplio rango de servicios ecosistémicos de regulación, aprovisionamiento, soporte y culturales (Millennium Ecosystem Assessment, 2005). Sin embargo, uno de los servicios más importantes es la captura y almacenamiento de carbono; la cual se lleva a cabo a través de las actividades biológicas y fisiológicas de este ecosistema (Zhu & Yan, 2022).

Los manglares se consideran el ecosistema costero más efectivo para la captura de carbono, teniendo en cuenta que a pesar de representar el 0.7% de la superficie forestal tropical mundial, pueden almacenar hasta 2.5 veces las emisiones mundiales anuales de gases de efecto invernadero (Gao et al., 2016; Zhu & Yan, 2022). Esto quiere decir que, protegiendo y restaurando los ecosistemas de manglar, el nivel de CO₂ en la atmósfera podría reducirse aproximadamente un 7% (Nellemann et al., 2009).

Pese a la importancia que los ecosistemas manglar representan a nivel mundial, estos han sido afectados por actividades antrópicas (Giri et al., 2011). De acuerdo con el Global Mangrove Watch la extensión mundial de bosques de manglar se ha reducido 6057.45 km² entre los años 1996 y 2016, siendo el 2007 el año donde se presentó una mayor disminución neta de 3876.44 km². Es importante tener en cuenta que al talar los bosques de manglar dejan de absorber y almacenar el CO₂ de la atmósfera, lo que lleva a un aumento significativo de las emisiones de gases de efecto invernadero, además, liberan el carbono retenido en sus suelos (Spalding & Leal, 2021).

Para abordar la deforestación y degradación de los ecosistemas de manglar, se han establecido acuerdos y programas con el objetivo de gestionar y detener la pérdida de estos bosques y contrarrestar el calentamiento global. Un ejemplo de ello es el Objetivo de Desarrollo Sostenible número 14 (ODS 14), el cual se enfoca en "Conservar y utilizar en forma sostenible los océanos, los mares y los recursos marinos para el desarrollo sostenible". Específicamente, la meta 14.2 busca "gestionar y proteger sosteniblemente los ecosistemas marinos y costeros para evitar efectos adversos importantes" (Naciones Unidas, 2015). Además, se encuentra el Programa de Reducción de Emisiones por Deforestación y Degradación Forestal (REDD+), el cual busca que los gobiernos nacionales se comprometan con acciones climáticas en el sector forestal.

En particular, Colombia ha otorgado una creciente importancia a los ecosistemas de manglar, no solo al ser signatario de los acuerdos y programas mencionados anteriormente, asumiendo así una responsabilidad a nivel internacional, sino también al establecer leyes y normativas orientadas a la protección y restauración de este ecosistema estratégico. Un ejemplo destacado de estos esfuerzos es la Ley 2243 de 2022, cuyo propósito es el de garantizar la protección de los ecosistemas de manglar, planificar su gestión y aprovechamiento, así como impulsar la conservación y restauración en áreas afectadas.

Es crucial contar con estimaciones de biomasa y carbono en los manglares para respaldar los acuerdos y programas de mitigación del cambio climático, así como para tomar decisiones en la planificación y ejecución de acciones de conservación y restauración. Las metodologías tradicionales, como la recolección de datos en campo, presentan desafíos como el difícil acceso, los altos costos y la limitación de estudios a nivel nacional o global (Howard et al., 2014). En contraste, los métodos basados en teledetección permiten mapear y monitorear los manglares, identificando atributos como especies, biomasa y reservas de carbono a mayor escala (Pham, Yokoya, et al., 2019).

En Colombia, se ha observado una menor cantidad de investigaciones sobre manglares en comparación con los estudios a nivel mundial (Castellanos-Galindo et al., 2021); estos estudios han tendido a enfocarse en investigar los patrones de distribución de organismos y comunidades dentro de los manglares, y la mayoría de ellos se han centrado en la costa

Caribe, mientras que la costa Pacífica ha sido menos estudiada en este aspecto. Además, Bolívar et al. (2018) señala la escasa disponibilidad de información pública sobre estudios de biomasa en los manglares colombianos, lo que dificulta su participación en los esquemas de mitigación del cambio climático.

Por otro lado, los estudios que estiman la biomasa y el carbono en los manglares de Colombia utilizando tecnologías de sensoramiento remoto son aún más escasos. La mayoría de las investigaciones, que utilizan esta tecnología, se han centrado en estimaciones a nivel global (Donato et al., 2011; Hutchison et al., 2014; Simard et al., 2019; Twilley et al., 1992), pero a nivel nacional los estudios son pocos. Por ejemplo, Bolívar et al. (2018) quien estimó la biomasa y el carbono por encima del suelo en la costa Caribe y Pacífica de Colombia, y la investigación de Selvaraj & Gallego-Pérez (2023) que estimaron la biomasa aérea únicamente en la Costa Pacífica Colombiana.

En conclusión, existe una necesidad apremiante de recopilar datos sobre la biomasa y el carbono en los manglares colombianos, especialmente en los manglares del Pacífico a escala local. Asimismo, es necesario que el país avance en el desarrollo e implementación de técnicas de estimación de biomasa y carbono almacenado utilizando sensores remotos, ya que esta herramienta permite buenas estimaciones con la posibilidad de abarcar una mayor área y realizar un monitoreo efectivo de los manglares. La falta de información sobre la biomasa y el carbono almacenado en los manglares dificulta la implementación de estrategias de gestión sostenible en estos ecosistemas.

1.2 Justificación

Dada la creciente importancia que Colombia ha atribuido a los ecosistemas de manglar y considerando la orientación de la Ley 2243 del 2022 hacia la protección de estos ecosistemas mediante la planificación de su manejo y aprovechamiento, esta investigación se alinea con los avances que la nación persigue. Contribuye a la evaluación del estado de conservación de los manglares al proporcionar información de la salud y conservación de este ecosistema a través de la medición de biomasa y carbono almacenado. Asimismo, ofrece información base esencial para la toma de decisiones.

Adicionalmente, los datos presentados en esta investigación contribuyen al avance, que se han venido llevando a cabo, de estudios en caracterización de manglar en el país. Estos datos pueden proveer información valiosa a la Corporación Autónoma Regional de Nariño (Corponariño), la cual está obligada a realizar la caracterización, diagnóstico y zonificación de los manglares, según lo establece el artículo 10 de la Resolución 1263 de 2018. Asimismo, estos datos también pueden contribuir con información para el Sistema de Información para la Gestión de los Manglares en Colombia (SIGMA). A nivel internacional, esta información será crucial para que Colombia informe sobre los avances en la reducción de gases de efecto invernadero y la conservación de los manglares, en el marco de los compromisos adquiridos en los Objetivos de Desarrollo Sostenible.

Por otra parte, esta investigación representa una contribución al progreso del desarrollo e implementación de metodologías para la estimación de biomasa y carbono almacenado en los bosques de manglar de Colombia, mediante el empleo de sensores remotos. La propuesta metodológica de este estudio se distingue por su carácter innovador al introducir un nuevo modelo fundamentado en algoritmos de aprendizaje automático. Este modelo utiliza una combinación única de factores.

En primer lugar, emplea imágenes de alta resolución espacial provenientes del satélite WorldView-2; aunque investigaciones previas hayan utilizado estas imágenes para estimar la biomasa aérea en manglares, es crucial señalar que dichos estudios se llevaron a cabo en distintas partes del mundo y no en Colombia (Jiang et al., 2021; Qiu et al., 2019; Y. Zhu et al., 2015, 2020). El cambio de país o región en el que se realiza un estudio implica variaciones en la composición del bosque y sus especies, lo cual afecta la reflectancia de la cobertura y, por ende, la respuesta del modelo.

En segundo lugar, implementa índices de vegetación que han demostrado ser efectivos en estudios previos centrados en el mapeo de características biofísicas en bosques de manglar. A pesar de que estos índices han sido previamente empleados en investigaciones anteriores, la combinación específica elegida en esta tesis es única. Las variables explicativas no operan de manera independiente; su comportamiento puede mejorar o deteriorar los resultados del modelo al combinarse o al ser excluidas unas de otras (Bishop, 2008).

Por último, el modelo incorpora datos capturados en campo, específicamente en la zona de estudio. Además, la biomasa aérea se estima con ecuaciones alométricas construidas específicamente para las especies presentes en la región del Pacífico (Yepes et al., 2011), proporcionando así un enfoque más ajustado a las características particulares de este entorno. Aunque la mayoría de estos factores se hayan utilizado en otros estudios, la singularidad de este estudio radica en la combinación única de factores empleada para construir el modelo.

Es importante resaltar que esta investigación es una de las pocas en la costa Pacífica que utiliza sensoramiento remoto y, hasta el momento, no se ha llevado a cabo un estudio que utilice imágenes de alta resolución WV-2 y técnicas de machine learning para estimar tanto la biomasa aérea (AGB, por sus siglas en inglés) como la biomasa subterránea (BGB, por sus siglas en inglés), así como el carbono almacenado en estos compartimientos. Además, al considerar la estimación de la biomasa viva por encima y por debajo del suelo del bosque de manglar, se obtiene una estimación representativa de biomasa (Palacios- penaranda, 2017).

Finalmente se eligió Tumaco como caso de estudio ya que en el Pacífico nariñense se encuentran los manglares más altos y mejor conservados del continente americano (Simard et al., 2019). Sin embargo, esta región ha sido la menos investigada en temas relacionados con los manglares (Castellanos-Galindo et al., 2021).

1.3 Pregunta problema

Partiendo de la necesidad de generar información sobre la cantidad de biomasa y carbono que albergan los ecosistemas de manglar del Pacífico y de generar herramientas que permitan agilizar la estimación de estos, esta tesis pretende dar respuestas al siguiente interrogante: ¿Qué modelo permitiría estimar de manera más adecuada la biomasa viva y el carbono almacenado en los bosques de manglar ubicados en la costa del municipio de Tumaco-Nariño?

1.4 Objetivos

A continuación, se presenta el objetivo general y los objetivos específicos de la presente propuesta de tesis.

1.4.1 Objetivo general

Proponer un modelo que permita estimar la biomasa viva (belowground y aboveground) y el carbono almacenado en los bosques de manglar ubicados en la costa del municipio de Tumaco, Nariño.

1.4.2 Objetivos específicos

- Calcular la superficie de bosque de manglar ubicado en la costa del municipio de Tumaco, Nariño.
- Estimar la biomasa viva (belowground y aboveground) y el carbono almacenado en los bosques de manglar ubicados en la costa del municipio de Tumaco, Nariño.
- Evaluar el modelo utilizado para la estimación de la biomasa viva (belowground y aboveground) y el carbono almacenado en los bosques de manglar ubicados en la costa del municipio de Tumaco, Nariño.

2. Capítulo 2: Marco Teórico y Estado del arte

En este capítulo se presentan los elementos teóricos, conceptos básicos y fundamentación científica que sustentan la presente investigación, además, los antecedentes e investigaciones previas.

2.1. Marco teórico

2.1.1. Bosques de Manglar

Los bosques de manglar son ecosistemas costeros que se encuentran ubicados en zonas tropicales y subtropicales entre los 25°N y 25°S (Kauffman & Donato, 2012). Estos ecosistemas están formados principalmente por árboles de manglar que son especies que han desarrollado una diversidad de estrategias para poder subsistir, una de estas es la capacidad de expulsar la sal de sus tejidos lo que les permite sobrevivir en áreas con alta concentración salina; además, sus raíces grandes y sobresalientes les permiten aferrarse en zonas lodosas y canalizar el aire hacia el suelo (Spalding & Leal, 2021). Gracias a estas adaptaciones estos bosques pueden encontrar en zonas intermareales, a lo largo de las riberas de los ríos y lagunas, en deltas y estuarios (Darmawan et al., 2015).

Según el Global Mangrove Watch (GMW) Version 3.0 en el 2020 existían aproximadamente 147.000km² de bosque de manglar, siendo el continente con mayor área de manglar el Sudeste Asiático (48.222km²) y Sur América el tercero (20.378 km²). Por su parte Colombia registra un área de 2.807 km² de hábitat de manglares. Las especies de manglar registradas en el Caribe colombiano son: *Avicennia germinans* (mangle negro), *Laguncularia racemosa* (mangle blanco), *Rhizophora mangle* (mangle rojo), *Pelliciera rhizophorae* (mangle piñuelo) y *Conocarpus erectus* (mangle botoncillo), adicional a estas en el Pacífico colombiano se encuentran: *Rhizophora harrisonii* (mangle caballero), *Rhizophora. racemosa* (mangle rojo) y *Mora oleífera* (mangle nato) (Navarrete-Ramírez & Rodríguez-Rincón, 2014).

Los manglares se pueden clasificar en cuatro asociaciones de acuerdo a las características físicas, climáticas e hidrológicas del entorno en el que existen: (1) manglares costeros o de borde, están influenciados directamente por las mareas y son inundados periódicamente por aguas salobres; (2) manglares ribereños o estuarinos, se desarrollan en estuarios asociados a ríos; (3) manglares de cuenca, crecen en llanuras costeras con bajo drenaje; y (4) manglares enanos o matorrales, se desarrollan en terrenos que presentan limitaciones para su crecimiento (Cantera et al., 1999; Kauffman & Donato, 2012). Además, dentro de cada asociación hay ciertas especies predominantes, por ejemplo, en la Costa Pacífica Colombiana se observa que en los manglares de borde hay prevalencia de las especies *Rhizophora sp.* y *Pelliciera rhizophorae* y en los manglares ribereños *Rhizophora sp.* *Laguncularia racemosa*, y *Avicenia germinans* (INVEMAR et al., 2006).

Los ecosistemas de manglar son vitales para la industria pesquera ya que sirven de guardería para los peces, lo que permite sustentar a más del 80% de los pescadores artesanales en muchos países (Spalding & Leal, 2021). Además, al estar ubicados donde el mar se encuentra con la tierra, reducen las inundaciones y actúan como barreras naturales contra las olas y el viento; además ayudan con la conformación del terreno en las costas (FAO, 2007). Asimismo, uno de los servicios ecosistémicos de mayor importancia en la lucha contra el calentamiento global es la captura y almacenamiento de carbono.

2.1.2. Almacenamiento de carbono en bosques de manglar

El proceso de almacenamiento de carbono en los manglares inicia con la captación de carbono, en forma de CO_2 , a través de la fotosíntesis. Cuando la captación es mayor que la respiración y la descomposición, el CO_2 es almacenado en forma de biomasa en hojas, madera y raíces, aumentando así las reservas de carbono en la biomasa viva a medida que el árbol crece (Spalding & Leal, 2021). Además, los manglares almacenan grandes cantidades de carbono en el suelo a través de la hojarasca y restos de madera (Thuy et al., 2020).

Los bosques de manglar pueden contener gran cantidad de carbono almacenado durante largos períodos de tiempo ya que su suelo está constantemente inundado por las mareas, por lo tanto, el proceso de descomposición de la materia orgánica se desacelera en un

ambiente anaeróbico permitiendo que los residuos vegetales se conserven durante mucho tiempo dando como resultado la acumulación de un suelo rico en carbono durante cientos de años (Hu et al., 2020; McLeod et al., 2011; Spalding & Leal, 2021; Zhu & Yan, 2022).

El carbono en los ecosistemas de manglar se puede almacenar en distintas partes: 1) hojas y tallos como biomasa aérea (AGB, por sus siglas en inglés), 2) raíces como biomasa subterránea (BGB, por sus siglas en inglés), 3) madera muerta y hojarasca como biomasa muerta sobre el suelo y 4) el carbono orgánico del suelo (Howard et al., 2014). Según la investigación de Palacios-Peñaranda et al. (2019) la biomasa viva sobre el suelo y debajo del suelo representa el 80% del carbono almacenado en los ecosistemas de manglar, el 20% restante equivale a la hojarasca, madera caída y el suelo.

Los bosques de manglar no son los únicos ecosistemas capaces de capturar y almacenar carbono, sin embargo, debido a su rápida tasa de crecimiento y acumulación de materia orgánica, son considerados como el ecosistema más efectivo para la captura de carbono (Donato et al., 2011). Esto los convierte en un ecosistema de alta importancia para la purificación de la atmósfera, la conservación en el equilibrio de CO₂ y la mitigación del calentamiento global (Gao et al., 2016; Zhu & Yan, 2022). Por otra parte, la acumulación de materia orgánica en los sedimentos del suelo del manglar contribuye a la estabilidad de las costas, protegiéndolas contra la erosión y las inundaciones, además, mejora la capacidad del ecosistema para retener agua y nutrientes, lo que beneficia a la salud y productividad de éste (Beeston et al., 2020).

Aunque en general el bosque de manglar es un sumidero de carbono altamente eficiente, hay grandes variaciones en su contenido de carbono debido a las diferencias estructurales y florísticas entre los bosques; por ejemplo, la asociación de manglares enanos aunque presenta una gran densidad de individuos estos son de bajo porte (altura < 1,5 m), lo que genera un menor contenido de carbono (Monsalve & Ramírez, 2015). Teniendo en cuenta que el contenido de carbono está directamente relacionado con la biomasa, el estudio de Kauffman & Cole (2010) mostró que la biomasa aérea varía en manglares ribereños y costeros con cantidades mayores a 500 Mg/ha y en manglares enanos con cantidades de 8 Mg/ha, en la región Indo-Pacífico.

2.1.3. Sensores remotos

Los sensores remotos son sistemas o instrumentos que permiten recolectar información meteorológica, oceanográfica, entre otros datos, sin la necesidad de estar físicamente presente en el lugar donde se va a recolectar la información. Generalmente los sensores remotos funcionan mediante la captación y registro de la radiación electromagnética emitida o reflejada por la superficie terrestre; pueden detectar ondas desde la luz visible hasta las microondas y los rayos infrarrojos y ultravioletas (Khorram et al., 2012). Su principal ventaja radica en la capacidad de obtener información sobre grandes áreas de manera rápida y eficiente, lo que permite realizar análisis y seguimientos a escala regional o global (Pham, Yokoya, et al., 2019).

La clasificación principal de los sensores remotos se basa en la forma en que funcionan los instrumentos; se consideran activos cuando cuentan con su propia fuente de emisión de luz, como el Radar de Apertura Sintética (SAR, por sus siglas en inglés) y el LiDAR; por otro lado, se consideran pasivos cuando dependen de una fuente de luz externa, es decir, de los rayos solares, como ocurre con las cámaras multiespectrales (Kogut, 2021). Dentro de esta clasificación, existen diferentes resoluciones de imagen, que se refieren al tamaño del píxel en una imagen, así como resoluciones espectrales, que hacen referencia a la cantidad y anchura de las bandas del espectro electromagnético que el sensor puede capturar, y resolución temporal, que es la frecuencia con la que el sensor captura imágenes (Khorram et al., 2012).

Existen diversos campos donde se pueden utilizar los sensores remotos, como la cartografía, la gestión del medio ambiente, la agricultura, la monitorización de desastres naturales, entre otros (Khorram et al., 2012). También son adecuados para mapear y monitorear ecosistemas de manglar e identificar sus atributos como tipo de especie, biomasa y reservas de carbono (Pham, Yokoya, et al., 2019). Para llevar a cabo estudios de mapeo, clasificación de especies y estimación de biomasa y carbono en ecosistemas de manglar se han utilizado muchos tipos de datos provenientes de sensores remotos. En la Tabla 1 se presenta los más usados.

Tabla 1. Resumen de los datos provenientes de sensores remotos utilizados para estudios en ecosistemas de manglar.

Captación	Tipo	Satélite	Imagen	Resolución espacial (m)	Capacidad de revisión (días)	Año de lanzamiento
Sensores remotos pasivos	Ópticos de alta resolución	Quick Bird	Pancromática	0.6	1.5-3	2000
			Multiespectral	2.4		
		IKONOS	Pancromática	1	1.5-3	1999
			Multiespectral	4		
		ALOS	PRISM	4	2	2006
			AVNIR	10		
		SPOT 4	Pancromática	10	2-3	1998
			Multiespectral	10		
		SPOT 5	Pancromática	5	2-3	2002
			Multiespectral	10		
	WorldView-2	Pancromática	0.46	1.1	2009	
		Multiespectral	1.85			
	GeoEye-1	Pancromática	0.41	2-3	2008	
		Multiespectral	1.65			
	KOMPSAT-2	Pancromática	1	2-3	2007	
Multiespectral		4				
Ópticos de resolución media/baja	Landsat 5	Multiespectral	30	16	1984-2013	
	Landsat 7	Pancromática	15	16	1999	
		Multiespectral	30			
	Landsat 8	Pancromática	15	16	2013	
		Multiespectral	30			
	Sentinel-2	Multiespectral	10,20,60	5-10	2015	
	IRS 1D	Pancromática	10	25	1997	
		Multiespectral	23.5			
EO-1	Hiperespectral	30	16	2000		
	Multiespectral	30				
Sensores remotos activos	SAR	ALOS	PALSAR	10	46	2006
			ALOS-2	Spotlight	1-3	16
		Stripmap		3,6,10		
		RADARSAT-2	Spotlight	1	24	2007
			Stripmap	3		
		Sentinel-1	Interferometric Wide Swath	5	12	2014
Stripmap	5					
	LiDAR	Ground Aerial	UAV	0.1	Movilizado a la orden	Desde 2000

Adaptado de Pham, Xia, et al. (2019)

Dentro del tema de sensores remotos es importante abordar los diferentes métodos que se han venido utilizando para el procesamiento de los datos capturados para llevar a buen término el estudio en los bosques de manglar. En la presente investigación se van a tener en cuenta dos tipos de estudios en los ecosistemas de manglar: el mapeo y la estimación de biomasa y carbono; en cada uno de estos estudios se van a mencionar las metodologías más utilizadas.

2.1.4. Técnicas para mapeo de manglares con sensores remotos

La clasificación de imágenes en el campo de la teledetección es el proceso de asignar píxeles o unidades básicas de una imagen a clases, esto permite caracterizar y a la vez cuantificar las coberturas que se pueden encontrar en un terreno (Perumal & Bhaskaran, 2010). Existen diversas técnicas para mapear coberturas de manglar haciendo uso de la teledetección, entre estas se encuentran las técnicas visuales como la Interpretación visual de imágenes satelitales. También, hay técnicas basadas en algoritmos de clasificación como la clasificación supervisada y no supervisada que se explican a continuación.

2.1.4.1. Clasificación no supervisada

La clasificación no supervisada es un método en el cual el algoritmo identifica patrones o estructuras en los datos, sin utilizar información previa sobre las categorías a las que podrían pertenecer, lo que permite la clasificación de los datos sin la necesidad de etiquetas preexistentes; los métodos más comunes son la Técnica Iterativa de Análisis de Datos Autoorganizados (ISODATA, por sus siglas en inglés) y los índices de vegetación (Pham, Xia, et al., 2019). La técnica de clasificación no supervisada es útil en situaciones donde no se dispone de información previa de las clases, sin embargo, la interpretación de los grupos resultantes debe requerir conocimiento experto y análisis adicional para asignar significado a cada clase y comprender su relevancia en el contexto del problema analizado (Pham, Xia, et al., 2019).

2.1.4.2. Clasificación supervisada

La clasificación supervisada es un método en el cual el algoritmo se debe entrenar mediante un conjunto de datos previamente etiquetados, lo que le permite al algoritmo asociar las entradas con las salidas deseadas, lo que facilita su capacidad para generalizar ese conocimiento a nuevos datos no etiquetados (Perumal & Bhaskaran, 2010).

Existen varios tipos de algoritmos de clasificación, estos pueden ser paramétricos y no paramétricos (Toosi et al., 2019); dentro de los paramétricos se encuentran algoritmos como Maximum Likelihood Classification (MCL), Parallelepiped Classifier, Minimum Distance Technique (Nguyen et al., 2020; Perumal & Bhaskaran, 2010), y dentro de los no paramétricos Support Vector Machine (SVM), Random Forest (RF), Artificial Neural Network (ANN) Classifier (Toosi et al., 2019). Los métodos más utilizados son Maximum Likelihood Classification y Decision Tree (Pham, Xia, et al., 2019).

El proceso de clasificación supervisada en imágenes implica los siguientes pasos: (1) la selección de muestras de entrenamiento donde se eligen muestras de referencia para cada clase de interés, utilizando información de campo u otras fuentes de datos confiable; (2) preparación de variables explicativas donde se extraen características espectrales, texturales o espaciales de las imágenes satelitales en las ubicaciones de las muestras de entrenamiento; (3) el entrenamiento del clasificador utilizando las muestras de entrenamiento y sus características correspondientes; (4) la clasificación de la imagen completa para asignar a cada entrada una clase específica (Congalton & Green, 2008).

Es importante resaltar que no todos los métodos de mapeo se hacen por píxeles, también hay análisis de imágenes basado en objetos (OBIA, por sus siglas en inglés). Esta es una técnica que analiza grupos de píxeles vecinos como objetos o segmentos en lugar de utilizar los enfoques de clasificación basados en píxeles (Pham, Xia, et al., 2019). En algunos casos los resultados obtenidos con el enfoque OBIA son similares o superan a los obtenidos con el enfoque basado en píxeles cuando se pretende detectar comunidades de especies de manglar (Pham, Yokoya, et al., 2019).

2.1.5. Evaluación de la exactitud del mapeo de manglares con sensores remotos

Evaluar la clasificación del mapa es un paso tan importante como hacer la misma clasificación, ya que esto permite al usuario conocer qué tan confiable es el resultado que se obtuvo para poder utilizarlo en futuros desarrollos; la evaluación de los mapas se realiza mediante la comparación con datos de referencia, que pueden ser obtenidos a través de muestreos de campo, imágenes de alta resolución o información existente previamente validada (Rwanga et al., 2017).

Hay dos tipos de evaluación de la exactitud del mapa: posicional y temático. La exactitud posicional mide qué tan lejos está una característica espacial en un mapa de su ubicación real y la precisión temática mide si las etiquetas de las características mapeadas son diferentes de la etiqueta de la característica real (Congalton & Green, 2008). Dado que la mayoría de las imágenes adquiridas suelen venir previamente corregidas geométricamente, lo que implica que ya cuentan con una evaluación posicional, esta subsección se centrará en la evaluación de la exactitud temática.

Para llevar a cabo una buena evaluación de la exactitud temática se deben tener en cuenta dos pasos fundamentales: (1) el diseño e implementación de procedimientos de muestreo imparcial, y (2) el análisis comparativo entre el mapa obtenido y datos de referencia (Congalton & Green, 2008). En el primer paso se define la metodología de muestreo, la cantidad de datos a recolectar y el esquema de clasificación; Congalton & Green (2008) sugieren recolectar un mínimo de 50 muestras para cada clase del mapa para mapas de menos de 405 hectáreas con menos de 12 clases. Por otra parte, el esquema de clasificación debe definir el número de clases en las que se clasificará la imagen y las reglas que se deben cumplir para que una unidad de muestreo pertenezca a cierta etiqueta (Congalton & Green, 2008). En el segundo paso se deben utilizar herramientas estadísticas para llevar a cabo el análisis comparativo, algunas de estas se expondrán a continuación.

2.1.5.1. Matriz de error

La matriz de error es un conjunto de números organizados en filas y columnas. Su propósito es representar la distribución de unidades de muestra asignadas a una categoría específica en una clasificación, en relación con la cantidad de unidades de muestra asignadas a esa misma categoría en otra clasificación (Congalton & Green, 2008). Usualmente en las columnas se registra la información de la clasificación de los datos de referencia y en las filas se registran los datos clasificados generados a partir de la imagen de detección remota (Figura 1).

La finalidad de la matriz de error es comparar la información arrojada por el mapa elaborado con la información real en campo o datos de referencia, por lo tanto, muestra el número de aciertos y errores en la clasificación, desglosados por clase (Rwanga et al., 2017). Por otra parte, la matriz de error permite calcular métricas de precisión, como la exactitud global, de usuario y de productor.

2.1.5.2. Exactitud global, del productor y del usuario

La exactitud global, del usuario y del productor son tres métricas ampliamente usadas para evaluar la precisión de un sistema de clasificación (Congalton & Green, 2008). La exactitud global da razón de la clasificación en general, mientras que la exactitud del productor y del usuario representan las precisiones de las categorías de clasificación de forma individual. Las ecuaciones (1), (2), (3) fueron tomadas de Congalton & Green (2008). En la Figura 1 se presenta un ejemplo matemático de la matriz de error en donde las columnas registran los datos reales y las filas los datos clasificados; los valores resaltados en amarillo corresponden a los datos correctamente clasificados por clase, en azul corresponden a la suma de los datos de referencia por cada clase, en magenta corresponden a la suma de los datos clasificados por cada clase y finalmente el valor en verde es el número total de datos en el mapa.

Figura 1. Ejemplo matemático de la matriz de error.

		Datos de referencia (j)				Σ Fil
		1	2	3	k	n_{i+}
Datos Clasificados (i)	1	n_{11}	n_{12}	n_{13}	n_{1k}	n_{1+}
	2	n_{21}	n_{22}	n_{23}	n_{2k}	n_{2+}
	3	n_{31}	n_{32}	n_{33}	n_{3k}	n_{3+}
	k	n_{k1}	n_{k2}	n_{k3}	n_{kk}	n_{k+}
Σ Col	n_{+j}	n_{+1}	n_{+2}	n_{+3}	n_{+k}	n

Adaptado de Congalton & Green (2008).

La exactitud global del mapa o de la imagen clasificada es una medida que indica el porcentaje de aciertos en la clasificación total; esta medida permite comparar diferentes clasificaciones entre sí (Rwanga et al., 2017). La exactitud global se calcula dividiendo el

número total de datos correctamente clasificados entre el número total de datos en el mapa (1).

$$Exactitud\ global = \frac{\sum_{i=1}^k n_{ii}}{n} \quad (1)$$

La exactitud del productor mide la probabilidad de que un elemento que realmente pertenece a una clase particular sea correctamente clasificado por el sistema (Congalton & Green, 2008). Se calcula dividiendo el número de elementos correctamente clasificados en una clase por el número total de elementos que realmente pertenecen a esa clase (2).

$$Exactitud\ del\ productor = \frac{n_{jj}}{n_{+j}} \quad (2)$$

La exactitud de usuario mide la probabilidad de que un elemento que el sistema clasifica como una clase particular sea realmente de esa clase (Congalton & Green, 2008). Se calcula dividiendo el número de elementos correctamente clasificados en una clase por el número total de elementos que el sistema clasificó como esa clase (3).

$$Exactitud\ del\ usuario = \frac{n_{ii}}{n_{i+}} \quad (3)$$

2.1.5.3. Coeficiente de Kappa

El coeficiente kappa es una medida estadística utilizada para evaluar la precisión de la clasificación de imágenes; se utiliza para determinar en qué medida la clasificación ha mejorado en comparación con una asignación al azar, teniendo en cuenta tanto los aciertos como los errores (Rwanga et al., 2017). Este coeficiente se calcula dividiendo la diferencia entre la concordancia observada, que es la proporción de elementos que el sistema de clasificación y la referencia clasifican de la misma manera (p_o), y la concordancia esperada (p_c), que se calcula como la probabilidad de que el sistema de clasificación y la referencia

coincidan por casualidad (4) (Congalton & Green, 2008). Los resultados del coeficiente kappa pueden variar entre -1 y 1, donde 1 indica una coincidencia perfecta entre la clasificación observada y la clasificación esperada y -1 indica una coincidencia peor que la concordancia esperada por azar (Congalton & Green, 2008).

$$Kappa = \frac{p_o - p_c}{1 - p_c} \quad (4)$$

2.1.6. Técnicas para la estimación de biomasa con sensores remotos

A lo largo de los años se han venido utilizando diversas técnicas para estimar la biomasa contenida en los bosques, uno de los enfoques más antiguos son los métodos de campo, en los cuales se recopilan muestras de vegetación directamente en el sitio y posteriormente se aplican técnicas de pesaje o volumetría para determinar la biomasa total (Yepes et al., 2011). Con el avance de la investigación, se ha desarrollado y utilizado cada vez más el enfoque de las ecuaciones alométricas, que permiten calcular la biomasa de manera indirecta y no destructiva, estas ecuaciones buscan relaciones matemática adecuadas entre las variables predictoras de medición directa, como la altura del individuo, y las variables de respuesta, utilizando una muestra representativa de individuos (Yepes et al., 2016). Además, se han empleado métodos de teledetección que aprovechan datos obtenidos por sensores remotos, como imágenes satelitales, para estimar la biomasa; estas técnicas ofrecen una perspectiva más amplia y eficiente al permitir la estimación de la biomasa en áreas extensas y de difícil acceso (Lu et al., 2016).

Dentro de los métodos de teledetección se encuentra la técnica de estimación de biomasa utilizando imágenes LiDAR. Estas imágenes permiten la caracterización tanto vertical como horizontal de las estructuras del dosel vegetal, al vincular esta información con ecuaciones alométricas es posible calcular la biomasa en áreas extensas y con resultados altamente precisos (Hickey et al., 2018; Lu et al., 2016). Sin embargo, es importante tener en cuenta que el uso de datos LiDAR tiene algunas limitaciones. En primer lugar, estos datos suelen ser más costosos en comparación con otras fuentes de información, además,

su disponibilidad puede estar restringida a áreas geográficas específicas, lo que limita su aplicabilidad a nivel global (Lu, 2006).

Los modelos de regresión son otra metodología utilizada en la teledetección para estimar la biomasa. Estos modelos se fundamentan en la correlación entre los valores de las variables explicativas y la biomasa, utilizando una muestra de datos de entrenamiento, posteriormente, se aplican técnicas estadísticas para ajustar los coeficientes del modelo y utilizarlos en las imágenes satelitales, con el propósito de estimar la biomasa a una escala más amplia (Quang et al., 2022). Algunas ventajas de esta técnica es que pueden utilizarse para realizar predicciones en áreas no muestreadas (Wang et al., 2019).

Es importante destacar que las variables explicativas, en los modelos de regresión, pueden provenir de diversos datos capturados por sensores remotos, como imágenes ópticas, de radar, índices de vegetación y texturas (Sinha et al., 2015). También es válido combinar diferentes tipos de variables explicativas, por ejemplo, se ha demostrado que es efectivo que los datos LiDAR o SAR se usen con datos ópticos para mapear la altura del dosel y la biomasa de los bosques de manglares (Pham, Yokoya, et al., 2019).

Los modelos de regresión pueden ser paramétricos y no paramétricos (Toosi et al., 2019). Los modelos paramétricos ya tienen una estructura predefinida con parámetros fijos; estos modelos se basan en suposiciones específicas sobre la distribución estadística de los datos, ejemplos comunes de modelos paramétricos son la regresión lineal y la regresión logística (Chave et al., 2005; Lu et al., 2016). Por otro lado, los modelos no paramétricos son más flexibles, no están restringidos por suposiciones sobre la forma de la distribución de los datos y no requieren una estructura fija, permitiendo capturar patrones complejos y no lineales, ejemplos de modelos no paramétricos son los desarrollados con aprendizaje de máquina (Jachowski et al., 2013; Lu et al., 2016).

Tanto los modelos paramétricos como los no paramétricos tienen ventajas y limitaciones. En el caso de los modelos paramétricos, la ventaja es su fácil interpretabilidad y eficiencia computacional, sin embargo, pueden presentar limitaciones al no ajustarse adecuadamente a datos complejos (Hastie et al., 2009). Por otro lado, los modelos no paramétricos tienen la capacidad de abordar relaciones complejas y adaptarse a diferentes tipos de datos, pero suelen requerir un mayor tamaño de muestra y tener una mayor complejidad computacional (Awad & Khanna, 2015).

Cuando se aplican estos modelos a problemas de estimación de biomasa utilizando datos de sensores remotos ópticos de alta resolución, se ha observado que las redes neuronales y los vectores de soporte de máquina proporcionan una buena modelación, también, se ha registrado que los métodos no paramétricos ofrecen mejores resultados de ajuste al utilizar imágenes multiespectrales y datos SAR (Pham, Yokoya, et al., 2019).

2.1.7. Evaluación de modelos de estimación de biomasa y carbono con sensores remotos

Los modelos no representan de manera perfecta la realidad, siempre existe una diferencia entre el fenómeno representado en el modelo y el fenómeno observado en el mundo real, cuya diferencia puede representarse en medidas del error. La calidad del modelo se determina por la magnitud de estos errores, siendo menor cuando el modelo se acerca más a la realidad. Por lo tanto, la calidad de los modelos utilizados para estimar la biomasa tiene un impacto directo en la fiabilidad de los resultados.

Para medir la calidad de los modelos, es esencial emplear métricas de evaluación adecuadas. Estas métricas son medidas numéricas que se utilizan para cuantificar y medir diversos aspectos del rendimiento del modelo, algunas de estas son: el error absoluto medio (MAE, por sus siglas en inglés), el error cuadrático medio (MSE, por sus siglas en inglés), el error cuadrático medio de la raíz (RMSE, por sus siglas en inglés) y el coeficiente de determinación (R^2) (FAO et al., 2018).

Por otra parte, existen métodos de evaluación del rendimiento del modelo, que son técnicas utilizadas para llevar a cabo la evaluación; entre ellos se encuentran la validación cruzada y la partición de datos (Awad & Khanna, 2015). Por lo tanto, las métricas de evaluación se aplican utilizando los métodos de evaluación correspondientes para obtener una estimación precisa del rendimiento del modelo en la tarea específica de estimación de biomasa en manglares. En las siguientes subsecciones se profundizará en estos aspectos, explicando en detalle las métricas de evaluación y el método de evaluación del rendimiento del modelo.

2.1.7.1. Validación cruzada

El rendimiento de los modelos generalmente se evalúa mediante la división de datos, donde el conjunto de muestra se divide en dos subconjuntos. Un subconjunto se utiliza

para calibrar el modelo de predicción, mientras que el otro subconjunto se utiliza para la validación, sin embargo, en casos donde los conjuntos de datos son escasos, la división puede resultar ineficiente, ya que no se aprovecha completamente la información del conjunto de datos tanto para la calibración como para la validación (Arlot & Celisse, 2010). Por esta razón, se recurre al método de validación cruzada.

La validación cruzada presenta diferentes métodos de funcionamiento, siendo el más eficiente para conjuntos de datos limitados el método k-fold; este enfoque garantiza que todas las observaciones del conjunto de datos original tengan la oportunidad de aparecer tanto en el conjunto de entrenamiento como en el conjunto de prueba (Arlot & Celisse, 2010). El procedimiento implica dividir el conjunto de datos en k submuestras, de igual tamaño, mutuamente excluyentes, luego el modelo se entrena k veces donde en cada iteración se utiliza una de las k submuestras como conjunto de prueba, mientras que las k-1 submuestras restantes se emplean para entrenar el modelo; los resultados de las k iteraciones se promedian para estimar la precisión como una única medida (Awad & Khanna, 2015).

2.1.7.2. Métricas de evaluación

A continuación, se detallan las métricas de evaluación empleadas en este estudio. En donde \hat{y}_i, y_i representan respectivamente los valores estimados y observados, n es el número total de parcelas a ser validadas y \bar{y}_i es la media de las observaciones de biomasa para las ecuaciones 5, 6, 7, 8.

- **Error absoluto medio (MAE):** es una medida de la precisión promedio de las estimaciones; se calcula como la media de las diferencias absolutas entre los valores observados y los valores estimados (5)(5). Un MAE más bajo indica una mayor precisión en las estimaciones, ya que las diferencias entre los valores reales y estimados son menores en promedio (FAO et al., 2018).

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (5)$$

- **Error cuadrático medio (MSE):** proporciona una medida de la precisión y varianza de las estimaciones; se calcula como la media de los cuadrados de las diferencias

entre los valores observados y los valores estimados (6). El MSE amplifica los errores más grandes, por lo que valores más altos indican una mayor dispersión en las estimaciones (FAO et al., 2018).

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (6)$$

- **Error cuadrático medio de la raíz (RMSE):** es la raíz cuadrada del MSE y se utiliza para proporcionar una medida de la precisión en la misma unidad que la variable de interés (7)(7). Al igual que el MSE, valores más bajos indican una menor dispersión y mayor precisión en las estimaciones (FAO et al., 2018).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (7)$$

- **Coefficiente de determinación (R²):** es una medida que indica qué proporción de la variabilidad total en los datos puede ser explicada por el modelo. Representa la bondad de ajuste del modelo (8). Un valor de R² próximo a 1 sugiere que el modelo explica la mayor parte de la variabilidad presente en los datos, en contraste, valores más cercanos a 0 indican que el modelo presenta un ajuste insatisfactorio (FAO et al., 2018).

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2} \quad (8)$$

2.1.8. Aprendizaje de máquina aplicado en investigaciones en bosques de manglar

El aprendizaje de máquina es una rama de la inteligencia artificial que se dedica a desarrollar sistemas con la capacidad de aprender de manera automática (Bishop, 2008). Esta metodología se ha venido utilizando en las investigaciones en bosques de manglar

para mapear, clasificar especies y estimar biomasa; los algoritmos que presentan mayor tendencia de uso son: Máquinas de Vectores de Soporte, Redes Neuronales Artificiales y Bosques Aleatorios (Pham, Xia, et al., 2019). Es importante resaltar que los métodos de aprendizaje de máquina proporcionan una mejor precisión en la estimación de la biomasa en manglares, en comparación con los métodos paramétricos tradicionales (Pham, Yokoya, et al., 2019).

En esta subsección se describirán los métodos Random Forest y Support Vector Machine ya que serán empleados en el presente estudio de tesis. Estos métodos han demostrado resultados satisfactorios en investigaciones previas centradas en la estimación de la biomasa aérea en bosques de manglar (Hu et al., 2020; Quang et al., 2022; Selvaraj & Gallego-Pérez, 2023; Y. Zhu et al., 2020).

2.1.8.1. Máquinas de Vectores de Soporte

Las Máquinas de Vectores de Soporte (SVM, por sus siglas en inglés) son una técnica de aprendizaje automático basada en kernel que se utiliza para la clasificación y regresión de datos; cuando SVM se utiliza para problemas de regresión se le denomina Regresión de Vector de Soporte (SVR, por sus siglas en inglés) (Awad & Khanna, 2015).

El objetivo de SVR es encontrar una función que se ajuste a los datos de entrenamiento con un margen de error aceptable, para esto crea n-hiperplanos a través de un espacio espectral n-dimensional y cada hiperplano separa datos numéricos basados en una función Kernel; cuando los datos no pueden ser linealmente separables se utiliza una técnica llamada "kernel trick" la cual transforma los datos en un espacio de mayor dimensión donde sí pueden ser separables linealmente (Awad & Khanna, 2015). Para definir la forma del hiperplano, minimizar la función de pérdida y encontrar los mejores coeficientes para la función de regresión el algoritmo optimiza parámetros como gamma, costo y épsilon; una vez que se ha entrenado el modelo se puede utilizar para hacer predicciones en nuevos datos (FAO et al., 2018).

Este algoritmo es eficaz en espacios de altas dimensiones, es decir, en problemas que contienen una gran cantidad de variables explicativas con pocos datos de entrenamiento, incluso sigue siendo eficaz en los casos en que el número de dimensiones es mayor que el número de muestras; sin embargo, si el número de factores es mucho mayor que el

número de muestras es importante evitar el ajuste excesivo del modelo (Pedregosa et al., 2011).

2.1.8.2. Bosques Aleatorios

Los Bosques Aleatorios (RF, por sus siglas en inglés) son un método de aprendizaje automático que utiliza múltiples árboles de decisión para conocer la relación entre la variable independiente y sus predictores y así obtener resultados de predicción más robustos y precisos (Breiman, 2001). En el caso específico del presente trabajo RF puede ser utilizado como clasificador, para el mapeo de áreas de manglar, y como regresión, para la estimación de biomasa y carbono almacenado en los bosques de manglar.

En el algoritmo de bosque aleatorio se construyen múltiples árboles de decisión independientes utilizando una técnica llamada bagging que selecciona aleatoriamente las muestras de entrenamiento y dentro del bagging, se utiliza la técnica del bootstrap que genera las muestras aleatorias con reemplazo del conjunto de datos original, procurando así que cada conjunto de datos de entrenamiento tenga la misma cantidad de valores (Luo et al., 2016). Algunos datos no son seleccionados para el entrenamiento de los árboles de decisión, a estos se les denomina out-of-bag (OOB); estos se utilizan como una medida interna de validación y evaluación del rendimiento del modelo, ya que cada árbol se prueba con los datos OOB y se calcula una medida de rendimiento, como la precisión en clasificación o el error cuadrático medio en regresión (Cutler et al., 2007).

Cada árbol se ajusta con un conjunto de datos de entrenamiento independiente, pero en cada nodo, solo una pequeña cantidad de variables seleccionadas aleatoriamente están disponibles para la partición binaria (Cutler et al., 2007). En el caso de la clasificación la predicción final se obtiene mediante la mayoría de votos de los árboles para la clasificación de cada instancia, mientras que, para la regresión en lugar de votar por una clase, se promedian las predicciones de todos los árboles para obtener un valor numérico final (FAO et al., 2018). Dos parámetros importantes en el ajuste de los modelos RF son el número de variables a probar en cada división (*mtry*, por sus siglas en inglés) y el número de árboles (*ntree*, por sus siglas en inglés), el parámetro *mtry* controla la cantidad de factores de predicción que se utilizan aleatoriamente en cada árbol, mientras que el parámetro *ntree* controla la cantidad de árboles generados (FAO et al., 2018).

Es importante señalar que RF es robusto frente a la redundancia de datos, la no linealidad y es capaz de manejar una amplia gama de predictores con diferentes propiedades y rangos de valores (Forkuor et al., 2020), además reduce el riesgo de sobreajuste del algoritmo dada la selección aleatoria de subconjuntos y factores de predicción (Zhu et al., 2017).

2.2. Estado del arte

2.2.1. Estimación del carbono almacenado en bosques de manglar con sensores remotos

En los últimos 60 años, los temas de investigación que más se han trabajado con sensores remotos en bosques de manglar han sido el mapeo de la distribución, los parámetros biofísicos y la caracterización de procesos ecosistémicos (Wang et al., 2019). Los primeros estudios estaban enfocados en el mapeo de áreas de manglar (Aldrich et al., 1959), estos no presentaban evaluaciones de exactitud y estaban reducidos a zonas locales, sin embargo, con la evolución de la tecnología y el acceso a imágenes satelitales de mayor resolución y con un espectro electromagnético amplio, los estudios se fueron enfocando en caracterizaciones más detalladas como la clasificación de especies, la estimación de biomasa y el almacenamiento de carbono, además de trabajar en escala global (Wang et al., 2019).

El primer estudio que se enfocó en mapear el almacenamiento de carbono en un bosque de manglar fue llevado cabo por (Wicaksono et al., 2011), en Indonesia, donde se utilizaron imágenes Landsat 7, datos en campo y se probaron diferentes índices de vegetación para estimar el carbono aboveground y belowground a través de un modelo de regresión lineal, el mejor modelo fue en el que se utilizó el índice de supervisión ambiental global (GEMI, por sus siglas en inglés) donde dio como ajuste para carbono sobre el suelo (AGC, por sus siglas en inglés) $R^2 = 0.3403$ y para carbono debajo del suelo (BGC, por sus siglas en inglés) $R^2 = 0.1992$.

En la Tabla 2 se resumen los estudios que se han llevado a cabo específicamente en la estimación de carbono en los ecosistemas de manglar con sensoramiento remoto. Se han utilizado diferentes imágenes y modelos. Algunos estudios se enfocan directamente en estimar biomasa y luego utilizar un factor de conversión para estimar el carbono (Bindu et

al., 2020; Hickey et al., 2018; Pham et al., 2020; Thuy et al., 2020) mientras que otros modelos estiman directamente carbono (Anand et al., 2020; Mukhtar et al., 2021; Wicaksono et al., 2011, 2016).

Por otra parte, está la investigación de Hickey et al. (2018) quien utilizó imágenes de un sensor óptico (Landsat 8 OLI) para determinar la distribución horizontal o el área de una zona de manglar en Australia, luego con datos LiDAR determinó la altura del dosel del bosque dentro del área de estudio y sustituyó estas alturas en una ecuación alométrica para estimar biomasa y carbono. La evaluación del modelo lo hizo con el Coeficiente kappa de Cohen, dando como resultado 0.83 ($p < 0.0005$) para el área de manglar y para la estimación de la biomasa aérea d.f. = 20, $t = 1.0$, $p=0.3$.

La gran mayoría de estudios utilizan imágenes multiespectrales con resolución mayor a 15 m. Bindu et al. (2020) utilizó una imagen de resolución menor a 5 m (IRS 1D LISS IV MX) junto con datos de campo y el índice de vegetación de diferencia normalizada (NDVI, por sus siglas en inglés) se estimó AGC y BGC a través de un modelo de regresión lineal, donde dio como ajuste $R^2 = 0.710$.

Varios estudios se han enfocado en la evaluación de diferentes índices de vegetación y modelos de regresión. Por ejemplo Mukhtar et al. (2021) utilizando una imagen Landsat 8 tuvo como mejor resultado el índice de manglares con un modelo de regresión exponencial ($R^2=0.566$) Otro estudio es el de Anand et al. (2020) quien logró un $R^2=86.98\%$ utilizando el Índice de Vegetación Mejorado (EVI, por sus siglas en inglés) a través de un modelo de regresión en base radial.

En cuanto a locación de los estudios, la mayoría se han llevado a cabo en los manglares del continente asiático (Wang et al., 2019). En Colombia son pocos los estudios que usan sensoramiento remoto para la estimación de carbono; uno de ellos es el de Bolivar et al. (2018), el cual se desarrolla en los manglares de la Costa Pacífica Colombiana y utilizan variables bioclimáticas y el Índice de vegetación mejorado a través de distintos modelos de regresión lineal múltiple; el mejor modelo tuvo como resultado $R^2= 0.4675$.

Además, una revisión de Castellanos-Galindo et al. (2021) en la que se buscó información sobre las diferentes investigaciones llevadas a cabo en los manglares de Colombia concluyó que los estudios sobre manglares es un tema joven en Colombia, iniciado en año 2000. Además, concluye que las investigaciones en Colombia se han centrado en la costa

Caribe ya que ha sido esta la que mayor degradación ha presentado por el desarrollo económico.

Tabla 2. Estudios de la estimación de carbono almacenado en ecosistemas de manglar con el uso de sensores remotos.

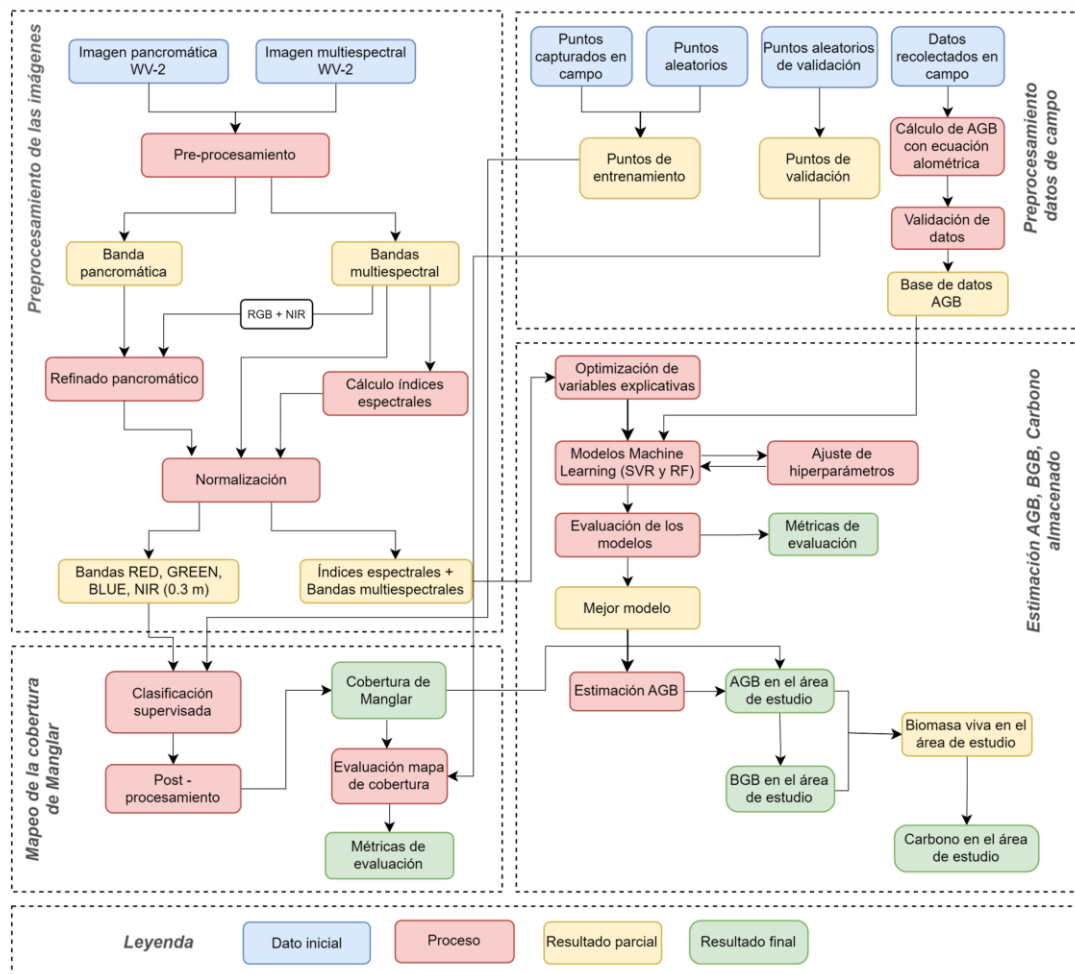
Referencia	Ubicación del bosque de manglar	Lugar de almacenamiento	Imagen satelital	Cálculo de biomasa		Cálculo de carbono	
				Método	Evaluación del método	Método	Evaluación del método
(Wicaksono et al., 2011)	INDONESIA- Islas Karimunjawa	Aboveground y Belowground	Landsat 7 ETM+	Ecuación alométrica de biomasa	-	Regresión lineal (Almacenamiento de carbono vs índice de vegetación)	Mejor modelo con el índice GEMI: (AGC) R ² = 0.3403 (BGC) R ² = 0.1992
(Wicaksono et al., 2016)	INDONESIA- Islas Karimunjawa	Aboveground y Belowground	ALOS AVNIR-2	Ecuación alométrica de biomasa	-	Modelo empírico (índices de vegetación, bandas PC)	(AGC) R ² = 0.681 (BGC) R ² = 0.603
(Takeuchi et al., 2016)	VIETNAM - Quang Ninh, Ca Mau y Kien Giang	Aboveground	ALOS PALSAR	Regresión retrodispersión vs altura del árbol	- HH: RMSE= 2.2 - HV: RMSE=2.0	Factor de conversión 0.5	-
(Hickey et al., 2018)	AUSTRALIA - Pilbara	Aboveground y Belowground	Landsat 8 OLI	- NDVI se utilizó para calcular el área de manglar - Datos LiDAR se utilizaron para conocer la altura del bosque de manglar y utilizar este dato en una ecuación alométrica	- Área de manglar: Cohen's kappa = 0.83 (p < 0.0005) - AGB: (d.f. = 20, t = 1.0, p=0.3)	Factor de conversión 0.475	-
(Bolivar et al., 2018)	COLOMBIA - Costa Caribe y Pacífica	Aboveground	Se utilizaron como variables de entrada las variables bioclimáticas de worldclim y Enhanced Vegetation Index (EVI) de MOD13Q1 MODIS	Modelo propuesto por Twilley et al. (1992)	R ² = -0.020	Factor de conversión 0.5	-
				Modelo propuesto por Hutchison et al. (2014)	R ² = 0.023		
				Regresión lineal múltiple (evaluaron 48 modelos diferentes)	Los 3 mejores modelos: (6) R ² = 0.4507 (4) R ² = 0.4675 (3) R ² = 0.4507		

Referencia	Ubicación del bosque de manglar	Lugar de almacenamiento	Imagen satelital	Cálculo de biomasa		Cálculo de carbono	
				Método	Evaluación del método	Método	Evaluación del método
(Bindu et al., 2020)	INDIA - Kerala	Aboveground y Belowground	IRS 1D LISS IV MX	Regresión (NDVI vs AGB)	R ² = 0.710	Factor de conversión 0.475	-
(Thuy et al., 2020)	VIETNAM - Thai Binh	Aboveground	Sentinel 2A	Regresión (Biomasa vs AGB)	R ² = 0.6653	Regresión (Almacenamiento de carbono vs Biomasa)	R ² = 0.999
				Regresión (Biomasa vs AGB)	R ² = 0.9993		
(Anand et al., 2020)	INDIA - Parque Nacional de Bhitarkanika	Aboveground	EO-1 Hyperion hyperspectral sensor	Ecuación alométrica de biomasa	-	Regresión lineal, polinómica, logarítmica, función de base radial y función sigmoideal (NDVI vs Carbono)	Mejor modelo función de base radial: R ² = 84.1%
						Regresión lineal, polinómica, logarítmica, función de base radial y función sigmoideal (EVI vs Carbono)	Mejor modelo función de base radial: R ² = 86.98%
(Mukhtar et al., 2021)	INDONESIA - Mandeh Bay	Aboveground	Landsat 8	Ecuación alométrica de biomasa	-	Regresión lineal, cuadrática, exponencial y de potencia (índice de vegetación vs carbono)	Mejor modelo regresión exponencial usando Mangrove Index: R ² = 0.566

3. Capítulo 3: Materiales y métodos

La figura 2 presenta el resumen de la metodología empleada en el presente estudio la cual será explicada a detalle en las siguientes secciones. El proceso se inicia con el preprocesamiento de las imágenes y los datos de campo, para posteriormente dividirse en dos secciones principales: el mapeo de la cobertura de manglar y la estimación de la biomasa viva (AGB y BGB) y el carbono almacenado.

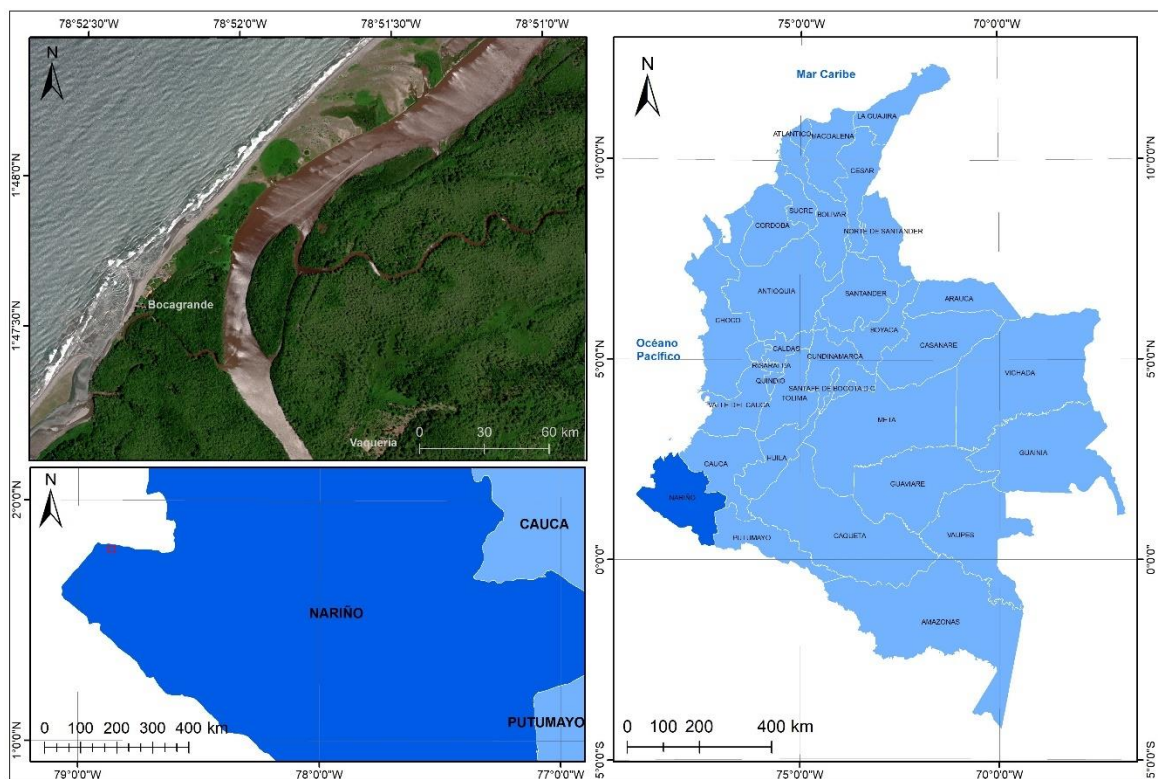
Figura 2. Esquema metodológico general del flujo de trabajo.



3.1. Área de estudio

La zona de estudio se encuentra ubicada en el área de jurisdicción del Consejo Comunitario Bajo Mira y Frontera del municipio de San Andrés de Tumaco en el departamento de Nariño (Figura 3). Los bosques de manglar del departamento de Nariño abarcan cerca de 150.000 hectáreas, dentro de la cual se pueden encontrar las especies de Mangle: rojo (*Rhizophora mangle*), concha (*Rhizophora harrisoni*), piñuelo (*Pelliciera rhizophorae*), nato (*Mora oleifera*), blanco (*Laguncularia racemosa*), negro (*Avicennia germinans*) y botón (*Conocarpus erecta*) (Escobar, 2010). De acuerdo con la investigación de Simard et al. (2019) se pueden encontrar mangles con alturas de hasta de 54.3 m. Por otra parte, la zona de estudio se caracteriza por tener un clima del tipo tropical húmedo, con altos niveles de precipitación en los periodos de diciembre-enero y abril-mayo, donde se presentan valores cercanos a los 400 mm/mes; la temperatura ambiente promedio oscila entre 23 °C en agosto y 26 °C en el periodo abril-mayo (Dimar, 2020).

Figura 3. Ubicación de la zona de estudio en el municipio de Tumaco-Nariño.



3.2. Adquisición y preprocesamiento de datos de campo

3.2.1. Recolección de datos en campo

La recolección de datos en campo se realizó en la zona de interés, ubicada en el municipio de San Andrés de Tumaco – Nariño (Figura 3) durante la última semana de junio y noviembre del 2022. El propósito principal del trabajo de campo fue levantar parcelas con información estructural de los mangles, para posteriormente calcular AGB mediante ecuación alométrica; además, se registraron las coordenadas correspondientes a diversas coberturas con las cuales se creó una base de datos de entrenamiento para la clasificación supervisada que permitió delimitar el área de manglar en la zona de estudio.

Para muestrear el contenido de AGB se levantaron 33 parcelas en el área de estudio, siguiendo la técnica de muestreo aleatorio simple (SRS, por sus siglas en inglés) ya que esta estrategia permite obtener una mayor representatividad de la heterogeneidad del bosque de manglar (Kershaw et al., 2016). Para lograr la aleatoriedad de las parcelas se creó un shapefile de puntos aleatorios, con una distancia de 40 metros entre estos, ubicados dentro del bosque de manglar; para esto se utilizó la herramienta Create Random Points del software ArcGIS 10.8 y la zona del manglar se delimitó con la información espacial proveniente del Reloj Mundial de Manglares versión 2.5 (Global Mangrove Watch, GMW).

Siguiendo la metodología de Kauffman y Donato (2012) las parcelas fueron circulares de 7 metros de radio, dentro de las cuales se muestrearon todos los mangles con un $DAP \geq 5m$ y únicamente dentro de los primeros 2 metros de radio se muestrearon los mangles con un $DAP < 5m$. Cada individuo fue identificado por especie, se le midió el diámetro a la altura del pecho (DAP) y la altura total (h). El DAP se midió a 1.30 m desde la raíz hasta el troco principal, cuando no fue posible llevar a cabo este tipo de medición por las características del individuo o del terreno se midió el DAP 30 cm sobre la última raíz fúlcrea. La altura se tomó con un Clinómetro Electrónico ECII D (GAFMEX, 2023). Finalmente, con un GPS Garmin se georreferenció el centro de cada parcela, además, a lo largo del trayecto se fueron georreferenciando puntos de diferentes coberturas.

3.2.2. Cálculo de biomasa aérea (AGB)

Los valores de biomasa aérea se estimaron haciendo uso de la ecuación alométrica propuesta por Saldarriaga et al. (2011) y establecida en el “Protocolo para la estimación nacional y subnacional de biomasa-carbono en Colombia” (Yepes et al., 2011) para la biomasa aérea (9) .

$$\ln(AGB) [kg] = -3.091 + 2.156 \ln(D) + 0.784 \ln(H) + 0.359 \ln(\rho) \quad (9)$$

Siendo las variables D: diámetro a la altura del pecho (cm), H: altura total (m) y ρ : densidad de la madera (g/cm^3). Los datos de densidad, en masa seca, utilizados para cada especie se encuentran en la Tabla 3. Es importante resaltar que se calculó AGB para cada parcela y se expresó en (t/ha).

Tabla 3. Densidad de la madera utilizada para calcular AGB.

Especie	Densidad [g/cm ³]	Fuente
<i>Avicennia germinans</i>	0.750	(Zanne et al., 2009)
<i>Laguncularia racemosa</i>	0.620	(Zanne et al., 2009)
<i>Rhizophora mangle</i>	0.880	(Zanne et al., 2009)
<i>Pelliciera rhizophorae</i>	0.540	(Monsalve & Ramírez, 2015)

3.2.3. Validación de los datos de AGB

Con el objetivo de conocer la distribución de los datos y determinar su normalidad, se llevó a cabo un análisis de estadística descriptiva para los valores de biomasa aérea por parcela. En este análisis se calcularon la media, la desviación estándar y los cuartiles. Además, se realizaron pruebas de normalidad utilizando la prueba de Shapiro-Wilk, el cual evalúa la hipótesis nula en la cual los datos provienen de una distribución normal (SciPy Community, 2024). Asimismo, se generó un gráfico de probabilidad con el objetivo de visualizar la distribución de los datos en comparación con una distribución normal (SciPy Community, 2024).

3.3. Preprocesamiento de Imágenes satelitales.

3.3.1. Características de la imagen satelital utilizada.

Se adquirió una imagen satelital Worldview-2 modo Bundle, esta tiene una resolución espacial de 1.2 m para las 8 bandas: Coastal, Blue, Green, Yellow, Red, RedEdge, NIR1, NIR2 (Tabla 4) y una resolución espacial de 30 cm para la banda del pancromático. Las imágenes se entregaron corregidas radiométrica y geoméricamente en el nivel de procesamiento Map Ready, con una precisión de localización 10 m CE90. Las características de la imagen se encuentran en la Tabla 5.

Por otra parte, es importante tener en cuenta el nivel de la marea al momento de capturar la imagen ya que esto incide en la reflectividad de la superficie (Nguyen et al., 2020; Perea-Ardila et al., 2021), por esta razón la imagen adquirida fue capturada el 23 de marzo del 2021 en marea baja.

Tabla 4. Resolución espectral de las bandas Worldview 2.

Banda	Nombre	Rango de longitudes de onda (nm)
Pan	Pancromática	450-800
B1	Coastal	400-450
B2	Blue	450-510
B3	Green	510-580
B4	Yellow	585-625
B5	Red	630-690
B6	Red Edge	705-745
B7	Near-IR1	770-895
B8	Near-IR2	860-1040

Tabla 5. Característica de la imagen satelital Worldview2.

Satélite	ID	Fecha	Coordenadas	Imagen	Res (m)
WorldView2	21MAR23155718-P3DS-050101784030_01_P001	2021-03-23	nwLat = 1.80772300 nwLong = -78.87861700	Pancromática	0.30
	21MAR23155718-M3DS-050101784030_01_P001		seLat = 1.78241600 seLong = -78.84737500	Multiespectral	1.2

3.3.2. Preprocesamiento de la imagen satelital

En los píxeles de la imagen adquirida el valor registrado de la cantidad de radiación espectral estaba en números digitales (DN), por lo tanto se convirtieron en valores de radiación espectral en la parte superior de la atmósfera (L_{TOA}) usando la ecuación proporcionada en el withe paper de MAXAR Technologies (2021) (10).

$$L_{TOA} = GAIN * DN \left(\frac{abscalfactor}{effectivebandwidth} \right) + OFFSET \quad (10)$$

Donde *abscalfactor* y *effectivebandwidth* se encuentran en los metadatos de la imagen, DN es el valor contenido en el píxel y GAIN y OFFSET son los factores de ajuste dependientes de la banda de calibración radiométrica absoluta. Luego se convirtió la imagen con valores L_{TOA} a valores de reflectancia (ρ_{SUP}) usando el método QUAC ya que no requiere ningún parámetro atmosférico como entrada (11), esta metodología recupera la reflectancia de la superficie dentro de la imagen en función del parámetro de GAIN y OFFSET (Basith & Prastyani, 2020).

$$\rho_{SUP} = GAIN(L_{TOA} - OFFSET) \quad (11)$$

Además, se le hizo una expansión de contraste a la imagen adquirida, también llamada normalización, para que el histograma gane amplitud y mejore el rendimiento del modelo de aprendizaje de máquina; se utilizó la ecuación (12) para llevar a cabo la normalización (Gonzalez & Woods, 2008; Saeys et al., 2007). Todos los cálculos se hicieron utilizando la

herramienta “raster calculator” de ArcGis 10.8. La imagen se reproyectó al nuevo sistema de proyección colombiana CTM12.

$$V_N = \frac{VALOR - VALOR_MIN}{VALOR_MAX - VALOR_MIN} \tag{12}$$

Por otra parte, se aplicó un refinado pancromático a las bandas Red, Green, Blue y NIR2 para mejorar la resolución de estas y por consecuencia la clasificación supervisada. Estas bandas quedaron con la resolución de la banda pancromática que es de 0.3 m, se utilizó la herramienta “Compute Pansharpen” de ArcGis 10.8.

3.3.3. Cálculo de índices espectrales

Las bandas de las imágenes multiespectrales pueden verse afectadas por factores ambientales externos, lo que puede tener un impacto en los modelos si se utilizan únicamente como variables explicativas en la estimación de AGB (Ahmad et al., 2021). Sin embargo, estos impactos pueden minimizarse mediante la combinación de estas bandas para calcular diferentes índices espectrales; varios estudios han demostrado que la inclusión de índices espectrales mejora el rendimiento de los modelos de estimación de AGB (Forkuor et al., 2020; Zhu et al., 2015).

Se llevó a cabo una revisión bibliográfica para seleccionar los índices espectrales más apropiados en la estimación y mapeo de la biomasa en bosques de manglar utilizando imágenes de Worldview 2. Se eligieron aquellos índices que demostraron mejores resultados, estos se detallan en la Tabla 6. La herramienta "raster calculator" del software ArcGIS 10.8 se utilizó para calcular cada uno de estos índices.

Tabla 6. Índices espectrales derivados de la imagen multiespectral Worldview-2.

Nombre	Función	Ecuación	Fuente
WorldView-Water (WV-WI)	Detecta cuerpos de agua como ríos, lagos y embalses.	$\frac{B8 - B1}{B8 + B1}$	(MAXAR Technologies, 2020)
WorldView-Vegetation (WV-VI)	Evaluar la salud y la cantidad de vegetación en un área	$\frac{B8 - B5}{B8 + B5}$	

Nombre	Función	Ecuación	Fuente
WorldView-Soil (WV-SI)	Identifica áreas de suelo desnudo o diferencia tipos de suelo	$\frac{B4 - B3}{B4 + B3}$	
WorldView-Built-up (WV-BI)	Diferencia estructuras y áreas urbanas de otros tipos de cobertura del suelo	$\frac{B6 - B1}{B6 + B1}$	
RedEdge Normalized Difference Vegetation Index (NDVI65)	Contenido de follaje del dosel	$\frac{B6 - B5}{B6 + B5}$	(Zhu et al., 2017)
RedEdge Simple Ratio Index (SR65)	Concentración de clorofila y contenido de humedad	$\frac{B6}{B5}$	
RedEdge Normalized Difference Vegetation Index (NDVI651)	Monitoreo forestal y detección del estrés de la vegetación	$\frac{B6 - B1}{B5 + B1}$	
RedEdge Simple Ratio Index (SR61)	Estado de la vegetación y estructura del dosel	$\frac{B6}{B1}$	
Normalized Difference Vegetation Index (NDVI84)	Estado de la vegetación y estructura del dosel	$\frac{B8 - B4}{B8 + B4}$	
Simple Ratio Index (SR84)	Estado de la vegetación y estructura del dosel	$\frac{B8}{B4}$	
Simple Ratio Index (SR85)	Estado de la vegetación y estructura del dosel	$\frac{B8}{B5}$	
Normalized difference Water Index (NDWI)	Detectar la cantidad de agua en la vegetación	$\frac{B3 - B8}{B3 + B8}$	
Normalized Difference Vegetation Index (NDVI75)	Evaluar la densidad y el vigor de la vegetación	$\frac{B7 - B5}{B7 + B5}$	(Zhu et al., 2015)
Simple Ratio Index (SR75)	Evaluar la densidad y el vigor de la vegetación	$\frac{B7}{B5}$	

Nombre	Función	Ecuación	Fuente
Difference Vegetation Index (DVI75)	Evaluar la densidad y el vigor de la vegetación	$B7 - B5$	
Difference Vegetation Index (DVI85)	Evaluar la densidad y el vigor de la vegetación	$B8 - B5$	
Optimized Soil-Adjusted Vegetation Index (OSAVI)	Evaluar la densidad y la salud de la vegetación en áreas con suelos desnudos o con vegetación escasa	$\frac{B8 - B5}{B8 + B5 + 0.16}$	(Rondeaux et al., 1996)
Enhanced Vegetation Index 2 (EVI2)	Estimar la densidad y el vigor de la vegetación, especialmente en áreas con alta cobertura vegetal.	$2.5 \left(\frac{B8 - B5}{B8 + (2.4 * B5) + 1} \right)$	(Jiang et al., 2008)
Soil-Adjusted Vegetation Index (SAVI)	Evaluar la densidad y la salud de la vegetación en áreas con suelos desnudos o con vegetación escasa	$1.75 \left(\frac{B8 - B5}{B8 + B5 + 0.75} \right)$	(Huete, 1988)
Combined Mangrove Recognition Index (CMRI)	Bosques de manglar	$WV_VI - NDWI$	(Gupta et al., 2018)

3.4. Mapeo de la cobertura del bosque de manglar

En el presente trabajo se utilizó el método Maximum Likelihood para la clasificación supervisada; esto se realizó en el programa ArcGis 10.8.

3.4.1. Datos de entrada para la clasificación de cobertura de manglar.

La calidad de una clasificación supervisada depende de la calidad de los datos de entrenamiento que se ingresen al algoritmo (Perumal & Bhaskaran, 2010); para esta investigación se utilizaron como datos de entrada 145 puntos recolectados en campo durante el año 2022, adicional a esto se crearon 230 puntos aleatorios, distanciados 8m, con ayuda de la herramienta “random points” de ArcGis 10.8. Tanto a los puntos

recolectados en campo como a los puntos aleatorios se les construyó un buffer de 5m con el fin de que abarcaran más píxeles para alimentar el algoritmo.

Por otra parte, las bandas que se utilizaron para la clasificación fueron RGB (Red, Green and Blue) y NIR2 (Near infrared) de Worldview 2. En el estudio de Santos & Freire (2015) el uso de bandas que capturan radiación electromagnética tanto visible como no visible mejora la exactitud de la clasificación de diferentes coberturas. Además, la adición de la banda NIR2 de Worldview 2 al conjunto de bandas tradicionales RGB mejora los resultados de la clasificación; ya que el rango espectral de NIR2 (8860-900 nm) es menos sensible a los efectos atmosféricos, que la banda NIR1 (7770-895 nm), y está destinada al análisis de la vegetación (Digital Globe, 2010).

3.4.2. Esquema de clasificación

Los esquemas de clasificación son un medio para organizar la información espacial de manera ordenada y lógica y permiten reducir el número total de elementos considerados a un número razonable (Congalton & Green, 2008). El esquema de clasificación que se definió en este proyecto tiene la finalidad de diferenciar la zona de manglar de otras zonas; además, este esquema tuvo en cuenta lo dispuesto en la metodología Corine Land Cover Adaptada para Colombia (Ideam, 2010). A continuación se describe el esquema de clasificación.

- Manglar: Se refiere a las zonas con vegetación de manglar, las cuales se caracterizan por tener un estrato más o menos continuo y suelen ubicarse comúnmente en las franjas cercanas a cuerpos de agua, especialmente los lóticos.
- Sin cobertura vegetal: en esta se unieron dos etiquetas, suelo desnudo y urbanización, ya que facilitaba la clasificación al representar zonas muy reducidas con características espectrales similares. Dentro de esta etiqueta hacen parte las áreas donde no hay presencia de cobertura vegetal como: zonas arenosas y zonas de tala; y zonas urbanizadas como: viviendas, escuelas, hospitales, infraestructuras de comunicación, áreas asfaltadas, mercados o industrias, escombreras, vertederos e infraestructuras vacacionales.
- Zonas húmedas: en esta se unieron dos etiquetas, sedimentos expuestos en baja mar y superficies de agua, ya que facilitaba la clasificación al representar zonas con características espectrales similares. Esta categoría engloba las extensiones formadas

por llanuras litorales que quedan al descubierto durante las fases de bajamar y áreas como ríos, canales, mares y océanos.

- Otra vegetación: Se refiere a las zonas que están predominantemente cubiertas por vegetación arbustiva, caracterizada por un dosel irregular y la presencia de arbustos, palmas, enredaderas y vegetación de baja altura como: helechal, herbazal abierto arenoso, vegetación secundaria, árboles aislados y otro tipo de vegetación diferente al Mangle.

3.4.3. Post procesamiento de la imagen.

Para eliminar el ruido que genera el efecto “sal y pimienta” y mejorar la calidad de la imagen clasificada se llevaron a cabo dos procesos. Primero se eliminaron los píxeles aislados aplicando un filtro mayoritario con una ventana de 3*3 con la restricción de que la mitad de las celdas deben tener el mismo valor y ser contiguas para que ocurra un reemplazo (McCarthy & Halls, 2014; Rahman et al., 2019); luego se suavizaron los límites de clase. Estos procesos se llevaron a cabo en el software ArcGis 10.8, con las herramientas "Majority Filter" y "Boundary Clean" respectivamente.

3.4.4. Evaluación de la exactitud del mapa de cobertura

La "regla de oro" de Congalton & Green (2008) para la accuracy assessment de mapas dice que se deben recolectar mínimo 50 muestras por clase para mapas menores a 4000 km². Por ende, para este estudio, se crearon 200 puntos aleatorios, por clase, con el propósito de evaluar la precisión del mapa clasificado. Siguiendo la recomendación de Olofsson et al. (2014), en donde manifiesta la necesidad de evaluar los mapas con recursos de mayor resolución espectral o espacial a la utilizada en la clasificación, estos puntos fueron interpretados visualmente utilizando imágenes Google Earth (Google & Maxar Technologies, 2021) y Bing Maps (Microsoft Bing & Maxar Technologies, 2022), correspondientes a los años 2021 y 2022 respectivamente, y considerando la experiencia previa adquirida en el terreno. Posteriormente, se utilizó el plugin "Class Accuracy", disponible para QGIS (Bunting, 2020), para generar la matriz de error y posteriormente calcular la precisión general, el coeficiente kappa, la precisión del productor y la del usuario, junto con sus respectivos intervalos de confianza, tal como se explica en el libro

de Congalton & Green (2008). Para llevar a cabo la evaluación del mapa desarrollado en este trabajo se utilizó el software QGIS 3.28 (QGIS, 2023).

3.5. Estimación de AGB en el bosque de manglar.

3.5.1. Optimización de variables explicativas para modelos de AGB.

Con el objetivo de simplificar los modelos, mitigar la multicolinealidad y mejorar su rendimiento, se implementó una optimización de las variables explicativas. Este proceso involucró diversos criterios para reducir la cantidad de variables. Como consecuencia, se generaron tres conjuntos distintos de variables explicativas. En la Tabla 7 se presentan las diferentes agrupaciones de variables explicativas utilizadas para entrenar los modelos.

En el primer grupo se emplearon todas las variables elegidas previamente (Tabla 6), un total de 28 variables explicativas, que comprendían 8 bandas de la imagen multiespectral de Worldview 2, previamente sometidas a un preprocesamiento, y 20 índices espectrales.

Para elegir las variables del segundo grupo, se llevó a cabo un análisis de componentes principales (PCA, por sus siglas en inglés) utilizando el paquete “stats” del software R Studio (Bolar, 2019). En este proceso, se optó por elegir aquellas variables cuya contribución a los dos primeros componentes superara el promedio esperado, en el supuesto de una distribución uniforme de contribuciones.

Finalmente, dado que la prueba de Spearman evidenció una alta correlación entre la mayoría de las variables (>0.7) del segundo grupo, se llevó a cabo una tercera selección para reducir la multicolinealidad de los modelos. En este proceso, se consideró la importancia de las variables mediante la pureza de los nodos en el modelo de Random Forest, además de aplicar el criterio propio del investigador.

Tabla 7. Agrupación de variables explicativas para entrenamiento de los modelos de AGB.

Variables	Corrida 1	Corrida 2	Corrida 3	Variables	Corrida 1	Corrida 2	Corrida 3
YELLOW	X			NDVI75	X	X	
REDEGE	X	X		NDVI84	X	X	X
RED	X			NDWI	X	X	
NIR1	X	X		OSAVI	X	X	
NIR2	X	X		SAVI	X	X	
GREEN	X			SR61	X	X	X
COASTAL	X			SR65	X		
BLUE	X			SR75	X		
CMRI	X	X	X	SR84	X		
DVI75	X			SR85	X		
DVI85	X			WVBI	X	X	
EVI2	X			WVSI	X		
NDVI65	X	X		WVVI	X	X	
NDVI651	X			WVWI	X	X	

3.5.2. Modelos para la estimación de AGB

Dado que los datos no seguían una distribución normal, se optó por utilizar modelos no paramétricos. Se evaluaron dos algoritmos, Random Forest y Support Vector Regression, para modelar el AGB en los bosques de manglar utilizando el paquete "caret" en R Studio (Kuhn et al., 2023). Estos algoritmos fueron seleccionados debido a sus buenos resultados en investigaciones anteriores (Hu et al., 2020; Selvaraj & Gallego-Pérez, 2023; Y. Zhu et al., 2020).

Cada algoritmo se entrenó individualmente con diferentes conjuntos de variables explicativas (Tabla 7). Se utilizaron 33 datos de AGB calculados a partir de mediciones en campo, y se incluyeron datos con AGB igual a 0 t/ha en cuerpos de agua. La inclusión de estos valores de AGB cero se llevó a cabo con el objetivo de mejorar el entrenamiento del modelo, asegurando que, en áreas con cuerpos de agua, la predicción de biomasa aérea fuera cero (Hu et al., 2020).

Para compensar la disparidad entre las dimensiones de las parcelas en campo y el área de los píxeles en las imágenes, se implementó un buffer de 7 metros de radio en los puntos de biomasa aérea. Este radio refleja la dimensión real de las parcelas en el terreno. Dicho buffer se empleó para extraer los valores de las variables explicativas asociadas a cada parcela, los cuales fueron luego promediados. Este enfoque sigue la metodología utilizada por Hu et al. (2020) y Mutanga et al. (2012) en sus respectivos estudios.

3.5.3. Ajuste de hiperparámetros

En el contexto del algoritmo RF, se llevó a cabo una optimización de los hiperparámetros *mtry* y *ntree*. Esta optimización se realizó mediante una exploración de valores dentro de un rango predefinido, que fue seleccionado automáticamente por el paquete utilizado. En el caso de *mtry*, se evaluaron valores desde dos hasta el número máximo de variables explicativas que se introdujeron en el modelo. En cuanto a *ntree*, se exploraron valores en el rango de 0 a 500. La métrica de evaluación utilizada para determinar el mejor hiperparámetro fue el RMSE.

Para el algoritmo SVR, se optimizaron los hiperparámetros *sepsilon* y *cost* utilizando una combinación de valores 0.01, 0.1, 1, 10 para ambos. Se mantuvieron los valores predeterminados para *Sigma*. Se utilizó el kernel radial, ya que permite mapear los datos en un espacio de mayor dimensionalidad y construir un hiperplano de separación no lineal (Awad & Khanna, 2015), lo cual era apropiado debido a la falta de una relación lineal entre los datos de entrenamiento y las variables explicativas.

3.5.4. Evaluación de los modelos de AGB

Debido a la escasez de datos disponibles, se optó por utilizar la validación cruzada para evaluar los modelos, evitando así la necesidad de dividir la base de datos y permitiendo el entrenamiento con la totalidad de los datos disponibles. Se realizaron 16 k-folds, lo que asegura que en cada corrida se utilicen al menos 2 datos para evaluar el modelo, lo cual proporciona una mayor variabilidad en los datos de prueba.

Como métrica para evaluar la calidad de los modelos, se seleccionó el error cuadrático medio de la raíz, ya que está expresado en las mismas unidades que la variable explicativa, lo que facilita una mejor comparación, y es ampliamente utilizada en la mayoría de los

artículos científicos (Lu et al., 2016; Simard et al., 2019; Vu et al., 2014). También se calculó el coeficiente de determinación y el error absoluto medio

El mejor modelo obtenido se utilizó para predecir los valores de AGB en toda el área de estudio. Para ello, se realizó una predicción en la imagen completa y posteriormente se recortó utilizando el polígono correspondiente al área de manglar obtenido del mapa de cobertura.

3.6. Estimación de BGB y el carbono almacenado en el bosque de manglar.

Se utilizó una estimación indirecta para calcular BGB y el contenido de carbono en el bosque de manglar, basándose en el contenido de AGB previamente calculado utilizando el mejor modelo. Para el cálculo del BGB, se empleó la ecuación alométrica propuesta por (Hutchison et al., 2014) (13). Siendo las variables AGB: aboveground biomass [t/ha] y BGB: belowground biomass [t/ha].

$$BGB\left[\frac{t}{ha}\right] = 0.073 AGB^{1.32} \quad (13)$$

Por otra parte, el carbono se calculó empleando lo propuesto por Howard et al. (2014) en el documento “Métodos para evaluar las reservas de carbono y los factores de emisión en manglares, marismas y praderas marinas” (14).

$$Carbon\left[\frac{tC}{ha}\right] = (0.5 * AGB) + (0.39 * BGB) \quad (14)$$

4. Capítulo 4: Resultados y Discusión

4.1. Características de los datos de AGB recolectados en campo.

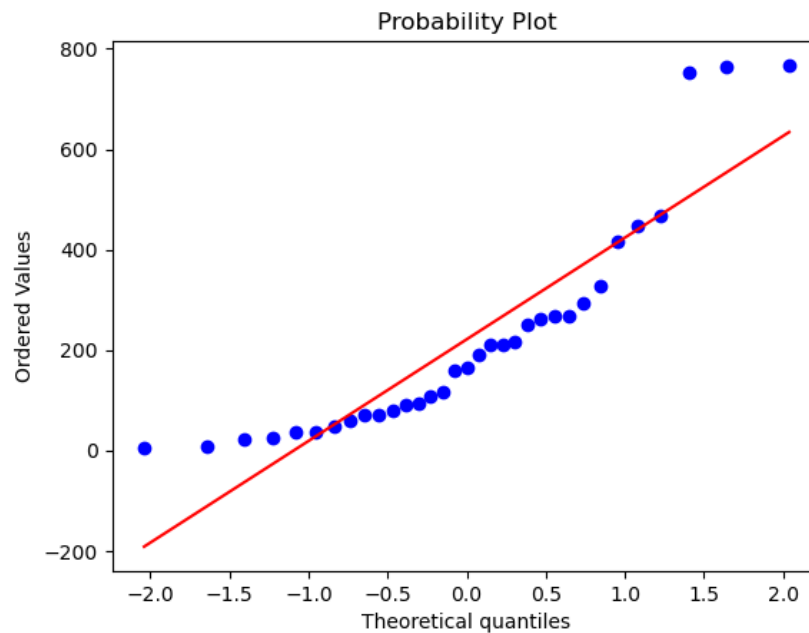
Se llevaron a cabo mediciones en un total de 33 parcelas en el campo, obteniendo un promedio de biomasa aérea de 230.82 t/ha, con una desviación estándar de 223.92 t/ha. Estas parcelas exhiben una notable variabilidad en su contenido de AGB, presentando un valor máximo de 787.76 t/ha y un valor mínimo de 6.02 t/ha (Tabla 8). Esta variabilidad era previsible debido a la presencia de diversas configuraciones de bosques de manglar en la zona de estudio. Por ejemplo, algunas parcelas están dominadas por *Laguncularia racemosa* y otras por *Rhizophora mangle*, lo que naturalmente resulta en una variabilidad en los valores de biomasa aérea.

Por otro lado, al realizar la prueba de normalidad Shapiro-Wilk, se obtuvo un estadístico $W=0.821$, lo que sugiere que los datos podrían ajustarse a una distribución normal. Sin embargo, el valor $p=0.000$ resultó ser menor que el nivel de significancia seleccionado, que fue 0.05, lo que indica que existe evidencia estadística para rechazar la hipótesis nula de normalidad. Para complementar la prueba de Shapiro-Wilk con un análisis visual de los datos, se generó un gráfico de probabilidad (Figura 4), el cual compara los cuantiles ordenados de la muestra con los cuantiles esperados de una distribución normal y traza una línea de mejor ajuste para los datos, facilitando así la visualización de la concordancia entre los datos muestreados y la distribución teórica (SciPy Community, 2024). En la figura 4 se observan desviaciones en los puntos con respecto a la línea de ajuste, lo que sugiere posibles desviaciones de la normalidad. Considerando ambos análisis, se concluyó que los datos no siguen una distribución normal, por lo que se deben emplear enfoques no paramétricos.

Tabla 8. Estadística descriptiva de las parcelas de AGB (t/ha) levantadas en campo.

Media	Desviación estándar	Valor mínimo	Valor máximo	Cuartiles		
				25%	50%	75%
230.82	223.92	6.02	787.76	72.99	162.39	276.09

Figura 4. Gráfico de probabilidad normal para las parcelas de AGB (t/ha) levantadas en campo.



La Tabla 9, exhibe una comparativa entre el presente estudio e investigaciones anteriores que han evaluado la biomasa aérea en campo de bosques de manglar en diversas zonas del Pacífico Colombiano. En general, este estudio sobresale con un promedio de AGB superior y una alta desviación estándar 230.820 ± 223.92 t/ha, casi en paridad con la media. La proximidad entre la desviación estándar y la media guarda relación con el estudio realizado por Monsalve & Ramírez (2015) en Bahía Málaga, el cual registró un promedio y desviación estándar de AGB de 143.7 ± 101 t/ha. Esta variabilidad podría atribuirse a la heterogeneidad inherente al bosque de manglar, como indican diversas investigaciones previas (Castellanos-Galindo et al., 2021; Gross et al., 2014; Rovai et al., 2021). Adicionalmente, Palacios-Peñaranda et al. (2019) llevó a cabo un muestreo en

áreas del Pacífico adyacentes al municipio de Buenaventura, siendo el valor máximo registrado de AGB 142.46 t/ha.

Por otra parte, en proximidad a la zona de estudio de esta investigación, el trabajo de Castellanos-Galindo et al. (2021) reveló un promedio de AGB de 626.4 t/ha, excediendo la cifra obtenida en esta tesis de investigación. Sin embargo, los autores enfatizaron en que su trabajo de campo se enfocó únicamente en zonas donde previamente se habían documentado los ejemplares de manglar más altos de Colombia (Simard et al., 2019), lo cual influyó en los valores promedio de AGB, elevándolos más allá de lo que habrían sido si las parcelas de campo hubiesen abarcado la diversidad del bosque de manglar. No obstante, es fundamental resaltar que en esta tesis de investigación se presentaron parcelas con valores de AGB cercanos al promedio encontrado por Castellanos-Galindo et al. (2021).

Tabla 9. Above Ground Biomass estimado en campo en diferentes zonas del Pacífico Colombiano.

Work	Location	Mean AGB t /ha \pm SD
(Palacios-Peñaranda et al., 2019)	Pacifico- Quebrada Valencia	142.46
	Pacifico -San Pedro	69.49
	Pacifico -Pianguita	109.18
(Monsalve & Ramírez, 2015)	Pacífico - Bahía Málaga	143.7 \pm 101.7
(Castellanos-Galindo, et al., 2021)	Pacífico - Costa norte departamento de Nariño	626.4
This study	Pacífico - BocagrandeTumaco	230.820 \pm 223.92

4.2. Selección de variables explicativas

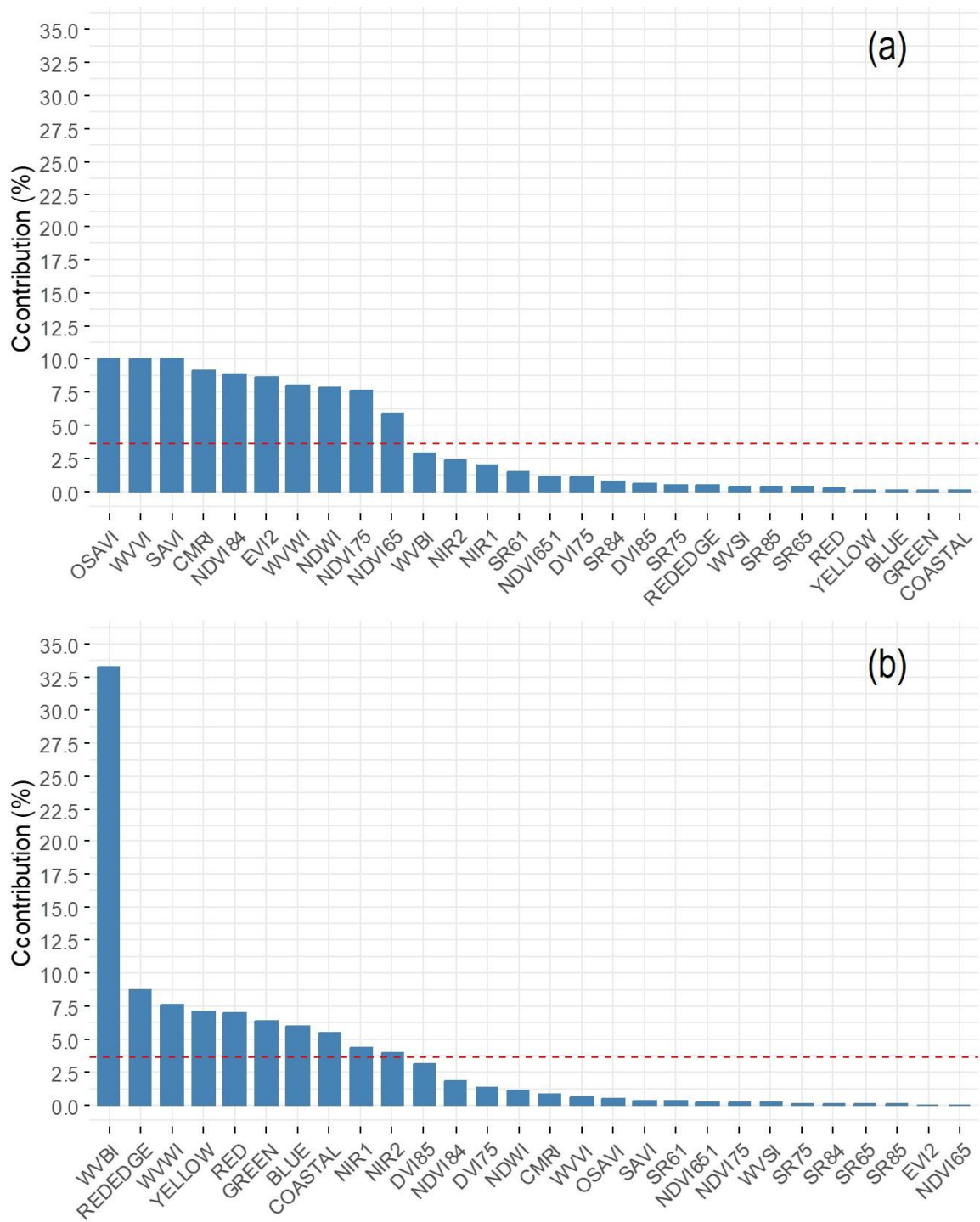
El segundo grupo de variables se construyó con base en un Análisis de Componentes Principales en el cual se identificaron dos componentes que explicaron la mayor parte de la varianza total (98.6%). El primero explicó el 95.5% de la varianza, mientras que el

segundo explicó el 3.1% de la varianza. Para la selección de las variables que presentaron la mayor contribución a cada componente, se empleó un Gráfico de Contribuciones de Variables que ilustra el porcentaje de contribución de cada variable a cada componente (Figura 5). En este gráfico, la línea punteada refleja el promedio de las contribuciones al componente; cuando la contribución de una variable supera esta línea de referencia, indica que su aporte es mayor al esperado si todas las variables contribuyeran de manera uniforme, por lo tanto, podría considerarse como importante en ese componente (Alboukadel, 2023). Las variables que superaron esta línea de referencia fueron: EVI2, WVVI, OSAVI, SAVI, CMRI, NDVI84, NDWI, NDVI65, NDVI75 y NIR2, NIR1, REDEEDGE, WVWI, WVBI y SR61.

La selección de variables para el tercer grupo se basó en la importancia de las variables del segundo grupo, evaluada mediante la pureza de los nodos en el modelo de Random Forest (Figura 7), y el criterio del investigador. En este proceso, se procedió a descartar variables en orden descendente considerando las características y propiedades específicas de cada una. Las variables seleccionadas fueron SR61, CMRI y NDVI84.

La elección de SR61 se justifica debido a que su cálculo implica el uso de las bandas REDEEDGE y COASTAL, reconocidas por su utilidad en estudios de biomasa (Santos & Freire, 2015; Zhu et al., 2017). Además, la banda REDEEDGE minimiza la influencia de la absorción atmosférica, del agua y del fondo del suelo (Mutanga et al., 2012). En cuanto a NDVI84, este índice se deriva de NIR2, menos susceptible a los efectos atmosféricos, lo que lo hace especialmente idóneo para el análisis de la vegetación. La inclusión de la banda YELLOW en NDVI84 resulta útil en la cartografía de la vegetación al detectar componentes de "color amarillento" en los objetivos (Santos & Freire, 2015). Por último, CMRI es un índice especial para el reconocimiento de manglares y se deriva de WVVI y NDWI, siendo este último sensible a los cambios en el contenido de agua líquida de las coberturas vegetales (Gao, 1996).

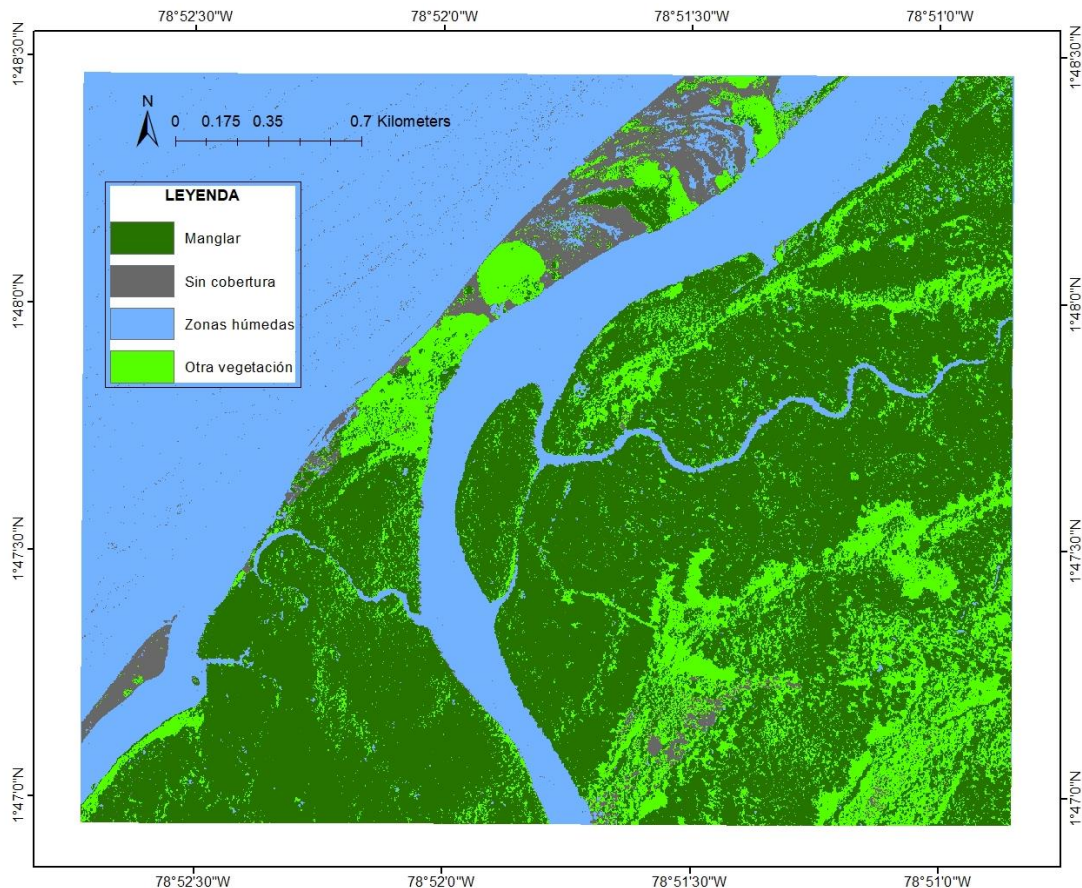
Figura 5. Contribuciones de variables al componente: (a) Componente 1, (b) Componente 2.



4.3. Caracterización y cuantificación de la cobertura del bosque de manglar

En relación con la cuantificación de las diferentes coberturas en la zona de estudio, se determinó que la cobertura predominante es la de manglar, ocupando un área de 437.32 ha. Le sigue en extensión las zonas húmedas, con 378.75 ha. Por otro lado, las áreas con menor cobertura corresponden a otra vegetación y a zonas sin cobertura, con dimensiones de 128.4 ha y 36.9 ha respectivamente (Figura 6). Adicionalmente, la precisión global de la imagen clasificada alcanzó un valor del $87\% \pm 9\%$, esto indica que existe una probabilidad entre el 78% y el 96% de que un píxel clasificado en una cierta categoría coincida con la realidad en el terreno; y el valor Kappa fue de 0.83. Ambos resultados sugieren una buena concordancia entre las clasificaciones realizadas en la imagen y las condiciones reales presentes en el terreno.

Figura 6. Mapa clasificación supervisada para el área de estudio.



En lo que respecta a la categoría de manglar, la exactitud del productor se cifró en 92% ± 5%, esto significa que el modelo acierta en un rango del 87% al 97% al clasificar correctamente las áreas como manglar cuando el modelo decide que son manglar (Tabla 10). La varianza baja de 0.001 refuerza la consistencia y confiabilidad de las clasificaciones del modelo en esta categoría. Por otro lado, la exactitud del usuario se estima en 86% ± 5%, este resultado indica que el modelo tiene un rango de acierto del 81% al 91% al clasificar adecuadamente las áreas que en realidad son manglar. La varianza baja, nuevamente de 0.001, destaca la coherencia y seguridad en las clasificaciones del modelo en este ámbito. En resumen, ambos resultados indican que el modelo es altamente competente en la clasificación de áreas como "manglar", tanto desde la perspectiva del productor como del usuario.

De acuerdo con lo señalado por Lu et al. (2016), en lo que respecta a las estrategias de planificación y gestión forestal a niveles regionales, es deseable contar con una precisión superior al 90%. En casos de planificación y gestión a nivel nacional y global, una precisión del 80% puede ser aceptable. En este estudio, se abordó un área a escala local, y el modelo logró alcanzar una precisión general del 87% al considerar todas las categorías de cobertura y específicamente para la cobertura de manglar, el modelo logró una precisión del 92%. Por lo tanto, este estudio se encuentra dentro del rango de precisión requerido para la toma de decisiones en este contexto.

Tabla 10. Precisión del productor y del usuario de la clasificación de la cobertura de manglar para el área de estudio ubicada en el Municipio de San Andrés de Tumaco.

Classification scheme	Producer's accuracy			User's accuracy		
	probability	Variance	Confidence interval	probability	Variance	Confidence interval
Otra vegetación	0.62	0.005	0.14	0.76	0.001	0.06
Manglar	0.92	0.001	0.05	0.86	0.001	0.05
Sin cobertura vegetal	0.57	0.017	0.26	0.83	0.001	0.05
Zona húmeda	1	0.000	0.00	0.96	0.0002	0.03

Al comparar los resultados obtenidos en este estudio con otras investigaciones similares que se centraban en delimitar la extensión de áreas de manglar, se puede concluir que este trabajo ha logrado una clasificación destacada. Por ejemplo, el análisis realizado por

Perea-Ardila et al. (2019) en la Bahía de Buenaventura, empleando imágenes del satélite Sentinel 2A y el clasificador Maximum Likelihood, arrojó un valor Kappa de 0.70. Por otra parte, en estudios a escala global, como la versión 2.5 del Global Mangrove Watch Bunting et al. (2022), alcanzó un valor de Kappa de 0.9 y una precisión general de $95\% \pm 1.5\%$, utilizando imágenes del satélite Sentinel-2 y un clasificador XGBoost.

4.4. Rendimiento de los modelos de AGB.

Generalmente el comportamiento de los modelos ejecutados con el algoritmo Random Forest mejoraron sus métricas a medida que se redujeron el número de variables (Tabla 11). El mejor modelo del algoritmo RF es el de la tercera ejecución, que utiliza la menor cantidad de variables explicativas. Aunque en la segunda ejecución el RMSE es más bajo que en la tercera ejecución, al considerar la desviación estándar (SD, por sus siglas en inglés), se revela que esta última es menor en la tercera corrida. Esto sugiere que, aunque el RMSE sea ligeramente superior en la tercera ejecución, el modelo es más confiable, ya que muestra una menor variabilidad en los resultados de la métrica. Además, al tener en cuenta el coeficiente de determinación (R^2), se observa que la reducción de variables mejora la capacidad del modelo para explicar una mayor variabilidad.

Por otro lado, los modelos implementados con el algoritmo Support Vector Regression experimentaron un deterioro en sus métricas de evaluación a medida que se redujeron las variables (Tabla 11). El mejor modelo sería el de la primera ejecución, que utiliza todas las variables. Sin embargo, es importante señalar que la segunda ejecución presentó el valor más bajo de R^2 y la mayor SD y el MAE más alto, lo que hace que la segunda corrida sea considerada como el modelo menos preciso de SVR.

Tabla 11. Métricas de evaluación para modelos de AGB (t/ha).

Run	Model	RMSE (t/ha)	RMSE SD (t/ha)	R ²	R ² SD	MAE (t/ha)	MAE SD (t/ha)
1	RF	144.52	103.37	0.70	0.33	93.556	65.22
	SVR	134.20	112.02	0.72	0.28	88.79	65.78
2	RF	137.66	107.24	0.75	0.31	87.97	67.48
	SVR	137.86	120.12	0.65	0.32	93.26	71.03
3	RF	140.68	98.76	0.78	0.28	84.39	60.63
	SVR	140.25	117.94	0.70	0.31	88.92	68.67

En comparación con los mejores modelos entrenados con cada algoritmo, RF mostró un mejor rendimiento. Además, al considerar las proyecciones generadas en el área de estudio, registró valores de AGB de hasta 600 t/ha, lo cual se acerca más a la realidad de las parcelas muestreadas en campo, donde había parcelas con valores de hasta más de 700 t/ha. El rendimiento superior del algoritmo RF se debe a su eficacia para abordar datos atípicos de manera efectiva (Breiman, 2001). En la base de datos de AGB, se identificaron valores atípicos que, aunque reflejaban las características reales de los manglares, eran considerados atípicos en el contexto del conjunto de datos. RF tiene la capacidad de lidiar con datos estos datos ya que promedia o vota entre múltiples árboles construidos de manera independiente, por lo tanto, la influencia de ese valor atípico en el resultado final se atenúa, ya que otros árboles pueden tener en cuenta la variabilidad natural de los datos y no verse tan afectados por valores extremos.

Es pertinente resaltar que la optimización de variables condujo a mejoras en el modelo entrenado con RF, subrayando así la importancia crítica de las variables explicativas para este algoritmo, donde su selección cuidadosa contribuye a la mejora general del modelo. Sin embargo, en el caso del modelo con SVR, la reducción en el número de variables resultó en un rendimiento inferior. Este resultado podría deberse a dos factores. En primer lugar, la selección final de variables se basó, en parte, en la importancia de las variables para el modelo de RF, lo cual podría introducir un sesgo hacia el modelo SVR. Un segundo aspecto es la capacidad de SVR para manejar conjuntos de datos con muchas variables explicativas, incluso cuando estas superan en número a los datos de entrenamiento (Bishop, 2008).

4.4.1. Importancia de las variables al modelo de AGB con el algoritmo Random Forest .

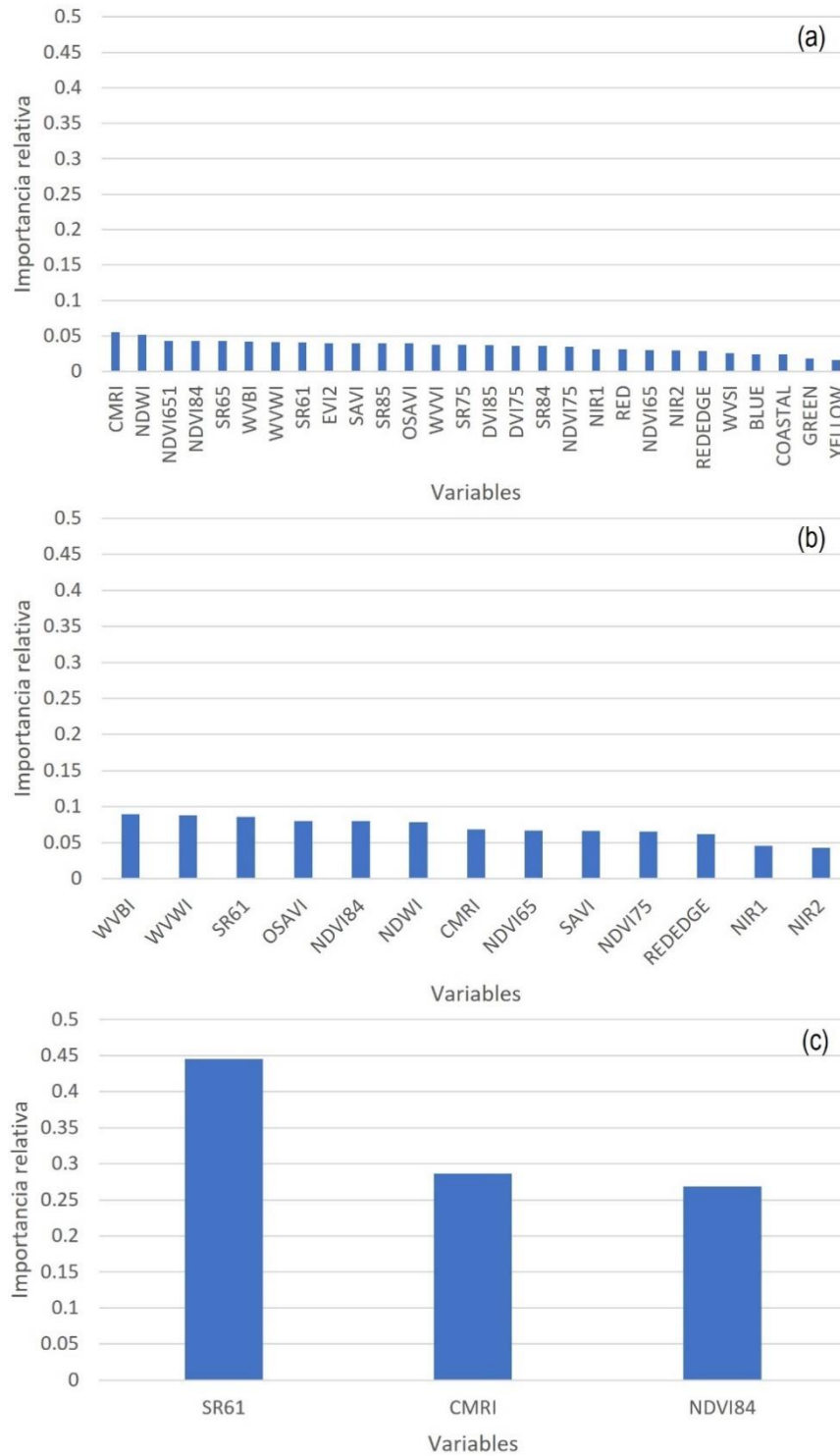
La Figura 7 exhibe la importancia relativa de las variables explicativas en cada ejecución del algoritmo RF. Esta importancia está en función de la pureza de los nodos en cada árbol de decisión; es decir, si una variable presenta mayor incremento de la pureza de los nodos, denota su importancia en el modelo y su influencia en la mejora de la precisión de las predicciones en cada nodo (Cutler et al., 2007). Como se mencionó previamente, esta importancia se relativizó para expresar los valores de manera proporcional, facilitando la identificación de aquellas que tienen un mayor impacto relativo en la mejora de la pureza en el contexto del modelo.

En la primera corrida, las cinco variables de mayor importancia fueron CMRI, NDWI, NDVI657, NDVI84 y SR65. En la segunda corrida, destacaron WVBI, WVWI, WR61, OSAVI y NDVI84. Y en la última corrida, que presentó mejor capacidad predictiva, la variable más importante fue SR61. En contraste, las cinco variables menos relevantes para la primera corrida fueron YELLOW, GREEN, COASTAL, BLUE y WWSI; para la segunda corrida fueron NIR2, NIR1, REDEGE, NDVI75 y SAVI, y en la última corrida, NDVI84 ocupó la posición de menor importancia.

De estos resultados se puede resaltar que, no se observó una tendencia consistente donde una variable siempre obtuviera la mayor importancia. Las variables presentaron importancias diversas según las agrupaciones, reafirmando que las variables explicativas no operan de manera independiente; su comportamiento puede mejorar o deteriorar los resultados del modelo al combinarse o al ser excluidas unas de otras (Bishop, 2008).

Además, en su mayoría, los índices de vegetación superaron en importancia a las bandas. Este resultado concuerda con estudios previos que han destacado la superioridad de los índices de vegetación derivados de las bandas en la modelización de la Biomasa Aérea en comparación con las bandas sin procesar, ya que tienen la capacidad de mitigar los efectos de la reflectancia causada por condiciones ambientales y sombras (Ahmad et al., 2021; Forkuor et al., 2020; Y. Zhu et al., 2015).

Figura 7. Importancia de las variables explicativas para el modelo Random Forest: (a) Primera corrida con todas las variables, (b) Segunda corrida con variables del primer filtro, (c) Tercera corrida con variables del segundo filtro.



4.4. Cuantificación de AGB, BGB y Carbono almacenado en el bosque de manglar.

El promedio de la biomasa aérea en la zona de estudio se sitúa en 192.50 ± 102.78 t/ha, con un mínimo registrado de 33.19 t/ha y un máximo de 605.33 t/ha (Tabla 12). Estos valores difieren de los valores mínimos y máximos observados en campo, 6.02 t/ha y 787.76 t/ha, respectivamente, indicando una propensión del modelo a sobreestimar la biomasa cuando es baja y subestimarla cuando es alta, una tendencia que se ha observado en otros estudios basados en machine learning (Hu et al., 2020; Lu et al., 2016; Mutanga et al., 2012). Este comportamiento puede atribuirse a limitaciones en las imágenes ópticas, que muestran saturación por la dispersión en el dosel debido a la densa vegetación (Zhu et al., 2020) o a la presencia de pocos datos en los valores extremos, que actuaban como valores atípicos en la base de datos. Por otra parte, la desviación estándar del promedio de la biomasa aérea es elevada, reflejando la heterogeneidad de la zona de estudio. La Figura 8 ilustra la distribución de la biomasa aérea en dicha área.

Es importante destacar que en general los modelos entrenados con algoritmos de aprendizaje de máquina buscan aprender patrones y relaciones del conjunto de datos de entrenamiento, permitiendo así realizar predicciones en datos no observados. Sin embargo, estos modelos tienen limitaciones inherentes y no logran capturar todas las variaciones y complejidades presentes en un conjunto de datos. Por lo tanto, es normal y esperado encontrar diferencias entre las predicciones del modelo y los valores de entrenamiento. Es precisamente en este contexto que las métricas de evaluación, como RMSE, MAE y R^2 , proporcionan información sobre el rendimiento del modelo y su capacidad de generalización.

En lo que respecta a la biomasa subterránea y el carbono almacenado en AGB y BGB, se obtuvo un promedio de 79.95 ± 56.85 t/ha y 127.43 ± 73.49 t/ha respectivamente (Tabla 12). Estos resultados están estrechamente vinculados a la precisión del modelo empleado para estimar la biomasa aérea, ya que los cálculos de BGB y carbono se basan en los resultados de AGB. La distribución de los valores de BGB y carbono en el área de estudio se presenta en las Figuras 9 y 10, respectivamente.

Tabla 12. Cuantificación de AGB, BGB y Carbono almacenado en los bosques de manglar del área de estudio.

	Mean \pm SD	Min	Max
AGB [t/ha]	192.50 \pm 102.78	33.19	605.33
BGB [t/ha]	79.95 \pm 56.85	7.46	343.20
Carbono [tC/ha]	127.43 \pm 73.49	19.56	436.51

Al comparar los valores de AGB, BGB y carbono es importante considerar las discrepancias que se presentan con otros estudios. Esto implica considerar las variaciones asociadas a los distintos sectores en los que se recopilan los datos, junto con aspectos como el año de captura de dichos datos y las ecuaciones alométricas empleadas para su estimación. Aunque estas variabilidades pueden inducir cambios en los valores, es pertinente realizar comparaciones de resultados para determinar la coherencia de los hallazgos y su consistencia con otras investigaciones. La Tabla 13 presenta una comparativa de los resultados de esta tesis con otras investigaciones en Colombia y el Pacífico Colombiano.

Dentro del ámbito de modelos globales que calculan biomasa y carbono en ecosistemas de manglar en Colombia, se encuentran estudios como el de Hutchison et al. (2014) quien estima AGB y BGB mediante un modelo que incorpora datos bioclimáticos y de campo; Simard et al. (2019) que estima AGB, BGB y carbono utilizando datos de altura de la misión Shuttle Radar Topography de la NASA; y Hu et al. (2020) que genera un mapa de AGB para el año 2004 con resolución de 250 metros, combinando sensores ópticos y de radar con datos de campo. Estos estudios informan valores de AGB de 253.2 t/ha, 129.5 t/ha y 166.95 \pm 66.41 t/ha, respectivamente. Por lo tanto, el valor de AGB del presente estudio se encuentra dentro de estos rangos nacionales. En relación con la estimación de BGB, Hutchinson reporta 108.1 t/ha, cifra superior a la encontrada en esta investigación.

En la Región Pacífica, se destacan investigaciones como la de Selvaraj & Gallego-Pérez (2023) que mapeó AGB mediante el uso de sensores ópticos y de radar, junto con datos de campo, obteniendo un promedio de 181.236 \pm 28.939 t/ha. En paralelo, Bolivar et al. (2018) estimaron el carbono almacenado en la AGB mediante variables bioclimáticas y datos de campo, reportando un promedio de AGB de 105.58 \pm 83.13 t/ha. Dado que el

área de investigación de esta tesis está ubicada en la Región Pacífica, los resultados para la biomasa aérea son cercanos, aunque ligeramente superiores, a los valores informados para toda la región.

En el ámbito de investigaciones locales en el Pacífico Colombiano, se encuentra el estudio de Perea Ardila & Murillo Sandoval (2022) llevado a cabo en el Parque Nacional Natural de Sanquianga. Este estudio, que empleó imágenes ALOS-PALSAR y una ecuación alométrica basada en retrodispersión HV, arrojó estimaciones de AGB y el carbono con valores de 393.15 t/ha y 196.58 t/ha, respectivamente. Estos resultados superan lo hallado en esta tesis, sin embargo, esto se debe a que los bosques de manglar en el Parque Nacional Natural Sanquianga son reconocidos como algunos de los más altos del mundo (Simard et al., 2019), lo que influye de manera directa en la en la biomasa aérea. Otro estudio en el contexto local es el de Monsalve & Ramírez (2015) desarrollado con trabajo de campo en Bahía Málaga, que arrojó resultados para AGB de 143.7 ± 101.7 , para BGB de 13.87 ± 7.38 , y para el carbono almacenado en BGB y AGB de 77.9. Estos resultados, son inferiores a los obtenidos en la presente investigación.

Finalmente, se utilizó el coeficiente de determinación para comparar el rendimiento de este modelo con otros llevados a cabo mediante sensoramiento remoto, ya que esta métrica fue evaluada en todos los casos. El coeficiente de determinación, que mide la bondad de ajuste del modelo, reveló resultados satisfactorios en esta tesis, alcanzando un R^2 de 0.78 y superando el ajuste de modelos globales (Hu et al., 2020; Hutchison et al., 2014; Simard et al., 2019). Para investigaciones de alcance regional, Selvaraj & Gallego-Pérez (2023) demostró un mejor ajuste (0.885) y en estudios locales, Perea-Ardila & Murillo-Sandoval (2022) presentó un R^2 superior (0.89). Es fundamental destacar que este coeficiente de determinación corresponde a la ecuación de Nesha et al. (2020) calculada para estimar AGB en función de la polarización HV del satélite ALOS-PLASAR, la cual fue utilizada en el estudio de Perea-Ardila & Murillo-Sandoval (2022) para determinar AGB. En términos generales, esta investigación exhibió resultados de ajuste destacados, siendo una herramienta útil y precisa en el contexto local.

Tabla 13. Estudios de Biomasa aérea (AGB), Biomasa subterránea (BGB) y Carbono en Colombia y el Pacífico Colombiano con sensoramiento remoto.

Work	Zona	Mean AGB \pm SD (t/ha)	Mean BGB \pm SD (t/ha)	Mean Carbon \pm SD (tC/ha)	R ²
(Hutchison et al., 2014) ^b	Colombia	253.2	108.1	-	0.27
(Simard et al., 2019) ^b	Colombia	129.5	-	370.29*	0.67
(Hu et al., 2020) ^b	Colombia	166.95 \pm 66.41	-	-	0.48
(Selvaraj & Gallego-Pérez, 2023) ^b	Colombian Pacifico Coast	181.236 \pm 28.939	-	-	0.885
(Bolívar et al., 2018) ^b	Colombian Pacifico Coast	105.58 \pm 83.13	-	-	0.47
(Perea Ardila & Murillo Sandoval, 2022) ^b	Pacífico - Sanquianga National Natural Park	393.15	-	196.58***	0.89
(Monsalve & Ramírez, 2015) ^a	Pacífico - Bahía Málaga	143.7 \pm 101.7	13.87 \pm 7.38	77.9**	-
This Study^b	Colombia – Bocagrande Tumaco	192.50\pm102.78	79.95\pm56.85	127.43\pm 73.49**	0.78

*Total Carbon; **Carbon stored in above and below ground biomass, *** Carbon stored in above ground biomass ^aField data-based studies; ^bRemote sensing based studies

Figura 8. Mapa de distribución de la biomasa aérea [t/ha] de los bosques de manglar ubicados en el área de estudio.

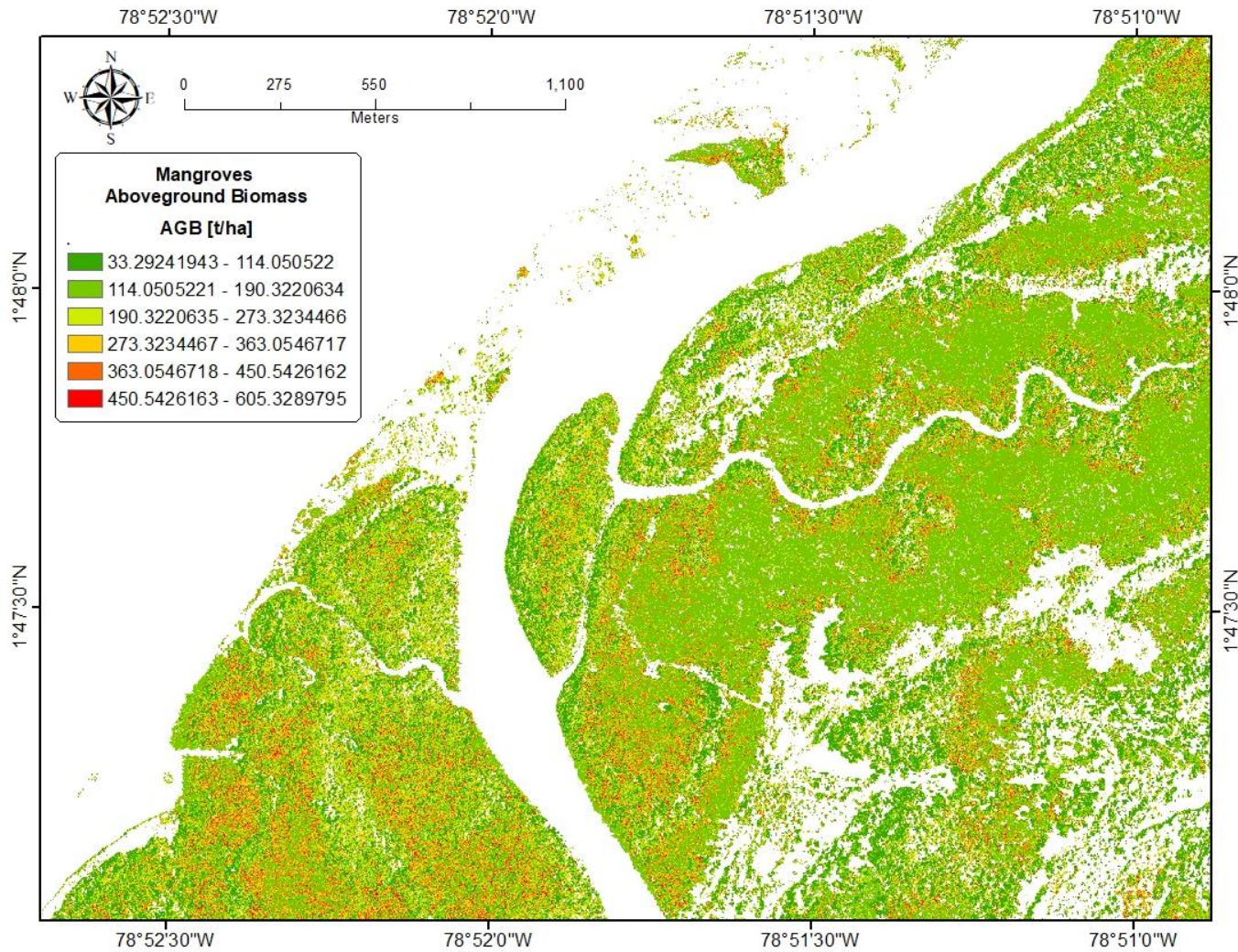


Figura 9. Mapa de distribución de la biomasa subterránea [t/ha] de los bosques de manglar ubicados en el área de estudio.

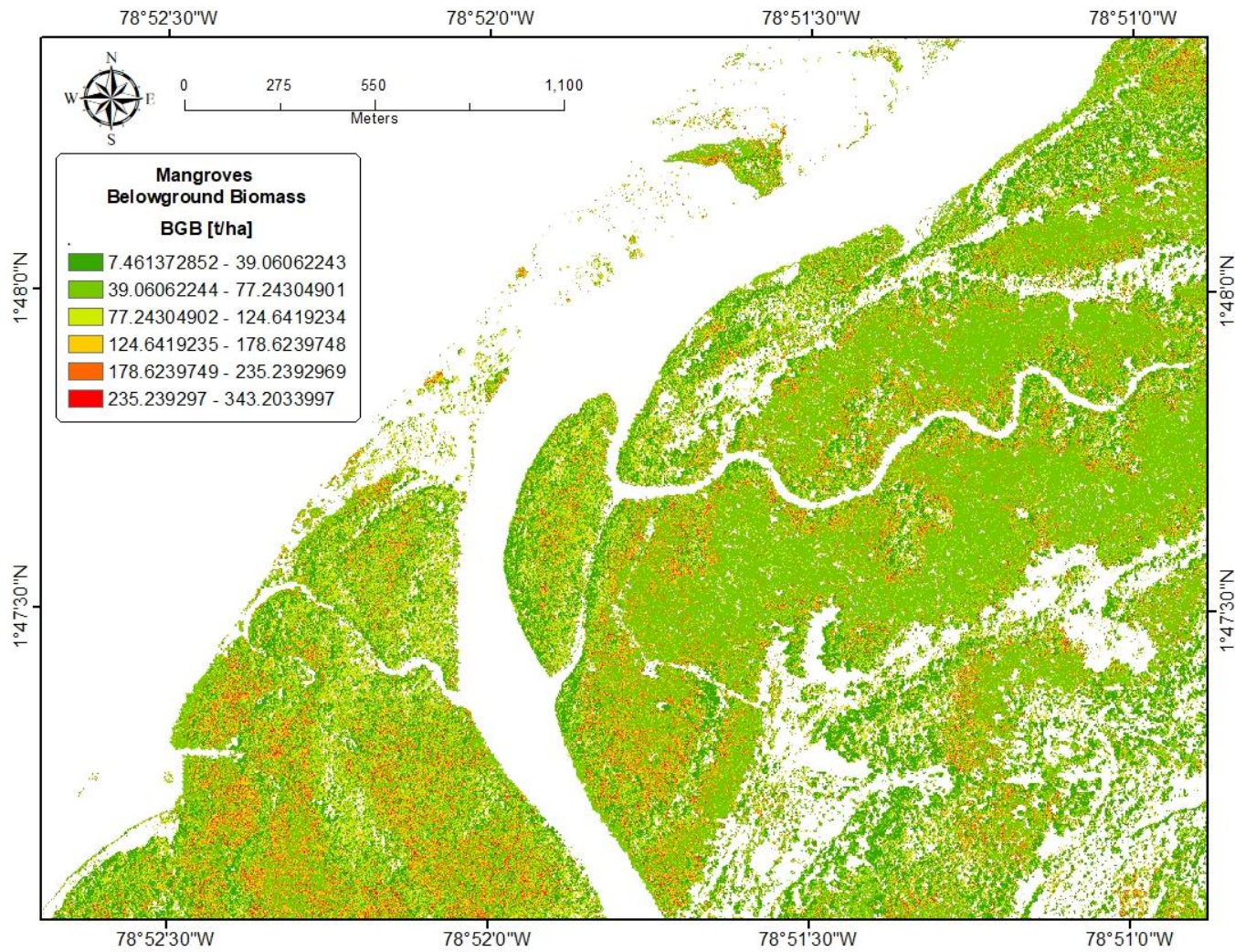
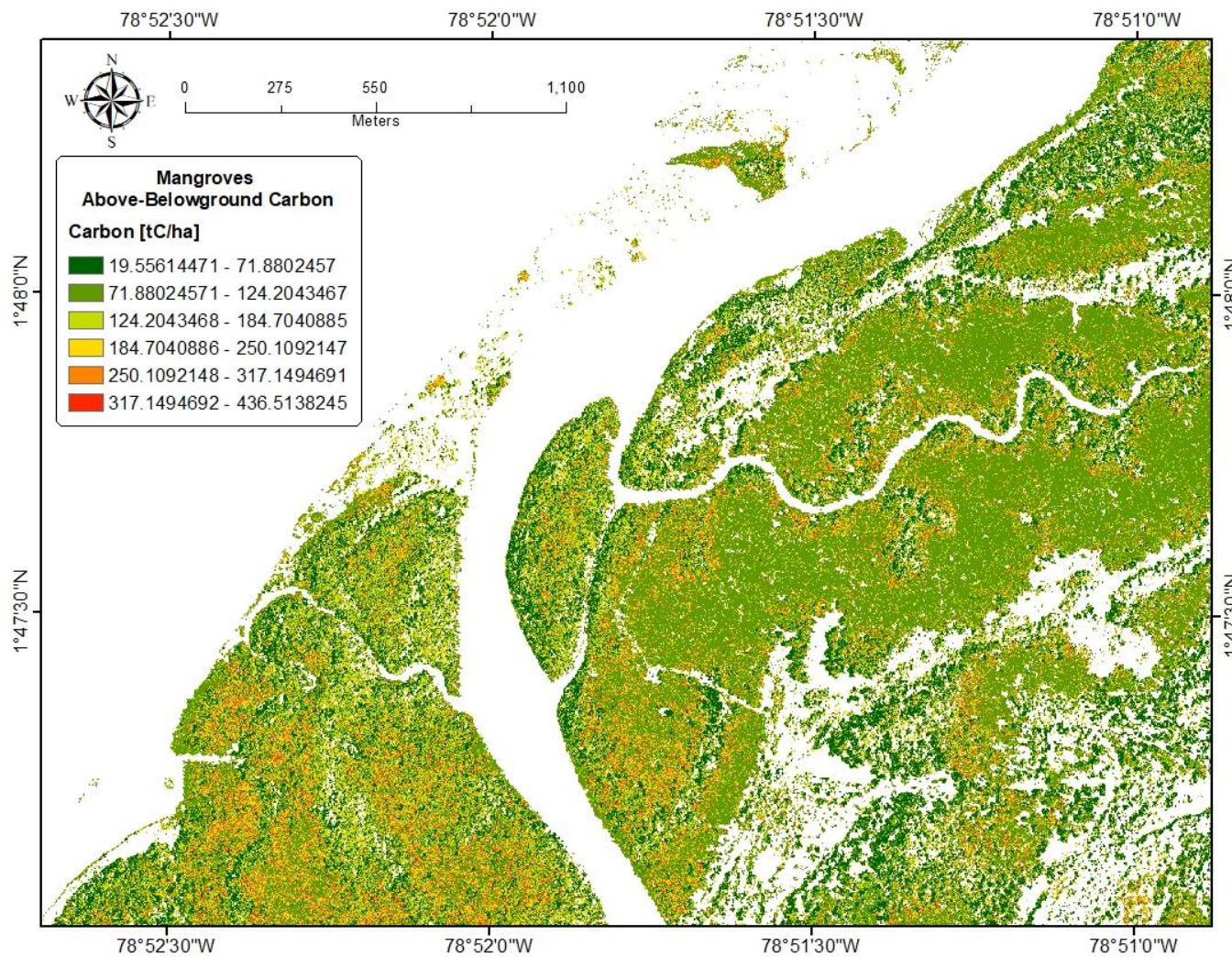


Figura 10. Mapa de distribución del carbono almacenado en la Biomasa aérea y subterránea [tC/ha] de los bosques de manglar ubicados en el área de estudio.



4.5. Posibles incertidumbres asociadas al modelo

En este proyecto, se abordó la incertidumbre del modelo a través del análisis de la desviación estándar de las métricas de evaluación, como el RMSE, R^2 y MAE, ya que permite evaluar cuán consistente es el rendimiento del modelo con diversos conjuntos de datos. Es importante resaltar que los mapas de BGB y de Carbono se derivaron a partir de los resultados del modelo AGB. Por lo tanto, las incertidumbres relacionadas con estos dos últimos mapas incluyen las del modelo AGB y las inherentes a las ecuaciones de conversión utilizadas.

Las métricas RMSE y MAE, relacionadas con el mejor modelo RF para estimar AGB, arrojaron valores de 140.68 ± 98.76 y 84.39 ± 60.63 , respectivamente. Estos valores presentaron una alta dispersión, destacando una desviación estándar que supera más de la mitad de su valor. Esta variabilidad podría atribuirse a múltiples fuentes, que incluyen las incertidumbres inherentes al modelo, las cuales se abordarán posteriormente, así como la limitada cantidad de muestras disponibles para la evaluación y entrenamiento del modelo. En contraste, el coeficiente de determinación presentó una desviación estándar baja de 0.78 ± 0.28 .

De acuerdo con Lu et al. (2016) y Wicaksono et al. (2016) las incertidumbres de los modelos de AGB con sensoramiento remoto se pueden agrupar de acuerdo a los errores asociados con: (1) la recolección de datos de campo, (2) los coeficientes y modelos de conversión como las ecuaciones alométricas, (3) los valores espectrales (4) las diferencias temporales y espaciales entre las mediciones de parcelas de campo y los datos obtenidos por teledetección.

A continuación, se exponen las incertidumbres relacionadas con los factores previamente mencionados. En primer lugar, en lo que respecta a la recolección de datos de campo, es posible que se presenten errores aleatorios hasta del 10% asociados a las mediciones de la altura de los árboles (Simard et al., 2008). A pesar de los esfuerzos realizados para emplear los instrumentos de medición más precisos disponibles y llevar a cabo una labor de campo exhaustiva, se debe reconocer que estas mediciones pueden estar sujetas a

errores inherentes vinculados a las peculiaridades de las circunstancias involucradas en este estudio.

En segundo lugar, las ecuaciones alométricas representan un aspecto crítico en la incertidumbre de los modelos. Por esta razón, resulta de suma importancia la elección de coeficientes y ecuaciones que, en la medida de lo posible, hayan sido calculados en la misma ubicación o en lugares de referencia similares a los del estudio en cuestión (Ahmad et al., 2021). En este trabajo se utilizó la ecuación propuesta por Saldarriaga et al. (2011) la cual fue obtenida de bosques de manglar del Chocó biogeográfico, región dentro de la cual queda ubicada el área de estudio del presente proyecto.

En lo que respecta a las incertidumbres vinculadas a los valores espectrales, diversos factores pueden introducir errores en las mediciones. Estos factores engloban desde plataformas desequilibradas, movimientos del escáner y malas condiciones atmosféricas, hasta la aplicación de métodos inadecuados de interpolación espacial para las correcciones geométricas y radiométricas, así como enfoques incorrectos para el procesamiento y análisis de imágenes (Congalton & Green, 2008). No obstante, es importante subrayar que, tal como se detalla en el capítulo metodológico de este estudio, las imágenes satelitales han sido procesadas de acuerdo con los estándares de calidad del nivel Map Ready. Además, se ha considerado cuidadosamente la influencia de las mareas para minimizar su impacto en la reflectancia de la imagen.

Finalmente, Wang et al. (2011) y Zhang et al. (2013) señalan que los posibles errores espaciales derivados de los desfases normales del GPS en la ubicación de las parcelas de muestreo no han influido de manera significativa en las estimaciones generales de carbono y biomasa forestal. Además, en este estudio, los errores temporales se consideran mínimos, dado que las parcelas se recolectaron un año después de la adquisición de las imágenes, y no se espera que la biomasa aérea del bosque de manglar experimente cambios sustanciales en tan corto plazo.

5. Conclusiones y recomendaciones

5.1. Conclusiones

El proceso metodológico propuesto en esta tesis ha resultado exitoso en la estimación de biomasa viva y carbono almacenado en el área de estudio. La investigación ha destacado que el modelo propuesto, que integra diversos elementos como imágenes de alta resolución, índices espectrales, datos de campo y algoritmos de machine learning, logra ofrecer estimaciones precisas de la biomasa en el área de estudio local, superando a modelos globales en términos de valores de ajuste.

Adicionalmente, se concluye que este trabajo ha alcanzado una clasificación destacada, evidenciada por una precisión global del $87\% \pm 9\%$ y un valor Kappa de 0.83. Estos resultados apuntan a una sólida concordancia entre las clasificaciones en la imagen y las condiciones reales del terreno en la zona de estudio, que abarca una extensión de manglar de 437.32 ha, con promedios de biomasa aérea, subterránea y carbono de 192.50 ± 102.78 t/ha, 79.95 ± 56.85 t/ha, 136.23 ± 79.73 tC/ha, respectivamente.

Es importante destacar que el algoritmo Random Forest ha demostrado un rendimiento superior al Support Vector Machine, resaltando su eficacia en explicar la variabilidad de los datos y sugiriendo su utilidad en este contexto. El uso de imágenes de alta resolución se revela como un factor crucial para la captación de detalles en un área de estudio local. Asimismo, la consideración de diversas métricas de evaluación y sus desviaciones estándar, junto con la visualización de los valores en el mapa, proporciona una comprensión completa de la precisión del modelo y su posible nivel de ruido.

Se observa también que los índices espectrales contribuyen más a la capacidad del modelo para explicar y predecir la biomasa aérea en comparación con las bandas de la imagen sin procesar, destacando su relevancia. Además, se confirma que la filtración de variables explicativas mejora el rendimiento del modelo, consolidando así los hallazgos y contribuciones de esta investigación.

5.2. Recomendaciones

Como recomendación para futuras investigaciones en este campo, se sugiere llevar a cabo un estudio sobre el impacto del uso de imágenes de alta resolución en la mejora de la precisión y exactitud de los modelos de predicción de la biomasa aérea en áreas de estudio regionales y/o nacionales. Dado que las imágenes de alta resolución, como las obtenidas a través de WorldView, no están disponibles al público y requieren una inversión significativa para su adquisición y procesamiento, sería esencial evaluar si esta inversión se traduce en beneficios tangibles en términos de precisión en la estimación de la AGB. Además, se recomienda explorar el efecto de las imágenes LiDAR o SAR en la mejora de la calidad de los modelos, ya que estas tecnologías ofrecen perspectivas prometedoras en la caracterización detallada de la biomasa y podrían complementar o potenciar los enfoques basados en imágenes de alta resolución.

Bibliografía

- Ahmad, A., Gilani, H., & Ahmad, S. R. (2021). Forest aboveground biomass estimation and mapping through high-resolution optical satellite imagery—a literature review. *Forests*, *12*(7), 914. <https://doi.org/10.3390/F12070914/S1>
- Alboukadel, K. (2023). *RDocumentation. Factoextra. fviz_contrib* (1.0.7). https://www.rdocumentation.org/packages/factoextra/versions/1.0.7/topics/fviz_contrib
- Aldrich, R. C., Bailey, W. F., & Heller, R. C. (1959). Large Scale 70 mm. Color Photography Techniques and Equipment and Their Application to a Forest Sampling Problem. *Photogrammetric Engineering*, *25*, 747–754.
- Anand, A., Pandey, P. C., Petropoulos, G. P., Pavlides, A., Srivastava, P. K., Sharma, J. K., & Malhi, R. K. M. (2020). Use of hyperion for mangrove forest carbon stock assessment in bhitarakanika forest reserve: A contribution towards blue carbon initiative. *Remote Sensing*, *12*(4). <https://doi.org/10.3390/rs12040597>
- Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, *4*, 40–79. <https://doi.org/10.1214/09-SS054>
- Awad, M., & Khanna, R. (2015). *Efficient Learning Machines. Theories, Concepts, and Applications for Engineers and System Designers*. Apress, Berkeley, CA.
- Basith, A., & Prastyani, R. (2020). *Evaluating ACOMP , FLAASH AND QUAC on WORLDVIEW-3 for Satellite Derived Bathymetry (SDB) in shallow water*. *46*(3), 151–158.
- Beeston, M., Cuyvers, L., & Vermilye, J. (2020). *Blue Carbon: Mind the Gap. October*, 16. <https://gallifrey.foundation/wp-content/uploads/2020/10/Blue-Carbon-Mind-the-Gap-V2.2.pdf>
- Bindu, G., Rajan, P., Jishnu, E. S., & Ajith Joseph, K. (2020). Carbon stock

- assessment of mangroves using remote sensing and geographic information system. *Egyptian Journal of Remote Sensing and Space Science*, 23(1), 1–9. <https://doi.org/10.1016/j.ejrs.2018.04.006>
- Bishop, C. (2008). *Pattern Recognition and Machine Learning*. Springer Verlag. <http://users.isr.ist.utl.pt/~wurmd/Livros/school/Bishop - Pattern Recognition And Machine Learning - Springer 2006.pdf>
- Bolar, K. (2019). *STAT V.0.1.0* (0.1.0). <https://cran.r-project.org/web/packages/STAT/index.html>
- Bolivar, J. M., Gutierrez-Velez, V. H., & Sierra, C. A. (2018). Carbon stocks in aboveground biomass for Colombian mangroves with associated uncertainties. *Regional Studies in Marine Science*, 18, 145–155. <https://doi.org/10.1016/j.rsma.2017.12.011>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324/METRICS>
- Bunting, P. (2020). *Class Accuracy* (V1.0.0). https://github.com/remotesensinginfo/classaccuracy/releases/tag/qgis3_v1.0.0
- Bunting, P., Rosenqvist, A., Hilarides, L., Lucas, R. M., Thomas, N., Tadono, T., Worthington, T. A., Spalding, M., Murray, N. J., & Rebelo, L. M. (2022). Global Mangrove Extent Change 1996–2020: Global Mangrove Watch Version 3.0. *Remote Sensing*, 14(15). <https://doi.org/10.3390/rs14153657>
- Cantera, J. R., Thomassin, B. A., & Arnaud, P. M. (1999). Faunal zonation and assemblages in the Pacific Colombian mangroves. In *Hydrobiologia* (Vol. 413).
- Castellanos-Galindo, G. A., Kluger, L. C., Camargo, M. A., Cantera, J., Mancera Pineda, J. E., Blanco-Libreros, J. F., & Wolff, M. (2021). Mangrove research in Colombia: Temporal trends, geographical coverage and research gaps. *Estuarine, Coastal and Shelf Science*, 248(October 2019). <https://doi.org/10.1016/j.ecss.2020.106799>
- Chave, J., Andalo, C., Brown, S., Cairns, M. A., Chambers, J. Q., Eamus, D., Fölster, H., Fromard, F., Higuchi, N., Kira, T., Lescure, J. P., Nelson, B. W., Ogawa, H., Puig, H., Riéra, B., & Yamakura, T. (2005). Tree allometry and

- improved estimation of carbon stocks and balance in tropical forests. *Oecologia*, 145(1), 87–99. <https://doi.org/10.1007/S00442-005-0100-X>
- Congalton, R. G., & Green, K. (2008). Assessing the accuracy of remotely sensed data: Principles and practices, second edition. In *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices, Second Edition*.
- Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random Forest for classification in ecology. *Ecology*, 88(11), 2783–2792. <https://doi.org/10.1890/07-0539.1>
- Darmawan, S., Takeuchi, W., Vetruta, Y., Wikantika, K., & Sari, D. K. (2015). Impact of topography and tidal height on ALOS palsar polarimetric measurements to estimate aboveground biomass of mangrove forest in Indonesia. *Journal of Sensors*, 2015. <https://doi.org/10.1155/2015/641798>
- Digital Globe. (2010). *DigitalGlobe Core Imagery Products Guide*. 8–11. <https://www.satimagingcorp.com/satellite-sensors/worldview-3>
- Dimar. (2020). *Compilación Oceanográfica de la Cuenca Pacífica Colombiana II* (Vol. 9). Editorial Dimar.
- Donato, D. C., Kauffman, J. B., Murdiyarso, D., Kurnianto, S., Stidham, M., & Kanninen, M. (2011). Mangroves among the most carbon-rich forests in the tropics. *Nature Geoscience*, 4(5), 293–297. <https://doi.org/10.1038/ngeo1123>
- Escobar, H. A. T. (2010). *Documento síntesis: Caracterización, Diagnóstico y Zonificación de los Manglares en el Departamento de Nariño*.
- FAO. (2007). The world's mangroves 1980-2005. In *FAO Forestry Paper* (Vol. 153).
- FAO, Yigini, Y., Olmedo, G. F., Reiter, S., Baritz, R., Viatkin, K., & Vargas, R. R. (2018). *Soil Organic Carbon Mapping Cookbook*. 2nd Edition, Rome. <https://fao-gsp.github.io/SOC-Mapping-Cookbook/mappingMethods.html#rf>
- Forkuor, G., Benewinde Zoungrana, J. B., Dimobe, K., Ouattara, B., Vadrevu, K. P., & Tondoh, J. E. (2020). Above-ground biomass mapping in West African dryland forest using Sentinel-1 and 2 datasets - A case study. *Remote Sensing of Environment*, 236(November 2019), 111496. <https://doi.org/10.1016/j.rse.2019.111496>

- GAFMEX. (2023). *Clinómetro Electrónico Ecii D Haglöf*.
<https://gafmex.com/clinometro-electronico-ecii-d-haglof/>
- Gao, B.-C. (1996). NDWI - A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water From Space. *REMOTE SENS. ENVIRON*, 72(12), 257–266.
- Gao, Y., Yu, G., Yang, T., Jia, Y., He, N., & Zhuang, J. (2016). New insight into global blue carbon estimation under human activity in land-sea interaction area: A case study of China. *Earth-Science Reviews*, 159, 36–46.
<https://doi.org/10.1016/J.EARSCIREV.2016.05.003>
- Giri, C., Ochieng, E., Tieszen, L. L., Zhu, Z., Singh, A., Loveland, T., Masek, J., & Duke, N. (2011). Status and distribution of mangrove forests of the world using earth observation satellite data. *Global Ecology and Biogeography*, 20(1), 154–159. <https://doi.org/10.1111/j.1466-8238.2010.00584.x>
- Gonzalez, R. C., & Woods, R. E. (2008). *Digital Image Processing Third Edition Library of Congress Cataloging-in-Publication Data on File*.
- Google, & Maxar Technologies. (2021). *Maps Data*. <https://earth.google.com/web/>
- Gross, J., Flores, E. E., & Schwendenmann, L. (2014). Stand structure and aboveground biomass of a *Pelliciera rhizophorae* mangrove forest, gulf of Montijo Ramsar Site, Pacific Coast, Panama. *Wetlands*, 34(1), 55–65.
<https://doi.org/10.1007/s13157-013-0482-1>
- Gupta, K., Mukhopadhyay, A., Giri, S., Chanda, A., Datta Majumdar, S., Samanta, S., Mitra, D., Samal, R. N., Pattnaik, A. K., & Hazra, S. (2018). An index for discrimination of mangroves from non-mangroves using LANDSAT 8 OLI imagery. *MethodsX*, 5, 1129–1139. <https://doi.org/10.1016/j.mex.2018.09.011>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. <https://doi.org/10.1007/978-0-387-84858-7>
- Hickey, S. M., Callow, N. J., Phinn, S., Lovelock, C. E., & Duarte, C. M. (2018). Spatial complexities in aboveground carbon stocks of a semi-arid mangrove community: A remote sensing height-biomass-carbon approach. *Estuarine, Coastal and Shelf Science*, 200, 194–201.
<https://doi.org/10.1016/j.ecss.2017.11.004>

- Howard, J., Hoyt, S., Isensee, K., Pidgeon, E., & Telszewski, M. (2014). Coastal Blue Carbon: Methods for Assessing Carbon Stocks and Emissions Factors in Mangroves, Tidal Salt Marshes, and Seagrass Meadows. *Conservation International, Intergovernmental Oceanographic Commission of UNESCO, International Union for Conservation of Nature. Arlington, Virginia, USA.*, 1–180. www.ioc.unesco.org
- Hu, T., Zhang, Y. Y., Su, Y., Zheng, Y., Lin, G., & Guo, Q. (2020). Mapping the Global Mangrove Forest Aboveground Biomass Using Multisource Remote Sensing Data. *Remote Sensing 2020, Vol. 12, Page 1690, 12(10)*, 1690. <https://doi.org/10.3390/RS12101690>
- Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment, 25(3)*, 295–309. [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X)
- Hutchison, J., Manica, A., Swetnam, R., Balmford, A., & Spalding, M. (2014). Predicting global patterns in mangrove forest biomass. *Conservation Letters, 7(3)*, 233–240. <https://doi.org/10.1111/conl.12060>
- Ideam. (2010). Leyenda nacional de coberturas de la tierra. Metodología CORINE Land Cover adaptada para Colombia, escala 1:100.000. In *Area: Vol. TH-62-04-1* (Issue 257).
- INVEMAR, UNIVALLE, & INCIVA. (2006). *Valoración de la biodiversidad marina y costera de Bahía Málaga (Valle del Cauca), como uno de los instrumentos necesarios para que sea considerada un área protegida.* http://www.invemar.org.co/redcostera1/invemar/docs/9860IF_BIOMALAGA2007.pdf
- Jachowski, N. R. A., Quak, M. S. Y., Friess, D. A., Duangnamon, D., Webb, E. L., & Ziegler, A. D. (2013). Mangrove biomass estimation in Southwest Thailand using machine learning. *Applied Geography, 45*, 311–321. <https://doi.org/10.1016/J.APGEOG.2013.09.024>
- Jiang, Y., Zhang, L., Yan, M., Qi, J., Fu, T., Fan, S., & Chen, B. (2021). High-Resolution Mangrove Forests Classification with Machine Learning Using Worldview and UAV Hyperspectral Data. *Remote Sensing 2021, Vol. 13, Page*

- 1529, 13(8), 1529. <https://doi.org/10.3390/RS13081529>
- Jiang, Z., Huete, A. R., Didan, K., & Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment*, 112(10), 3833–3845. <https://doi.org/10.1016/J.RSE.2008.06.006>
- Kauffman, J. B., & Cole, T. G. (2010). Micronesian mangrove forest structure and tree responses to a severe typhoon. *Wetlands*, 30(6), 1077–1084. <https://doi.org/10.1007/S13157-010-0114-Y/METRICS>
- Kauffman, J. B., & Donato, D. C. (2012). Protocols for the measurement, monitoring and reporting of structure, biomass and carbon stocks in mangrove forests. Working Paper 86. CIFOR, Bogor, Indonesia. *Working Paper 86, CIFOR, January*, 50.
- Kershaw, J. A., Ducey, M. J., Beers, T. W., & Husch, B. (2016). Sampling Designs in Forest Inventories. In *Forest Mensuration* (pp. 305–360). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118902028.ch10>
- Khorram, S., Nelson, S. A. C., Koch, F. H., & van der Wiele, C. F. (2012). *Remote Sensing*. <https://doi.org/10.1007/978-1-4614-3103-9>
- Kogut, P. (2021). *Teledetección: Tipos Y Aplicaciones De Los Sensores Remotos*. <https://eos.com/es/blog/teledeteccion/>
- Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., & Hunt, T. (2023). *Caret V.6.0-94* (6.0-94). <https://cran.r-project.org/web/packages/caret/index.html>
- Lu, D. (2006). The potential and challenge of remote sensing-based biomass estimation. *International Journal of Remote Sensing*, 27(7), 1297–1328. <https://doi.org/10.1080/01431160500486732>
- Lu, D., Chen, Q., Wang, G., Liu, L., Li, G., & Moran, E. (2016). A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems. In *International Journal of Digital Earth* (Vol. 9, Issue 1, pp. 63–105). <https://doi.org/10.1080/17538947.2014.990526>
- Luo, Y. M., Huang, D. T., Liu, P. Z., & Feng, H. M. (2016). An novel random forests and its application to the classification of mangroves remote sensing image.

- Multimedia Tools and Applications*, 75(16), 9707–9722.
<https://doi.org/10.1007/s11042-015-2906-9>
- MAXAR Technologies. (2020, July). *Multispectral reference guide*.
<https://resources.maxar.com/white-papers/multispectral-reference-guide>
- McCarthy, M. J., & Halls, J. N. (2014). Habitat mapping and change assessment of coastal environments: An examination of worldview-2, quickbird, and ikonos satellite imagery and airborne lidar for mapping barrier island habitats. *ISPRS International Journal of Geo-Information*, 3(1), 297–325.
<https://doi.org/10.3390/ijgi3010297>
- McLeod, E., Chmura, G. L., Bouillon, S., Salm, R., Björk, M., Duarte, C. M., Lovelock, C. E., Schlesinger, W. H., & Silliman, B. R. (2011). A blueprint for blue carbon: Toward an improved understanding of the role of vegetated coastal habitats in sequestering CO₂. *Frontiers in Ecology and the Environment*, 9(10), 552–560. <https://doi.org/10.1890/110004>
- Mejía-Rentería, J. C., Castellanos-Galindo, G. A., Cantera-Kintz, J. R., & Hamilton, S. E. (2018). A comparison of Colombian Pacific mangrove extent estimations: Implications for the conservation of a unique Neotropical tidal forest. *Estuarine, Coastal and Shelf Science*, 212(July), 233–240.
<https://doi.org/10.1016/j.ecss.2018.07.020>
- Microsoft Bing, & Maxar Technologies. (2022). *Bing Maps*.
<https://www.bing.com/maps?cc=es>
- Millennium Ecosystem Assessment. (2005). *Ecosystems and Human Well-being : Current State and Trends, Volume 1* (R. Hassan, R. Scholes, & N. Ash (eds.)). Island Press.
<http://www.millenniumassessment.org/en/products.global.condition.aspx>
- Monsalve, A., & Ramírez, G. (2015). *Caracterización de la estructura y contenido de carbono de los bosques de manglar en el área de jurisdicción del Consejo comunitario La Plata, Bahía Málaga, Valle del Cauca* (Vol. 57, Issue 4).
- Mukhtar, E., Raynaldo, A., & Novarino, W. (2021). Carbon stock mapping using mangrove discrimination indices in mandeh bay, west sumatra. *AACL Bioflux*,

- 14(1), 430–440.
- Mutanga, O., Adam, E., & Cho, M. A. (2012). High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. *International Journal of Applied Earth Observation and Geoinformation*, 18, 399–406. <https://doi.org/10.1016/j.jag.2012.03.012>
- Naciones Unidas, & Asamblea General. (2015). *Transformar nuestro mundo: la Agenda 2030 para el Desarrollo Sostenible*.
- Navarrete-Ramírez, S. M., & Rodríguez-Rincón, A. M. (2014). *Protocolo Indicador Condición Tendencia Bosques de Manglar. Indicadores de monitoreo biológico del Subsistema de Áreas Marinas Protegidas (SAMP): Vol. No. 67* (Invemar, GEF, & PNUD (eds.)). Serie de Publicaciones Generales del Invemar.
- Nellemann, C., Corcoran, E., Duarte, C. M., Valdés, L., De Young, C., Fonseca, L., & Grimsditch, G. (2009). Blue carbon: A Rapid Response Assessment. In *Environment*. http://www.grida.no/files/publications/blue-carbon/BlueCarbon_screen.pdf
- Nesha, M. K., Hussin, Y. A., van Leeuwen, L. M., & Sulistioadi, Y. B. (2020). Modeling and mapping aboveground biomass of the restored mangroves using ALOS-2 PALSAR-2 in East Kalimantan, Indonesia. *International Journal of Applied Earth Observation and Geoinformation*, 91, 102158. <https://doi.org/10.1016/J.JAG.2020.102158>
- Nguyen, H. H., Nghia, N. H., Nguyen, H. T. T., Le, A. T., Tran, L. T. N., Duong, L. V. K., Bohm, S., & Furniss, M. J. (2020). Classification Methods for Mapping Mangrove Extents and Drivers of Change in Thanh Hoa Province, Vietnam during 2005-2018. *Forest and Society*, 4(1), 225–242. <https://doi.org/10.24259/FS.V4I1.9295>
- Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., & Wulder, M. A. (2014). Good practices for estimating area and assessing accuracy of land change. In *Remote Sensing of Environment* (Vol. 148, pp. 42–57). Elsevier Inc. <https://doi.org/10.1016/j.rse.2014.02.015>
- Palacios- penaranda, M. L. (2017). *Evaluación del almacenamiento de carbono*

- como servicio ecosistémico en bosques de manglar de la costa pacífica colombiana.* [UNIVERSIDAD DEL VALLE].
<https://bibliotecadigital.univalle.edu.co/handle/10893/14485>
- Palacios-Peñaranda, M. L., Cantera-Kintz, J. R., & Peña-Salamanca, E. J. (2019). Carbon stocks in mangrove forests of the Colombian Pacific. *Estuarine, Coastal and Shelf Science*, 227(April 2018), 106299. <https://doi.org/10.1016/j.ecss.2019.106299>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., & Thirion, B. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. <http://scikit-learn.sourceforge.net>.
- Perea-Ardila, M. A., Oviedo-Barrero, F., & Leal-Villamil, J. (2019). Cartografía de bosques de manglar mediante imágenes de sensores remotos: estudio de caso Buenaventura, Colombia. *Revista de Teledetección*, 53, 73.
- Perea Ardila, M. A., Leal Villamil, J., & Oviedo Barrero, F. (2021). Caracterización espectral y monitoreo de bosque de manglar con teledetección en el litoral Pacífico Colombiano: Bajo Baudó, Chocó. *LA GRANJA: REVISTA DE CIENCIAS DE LA VIDA*, 34(C), 27–44. <https://doi.org/http://doi.org/10.17163/lgr.n34.2021.02> Edición
- Perea Ardila, M. A., & Murillo Sandoval, P. J. (2022). La ganancia de manglar y sus implicaciones en el reservorio de Carbono del Parque Nacional Natural Sanquianga en Colombia. *Ecosistemas*, 31(3), 2386. <https://doi.org/10.7818/ecos.2386>
- Perumal, K., & Bhaskaran, R. (2010). Supervised Classification Performance of Multispectral Images. *Journal of Computing*, 2(2), 124–129. <http://arxiv.org/abs/1002.4046>
- Pham, T. D., Le, N. N., Ha, N. T., Nguyen, L. V., Xia, J., Yokoya, N., To, T. T., Trinh, H. X., Kieu, L. Q., & Takeuchi, W. (2020). Estimating mangrove above-ground biomass using extreme gradient boosting decision trees algorithm with fused sentinel-2 and ALOS-2 PALSAR-2 data in can Gio biosphere reserve, Vietnam. In *Remote Sensing* (Vol. 12, Issue 5). <https://doi.org/10.3390/rs12050777>

- Pham, T. D., Xia, J., Thang Ha, N., Tien Bui, D., Nhu Le, N., & Tekeuchi, W. (2019). A review of remote sensing approaches for monitoring blue carbon ecosystems: Mangroves, sea grasses and salt marshes during 2010–2018. *Sensors (Switzerland)*, *19*(8). <https://doi.org/10.3390/s19081933>
- Pham, T. D., Yokoya, N., Bui, D. T., Yoshino, K., & Friess, D. A. (2019). Remote sensing approaches for monitoring mangrove species, structure, and biomass: Opportunities and challenges. *Remote Sensing*, *11*(3), 1–24. <https://doi.org/10.3390/rs11030230>
- QGIS. (2023, February 25). *QGIS 3.28*. <https://www.qgis.org/en/site/>
- Qiu, P., Wang, D., Zou, X., Yang, X., Xie, G., Xu, S., & Zhong, Z. (2019). Finer Resolution Estimation and Mapping of Mangrove Biomass Using UAV LiDAR and WorldView-2 Data. *Forest*, *10*(817). <https://doi.org/10.3390/f10100871>
- Quang, N. H., Quinn, C. H., Carrie, R., Stringer, L. C., Hue, L. T. Van, Hackney, C. R., & Tan, D. Van. (2022). Comparisons of regression and machine learning methods for estimating mangrove above-ground biomass using multiple remote sensing data in the red River Estuaries of Vietnam. *Remote Sensing Applications: Society and Environment*, *26*(November 2021), 100725. <https://doi.org/10.1016/j.rsase.2022.100725>
- Rahman, M. M., Lagomasino, D., Lee, S. K., Fatoyinbo, T., Ahmed, I., & Kanzaki, M. (2019). Improved assessment of mangrove forests in Sundarbans East Wildlife Sanctuary using WorldView 2 and TanDEM-X high resolution imagery. *Remote Sensing in Ecology and Conservation*, *5*(2), 136–149. <https://doi.org/10.1002/rse2.105>
- Rondeaux, G., Steven, M., & Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment*, *55*(2), 95–107. [https://doi.org/10.1016/0034-4257\(95\)00186-7](https://doi.org/10.1016/0034-4257(95)00186-7)
- Rovai, A. S., Twilley, R. R., Castañeda-Moya, E., Midway, S. R., Friess, D. A., Trettin, C. C., Bukoski, J. J., Stovall, A. E. L., Pagliosa, P. R., Fonseca, A. L., Mackenzie, R. A., Aslan, A., Sasmito, S. D., Sillanpää, M., Cole, T. G., Purbopuspito, J., Warren, M. W., Murdiyarso, D., Mofu, W., ... Riul, P. (2021). Macroecological patterns of forest structure and allometric scaling in mangrove

- forests. *Global Ecology and Biogeography*, 30(5), 1000–1013. <https://doi.org/10.1111/GEB.13268>
- Rwanga, S. S., Ndambuki, J. M., Rwanga, S. S., & Ndambuki, J. M. (2017). Accuracy Assessment of Land Use/Land Cover Classification Using Remote Sensing and GIS. *International Journal of Geosciences*, 8(4), 611–622. <https://doi.org/10.4236/IJG.2017.84033>
- Saeys, Y., Inza, I., & Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics (Oxford, England)*, 23(19), 2507–2517. <https://doi.org/10.1093/BIOINFORMATICS/BTM344>
- Saldarriaga, J. G., Duque, A. J., & Álvarez, E. (2011). *Modelos para la estimación de la biomasa y el carbono en diferentes tipos de bosque del choco biogeográfico, Colombia: manglar, guandal y bosques de colina localizados en los Consejos Comunitarios de Bajo Mira y Concosta a partir de los datos colectados e*. Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM), Ministerio de Medio Ambiente, Vivienda y Desarrollo Territorial (MAVDT), Fundación Natura, Fundación Gordon y Betty Moore – USAID.
- Santos, T., & Freire, S. (2015). Testing the Contribution of WorldView-2 Improved Spectral Resolution for Extracting Vegetation Cover in Urban Environments. *Canadian Journal of Remote Sensing*, 41(6), 505–514. <https://doi.org/10.1080/07038992.2015.1110011>
- SciPy Community. (2024). *SciPy documentation. SciPy v1.12.0 Manual*. <https://docs.scipy.org/doc/scipy/index.html>
- Selvaraj, J. J., & Gallego-Pérez, B. E. (2023). Estimating mangrove aboveground biomass in the Colombian Pacific coast: A multisensor and machine learning approach. *Heliyon*, 9, 2405–8440. <https://doi.org/10.1016/j.heliyon.2023.e20745>
- Simard, M., Fatoyinbo, L., Smetanka, C., Rivera-Monroy, V. H., Castañeda-Moya, E., Thomas, N., & Van der Stocken, T. (2019). Mangrove canopy height globally related to precipitation, temperature and cyclone frequency. *Nature Geoscience*, 12(1), 40–45. <https://doi.org/10.1038/s41561-018-0279-1>

- Simard, M., Rivera-Monroy, V. H., Mancera-Pineda, J. E., Castañeda-Moya, E., & Twilley, R. R. (2008). A systematic method for 3D mapping of mangrove forests based on Shuttle Radar Topography Mission elevation data, ICESat/GLAS waveforms and field data: Application to Ciénaga Grande de Santa Marta, Colombia. *Remote Sensing of Environment*, 112(5), 2131–2144. <https://doi.org/10.1016/j.rse.2007.10.012>
- Sinha, S., Jeganathan, C., Sharma, L. K., & Nathawat, M. S. (2015). A review of radar remote sensing for biomass estimation. *International Journal of Environmental Science and Technology* 2015 12:5, 12(5), 1779–1792. <https://doi.org/10.1007/S13762-015-0750-0>
- Spalding, M. D., & Leal, M. (2021). *The State of the World's Mangroves 2021*. Global Mangrove Alliance.
- Takeuchi, W., Darmawan, S., Nakazono, E., Vetrira, Y., Winarso, G., Dien, V. T., Oo, K. S., Wikantika, K., Sari, D. K., Darmawan, S., Nakazono, E., Vetrira, Y., Winarso, G., Dien, V. T., Oo, K. S., Wikantika, K., & Sari, D. K. (2016). Carbon stock calculation and forest change assessment toward REDD+ activities for the mangrove forests in Southeast Asia. *Institute of Industrial Science The University of Tokyo, Japan*, 2. <http://www.asienreisender.de/mangroveforest.html>
- Thuy, H. L. T., Tan, M. T., Van, T. T. T., Bien, L. B., Ha, N. M., & Nhung, N. T. (2020). Using sentinel image data and plot survey for the assessment of biomass and carbon stock in coastal forests of Thai Binh Province, Vietnam. *Applied Ecology and Environmental Research*, 18(6), 7499–7514. https://doi.org/10.15666/aeer/1806_74997514
- Toosi, N. B., Soffianian, A. R., Fakheran, S., Pourmanafi, S., Ginzler, C., & Waser, L. T. (2019). Comparing different classification algorithms for monitoring mangrove cover changes in southern Iran. *Global Ecology and Conservation*, 19, e00662. <https://doi.org/10.1016/J.GECCO.2019.E00662>
- Twilley, R. R., Chen, R. H., & Hargis, T. (1992). Carbon sinks in mangroves and their applications to carbon budget of tropical coastal ecosystems. *Water, Air, and Soil Pollution* 64:, 64, 265–288. <https://doi.org/10.1021/es983796z>

- Vu, T. D., Takeuchi, W., & Van, N. A. (2014). Carbon Stock Calculating and Forest Change Assessment Toward REDD+ Activities for The Mangrove Forest in Vietnam. *Transactions of the Japan Society for Aeronautical and Space Sciences, Aerospace Technology Japan*, 12(ists29), Pn_23-Pn_31. https://doi.org/10.2322/tastj.12.pn_23
- Wang, G., Zhang, M., Gertner, G. Z., Oyana, T., McRoberts, R. E., & Ge, H. (2011). Uncertainties of mapping aboveground forest carbon due to plot locations using national forest inventory plot and remotely sensed data. *Scandinavian Journal of Forest Research*, 26(4), 360–373. <https://doi.org/10.1080/02827581.2011.564204>
- Wang, L., Jia, M., Yin, D., & Tian, J. (2019). A review of remote sensing for mangrove forests: 1956–2018. *Remote Sensing of Environment*, 231(May). <https://doi.org/10.1016/j.rse.2019.111223>
- Wicaksono, P., Danoedoro, P., Hartono, H., Nehren, U., & Ribbe, L. (2011). Preliminary work of mangrove ecosystem carbon stock mapping in small island using remote sensing: above and below ground carbon stock mapping on medium resolution satellite image. *Remote Sensing for Agriculture, Ecosystems, and Hydrology XIII*, 8174, 81741B. <https://doi.org/10.1117/12.897926>
- Wicaksono, P., Danoedoro, P., Hartono, & Nehren, U. (2016). Mangrove biomass carbon stock mapping of the Karimunjawa Islands using multispectral remote sensing. *International Journal of Remote Sensing*, 37(1), 26–52. <https://doi.org/10.1080/01431161.2015.1117679>
- Yepes, A., Duque, Á., Navarrete, D., & Philips, J. (2011). *Protocolo para la estimación nacional y subnacional de biomasa de biomasa-carbono en Colombia*. Instituto de Hidrología, Meteorología y Estudios Ambientales-IDEAM. [https://www.researchgate.net/publication/269107473_What_is_governance/link/548173090cf22525dcb61443/download%0Ahttp://www.econ.upf.edu/~reynal/Civil wars_12December2010.pdf%0Ahttps://think-](https://www.researchgate.net/publication/269107473_What_is_governance/link/548173090cf22525dcb61443/download%0Ahttp://www.econ.upf.edu/~reynal/Civil_wars_12December2010.pdf%0Ahttps://think-)

- asia.org/handle/11540/8282%0Ahttps://www.jstor.org/stable/41857625
- Yepes, A., Zapata, M., Bolivar, J., Monsalve, A., Espinosa, S. M., Sierra-Correa, P. C., & Sierra, A. (2016). Ecuaciones alométricas de biomasa aérea para la estimación de los contenidos de carbono en manglares del Caribe Colombiano. In *Rev. Biol. Trop. (Int. J. Trop. Biol. ISSN (Vol. 64, Issue 2)*.
- Zanne, A. E., Lopez-Gonzalez, G., Coomes, D. A., Ilic, J., Jansen, S., Lewis, S. L., Miller, R. B., Swenson, N. G., Wiemann, M. C., & Chave, J. (2009). *Global wood density database*. Dryad. <http://hdl.handle.net/10255/dryad.235>
- Zhang, M., Lin, H., Zeng, S., Li, J., Shi, J., & Wang, G. (2013). Impacts of plot location errors on accuracy of mapping and scaling up aboveground forest carbon using sample plot and landsat tm data. *IEEE Geoscience and Remote Sensing Letters*, 10(6), 1483–1487. <https://doi.org/10.1109/LGRS.2013.2260719>
- Zhu, J. J., & Yan, B. (2022). Blue carbon sink function and carbon neutrality potential of mangroves. *Science of the Total Environment*, 822, 153438. <https://doi.org/10.1016/j.scitotenv.2022.153438>
- Zhu, Y., Liu, K., Liu, L., Myint, S. W., Wang, S., Cao, J., & Wu, Z. (2020). Estimating and Mapping Mangrove Biomass Dynamic Change Using WorldView-2 Images and Digital Surface Models. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 2123–2134. <https://doi.org/10.1109/JSTARS.2020.2989500>
- Zhu, Y., Liu, K., Liu, L., Myint, S. W., Wang, S., Liu, H., & He, Z. (2017). Exploring the potential of world view-2 red-edge band-based vegetation indices for estimation of mangrove leaf area index with machine learning algorithms. *Remote Sensing*, 9(10). <https://doi.org/10.3390/rs9101060>
- Zhu, Y., Liu, K., Liu, L., Wang, S., & Liu, H. (2015). Retrieval of Mangrove Aboveground Biomass at the Individual Species Level with WorldView-2 Images. *Remote Sensing 2015, Vol. 7, Pages 12192-12214*, 7(9), 12192–12214. <https://doi.org/10.3390/RS70912192>