

Desarrollo De Una Metodología Para La Captura Y Procesamiento De Información Sobre
El Estado De Los Bosques Utilizando UAVs

Arnold Guarín Ramirez

Trabajo de Grado para optar al título de
Magíster en Ingeniería de Telecomunicaciones

Director:

Homero Ortega Boada

Ph.D. Engineering Sciences

Codirector:

Hans Yecid Garcia Arenas

Ph.D.(C) Ingeniería Electrónica

Universidad Industrial de Santander
Facultad De Ingenierías Físicomecánicas
Escuela De Ingenierías Eléctrica, Electrónica
y de Telecomunicaciones

Bucaramanga

2021

Tabla de Contenido

	Pág.
Introducción	12
1. Objetivos	18
1.1 Objetivo General	18
1.2 Objetivos Específicos.....	18
2. Métodos De Estimación De Bancos De Carbono	19
2.1 Conceptos Clave	19
2.1.1 Imágenes Espectrales	19
2.1.2 Índices De Vegetación	21
2.1.3 Fotogrametría Con Uav.....	22
2.2 Métodos De Campo	22
2.3 Métodos Remotos	24
2.4 Metodología De Estimación De Carbono Propuesta	28
2.4.1 Fase De Entrenamiento	29
2.4.2 Fase De Entrenamiento	31
3. Proceso De Planificación, Recolección Y Preprocesamiento De La Información	32
3.1 Zona De Estudios.....	32
3.2 Expedición En Campo Para La Estimación De Biomasa Superficial	33
3.3 Equipo De Captura De Imágenes.....	37
3.4 Captura De Los Ortomosaicos De La Zona De Estudio	39
3.5 Preparación De Los Datos Capturados Con El Uav	42

4. Construcción Del Modelo Matemático Para La Estimación De Carbono	45
4.1 Modelo Propuesto Para La Estimación De Carbono A Partir De Imágenes R-G-B-Nir	45
4.1.1 Preparación De Los Datos De La Expedición De Campo	46
4.1.2 Metodología Para Relacionar El Cubo De Datos Y La Información De Campo	48
4.2 Organización De Los Datos	50
5. Entrenamiento Y Validación Del Modelo Matemático Para La Estimación De Carbono	53
5.1 Medición Del Error En Los Modelos.....	53
5.2 Red Neuronal Utilizada En El Proyecto	55
5.3 Metodologías De Estimación De Carbono Implementadas	58
5.4 Resultados Del Entrenamiento De Las Metodologías Implementadas	59
5.4.1 Entrenamiento Con La Red Neuronal Propuesta Usando El Cubo De Datos Después El Análisis De Pca	61
5.4.2 Entrenamiento Usando Regresión Lineal Con El Índice De Vegetación Ndvi	64
5.4.3 Entrenamiento Con La Red Neuronal Propuesta Usando El Índice De Vegetación Ndvi ..	67
5.4.4 Entrenamiento Con La Red Neuronal Propuesta Usando Imágenes Satelitales	69
5.4.5 Entrenamiento Usando Regresión Lineal Con El Índice De Vegetación Ndvi De Las Imágenes Satelitales.....	72
5.5 Resumen De Resultados	73
5.5.1 Todas Las Parcelas.....	73
5.5.2 Todos Los Árboles.....	74
5.5.3 Todos Los Píxeles	77
5.5.4 Por Parcela	79
5.6 Comparación De Resultados.....	81

6. Conclusiones	82
Trabajo Futuro	84
Bibliografía	86

Lista de Tablas

	Pág.
Tabla 1. Configuraciones de las misiones de vuelo.....	39
Tabla 2. Resumen de resultados con el método propuesto de redes neuronales para las imágenes de carbono <i>CQH</i> y <i>CGeneral</i>	63
Tabla 3. Resumen de resultados con el método de regresión lineal con NDVI para las imágenes de carbono <i>CQH</i> y <i>CGeneral</i>	66
Tabla 4. Resumen de resultados con el método propuesto de redes neuronales para las imágenes de carbono <i>CQH</i> y <i>CGeneral</i> usando la red neuronal propuesta con el índice de vegetación NDVI.....	70
Tabla 5. Resumen de resultados con el método propuesto de redes neuronales para las imágenes de carbono <i>CQH</i> y <i>CGeneral</i> usando un cubo de datos adquirido con una imagen satelital.....	71
Tabla 6. Resumen del promedio de resultados con el método de regresión lineal con NDVI para las imágenes de carbono <i>CQH</i> y <i>CGeneral</i> usando un cubo de datos adquirido con una imagen satelital.....	73
Tabla 7. Resumen del promedio de todos los resultados usando la organización de datos de todas las parcelas.....	74
Tabla 8. Resumen de los mejores resultados de RMSE% usando la organización de datos de todas las parcelas.....	75
Tabla 9. Resumen del promedio de todos los resultados usando la organización de datos de todos los árboles.....	76

Tabla 10. Resumen de los mejores resultados de RMSE% usando la organización de datos de todos los árboles.....	76
Tabla 11. Resumen del promedio de todos los resultados usando la organización de datos de todos los píxeles.....	78
Tabla 12. Resumen de los mejores resultados de RMSE% usando la organización de datos de todos los píxeles.....	79
Tabla 13. Resumen del promedio de todos los resultados usando la organización de datos de todos los píxeles.....	80
Tabla 14. Comparación de los mejores resultados de cada uno de los métodos de organización de datos.	81

Lista de Figuras

	Pág.
Figura 1. Cubo de información espectral con coordenadas espaciales (x,y) y longitud de onda λ	
.....	20
Figura 2. Captura consecutiva de ortofotos durante una misión de vuelo.	22
Figura 3. Diagrama de bloques para realizar mediciones remotas	25
Figura 4. Diagrama de bloques de la metodología de estimación de carbono propuesta.	29
Figura 5. Imagen de la zona de estudio visto desde Google Earth y su ubicación dentro del país.	
.....	33
Figura 6. Ubicación de las parcelas dentro de la zona de estudio realizada desde QGIS.	35
Figura 7. Diagrama de bloques con la descripción de las actividades realizadas en la expedición	
de campo.	36
Figura 8. Drone DJI Matrice 600 pro utilizado para capturar las fotografías de la zona de estudio.	
.....	37
Figura 9. Especificaciones técnicas del drone Matrice 600 pro.	37
Figura 10. Especificaciones técnicas de la cámara DJI Zenmuse X3.	38
Figura 11. Trazado de la ruta para las misiones de vuelo realizado desde la aplicación DJI GS Pro.	
.....	40
Figura 12. Ortomosaicos de la zona de estudio capturada desde los 100 metros de altura, a la	
izquierda se encuentra el ortomosaico RGB y a la derecha el ortomosaico BGNIR.	41
Figura 13. Ortomosaicos de la zona de estudio capturada desde los 130 metros de altura, a la	
izquierda se encuentra el ortomosaico RGB y a la derecha el ortomosaico BGNIR.	41
Figura 14. Bandas espectrales del ortomosaico RGBNIR. a. Red, b. Green, c. Blue y d. NIR.	42

Figura 15. Índices de vegetación estimados utilizando las bandas RGBNIR.	44
Figura 16. Cubo de datos generado para usar como fuente de datos para el proyecto.	44
Figura 17. Proceso para generar las imágenes de carbono.	48
Figura 18. Diagrama de bloques de un perceptrón.	49
Figura 19. Organización de datos de entrenamiento y validación para las pruebas. a. Todas las parcelas, b. Todos los árboles, c. Todos los píxeles y d. Por parcela.	52
Figura 20. Cada píxel del cubo de datos es un vector de 19 valores, los cuales se utilizan como entrada de la red neuronal.	55
Figura 21. Red neuronal utilizada en el proyecto, la cual cuenta con tres capas, una capa de entrada con 4 perceptrones, una capa oculta con 16 perceptrones y una capa de salida con un perceptrón.	57
Figura 22. Gráficas del RMSE% obtenido al estimar el CQH utilizando la red neuronal propuesta con las diferentes organizaciones de datos de entrenamiento.	62
Figura 23. Gráficas del RMSE% obtenido al estimar el CGeneral utilizando la red neuronal propuesta con diferentes organizaciones de datos de entrenamiento.	63
Figura 24. Gráficas del RMSE% obtenido al estimar el CQH utilizando una regresión lineal con el índice de vegetación NDVI.	65
Figura 25. Gráficas del RMSE% obtenido al estimar el CGeneral utilizando una regresión lineal con el índice de vegetación NDVI.	65
Figura 26. Gráficas del RMSE% obtenido al estimar el CQH utilizando la red neuronal propuesta con el índice de vegetación NDVI utilizando las diferentes organizaciones de datos de entrenamiento propuesta.	67

Figura 27. Gráficas del RMSE% obtenido al estimar el CGeneral utilizando la red neuronal propuesta con el índice de vegetación NDVI utilizando las diferentes organizaciones de datos de entrenamiento propuesta.	68
Figura 28. Gráficas del RMSE% obtenido al estimar el CQH y CGeneral utilizando la red neuronal propuesta con un cubo de datos generado con una imagen satelital.	71
Figura 29. Gráficas del RMSE% obtenido al estimar el CQH y CGeneral utilizando una regresión lineal del índice de vegetación NDVI con un cubo de datos generado con una imagen satelital.	72

Resumen

Título: Desarrollo de una metodología para la captura y procesamiento de información sobre el estado de los bosques utilizando UAVs *

Autor: Arnold Guarín Ramírez **

Palabras Clave: Biomasa Superficial, Bancos De Carbono, Redes Neuronales, Procesamiento De Imágenes, Imágenes Espectrales, UAV, Monitoreo Forestal.

Descripción: Colombia es un pulmón para el mundo y eso representa un enorme compromiso para controlar la tala ilegal de bosques, para ello cuenta con una plataforma de monitoreo forestal llamada SBYC que entrega reportes sobre el estado de los bosques trimestralmente. Sin embargo, zonas en Colombia requieren un seguimiento más frecuente para tomar acciones tempranas. Por esta razón, este trabajo se ha dedicado al desarrollo una metodología que permita entregar información forestal en cortos periodos de tiempo, usando como insumo datos capturados desde vehículos aéreos no tripulados (UAV). Para el estudio se eligió un bosque del departamento de Santander en el cual predomina la especie *Quercus Humboldtii*, también conocida como el roble blanco. Se realizó una expedición con el apoyo de expertos en estudios forestales que permitió caracterizar la biomasa del bosque, la cual refleja la capacidad del bosque para absorber carbono de la atmósfera. Adicionalmente se involucró el uso de un UAV dotado con cámaras espectrales de bandas R, G, B y NIR para capturar imágenes espectrales del bosque para obtener índices de vegetación. Finalmente se entrenó una red neuronal usando toda la información recopilada para construir un modelo para estimar la biomasa del bosque. Se pudo demostrar que la metodología desarrollada permite lograr una ganancia en las estimaciones representada no solo en tiempo sino también en una disminución de 11.21% del error de las estimaciones si se compara con la metodología que se apoya en información satelital.

* Trabajo de investigación

** Escuela de ingenierías eléctrica, electrónica y telecomunicaciones, Universidad Industrial de Santander. Director, Homero Ortega Boada. Co-director, Hans Yecid García Arenas

Abstract

Title: Development of a methodology for the capture and processing of information on forests health using UAVs*

Author: Arnold Guarín Ramírez **

Key Words: Aboveground Biomass, Neural Networks, Image Processing, Spectral Images, Carbon Pools, UAV, Forestry.

Description: Colombia is one of the world's lungs and that represents a huge commitment to control illegal logging on forests, to make this happen, it has a forest monitoring platform called SBYC that delivers reports on the state of forests on a quarterly basis. However, certain areas in Colombia require more frequent monitoring to take early action. For this reason, this work has been dedicated to the development of a methodology that allows delivering forest information in short periods of time, using data captured from unmanned aerial vehicles (UAV) as input. For the study, a forest in the department of Santander was chosen where the species *Quercus Humboldtii* predominates. An expedition was carried out with the support of experts in forestry that allowed characterizing the biomass of the forest, which reflects the capacity of the forest to absorb carbon from the atmosphere. Additionally, the use of a UAV equipped with R, G, B and NIR band spectral cameras was involved to capture spectral images of the forest to obtain vegetation indices. Finally, a neural network was trained using all the information collected to build a model to estimate the biomass of the forest. It was possible to demonstrate that the methodology developed allows achieving a gain in the estimates represented not only in time but also in a decrease of 11.21% in the error of the estimates when compared to the methodology that is based on satellite information.

* Research Work

** Escuela de ingenierías eléctrica, electrónica y telecomunicaciones, Universidad Industrial de Santander. Advisor,

Homero Ortega Boada. Coadvisor, Hans Yecid García Arenas

Introducción

El cambio climático es quizá la problemática más discutida en el mundo durante los últimos años, debido a que el ser humano ha sido el principal causante del aumento de los niveles de dióxido de carbono (CO_2) en la atmósfera. Anteriormente el CO_2 aumentaba de forma natural y se estabilizaba naturalmente debido a la captación del carbono de la atmósfera a la superficie terrestre realizada por los componentes oceánicos (ecosistemas marítimos) y los componentes terrestres (bosques, manglares, otros ecosistemas) (Stocker, 2014). El proceso de captación del CO_2 ocurre como resultado del ciclo del carbono en la atmósfera, donde las plantas toman CO_2 durante la fotosíntesis y liberan cierto porcentaje, convirtiendo el CO_2 que logran absorber en biomasa viva para el ecosistema (Farquhar & Sharkey, 1982). Un mayor nivel de biomasa significa a su vez, mayor densidad de carbono acumulado en el ecosistema por área y esto se da principalmente en los bosques. Por esa razón un término equivalente a la biomasa es lo que se conoce como bancos de carbono que representan el invaluable recurso natural que permite absorber carbono de la atmósfera y mitigar el cambio climático. Desafortunadamente, las interacciones del ser humano con los diferentes ecosistemas han desestabilizado los ciclos de CO_2 , debido los manejos deficientes de los diferentes ecosistemas, en especial los forestales. De aquí surge la necesidad de realizar una administración forestal sostenible, y en el reto que esto impone, las tecnologías de información y comunicación (TIC) juegan un papel muy importante, debido a que son las principales herramientas utilizadas en la adquisición, almacenamiento y disposición de la información forestal (ITU, 2018). Por esta razón, la Unión Internacional de Telecomunicaciones (UIT) como organismo especializado de las Naciones Unidas para las TIC, ha abierto espacios para que las partes interesadas en el cambio climático intercambien conocimientos y experiencias,

con el objetivo de identificar las necesidades para apoyar la integración de las TIC en la lucha contra el cambio climático (ITU, 2018).

Durante la conferencia de las naciones unidas sobre el cambio climático del 2013 se decidió que los países deben establecer sistemas nacionales de monitoreo forestal, los cuales deben permitir cuantificar la emisión y absorción de gases de efecto invernadero, al igual que la variación de carbono capturado en la superficie de los bosques y las reservas forestales, con el fin de poder ejecutar un monitoreo forestal y preservar los recursos naturales (UNFCCC, 2015). En la actualidad Colombia cuenta una plataforma de monitoreo forestal llamada sistema de monitoreo de bosques y carbono (SMBYC), desarrollado por el instituto de hidrología, meteorología y estudios ambientales (IDEAM); la cual genera trimestralmente reportes de deforestación en el país (Gutiérrez, y otros, 2018). Para realizar un proceso de monitoreo forestal efectivo se deben definir las zonas naturales que capturan grandes cantidades de carbono de la atmósfera, las cuales son llamadas pozos o bancos de carbono, la convención marco de las naciones unidas sobre el cambio climático, reconoce 5 tipos de bancos de carbono: biomasa superficial (también conocida como biomasa aérea), biomasa subterránea, madera muerta, lecho y carbono orgánico de tierra (SOC) (UNFCCC, 2015).

Existen diferentes formas de estimar los bancos de carbono, las cuales se pueden separar en dos categorías: métodos de campo y métodos remotos. Los métodos en campo son aquellos en los que la captura de información se realiza por medio de expediciones a la zona de estudio, donde se hacen mediciones y se toman muestras de la zona para hacer las estimaciones de todos los bancos de carbono (MADS & IDEAM, 2018). Por otro lado, los métodos remotos son aquellos que utilizan información capturada mediante sensores como cámaras espectrales, radares o láseres, que no se encuentran instalados en la zona de estudio, estos sensores generalmente se encuentran

instalados en satélites o aeronaves, y tienen la capacidad de realizar mediciones en zonas de interés a grandes escalas (departamentales, nacionales o continentales) (Giri, 2016). Cada metodología tiene ventajas y desventajas, las cuales están relacionadas con el costo de adquisición (tiempo y dinero) y a la incertidumbre o error asociado a la estimación. Donde los métodos de campo entregan información detallada de los bancos de carbono en áreas de estudio puntuales, pero requieren de altas inversiones en tiempo y dinero. En comparación, los métodos remotos presentan incertidumbres mayores que las excursiones de campo, pero a cambio, los métodos remotos tienen la ventaja de que cubren áreas de estudio mayores, y no requieren de una interacción directa con la zona de estudio (Howard, y otros, 2014). Debido a los beneficios que ofrecen los métodos remotos, estos han sido estudiados constantemente a través de los años buscando reducir la incertidumbre que presentan. Se ha demostrado que es posible reducir la incertidumbre de las estimaciones generadas por el error inherente de los sensores, las condiciones climáticas u otros factores, si se complementa la información remota satelital con información de las expediciones de campo o con mediciones remotas realizadas desde plataformas de vuelo como aeroplanos o globos aerostáticos, usando sensores como láseres de detección de luz y alcance (LiDAR) y cámaras espectrales (Hall, 2012). La información obtenida mediante estas plataformas de vuelo es llamada de alta resolución, debido a que estas plataformas operan a alturas inferiores a los satélites, lo cual permite que la información medida capture un mejor nivel de detalle que la información medida desde un satélite. Este nivel de detalle se conoce como resolución espacial, y hace referencia al área espacial de la zona estudiada que es representada por cada píxel en la imagen capturada en la plataforma. La resolución espacial generalmente se presenta como unidad de longitud sobre píxel. La resolución de los datos satelitales en general oscila entre 250 y 30 metros cuadrados por píxel, y la información de alta resolución puede estar en el orden de 1 metro

cuadrado por píxel o menor. Sin embargo, en los últimos años el uso de vehículos aéreos no tripulados (UAV) como plataformas de vuelo ha sido una alternativa para obtener información remota, esto debido a que pueden entregar datos con resoluciones en el orden de los centímetros por píxel, a su bajo costo de operación y a la facilidad que presentan para acoplar sobre ellos diferentes tipos de sensores (Berie & Burud, 2018).

Actualmente, pese a que el SMBYC entrega reportes a escala nacional en Colombia, estos reportes son generados trimestralmente, lo cual hace que esta plataforma no sea apta para labores de control permanente. Por esta razón, el uso de tecnologías emergentes que apoyen las labores de monitoreo se ha convertido en la principal misión de los entes de control forestal en el país. Con esta problemática en mente, el centro de investigación en tecnologías de información y comunicación (CentroTIC), como centro enfocado en la investigación de tecnologías TIC y su uso en pro del desarrollo departamental, propuso el proyecto *Plataforma IoT para el desarrollo de servicios inteligentes de apoyo al monitoreo ambiental* y recibió apoyo de la VIE-UIS donde fue matriculado con código 1971. Dentro de los diferentes casos de uso propuestos para la plataforma ambiental, uno de ellos busca implementar una metodología que utilice UAV para hacer seguimiento a la evolución de los bosques, con el fin de tener una metodología que aproveche la alta resolución espacial de la información entregada, al igual que la facilidad de despliegue de estos vehículos para hacer seguimientos en zonas de interés con mayor frecuencia.

En este proyecto de investigación se busca desarrollar una metodología que permita estimar el carbono superficial en una zona boscosa por medio de imágenes capturadas desde un UAV, sin embargo, se limita a una zona piloto que pueda ser sobrevolada por un UAV en una sola misión de vuelo. Para ello primero se seleccionó una zona de estudio en la cual se encuentra un bosque de roble blanco ubicado en el municipio de Molagavita, vereda Potrero de Rodríguez, Santander.

Dentro de la zona de estudio, con la ayuda de estudiantes y profesores de ingeniería forestal de la UIS sede Málaga se realizó un estudio en campo para estimar el carbono superficial, luego se sobrevoló un UAV para obtener imágenes espectrales de la zona de estudio. Posteriormente, con las imágenes espectrales capturadas se estimaron índices de vegetación, con los cuales se armó un cubo de datos que fue utilizado como entrada a un modelo de redes neuronales, donde la información objetivo de la red son los valores de carbono superficial estimados en campo por los ingenieros forestales. Finalmente, se comparó los resultados de la estimación propuesta con los resultados de otros modelos de estimación de carbono superficial y con los datos medidos en campo, demostrando que es posible estimar el carbono superficial en bosques usando imágenes de UAV, y que estas estimaciones son mejores que las obtenidas usando imágenes satelitales. Esta metodología facilita futuros estudios dentro de la zona seleccionada, debido a que, al contar con un modelo de redes neuronales entrenado, es posible obtener estimaciones únicamente con imágenes de UAV.

Este libro está estructurado de la siguiente manera: el capítulo 1 presenta el objetivo general y los objetivos específicos del proyecto; el capítulo 2 consiste en una revisión del estado del arte donde se presentan los diferentes estudios realizados en estimación de bancos de carbono, al igual que la explicación de la metodología diseñada en el proyecto; el capítulo 3 presenta el proceso de planeación de la expedición de campo, la recolección de datos durante la expedición de campo, la misión de vuelo ejecutada con el UAV para obtener imágenes espectrales de la zona de estudio y el preprocesamiento realizado a los datos capturados; el capítulo 4 consiste en la construcción del modelo matemático para la estimación de carbono usando como insumo los datos obtenidos mediante UAV; el capítulo 5 presenta el entrenamiento del modelo matemático y la validación del

mismo contra otros modelos de estimación de carbono, al igual que los resultados de las pruebas realizadas; finalmente, el capítulo 6 presenta las conclusiones del trabajo de investigación

1. Objetivos

1.1 Objetivo General

Desarrollar una metodología para estimar bancos de carbono a partir de imágenes RGB-NIR de bosques capturadas mediante drones.

1.2 Objetivos Específicos

- Caracterizar la capacidad de absorción de carbono en una zona piloto que se usará como referencia para el estudio.
- Obtener índices de vegetación a partir de ortomosaicos RGB-NIR que permitan estimar la capacidad de absorción de carbono en bosques.
- Desarrollar una solución que permita estimar los bancos de carbono superficiales utilizando los índices de vegetación obtenidos a partir de ortomosaicos RGB-NIR.
- Realizar la validación de la metodología propuesta, con la información de subáreas caracterizadas dentro de la zona piloto.

2. Métodos de estimación de bancos de carbono

En este capítulo se presenta una revisión de las diferentes metodologías utilizadas para hacer estimaciones de bancos de carbono. Se abarcan los métodos de campo, los métodos remotos, y finalmente se describe la metodología utilizada en este proyecto de investigación, la cual utiliza información remota de UAV. Sin embargo, antes de empezar se definirán unos conceptos clave que son utilizados durante el trabajo: imágenes espectrales, índices de vegetación y la fotogrametría con UAV.

2.1 Conceptos clave

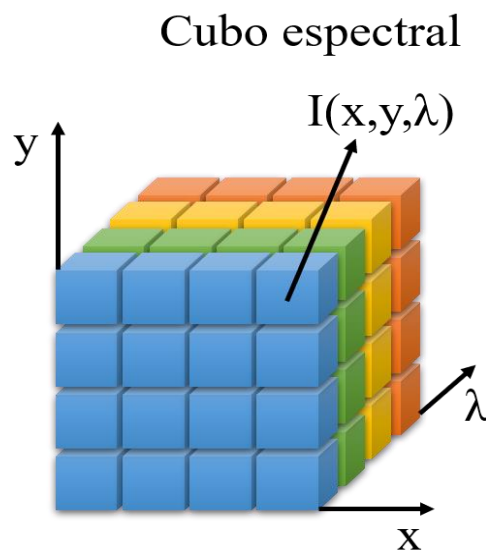
2.1.1 *Imágenes Espectrales*

Una imagen es una representación matricial de la integración de las dimensiones: espacio, longitud de onda, y tiempo; donde cada elemento de la matriz se conoce como vóxel (Schowengerdt, 2006). Es importante aclarar que la integración sobre el tiempo es usualmente pequeña (en el orden de los microsegundos o milisegundos), lo cual hace que en una imagen la dimensión temporal no sea relevante, mientras que la integración sobre el espacio y la longitud de onda, definen las características de la imagen. Si se representa un espacio tridimensional (x, y, λ) , definido sobre unas coordenadas espaciales (x, y) y una longitud de onda espectral (λ) , cada vóxel de una imagen dada puede ser representado como una integración sobre un volumen relativamente pequeño en ese espacio continuo (Schowengerdt, 2006). En la Figura 1 se observa una representación de una imagen espectral, la cual es vista como un cubo de información distribuida en las tres dimensiones (x, y, λ) . Las imágenes espectrales pueden ser clasificadas según los valores de λ que posean. Por ejemplo, las imágenes de las cámaras de los teléfonos móviles capturan la región visible del espectro electromagnético, específicamente, tres valores de λ , λ_1 , λ_2 y λ_3 .

uno correspondiente a la región del espectro rojo (R), otro correspondiente al verde (G) y otro a la región azul (B), formando colores secundarios a partir de las regiones RGB que permiten representar la mayor parte de los colores percibidos por el ser humano. Cada longitud de onda entrega información diferente del objeto del cual se refleja, mientras que los valores de λ referentes al espectro visible entregan información referente a los colores y las texturas de los objetos, otros valores de λ como los del espectro infrarrojo, entregan información referente a las temperaturas. Las imágenes pancromáticas (i.e., escala de grises) son otro ejemplo, cuyo valor de λ integra el todo rango visible del espectro mostrándolo como una imagen en escala de grises (Shaw & Burke, 2003). En este estudio se utilizará con frecuencia la banda del infrarrojo cercano, la cual son los valores de λ que se encuentran entre la banda del rojo y la banda el infrarrojo, esta banda se utiliza principalmente en estudios referentes al estudio de zonas naturales, como bosques o cultivos (Zhao, y otros, 2018).

Figura 1.

Cubo de información espectral con coordenadas espaciales (x, y) y longitud de onda λ



A partir de las imágenes espectrales se puede extraer información que puede ser utilizada en diferentes áreas, como la telemetría (Shaw & Burke, 2003), biología microscópica (Zimmermann, Rietdorf, & Pepperkok, 2003), imágenes biomédicas (Lu & Fei, 2014), entre otras. En este proyecto se hace énfasis en el uso imágenes espectrales para obtener parámetros que indiquen el estado de zonas forestales, también conocidos como índices de vegetación (Zhao, y otros, 2018).

2.1.2 Índices de Vegetación

Con la necesidad de implementar sistemas de monitoreo a nivel nacional, las imágenes espectrales capturadas desde satélites aeroespaciales han sido una de las principales fuentes de información para los sistemas de monitoreo. Sin embargo, es importante aclarar que una imagen espectral por sí sola no entrega información relevante para el seguimiento de los bosques, es necesario interpretar las imágenes espectrales mediante transformaciones matemáticas o análisis estadísticos. Con esta premisa nacen los índices de Vegetación, los cuales son parámetros que ayudan a monitorizar las variaciones de la estructura vegetal durante periodos de tiempo deseado. Estos parámetros son obtenidos mediante transformaciones algebraicas de dos o más bandas espectrales, con el fin de resaltar ciertas propiedades vegetales (Huete, y otros, 2002).

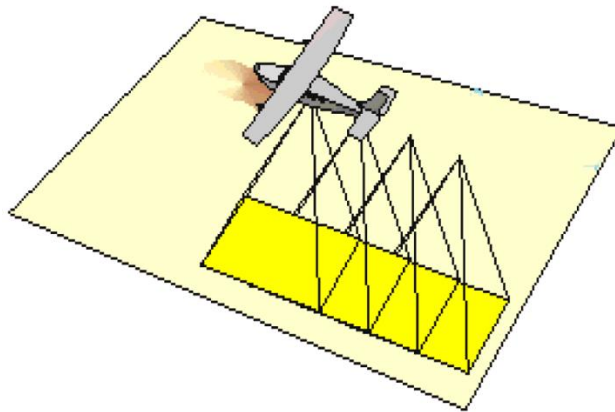
Existen diferentes índices de Vegetación, sin embargo, la mayor parte de estos son obtenidos a partir de las bandas RGB e infrarrojo cercano (NIR). Algunos de los IV más utilizados son: *Normalized Difference Vegetation Index* (NDVI) (Rouse, Hass, Schell, & Deering, 1973), *Ratio Vegetation Index* (RVI) (Pearson & Miller, 1972), *Normalized Red Green Difference Index* (NGRDI) (Gitelson, Kaufman, Stark, & Rundquist, 2002), el *Green NDVI* (GNDVI) (Gitelson, Kaufman, & Merzlyak, Use of a green channel in remote sensing of global vegetation from EOS-MODIS, 1996).

2.1.3 Fotogrametría con UAV

La fotogrametría con UAV es el término utilizado para referirse al proceso de capturar fotografías en una zona de interés utilizando UAV para obtener una representación digital del área sobrevolada (Eisenbeiß, 2009). Para ello se hace necesario programar una ruta sobre el área de estudio, que será recorrida por el UAV y capturará fotografías de forma consecutiva las cuales se conocen como ortofotos, estas fotografías son tomadas consecutivamente con un porcentaje de solapamiento con el fin de capturar diferentes ángulos de un mismo objetivo y poder construir un modelo de la zona de estudio deseada, este modelo se conoce como modelo digital (Kraus, 2011). A partir de este modelo digital es posible extraer una imagen de toda la zona de estudio, la cual se conoce como ortomosaico (Remondino, Barazzetti, Nex, Scaioni, & Sarazzi, 2011). En la Figura 2 se presenta una ilustración de la captura consecutiva de fotografías durante la misión de vuelo.

Figura 2.

Captura consecutiva de ortofotos durante una misión de vuelo.



2.2 Métodos de campo

La finalidad de las expediciones de campo es recolectar información relacionada directa o indirectamente con la biomasa de la zona de estudio con el fin de cuantificarla. La información

recolectada va desde mediciones de las características de la flora, hasta la extracción de muestras para un posterior análisis en laboratorio. Sin embargo, hay que tener en cuenta que se requiere tener control sobre la información recolectada, al igual que se necesita que la metodología sea replicable para medir en diferentes zonas de interés. Por estas razones, los entes de control ambiental se han encargado de desarrollar manuales para la recolección de información en campo (FAO, 2018; Howard, y otros, 2014; Goslee, y otros, 2015; Norwegian Agency for Development Cooperation & FECOFUN, 2010; MADS & IDEAM, 2018). Estos manuales explican las actividades para llevar a cabo las mediciones en campo, las cuales pueden ser generalizadas en:

- **La realización del plan de muestreo:** Los bosques naturales pueden llegar a tener grandes áreas de cobertura, lo cual hace que realizar un estudio exhaustivo en toda el área sea imposible. Por esta razón, se hace necesario desarrollar un plan que permita seleccionar zonas de muestra dentro del área de estudio, sobre las cuales se realizaran mediciones exhaustivas de forma que la información recolectada permita estimar el estado del área completa. Los manuales de campo definen los tamaños de cada zona de muestra, también llamada conglomerados o parcelas. En esta etapa también se definen los roles y actividades que desarrollaran los integrantes del equipo de trabajo durante la expedición, la información que se recolectará, al igual que las normas y reglas a seguir durante la expedición.
- **Mediciones en las zonas de muestreo:** Con las parcelas definidos, el siguiente paso consiste en realizar las mediciones planeadas. Los datos por recolectar en cada conglomerado varían con respecto al tipo de estudio requerido. En el caso de realizar mediciones para la estimación de carbono se mide: la posición de los árboles dentro de cada parcela, el diámetro del tronco del árbol a la altura del pecho

(DAP), la altura de los árboles, muestras botánicas como ramas de los árboles, secciones del tallo de un árbol, frutos, y muestras de la tierra.

- **Registro y seguimiento de las actividades:** Una vez realizadas las mediciones, el siguiente paso consiste en organizar toda la información recolectada, hacer los estudios respectivos a las muestras y disponer de la información para uso posterior de los interesados. Debido a que en gran parte de los casos la información recolectada pasa a manos de los entes de control, estos se encargan de especificar en los manuales de campo los modelos a seguir para tabular y almacenar la información recolectada, de esta forma, todas las expediciones manejan un mismo modelo de datos y un mismo lenguaje.

Las metodologías de campo permiten extraer información con la menor incertidumbre de cada conglomerado en la zona de estudio, pero hay que resaltar que estas metodologías son invasivas, lo que implica una alteración en el ecosistema o una destrucción parcial del mismo en la extracción de muestras (Chave, y otros, 2014). Desafortunadamente, existen casos en que las mediciones de campo no pueden ser realizadas, debido a que el bosque se encuentra dentro de una propiedad privada y el dueño del lugar no permite el acceso, o cuando el bosque se encuentra en una zona donde el acceso puede ser peligroso para el equipo de trabajo; bien sea por dificultades de movilidad dentro de la zona o por alguna condición en que el ecosistema se encuentre en ese momento.

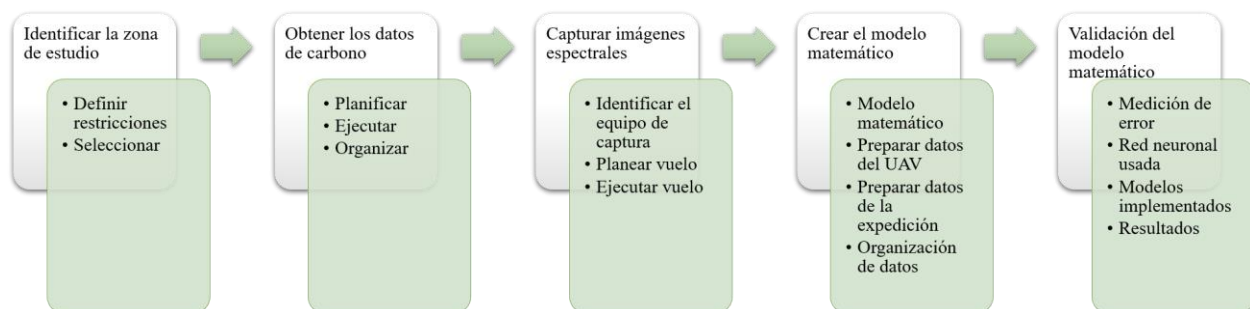
2.3 Métodos remotos

En general, los métodos remotos pueden ser representados como un proceso de 5 pasos: identificación de la zona de estudio, obtención de datos de carbono dentro de la zona de estudio,

realización de las mediciones remotas, desarrollo y entrenamiento de un modelo matemático que relacione las mediciones remotas y los datos de carbono, y la validación del modelo matemático. En la Figura 3 se presenta un diagrama de bloques con los pasos utilizados en el proceso de las mediciones remotas. Estos pasos son comunes en todas las metodologías encontradas en la literatura, teniendo en cuenta que se diferencian por la forma en que se obtienen los datos de carbono, por el tipo de datos remotos que utilizan o por el modelo matemático que hayan escogido (Ponce-Corona, y otros, 2019). Este tipo de métodos de estimación de biomasa son utilizados principalmente para realizar inventarios forestales en áreas de mayor cobertura a las posibles mediante expediciones de campo. Esta clase de estimaciones de biomasa utilizan información remota medida desde diferentes plataformas, junto con información recolectada mediante expediciones de campo y algoritmos de procesamiento que permitan relacionar matemáticamente la información recolectada en campo y la remota, para así generar mapas de biomasa (FAO, 2018).

Figura 3.

Diagrama de bloques para realizar mediciones remotas



La zona de estudio es el factor más importante en un estudio de biomasa por medio de datos remotos, esto se debe a que, dependiendo de la extensión de esta, se seleccionan las plataformas sobre las cuales se puede obtener información y por consiguiente los algoritmos para

relacionar los datos de campo con las mediciones realizadas remotamente (Byrd, y otros, 2018). Basados en la revisión realizada, en este trabajo se decidió separar las zonas de estudio en tres clases:

- **Zonas de estudio a grandes escalas:** son aquellas que buscan generar inventarios a escalas nacionales o departamentales a partir de información medida desde satélites, por medio de diferentes sensores como cámaras espectrales, radares de apertura sintética (SAR), láseres, entre otros (Byrd, y otros, 2018; Urbazaev, y otros, 2018; Sun, y otros, 2019). Actualmente en Colombia se utilizan las mediciones remotas con imágenes satelitales para hacer inventarios forestales nacionales, el procedimiento para trabajar con las imágenes satelitales es explicado en el manual creado por el instituto de hidrología, meteorología y estudios ambientales (IDEAM), que es uno de los entes de control del medio ambiente en el país (Cabrera E., G. Galindo, D.M. Vargas, 2011).
- **Zonas de estudio enfocadas a ecosistemas:** son aquellas que buscan estudiar bosques completos en ecosistemas situados a condiciones climáticas o territoriales particulares. En estos casos se utiliza la información satelital, y en algunos casos es posible utilizar otras plataformas para obtener la información como aeroplanos, globos aerostáticos, y recientemente UAV. (Galidaki, y otros, 2017; Zhang, y otros, 2019; Hickey, Callow, Phinn, Lovelock, & Duarte, 2018; Navarro, y otros, 2019; Pandey, Srivastava, Chetri, Choudhary, & Kumar, 2019; Gomez, White, & Wulder, 2016; Tejada, Görgens, Espírito-Santo, Cantinho, & Ometto, 2019).
- **Zonas de estudio puntuales:** son aquellas que buscan realizar estudios de biomasa en áreas pequeñas donde la resolución de la información satelital no permite

realizar un análisis adecuado del área. Este tipo de zonas normalmente se encuentran ubicadas en fincas o predios privados, donde los dueños de esas tierras desean hacer seguimientos a sus recursos forestales. En estos casos, los inventarios se realizan usando expediciones de campo, sin embargo, debido a la dificultad de este tipo de estudio, no se realiza de manera periódica. Es aquí donde el uso de información remota de alta resolución obtenida mediante aeroplanos, globos aerostáticos o UAV puede ayudar a complementar los estudios de la zona de interés. Sin embargo, este tipo de mediciones no están exentas de problemas, debido a que, según las características ecosistémicas o topográficas de la zona de estudio, la capacidad de captura de información puede verse limitada, dando lugar a malas estimaciones o a información que no puede ser utilizada. (Chen, y otros, 2017; Warfield & Leon, 2019; de Almeida, y otros, 2019; Jayathunga, Owari, & Tsuyuki, 2018; Swinfield, y otros, 2019; Domingo, Ørka, Næsset, Kachamba, & Gobakken, 2019; Ota, y otros, 2019).

Por otro lado, en la literatura se trabaja el uso de aprendizaje de maquina (ML) para relacionar los datos remotos con los datos obtenidos en la expedición de campo, para generar modelos que permiten mejorar las estimaciones de biomasa (Byrd, y otros, 2018; Sun, y otros, 2019; Urbazaev, y otros, 2018). También, algo importante a resaltar de las diferentes plataformas desde las cuales se puede extraer información remota, es que estas no son excluyentes entre sí, de hecho, gran parte de los trabajos encontrados en la literatura buscan mejorar las estimaciones de biomasa utilizando combinaciones de diferentes fuentes de información (Byrd, y otros, 2018; Galidaki, y otros, 2017; Sun, y otros, 2019; Hickey, Callow, Phinn, Lovelock, & Duarte, 2018; Urbazaev, y otros, 2018; Saarela, y otros, 2018) .

Finalmente, se hace necesario resaltar ciertas características de las mediciones remotas de alta resolución, debido a que en el proyecto estas juegan un papel importante. Las mediciones remotas de alta resolución dependiendo de la plataforma de vuelo sobre la cual se capture, pueden ser obtenida a periodos cortos de tiempo, ya que una plataforma como los UAV se puede obtener todos los días si el clima lo permite (Ota, y otros, 2019). Por otro lado, de los sensores más utilizados en la literatura sobre este tipo de plataformas son los LiDAR y las cámaras espectrales, esto se debe a que la información remota que se recolecta para este tipo de estudios normalmente son las imágenes espectrales e información referente a la altura de los árboles, ya que con las imágenes espectrales es posible calcular índices de vegetación, y la información de las alturas de los árboles está fuertemente relacionada a la biomasa (de Almeida, y otros, 2019). Sin embargo, debido a los altos costos de los sensores LiDAR, se han realizado estudios que buscan comparar los resultados obtenidos mediante los sensores LiDAR, con resultados obtenidos con técnicas de procesamiento de imágenes como la reconstrucción por movimiento (SFM); las cuales concluyen que los sensores LiDAR entregan buenas estimaciones, pero, dependiendo del tipo de bosque a estudiar, en caso de que este no sea muy frondoso y que las condiciones del clima permitan capturar fotografías claras e iluminadas, el SFM puede obtener resultados similares a los de LiDAR (Warfield & Leon, 2019; de Almeida, y otros, 2019; Jayathunga, Owari, & Tsuyuki, 2018; Swinfield, y otros, 2019).

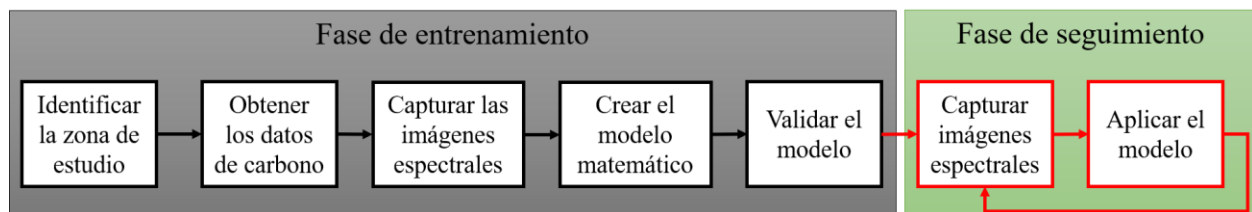
2.4 Metodología de estimación de carbono propuesta

La metodología propuesta en este trabajo de investigación sigue los cinco pasos descritos de los métodos remotos, sin embargo, estos fueron adaptados para que se ajusten a los requerimientos del proyecto. Teniendo en cuenta que el objetivo del proyecto es estimar bancos

de carbono usando imágenes espectrales obtenidas desde UAV en un área forestal particular, la metodología propuesta corresponde la fase de entrenamiento en un proceso de monitoreo de bosques, teniendo en cuenta que para completar el proceso de monitoreo de bosques se requiere de un seguimiento periódico. El monitoreo de bosques consta de dos grandes fases: una de entrenamiento y una de seguimiento. En la fase de entrenamiento se realiza la construcción y entrenamiento del modelo matemático que se encarga de estimar el carbono en base a las imágenes espectrales de UAV. En la fase de seguimiento se realiza el monitoreo continuo de la zona de estudio. Sin embargo, es importante aclarar que en este proyecto de investigación se cubre únicamente la fase de entrenamiento de la metodología propuesta, ya que un seguimiento continuo a la zona de estudio se sale de los objetivos del proyecto. En la Figura 4 se presenta un diagrama de bloques, en el cual se encuentran los pasos del monitoreo forestal en el que se incluye la metodología propuesta.

Figura 4.

Diagrama de bloques de la metodología de estimación de carbono propuesta.



2.4.1 Fase de entrenamiento

El primer paso de esta fase consiste en la identificación de la zona de estudio, en el cual se selecciona la zona de estudio. Para ello se identifican los equipos de medición con los que se cuenta y evaluar su comportamiento en la zona de estudio. En esta metodología las mediciones remotas se realizan por medio de UAV, por ende, se debe garantizar que la zona de estudio cuente con las

condiciones climáticas y topológicas necesarias para poder operar el UAV en ella. Al igual, que debe garantizarse un acceso seguro y sin inconvenientes a dicha zona para poder realizar la expedición de campo.

El segundo paso consiste en obtener los datos de carbono de la zona de estudio, estos son los utilizados para entrenar y validar el modelo matemático. Estos datos deben ser obtenidos mediante una expedición de campo a una fecha cercana a la de la captura de imágenes con el UAV.

El tercer paso es el de la captura de imágenes espectrales, en el cual se realiza la misión de vuelo con el UAV sobre la zona de estudio, por medio de la fotogrametría se genera información con las cámaras espectrales equipadas al UAV. Es importante asegurarse que las condiciones para realizar la misión de vuelo sean las indicadas, que no haya vientos fuertes y de preferencia en días en los que el cielo este despejado. Posteriormente, está la preparación de los datos para poder ser utilizados en el modelo matemático. Para esta metodología, las imágenes capturadas son de bandas *R*, *G*, *B* y *NIR*, con las cuales se calculan diferentes índices de vegetación, para finalmente generar un cubo de datos el cual servirá de entrada para el modelo matemático.

El cuarto paso consiste en crear un modelo matemático que permita relacionar las imágenes espectrales con los datos de carbono. Para encontrar un modelo adecuado en este proyecto se realizan pruebas con diferentes modelos matemáticos y se comparan los resultados de las estimaciones para encontrar el modelo que presente el menor error en la estimación.

El último paso de la fase de entrenamiento consiste en la validación del modelo matemático, para ello se proponen diferentes formas de organización de los datos de entrenamiento, los cuales son posteriormente separados en conjuntos de entrenamiento y de prueba, y así evaluar cual modelo matemático entrega los menores resultados usando un determinado método de organización de datos.

2.4.2 Fase de entrenamiento

Una vez terminada la fase entrenamiento, ya se cuenta con un modelo que está adaptado a la zona de estudio, por ende, es posible hacer misiones de vuelo futuras y procesar las nuevas imágenes en el modelo entrenado para estimar el carbono. Por esta razón, esta fase consta de dos pasos, una captura de imágenes espectrales y una aplicación del modelo entrenado sobre estas imágenes, de esta forma es posible hacer seguimiento a la evolución de la zona de estudio, siempre y cuando se garantice que las condiciones de la captura de imágenes espectrales sean las mismas a las de las imágenes de entrenamiento. Es importante aclarar que la metodología propuesta proyecta la fase de seguimiento como una implementación futura, la cual depende de negociaciones concretas para realizar un monitoreo permanente en la zona de estudio, por esta razón, en este proyecto de investigación esta fase no fue validada, ya que se sale de los objetivos del proyecto, sin embargo, en un trabajo futuro se proyecta la validación de esta fase.

3. Proceso de planificación, recolección y preprocesamiento de la información

Este capítulo presenta el desarrollo de los tres primeros pasos de la metodología propuesta: identificar la zona de estudio, obtener los datos de carbono y capturar las imágenes espectrales. El primer problema es la selección de la zona de estudio, para ello se muestran los criterios de selección y las características de la zona escogida. Posteriormente, se describe la expedición de campo realizada para obtener los datos de carbono de la zona de estudio. Luego se presenta el UAV y las cámaras espectrales utilizadas, seguido del proceso de captura de imágenes espectrales, y finalmente preprocesamiento de las imágenes capturadas para ser entregadas al modelo matemático.

3.1 Zona de estudios

Para de seleccionar la zona de estudio, se tuvieron en cuenta los siguientes criterios:

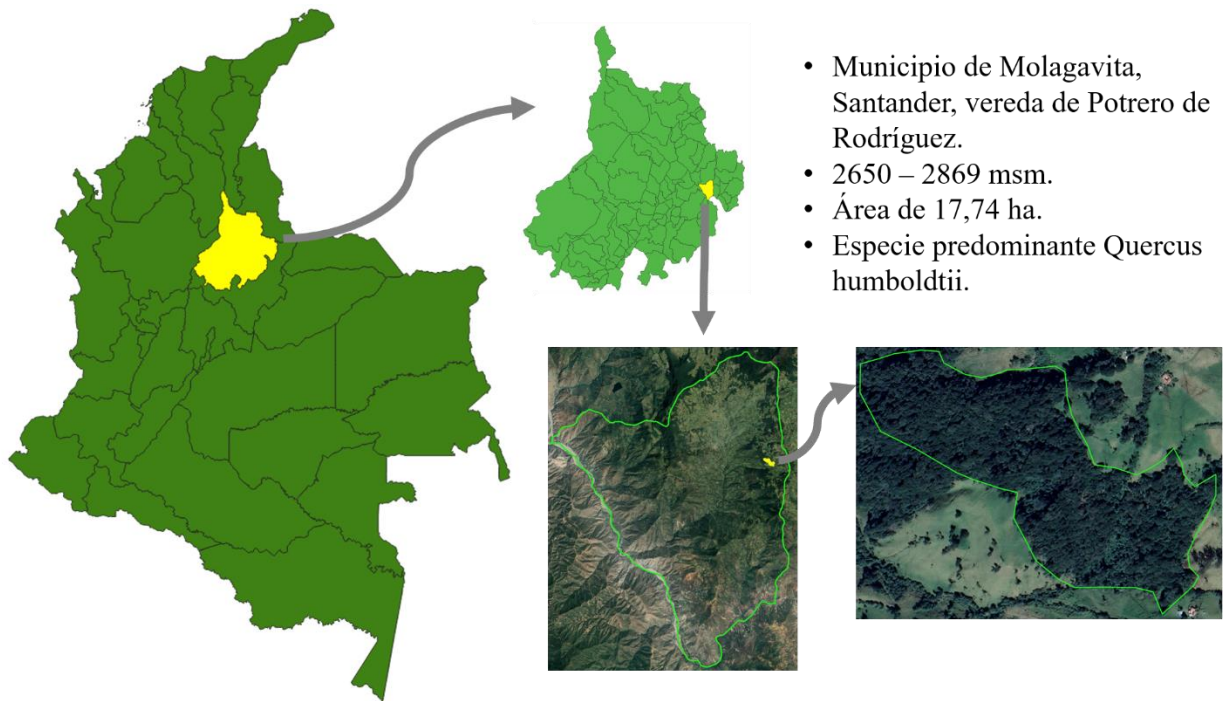
- Debe contar con un área boscosa notoria.
- Las especies de árboles que se encuentren dentro de la zona deben ser especies previamente estudiadas.
- La topografía de la zona de estudio debe permitir volar el UAV sobre ella, sin inconvenientes.
- El área de cobertura de la zona de estudios debe permitir ser transcurrida en una misión de vuelo del UAV.
- Debe ser de fácil acceso para las personas que realizarán la expedición de campo.

En base a los criterios anteriores, se seleccionó una zona que se encuentra ubicada en la república de Colombia, en el departamento de Santander, en el municipio de Molagavita dentro de la vereda de Potrero de Rodríguez. La zona de estudio está dentro de la finca los Cinchos y cuenta

con una temperatura promedio de 13°C y una precipitación promedio de 1550mm anuales. Con altitudes desde 2650 hasta 2869 metros sobre el nivel del mar. Esta zona tiene un área de 17,74 hectáreas y se encuentra delimitada por las coordenadas 6°39'45,05" - 6°39'5" sentido SN y 72°46'24,00" - 72°46'69" sentido EW. El bosque presente en esta zona de estudio corresponde a un bosque muy húmedo montano bajo, en el cual predomina la especie *Quercus Humboldtii* también conocidos como roble andino o roble colombiano. En la Figura 5 se encuentra una imagen de la zona de estudio obtenida de Google Earth junto con su ubicación dentro del país.

Figura 5.

Imagen de la zona de estudio visto desde Google Earth y su ubicación dentro del país.



3.2 Expedición en campo para la estimación de biomasa superficial

La expedición de campo a la zona de estudio fue realizada por dos estudiantes del programa de ingeniería forestal de la universidad Industrial de Santander sede de Málaga, como parte del

proyecto de grado “**Estimación de biomasa con métodos no destructivos en bosque de *Quercus humboldtii Bonpl*”**”, dirigida por la profesora M.Sc. Sandra Milena Díaz López.

Para realizar la expedición se siguieron las instrucciones del manual de campo para el inventario forestal (Ferreira, 2005), donde se indican los pasos necesarios para realizar el estudio de biomasa superficial en una zona forestal. El primer paso consiste en estimar el número de parcelas o subáreas necesarias para muestrear la zona de estudio. Para ello, los encargados de realizar las mediciones consideraron que un muestreo del 10% del área total es suficiente; por otro lado, como el manual define parcelas cuadradas de $25 \times 25\text{m}$ (0,0625ha) y teniendo en cuenta que el área total de la zona de estudio es de 17,74ha, se concluye que para cubrir el área de muestreo deseada se requieren de 28 parcelas.

El siguiente paso consiste en la ubicación de las parcelas en el área de estudio, para ello se utilizó QGIS, la cual es una herramienta software que permite la manipulación de datos georreferenciados. Con QGIS es posible usar la delimitación del área de estudio presentada en la Figura 5, y generar puntos aleatorios dentro de la misma, y así obtener las coordenadas de los puntos centrales donde se ubicará cada parcela. Posteriormente, se procede a ir a cada uno de los puntos y trazar el perímetro definido, teniendo en cuenta las correcciones necesarias en caso de que la parcela se encuentre ubicada en una pendiente. En la Figura 6 se encuentran las ubicaciones de cada una de las parcelas dentro de la zona de estudio.

Dentro de cada parcela se toman los siguientes datos a cada uno de los árboles: el diámetro del tronco a la altura del pecho (DAP) con una cinta métrica, la altura total hasta la copa del árbol con un Hipsómetro Suunto PM5-1520, el diámetro de la copa del árbol, la altura a la primera ramificación del árbol y las coordenadas espaciales donde está ubicado el árbol con un sistema de posicionamiento global (GPS) Garmin GPSMAP 64S one size. Posteriormente, utilizando las

ecuaciones obtenidas en la plataforma GlobAllomeTree para la especie *Quercus humboldtii*, se estima la biomasa superficial y el carbono almacenado en cada uno de los árboles dentro de cada parcela. Se utilizaron dos modelos en cada uno de los árboles para estimar la biomasa, dados por la ecuación (1) la cual es la que se usa en bosques donde únicamente se encuentra la especie de *Quercus humboldtii* debido a que es la que predomina en la zona de estudios, y la ecuación (2) se usa en bosques donde existen diversas especies, pero la especie predominante es la *Quercus humboldtii* respectivamente.

$$B_{QH} = (-5,864 + 0,906 * DAP)^2 \quad (1)$$

$$B_{General} = (-1,934 + 0,676 * DAP)^2 \quad (2)$$

Para obtener los valores de carbono superficial almacenado, los estudios realizados en bosques tropicales sugieren que el carbono está directamente relacionado con la biomasa presente en un bosque, la relación consiste en que el carbono es equivalente al 50% de la biomasa presente en el bosque (Yepes, y otros, 2016). De esta forma se presentan dos posibles estimaciones de carbono $C_{QH} = B_{QH}/2$ y $C_{General} = B_{General}/2$. En la Figura 7 se presenta un diagrama de bloques que resume los pasos de la expedición de campo.

Figura 6.

Ubicación de las parcelas dentro de la zona de estudio realizada desde QGIS.

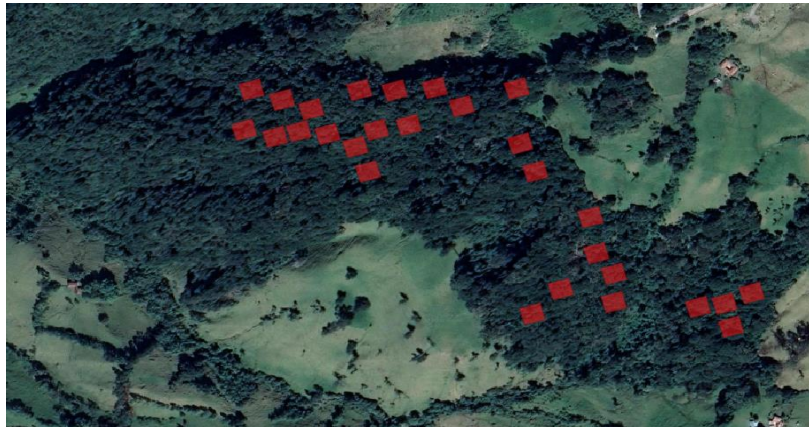
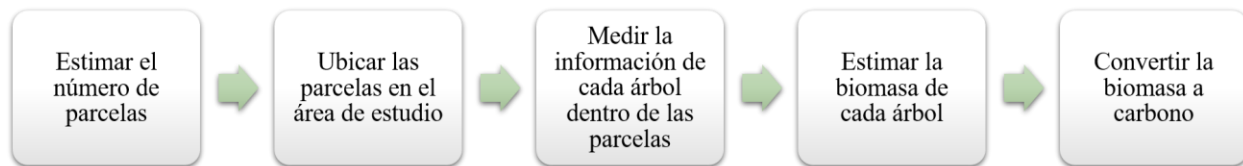


Figura 7.

Diagrama de bloques con la descripción de las actividades realizadas en la expedición de campo.



Finalmente, los datos de la expedición de campo se entregan en una base de datos, en la cual se encuentran datos de cada uno de los árboles medidos como:

- El identificador de la parcela donde se midió.
- El identificador del árbol dentro de la parcela.
- El identificador del árbol en todo el bosque.
- El diámetro del tallo del árbol.
- Las dos estimaciones de carbono.
- Las coordenadas geográficas del árbol.

3.3 Equipo de captura de imágenes

Como la metodología propuesta requiere de imágenes espectrales R-G-B-NIR, el equipo utilizado para la captura de información, debe ser capaz de utilizar cámaras que permitan obtener imágenes espectrales, también debe ser lo suficientemente robusto para volar sobre el área de estudio sin inconvenientes. Con estas condiciones en mente, para la captura de las imágenes de la zona de estudio se utilizó un dron de DJI llamado Matrice 600 Pro el cual forma parte del grupo de investigación RadioGIS de la UIS. Este es un equipo industrial de seis hélices caracterizado por su robustez. En la Figura 8 se encuentra una fotografía del UAV utilizado y en Figura 9 se encuentran las características relevantes del UAV.

Figura 8.

Drone DJI Matrice 600 pro utilizado para capturar las fotografías de la zona de estudio.



Figura 9.

Especificaciones técnicas del drone Matrice 600 pro.

Velocidad máxima	17 m/s	Tiempo de vuelo (sin carga)	32 min
Resistencia al viento	8 m/s	Tiempo de vuelo (carga máxima)	15 min
Velocidad de ascenso	5 m/s	Precisión de vuelo	Vertical : $\pm 0,5$ m Horizontal: $\pm 1,5$ m
Velocidad de descenso	3 m/s	Frecuencias de operación	5,725 - 5,825 GHz 2,400 - 2,483 GHz
Carga máxima	6 kg	Distancia máxima de transmisión	3,5 km (Libre de obstáculos e interferencias)

Las cámaras utilizadas fueron dos DJI Zenmuse X3, de las cuales una captura las bandas RGB y otra BGNIR. Es importante aclarar que ambas cámaras son de un mismo modelo, pero a una de ellas fue modificada por la empresa distribuidora para reemplazar la banda del rojo por la de infrarrojo cercano, capturando longitudes de onda entre los 680 y 800nm, y quitando las longitudes desde 568nm hasta los 680nm pertenecientes a la banda del rojo. Por otro lado, debido a la estructura donde se coloca la cámara en el drone, no es posible realizar un vuelo con el drone utilizando las dos cámaras al mismo tiempo, razón por la cual se hace necesario realizar dos vuelos en las mismas condiciones para capturar información con cada una de las cámaras. Las características relevantes del modelo de la cámara DJI Zenmuse X3 se encuentran en la Figura 10, en el cual se subraya la característica de la resolución que se utilizó con ambas cámaras durante el proyecto.

Figura 10.*Especificaciones técnicas de la cámara DJI Zenmuse X3.*

Tipo	CMOS	Óptica	20 mm
Píxeles efectivos	12,4 M	FOV diagonal	94 grados
Velocidades de obturación	8 ~ 1/8000 sec	Encoder	MPEG4/AVC/H.264
Rango ISO	100 - 3200	Resolución	4095 × 2160 @ 25/24fps 3860 × 2160 @ 30/25fps 1920 × 1080 @ 60/50/48/30/25/24fps 1280 × 720 @ 60/50/48/30/25/24fps

3.4 Captura de los ortomosaicos de la zona de estudio

Para obtener los ortomosaicos de la zona de estudio, se realizaron dos misiones de vuelo con el UAV a diferentes alturas, una a 100m y otra a 130m. En cada una de las misiones se trazó la ruta de vuelo a recorrer, la altura de vuelo, la periodicidad en la que se capturan las fotografías y las configuraciones de la cámara para la captura de las fotografías. Cada misión de vuelo se ejecutó dos veces debido a que era necesario capturar información con cada una de las cámaras. En la Tabla 1 se describen las configuraciones utilizadas en cada misión de vuelo y en la Figura 11 se presenta la ruta de vuelo utilizada en las dos misiones, donde el área azul hace referencia al terreno capturado por el drone durante la misión, y las líneas verdes indican la ruta que siguió el drone para capturar las fotografías.

Tabla 1.

Configuraciones de las misiones de vuelo.

Característica	Misión a 100m	Misión a 130m
Velocidad de vuelo	8,7 m/s	8,5 m/s

Distancia Recorrida	4429,3 m	3622,3 m
Intervalo de captura de fotos	3 seg	4 seg
Cantidad de fotos tomadas	188	113
Resolución espacial	4,3 cm/px	5,6 cm/px
Superposición frontal	80%	80%
Superposición lateral	60%	61%

Una vez capturadas las fotografías, estas son procesadas en Agisoft Photoscan para generar un modelo digital de la zona de estudio, desde el cual se puede extraer el ortomosaico. Una vez realizado el proceso con las imágenes de cada cámara, se obtienen los ortomosaicos presentados en la Figura 12 y la Figura 13, donde la imagen de la izquierda es el ortomosaico RGB y la de la derecha es el BGNIR en ambas imágenes. Sin embargo, al no considerar las variaciones climáticas el día que se realizó la misión de vuelo, se vio la necesidad de descartar las imágenes capturadas a 100m de altura, debido a que estas presentaron sombras causadas por nubes, por esta razón, estas imágenes no fueron utilizadas en el proyecto. Mientras que las de 130m presentaron mejores condiciones.

Figura 11.

Trazado de la ruta para las misiones de vuelo realizado desde la aplicación DJI GS Pro.

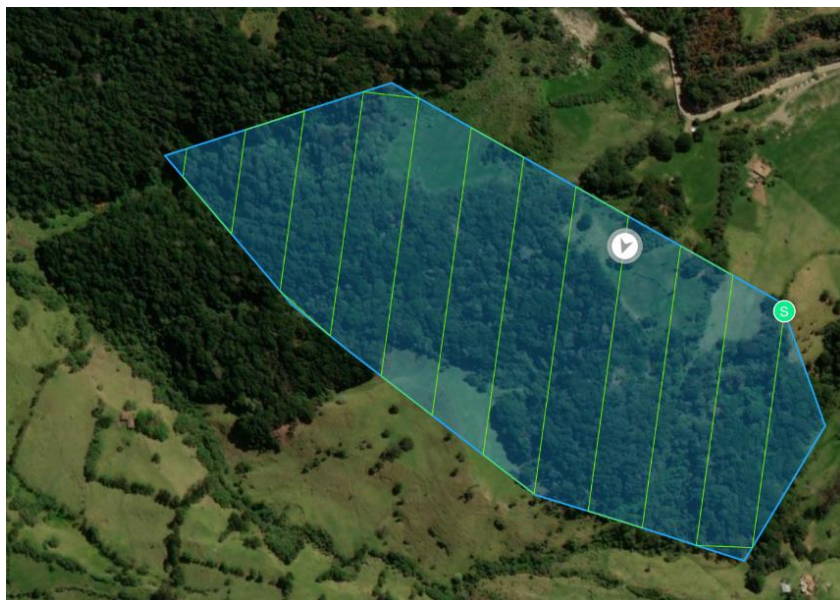
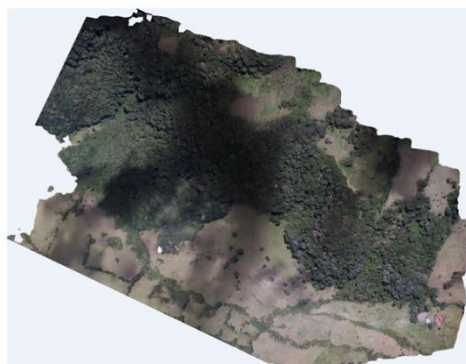
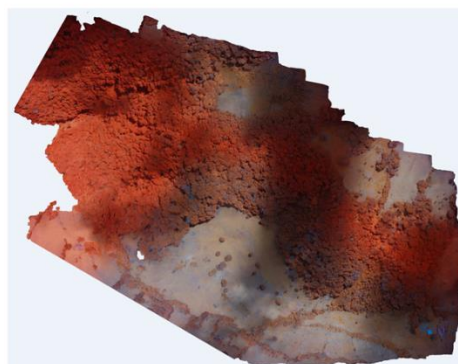


Figura 12.

Ortomosaicos de la zona de estudio capturada desde los 100 metros de altura, a la izquierda se encuentra el ortomosaico RGB y a la derecha el ortomosaico BGNIR.



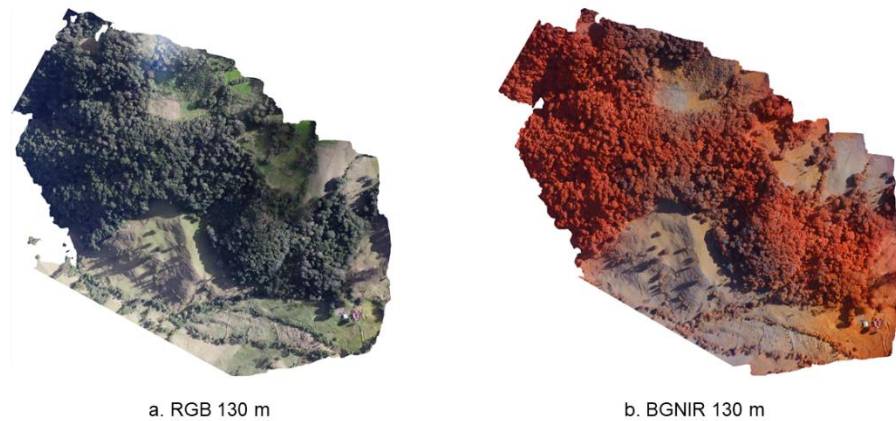
a. RGB 100 m



b. BGNIR 100 m

Figura 13.

Ortomosaicos de la zona de estudio capturada desde los 130 metros de altura, a la izquierda se encuentra el ortomosaico RGB y a la derecha el ortomosaico BGNIR.

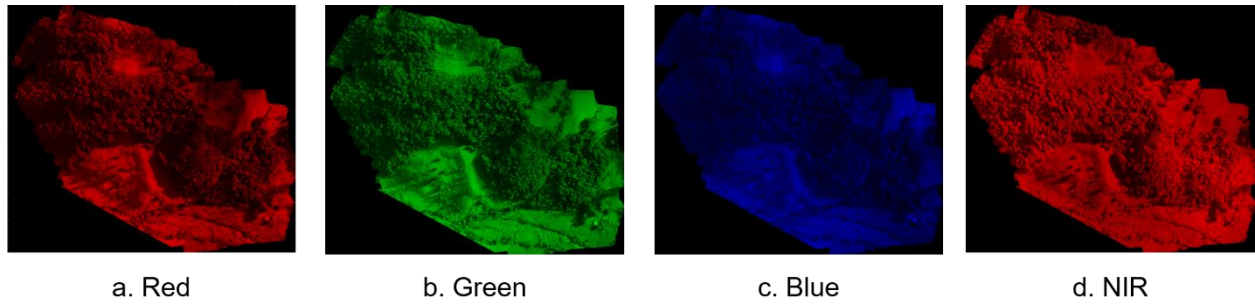


3.5 Preparación de los datos capturados con el UAV

El siguiente paso consistió en alinear los ortomosaicos, esto debido a que las fotografías capturadas no quedaron exactamente iguales pese a haber usado la misma misión de vuelo con ambas cámaras. Este proceso de alineación se conoce como registro de imágenes, y consiste en transformar diferentes conjuntos de datos a un solo plano coordinado (Zitova & Flusser, 2003). En general los procesos de registro consisten en cuatro pasos: detección de características, emparejamiento de características, estimación del modelo de transformación y transformación de la imagen (Zitova & Flusser, 2003). El registro de los ortomosaicos se realizó utilizando QGIS, donde se marcaron puntos de referencia visibles en ambos ortomosaicos para luego buscar que las dos imágenes quedaran alineadas en base a los puntos de referencia. En la Figura 14 se presentan los ortomosaicos en las diferentes bandas RGBNIR de la imagen a 130m posterior al proceso de registrado.

Figura 14.

Bandas espectrales del ortomosaico RGBNIR. a. Red, b. Green, c. Blue y d. NIR.



Con los ortomosaicos alienados, utilizando Python y la librería GDAL para la manipulación de imágenes georreferenciadas, se estimaron 15 índices de vegetación en base a las bandas espectrales R-G-B-NIR. Para estimar algunos de los índices de vegetación se requirió el calculo de dos parámetros: ρ_{RB} y η . Estos se pueden obtener mediante:

$$\rho_{RB} = R - \gamma (B - R), \quad (3)$$

$$\eta = \frac{2(NIR^2 - R^2) + 1,5NIR + 0,5R}{NIR + R + 0,5} \quad (4)$$

donde R, B y NIR son los valores de cada píxel en las bandas espectrales del rojo, azul e infrarrojo cercano respectivamente, y utilizando $\gamma = 1$, debido a que en la literatura explican que es el valor óptimo para esta constante (Rhyma, Norizah, Hamdan, Faridah-Hanum, & Zulfa, 2020). En la

Figura 15 se presentan los índices de vegetación utilizados junto a la ecuación para estimarlos.

Una vez calculados los índices de vegetación, se construyó un cubo de datos $\mathbf{D} \in R^{M \times N \times L}$, donde M y N son los píxeles que construyen el ancho y largo de la imagen, y L son las capas del cubo de datos, constituidas por índices de vegetación y bandas espectrales, donde $M = 10155$, $N = 9271$ y $L = 19$. El cubo de datos puede ser visto como una serie de matrices superpuestas una sobre otra en el mismo orden que se muestra en la

Figura 16.

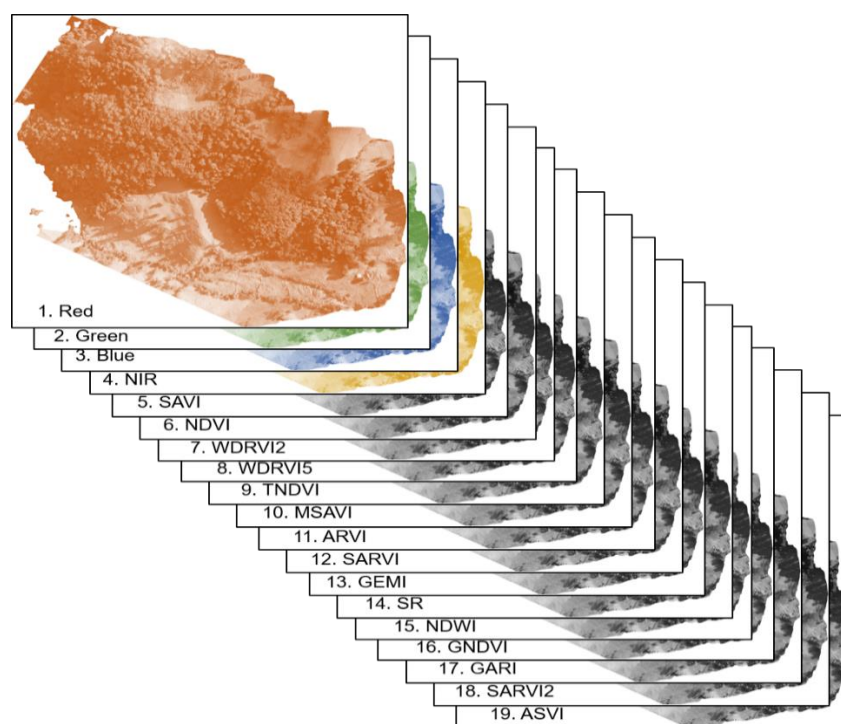
Figura 15.

Índices de vegetación estimados utilizando las bandas RGBNIR.

Índice de vegetación	Ecuación	Índice de vegetación	Ecuación
SAVI	$\frac{1,5(NIR-R)}{NIR+R+0,5}$	SARVI	$\frac{1,5(NIR-\rho_{RB})}{NIR+\rho_{RB}+0,5}$
NDVI	$\frac{NIR-R}{NIR+R}$	GEMI	$\frac{\eta(1-0,25\eta)-(R-0,125)}{1-R}$
WDRVI2	$\frac{0,2NIR-R}{0,2NIR+R}$	MSAVI	$\frac{(2NIR+1)^2-\sqrt{(2NIR+1)^2+8(NIR-R)}}{2}$
WDRVI5	$\frac{0,5NIR-R}{0,5NIR+R}$	GNDVI	$\frac{NIR-G}{NIR+G}$
TNDVI	$\sqrt{\frac{NIR-R}{NIR+R} + 0,5}$	GARI	$\frac{NIR-(G-(B-R))}{NIR+(G-(B-R))}$
SR	$\frac{NIR}{R}$	SARVI2	$\frac{2,5(NIR-R)}{1+NIR+6R-7,5B}$
ARVI	$\frac{NIR-\rho_{RB}}{NIR+\rho_{RB}}$	ASVI	$\frac{(2NIR+1)^2-\sqrt{(2NIR+1)^2+8(NIR-\rho_{RB})}}{2}$
NDWI	$\frac{G-NIR}{G+NIR}$		

Figura 16.

Cubo de datos generado para usar como fuente de datos para el proyecto.



4. Construcción del modelo matemático para la estimación de carbono

Con los datos de carbono obtenidos en la expedición de campo y el cubo de datos generado, la problemática ahora es ¿Cómo vincular esta información? La construcción de un modelo matemático debe permitir relacionar el cubo de datos con la información de la expedición de campo, sin embargo, hay que tener en cuenta factores como: ¿Que reglas debe cumplir el modelo matemático para resolver el problema?, ¿Que estrategias deben seguirse para realizar los entrenamientos del modelo? y ¿De qué manera se pueden organizar los datos para que el modelo entrenado entregue mejores resultados? El primer paso en este capítulo será explicar el trasfondo matemático utilizado para construir el modelo de estimación de carbono en base a un cubo matricial. Posteriormente, el procesamiento realizado a los datos de la expedición de campo para ser utilizados en el entrenamiento, usando un modelado de densidad de carbono por área. Luego, se explicará la metodología implementada para construir el modelo matemático, que son las redes neuronales. Finalmente, se describe la forma en que se organizaron los datos para realizar los entrenamientos para el modelo matemático.

4.1 Modelo propuesto para la estimación de carbono a partir de imágenes R-G-B-NIR

Para construir el modelo matemático que resuelva la problemática planteada de relacionar los datos obtenidos con el UAV y la información de la expedición de campo, hay que tener en cuenta la naturaleza de ambos tipos de datos. Los datos obtenidos con el UAV, después de ser arreglados tienen forma de un cubo matricial $\mathbf{D} \in R^{M \times N \times L}$, con donde $M = 10155$, $N = 9271$ y $L = 19$, donde M y N son los píxeles que construyen el ancho y largo de la imagen, y L es el número de bandas espectrales e índices de vegetación contenidos en el cubo de datos. Por otro lado, los datos de la expedición de campo se encuentran tabulados en un registro de datos, es

necesario realizar un procesamiento a estos datos de modo que sea posible obtener una matriz de carbono, que pueda utilizarse como base para el entrenamiento del modelo. Partiendo de que se tiene un cubo de datos \mathbf{D} y una matriz de carbono $\mathbf{C} \in R^{M \times N}$, el modelo matemático se plantea como una transformación T de tal forma que

$$T(\mathbf{D}) = \mathbf{C} \quad (5)$$

4.1.1 Preparación de los datos de la expedición de campo

Como se planteó anteriormente, los datos de la expedición de campo deben ser procesados de modo que pueda obtenerse una matriz de datos que representen el carbono de la zona estudio. Teniendo en cuenta que la información del cubo de datos está georreferenciada, al igual que los datos de la expedición de campo, es posible correlacionar la información espacial de cada conjunto de datos para realizar la estimación de la transformación T . Partiendo de que se cuenta con la ubicación, el diámetro del tallo y carbono para cada uno de los árboles medidos; el primer paso consiste en crear un mapa utilizando QGIS en el cual, a la coordenada de cada árbol le sea asignada el valor de carbono de ese árbol. Luego, usando los diámetros de cada tallo se crean áreas circulares en el mapa que representaran cada árbol y su respectiva área dentro de la zona de estudio. Sin embargo, las mediciones de la expedición condensan toda el área del árbol a un solo punto para obtener el carbono, por ende, se hace necesario distribuir esa medición dentro del área del árbol. La distribución se hace estimando una densidad de carbono para cada uno de los árboles de modo que $\rho_C = C_{\text{Árbol}}/A_{\text{Árbol}}$.

El siguiente paso consiste en convertir el mapa generado en QGIS a una matriz bidimensional de las mismas dimensiones M y N del cubo de datos, sin embargo, hay que tener en cuenta que la conversión del mapa a matriz no se hace perfectamente. La matriz de datos de carbono puede ser representada como una imagen en la cual cada elemento de la matriz equivale

a un píxel, donde los únicos píxeles donde hay información diferente de cero es donde se encuentran ubicados los árboles. Pero la conversión realizada no es perfecta, ya que los píxeles de la imagen no son capaces de recrear completamente la forma circular de los árboles. Teniendo en cuenta que la forma de los árboles en la imagen no es perfectamente circular y que solamente los píxeles donde están ubicados los árboles son diferentes a cero, se decide utilizar el área de cada píxel como base para calcular el carbono, de forma que al sumar todos los píxeles que representan un árbol, el resultado será el valor del carbono del árbol. La forma de calcular el carbono contenido en un píxel es multiplicando la densidad de carbono por área ρ_C por el área que representa cada píxel, que en este caso es de 81 cm^2 .

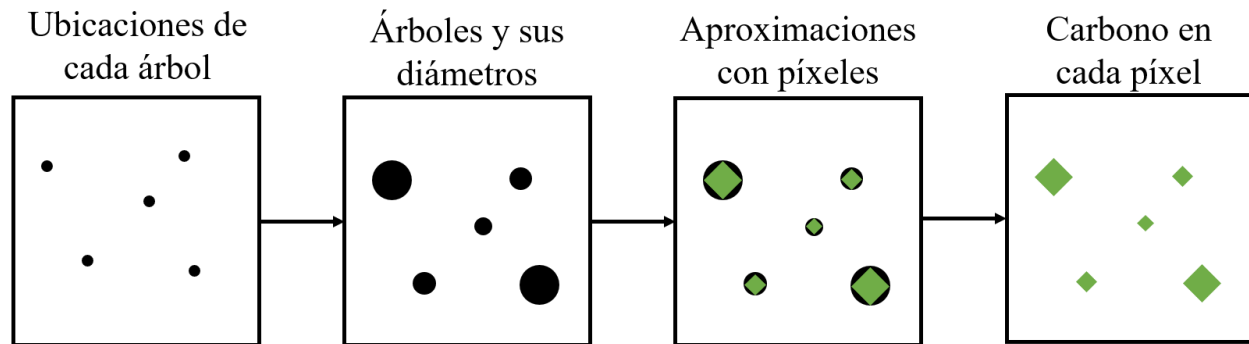
Teniendo en cuenta que el área de cada árbol no se representa perfectamente debido a la resolución espacial de cada píxel, se genera un error a la hora de generar la imagen de carbono con respecto a los valores recolectados en la expedición en campo. Teniendo en cuenta esto, se generan dos imágenes con información de carbono, una para C_{QH} y otra para $C_{General}$, donde los píxeles que representan la ubicación del tronco árboles tienen su respectivo valor de carbono y las zonas donde no existen troncos de árboles se asigna un valor de carbono igual a cero. Partiendo de que cada píxel tiene valor de carbono correspondiente al área espacial que representa, al hacer una sumatoria de todos los píxeles de la imagen de carbono, el resultado debe dar el valor del carbono total medido en la expedición en campo. Sin embargo, debido a las fuentes de error mencionadas se genera un error en el valor de carbono total medido en el área de estudio, con valores de 6,69% para la imagen de C_{QH} y de 6,61% para la imagen de $C_{General}$.

En la Figura 17 se presenta un diagrama mostrando un ejemplo del proceso para generar las imágenes de carbono donde primero se ubican los árboles dentro de la imagen, luego se dibujan

los diámetros de los árboles, posteriormente se aproximan los diámetros por medio de píxeles y finalmente se estima el carbono contenido en cada píxel.

Figura 17.

Proceso para generar las imágenes de carbono.



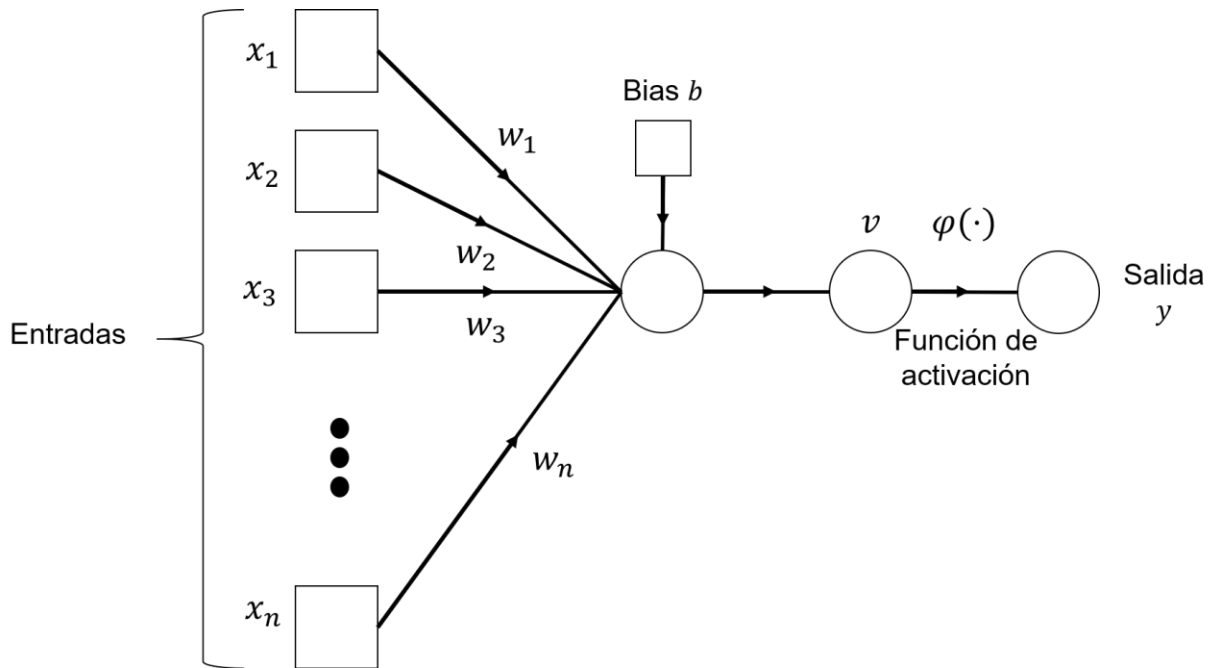
4.1.2 Metodología para relacionar el cubo de datos y la información de campo

Para solucionar el problema de encontrar la transformación T , se utilizan técnicas de inteligencia artificial (AI) (Russell & Norvig, 2016), las cuales pueden aprender a deducir una función f que relaciona los parámetros de entrada X y los valores de salida $f(X)$ (Alpaydin, 2020). Dentro del campo de AI se encuentra el aprendizaje de maquina (ML), el cual busca desarrollar una tarea específica gracias a entrenamientos previos. En ML es notoria la separación entre aprendizaje supervisado y aprendizaje no supervisado. El supervisado es aquel en el que se conoce que es lo que se quiere que la maquina aprenda, mientras que el no supervisado es aquel en el que se deja que la maquina encuentre por si misma que puede aprender (Russell & Norvig, 2016). Por la naturaleza del problema planteado en este proyecto, el tipo entrenamiento que se realiza es de tipo supervisado, donde los datos de la matriz de datos generada con los datos de campo \mathbf{C} son los datos a los que se quiere llegar, y los datos del cubo de datos \mathbf{D} son la entrada del modelo.

De los diferentes algoritmos que existen para aplicar ML, en este proyecto nos centramos en las redes neuronales, las cuales son una estructura inspirada en el sistema nervioso humano, donde se interconectan múltiples neuronas, que procesan información no lineal, compleja y de forma paralela (Goodfellow, Bengio, & Courville, 2016). Las redes neuronales se conforman interconectando múltiples perceptrones, los cuales son los elementos base de una red neuronal. Los perceptrones se modelan como una combinación lineal, a la cual a cada parámetro de entrada se le asigna un peso. Posterior a la combinación, se ubica de una función de activación que se encarga de añadir la no linealidad al modelo (Haykin, 2010). En la Figura 18 se observa un diagrama ilustrando un perceptrón, donde x_1, \dots, x_n son las variables del entrada al perceptrón, W_1, \dots, W_n son los pesos asignados a cada variable de entrada, b es el bias o la variable de compensación, v hace referencia a la suma de todas entradas multiplicadas por su respectivo peso y el bias, de forma que $v = x_1 * W_1 + \dots + x_n * W_n + b$, φ es la función de activación del perceptrón, la cual se encarga de añadir no linealidades al modelo, puede ser una función sigmoide, una tangente hiperbólica, entre otras.

Figura 18.

Diagrama de bloques de un perceptrón.



4.2 Organización de los datos

Teniendo en cuenta que la expedición de campo entrega diferentes formas de clasificar los árboles medidos, nos planteamos la siguiente pregunta: ¿Cómo podríamos organizar el entrenamiento de los datos, de modo que se pueda reducir la información a medir en la expedición de campo? En respuesta a esto, se plantean diferentes formas de organización de datos, basados en información como la ubicación o la parcela del árbol para entrenar el modelo de redes neuronales y evaluar qué porcentaje de datos permite obtener los mejores resultados. Las metodologías propuestas son:

- **Todas las parcelas:** Se aprovecha la separación por parcelas de la zona de estudio y se separan los datos de forma que se utilizan unas parcelas para entrenar y otras parcelas para verificar. Este tipo de organización busca evaluar el comportamiento de la red al utilizar parcelas enteras como bases de entrenamiento, de tal forma que

sea posible concluir que porcentaje de parcelas usadas como datos de entrenamiento entregan los mejores resultados. Con esta forma de organización de datos, se busca reducir la cantidad de parcelas a medir a la hora de realizar la expedición de campo.

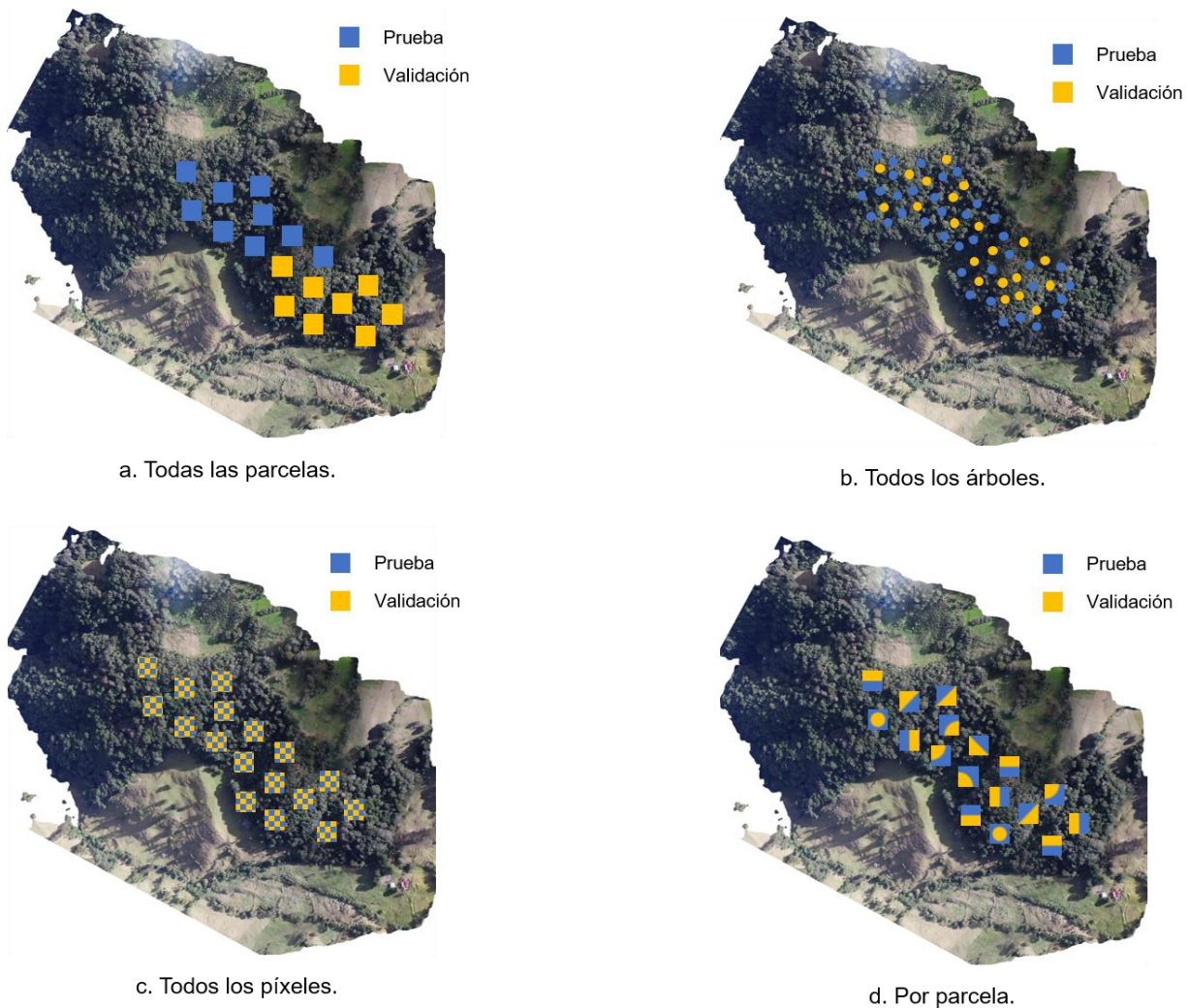
- **Todos los árboles:** Se utiliza la distinción de cada árbol medido para separarlos en un grupo de árboles de entrenamiento y otro grupo de árboles para evaluar, teniendo en cuenta que existen árboles de todas las parcelas en ambos grupos. Este tipo de organización busca evaluar si utilizar la información de un árbol completo como base de entrenamiento entrega resultados aceptables. Con esta forma de organización se busca reducir la cantidad de árboles a medir cuando se está realizando la expedición de campo.
- **Todos los píxeles:** Se utilizan píxeles aleatorios de la imagen de carbono para entrenar y para validar, sin embargo, se garantiza que haya píxeles de todas las parcelas en cada grupo. Esta organización es la forma clásica de selección de información de forma aleatoria.
- **Por parcela:** Se aprovecha la separación por parcelas y se crea un modelo de entrenamiento a cada parcela por separado, utilizando un porcentaje de los datos dentro de la parcela para entrenar y otro porcentaje para verificar. Con esta forma de organización de datos se busca evaluar si es mejor realizar un modelo de estimación para cada parcela por separado o para todo el bosque en general.

Es importante tener en cuenta que, pese a que se busca reducir la información medida en las expediciones de campo analizando estas formas de organización de datos, para poder validarlas se requiere de otra expedición de campo, la cual no puede ser ejecutada dentro de este proyecto y

por ende será planteado como un trabajo a futuro. En la Figura 19 se presenta una representación gráfica de cada una de las formas de organización para los datos de entrenamiento y validación.

Figura 19.

Organización de datos de entrenamiento y validación para las pruebas. a. Todas las parcelas, b. Todos los árboles, c. Todos los píxeles y d. Por parcela.



5. Entrenamiento y validación del modelo matemático para la estimación de carbono

En el capítulo 4 se planteó el modelo matemático que relaciona los datos de la expedición de campo con la información capturada con el UAV, junto con las formas de organizar los datos para el entrenamiento del modelo. Sin embargo, hasta el momento no se han realizado pruebas o entrenamientos que permitan evaluar el comportamiento del modelo para estimar carbono a partir de imágenes. La finalidad del paso de validación es evaluar que tan bien el modelo matemático permite estimar carbono, para ello se entrena la red neuronal que es un modelo propuesto, al igual que otros modelos encontrados en la literatura para estimar carbono. El entrenamiento para todos los modelos se realiza utilizando los métodos de organización de datos propuestos y comparando los resultados obtenidos en cada modelo. Este capítulo se estructura de la siguiente forma: primero se explicarán los parámetros utilizados para medir el porcentaje de error en las estimaciones realizadas en el modelo; luego se expone la estructura de red neuronal utilizada para construir el modelo matemático; posteriormente, se describen las metodologías de estimación de carbono utilizadas en el proyecto, donde se presentan modelos de la literatura con los cuales se compararán los modelos propuestos; después se muestran los resultados de las pruebas realizadas por cada una de las metodologías de estimación de carbono y finalmente un compendio de los mejores resultados.

5.1 Medición del error en los modelos

La validación del modelo matemático en este proyecto se realiza midiendo que tan erróneas son las estimaciones realizadas. Luego de obtener el valor del error, este se compara el obtenido con otros modelos de estimación encontrados en el estado del arte para encontrar el mejor en un determinado método de organización de datos. Teniendo esto en cuenta, el parámetro a utilizar

debe ser aplicable al tipo de datos con que se está trabajando, mostrando los resultados preferiblemente en porcentaje para una mayor interpretación. El parámetro utilizado para realizar la comparación es la raíz del error cuadrático medio en porcentaje (RMSE%), debido a que permite medir el error en las dos imágenes de carbono, la obtenida de la expedición de campo y la estimada con el modelo matemático, entregando sus resultados en porcentaje. El RMSE% se calcula utilizando la siguiente ecuación:

$$RMSE\% = \sqrt{\frac{mean((\hat{y} - y)^2)}{\max(y) - \min(y)}} \quad (6)$$

donde \hat{y} son los valores estimados por el modelo matemático, y y son los valores de carbono obtenidos de la expedición de campo usados para entrenar el modelo.

Para evaluar los modelos matemáticos en este proyecto, se hace necesario saber que tan precisa es la estimación realizada en cada modelo con cada forma de organización de datos, por esta razón, se crean 20 conjuntos de datos diferentes para cada una de las formas de organización de datos; de esta manera, un modelo matemático realiza 20 entrenamientos con diferentes conjuntos de datos de una misma forma de organización de datos, esperando que el valor de RMSE% no cambie drásticamente durante todos los entrenamientos. Por esta razón, se utiliza otro parámetro para evaluar los modelos matemáticos, el cual es el porcentaje de variabilidad del RMSE%, calculado como:

$$var\% = \frac{\Delta_{inf} + \Delta_{sup}}{RMSE} * 100 \quad (7)$$

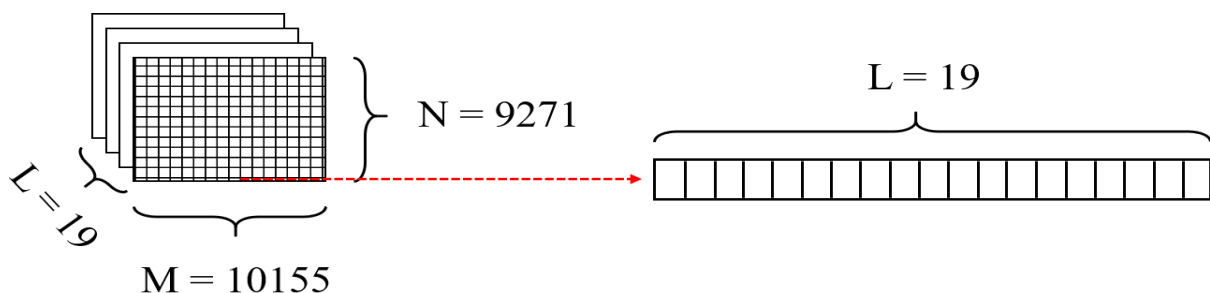
donde Δ_{inf} y Δ_{sup} son el límite inferior y superior del rango de valores obtenidos durante los entrenamientos de los 20 conjuntos de datos, y RMSE es el valor de RMSE medio en los entrenamientos de los 20 conjuntos de datos.

5.2 Red neuronal utilizada en el proyecto

Para construir la red neuronal que se encargará de resolver el problema planteado en el capítulo 4 hay que tener en cuenta como son los datos que ingresaran a la red, las arquitecturas utilizadas en otras investigaciones del estado del arte relacionadas a la estimación de carbono y que forma tendrá la información que saldrá de la red neuronal. Para definir la forma de los datos que entraran a la red neuronal, se parte del cubo de datos creado en el capítulo 3, el cual está conformado por 19 matrices de dimensiones $M = 10155$ y $N = 9271$, sin embargo, debido a que solamente se cuenta con una imagen de la zona de estudio, para operar la red neuronal se usan cada uno de los píxeles del cubo de datos, dando lugar a $M \times N$ vectores de 19 valores, donde cada uno de estos valores hacen referencia a las bandas espectrales R-G-B-NIR y a los índices de vegetación. De esta forma es válido decir que los datos que ingresaran a la red neuronal serán una serie de vectores, los cuales después de ser procesado entregarán un único valor de carbono por cada vector. En la Figura 20 se presenta la representación gráfica de cómo se seleccionan los datos de entrada a la red neuronal.

Figura 20.

Cada píxel del cubo de datos es un vector de 19 valores, los cuales se utilizan como entrada de la red neuronal.



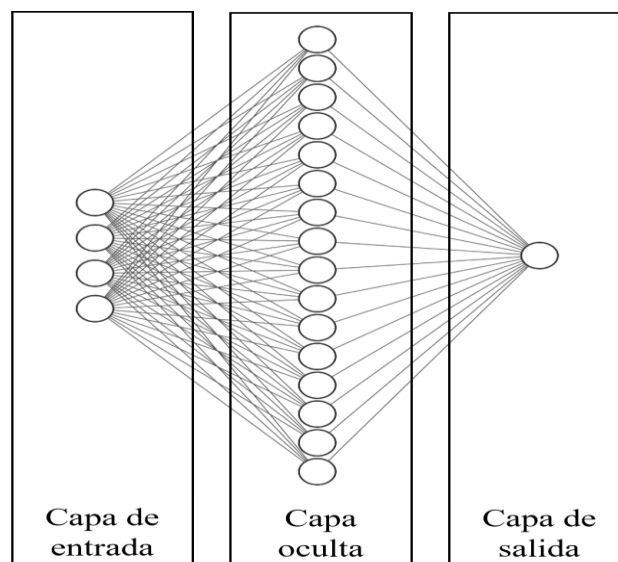
Pese a que es bueno tener grandes cantidades de datos para entrenar un modelo de redes neuronales, también es necesario encargarse de que los datos utilizados no entreguen información redundante o repetida, por esta razón, al cubo de datos se le somete a un análisis de componentes principales (PCA) (Wall, Rechtsteiner, & Rocha, 2003), el cual se encarga de reducir el número de elementos de cada vector que se ingresa a la red neuronal, dejando únicamente, aquellos elementos del vector que entregan mayor varianza en sus datos. De esta forma, la información del cubo de datos es normalizada y luego pasa por un análisis de PCA, donde se logra reducir los elementos del vector que ingresará a la red neuronal de 19 a 4 elementos, los cuales representan el 99,2566% de la varianza total de los datos. Desafortunadamente, debido a la naturaleza del proceso de PCA, es imposible discernir cuales fueron las características seleccionadas que contienen el porcentaje de la varianza de los datos.

Una vez definida la forma de los datos de entrada y salida de la red neuronal, el siguiente paso consiste en crear la arquitectura de la red que se utilizará en el proyecto, para ello se utilizaran como referencias arquitecturas encontradas en el estado del arte, las cuales constaban comúnmente de tres capas: una capa de entrada, una capa oculta y una capa de salida. Lo que diferencian las arquitecturas entre ellas es el número de perceptrones utilizados en cada una de las capas, donde en la capa de entrada se utiliza un perceptrón por cada característica presente en los datos de entrada, en la capa oculta se utiliza la cantidad de perceptrones requeridos que entreguen los resultados con menores niveles de error, finalmente en la capa de salida, se coloca un perceptrón por cada característica en los datos de salida. Con base a lo anterior, la capa de entrada y de salida está definida por los datos del proyecto, donde los datos de entrada luego del PCA cuentan con cuatro características por ende en la capa de entrada se cuenta con 4 perceptrones y los datos de salida son un único valor de carbono, por ende, la capa de salida es un único perceptrón. Para

encontrar los perceptrones de la capa oculta, se utilizaron los métodos de organización de píxeles con un 50% de datos para entrenar y un 50% para validar, luego se varió el número de perceptrones en la capa oculta de usando potencias de 2 como base, partiendo de 2 hasta 256, también se realizaron pruebas cambiando perceptrones de uno en uno, partiendo de 4 hasta 20. La razón por las cuales se utilizaron estos números para hacer las pruebas es debido a que son las cantidades que se encontraban más seguido para este tipo de análisis en el estado del arte. Después de realizar las pruebas se encontró que el número de perceptrones que entregaron los mejores resultados fueron 16 perceptrones en la capa oculta. De esta forma, se tiene que la red neuronal utilizada consta de una capa de entrada con cuatro perceptrones, una capa oculta con 16 perceptrones y una capa de salida con un perceptrón. En la Figura 21 se presenta un diagrama de la red neuronal con la cual se realizaron las pruebas en este proyecto.

Figura 21.

Red neuronal utilizada en el proyecto, la cual cuenta con tres capas, una capa de entrada con 4 perceptrones, una capa oculta con 16 perceptrones y una capa de salida con un perceptrón.



5.3 Metodologías de estimación de carbono implementadas

En la literatura para realizar estimaciones de carbono basadas en imágenes espectrales, pueden ser utilizadas las regresiones lineales cuando se cuenta con datos de una expedición de campo que cubren parte considerable de la zona de estudio, en este proyecto al tenerse una gran zona del bosque estudiada, es posible usar la regresión lineal sin inconvenientes. Por otro lado, en la literatura el índice de vegetación normalizado (NDVI) es estudiado constantemente debido a su fuerte relación con la biomasa y otras características de salud en las plantas (Carlson & Ripley, 1997). El índice NDVI es el más utilizado en la literatura para estudios de estimación de carbono y los autores lo usan constantemente como eje principal en sus estudios y análisis (Gao, y otros, 2020). Por esta razón, teniendo en cuenta la fuerte correlación del índice de vegetación NDVI con la biomasa, la cual a su vez está directamente relacionada con el carbono en los árboles, es posible realizar un modelo estadístico que permita relacionar el índice de vegetación con el carbono (Shi-Long, Fangjing-Yun, & Yu, 2004). También hay que tener en cuenta que una fuente importante de información en la literatura son las imágenes satelitales, que, debido a las condiciones presentes en este proyecto, se espera que entreguen resultados con altos niveles de error. En base a lo anterior, en este trabajo se implementan cinco metodologías de estimación de carbono, las cuales incluyen métodos de la literatura y la red neuronal propuesta, teniendo en cuenta que todas las metodologías se entrenan y validan usando las imágenes de carbono $C_{General}$ y C_{QH} . Las metodologías implementadas son:

- Un entrenamiento empleando la red neuronal propuesta, usando como datos de entrada la salida del PCA del cubo de datos.
- Una regresión lineal usando el índice de vegetación NDVI obtenido por medio de los datos capturados con el UAV.

- Un entrenamiento empleando redes neuronales utilizando el índice de vegetación NDVI obtenido por medio de los datos capturados con el UAV.
- Un entrenamiento utilizando redes neuronales, aplicando el modelo propuesto con un cubo de datos generado por medio de una imagen satelital capturada desde Landsat 8, donde se usaron los píxeles de la imagen que representaban la zona de estudio, sin embargo, la resolución espacial de estas imágenes es de $30\text{ m}^2/\text{px}$, lo cual hace que la zona de estudio sea representada únicamente con 28 píxeles. La imagen satelital utilizada, fue capturada en la fecha más cercana a la expedición de campo realizada en la zona de estudio. En este entrenamiento solo aplica la organización de los datos de todos los píxeles debido a que la resolución espacial de la imagen no permite aplicar otra metodología.
- Una regresión lineal usando el índice de vegetación NDVI obtenido por medio de una imagen satelital de Landsat 8. En este entrenamiento debido a que usa una imagen satelital de Landsat 8, se cuenta con las mismas restricciones del entrenamiento utilizando redes neuronales sobre la imagen Landsat.

5.4 Resultados del entrenamiento de las metodologías implementadas

Con los modelos de estimación de carbono definidos y los datos preparados para los entrenamientos, el siguiente paso consiste en entrenar los modelos y evaluar su desempeño estimando el carbono de las imágenes obtenidas por la expedición de campo de carbono $C_{General}$ y C_{QH} . Para realizar los entrenamientos se utilizaron las formas de organización de datos planteadas en el capítulo 4, donde en cada una se crearon dos grupos de datos, uno para entrenar el modelo y otro para validar el modelo; durante los entrenamientos el porcentaje de datos utilizados para

entrenar los modelos se varió entre 10% y 90%, generando 20 conjuntos de datos diferentes para cada porcentaje. Es importante aclarar que, para cada forma de organización de datos, el porcentaje de datos usado para entrenar se hace en función a la agrupación de la forma de organización; por ejemplo, en la metodología de todas las parcelas se usa un porcentaje de parcelas para entrenar y el restante para validar, en todos los árboles se usa un porcentaje del total de los árboles para entrenar y el restante para validar.

Los resultados de cada entrenamiento se muestran como gráficas de RMSE%, donde para la organización de datos: todas las parcelas, todos los píxeles y todos los árboles; el eje x corresponde al porcentaje de datos utilizados para probar el modelo, y en cada punto se presenta el valor medio de RMSE% obtenido, junto a la variabilidad de resultados obtenidos en los 20 entrenamientos de los conjuntos de dato. También se resalta en cada gráfica el que se considera como el mejor resultado, mostrando el valor de RMSE% y un Δ indicando los límites del rango de valores obtenidos en los 20 entrenamientos. Por otro lado, para la organización de datos por parcela, el eje x corresponde al número que identifica cada parcela (números entre 1 y 28), y el valor de RMSE% presentado, hace referencia al mejor resultado obtenido en esa parcela, teniendo en cuenta el valor medio de RMSE%. De la misma forma cada una de las gráficas de RMSE% presentan dos curvas: una curva de color azul llamada *Train* la cual muestra los valores de RMSE% al validar el modelo con los datos usados para entrenar el modelo matemático y una curva de color rojo llamada *Test* que muestra los valores de RMSE% al validar el modelo con los datos de validación que no fueron usados para entrenar el modelo.

5.4.1 Entrenamiento con la red neuronal propuesta usando el cubo de datos después el análisis de PCA

En la Figura 22 y Figura 23 se presentan los resultados de RMSE% con las imágenes de carbono C_{QH} y $C_{General}$ utilizando la red neuronal propuesta, posteriormente en la Tabla 2 se presenta el valor de RMSE% promedio de todos los porcentajes de datos utilizados en cada método de organización, junto con el porcentaje de variabilidad promedio de todos los porcentajes de datos en los resultados obtenidos. De estos resultados lo primero que resalta es que el valor de RMSE% más bajo, el cual lo entrega el método de organización de todos los píxeles, usando la imagen de $C_{General}$ con el 70% de los datos para entrenar el modelo. Posteriormente, al comparar cada método de organización de datos se tiene que: el método de todos los píxeles fue el que entregó los mejores resultados en ambas imágenes de carbono, no solo por el valor promedio de RMSE%, también porque este método presentó la variabilidad en los resultados más baja entre todos los métodos de organización, lo cual hace que sus resultados sean más confiables en comparación a los obtenidos con otras metodologías de organización de datos. Este método de organización presenta errores más bajos con un punto de error mínimo entre el 60 y 70% de los datos. El método de todos los árboles es el segundo en entregar errores bajos, ya que entrega variabilidades en los resultados menores a todas las parcelas y por parcela, sin embargo, este método presenta los mejores resultados usando bajos porcentajes de datos para entrenar, ya que al aumentar los datos el valor de RMSE% aumenta; el método de todas las parcelas presentó variabilidades elevadas al usar el 10 y 90% de los datos en la imagen de C_{QH} , lo cual hace que parezca que tiene una muy variabilidad elevada, sin embargo al calcular el promedio dejando por fuera los resultados de estos dos porcentajes de datos, el porcentaje de variabilidad se reduce a 21,0212%, el cual es mucho más bajo que el entregado por la imagen $C_{General}$. Por otro lado, la organización de datos por

parcela fue la que entregó los resultados con errores más grandes, también entregan resultados con variabilidades elevadas lo cual hace que estos resultados no sean confiables para realizar una estimación. Finalmente, de estos resultados se puede concluir que la imagen de C_{QH} es la que mejor puede ser modelada con la estimación de carbono propuesta, debido a que las variabilidades de los resultados con esta imagen son menores a las de la imagen de $C_{General}$, y al tener valores de RMSE% cercanos, la variabilidad se convierte en el factor decisivo.

Figura 22.

Gráficas del RMSE% obtenido al estimar el C_{QH} utilizando la red neuronal propuesta con las diferentes organizaciones de datos de entrenamiento.

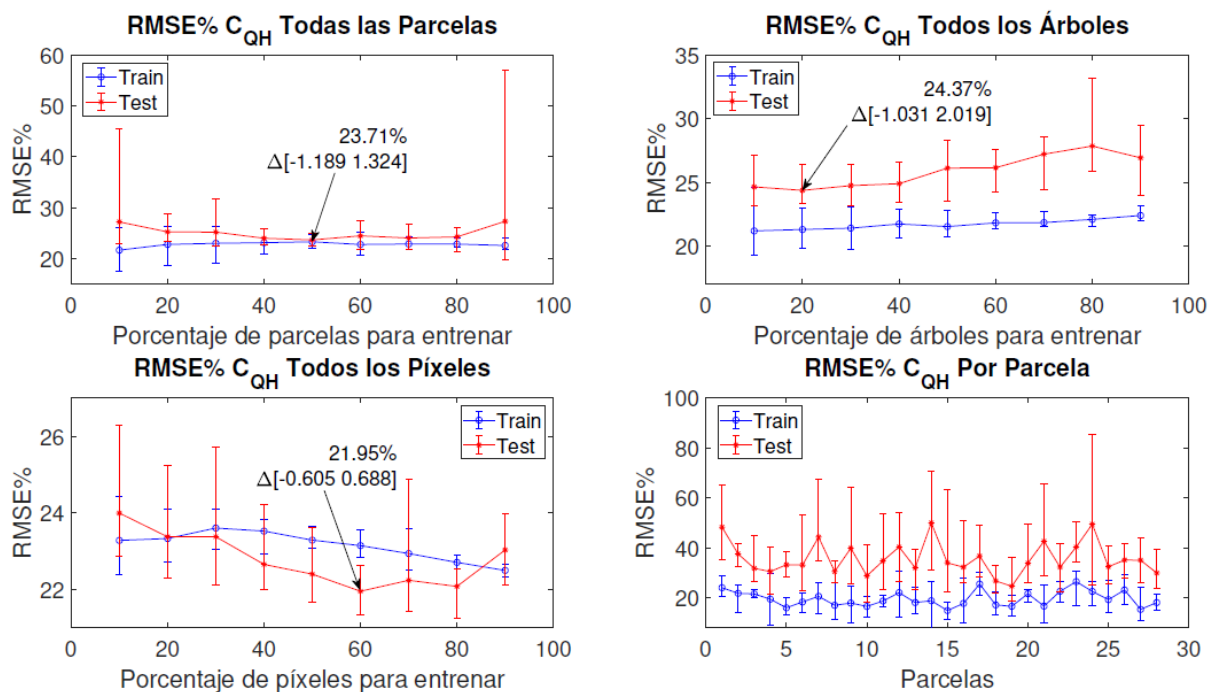
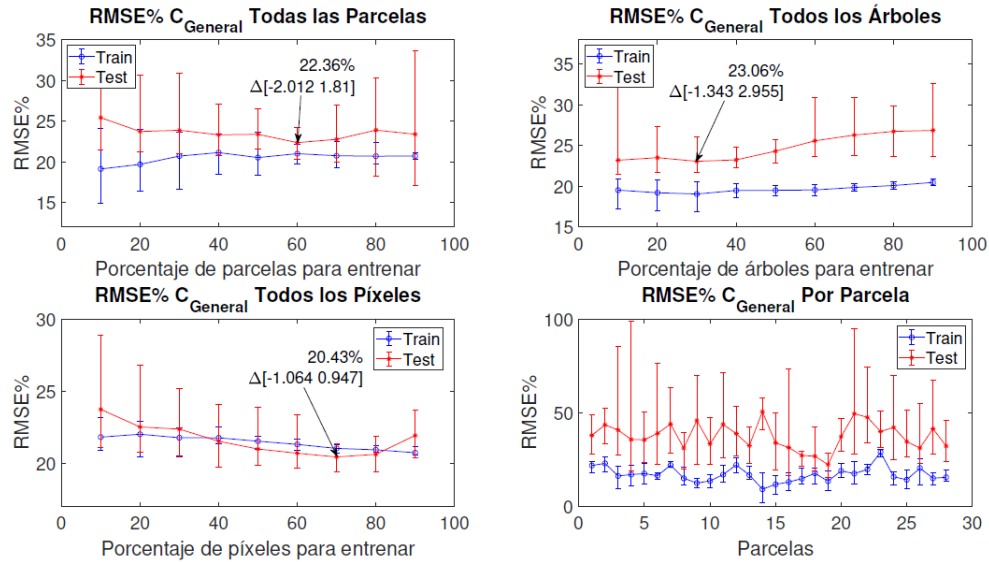


Figura 23.

Gráficas del RMSE% obtenido al estimar el $C_{General}$ utilizando la red neuronal propuesta con diferentes organizaciones de datos de entrenamiento.

**Tabla 2.**

Resumen de resultados con el método propuesto de redes neuronales para las imágenes de carbono C_{QH} y $C_{General}$.

Organización de	C_{QH}		$C_{General}$	
	RMSE%	Variabilidad	RMSE%	Variabilidad
datos	Promedio	Promedio	Promedio	Promedio
Todas las parcelas	25,0878	42,3956%	23,5712	37,5449%

Todos los árboles	25,8776	16,5602%	24,7526	26,4564%
Todos los píxeles	<u>22,7848</u>	<u>10,75%</u>	<u>21,6397</u>	<u>19,7435%</u>
Por parcela	35,6659	66,9347%	37,4936	89,2409%

5.4.2 Entrenamiento usando regresión lineal con el índice de vegetación NDVI

En

Figura 24 y Figura 25 se presentan los resultados de RMSE% con las imágenes de carbono C_{QH} y $C_{General}$ utilizando una regresión lineal con el índice de vegetación NDVI, también en la Tabla 3 se presentan los valores de RMSE% y variabilidad promedio de los entrenamientos con cada método de organización de datos. En estos resultados el RMSE% más bajo lo entrega el método de organización de datos de todos los píxeles usando el 50% de los datos para entrenar el modelo con la imagen de $C_{General}$. Comparando los resultados obtenidos, se tiene que el método todos lo árboles es el segundo mejor con respecto a los errores, ya que entrega variabilidades en los resultados menores a todas las parcelas y por parcela, sin embargo, esté método presenta los mejores resultados usando bajos porcentajes de datos para entrenar. El método de todas las parcelas presentó variabilidades elevadas usando el 90% de los datos en la imagen de C_{QH} , lo cual hace que parezca que tiene una muy variabilidad elevada, sin embargo, al calcular el promedio dejando por fuera los resultados de este porcentaje de datos, el RMSE% cambia a 23,7941% y porcentaje de variabilidad se reduce a 20,2530%. Por otro lado, la organización de datos por parcela fue la que entregó los resultados con errores más grandes, también entregan resultados con variabilidades elevadas lo cual hace que estos resultados no sean confiables para realizar una estimación.

Finalmente, de estos resultados se puede concluir que la imagen de $C_{General}$ es la que mejor puede ser modelada con la estimación de carbono usando una regresión lineal con el índice de vegetación NDVI, esto debido a que sus valores de RMSE% son más bajos que los de la imagen de C_{QH} y pese a que las variabilidades son mayores en $C_{General}$, al ser cercanas a las de C_{QH} , el valor del RMSE% se convierte en el factor decisivo.

Figura 24.

Gráficas del RMSE% obtenido al estimar el C_{QH} utilizando una regresión lineal con el índice de vegetación NDVI.

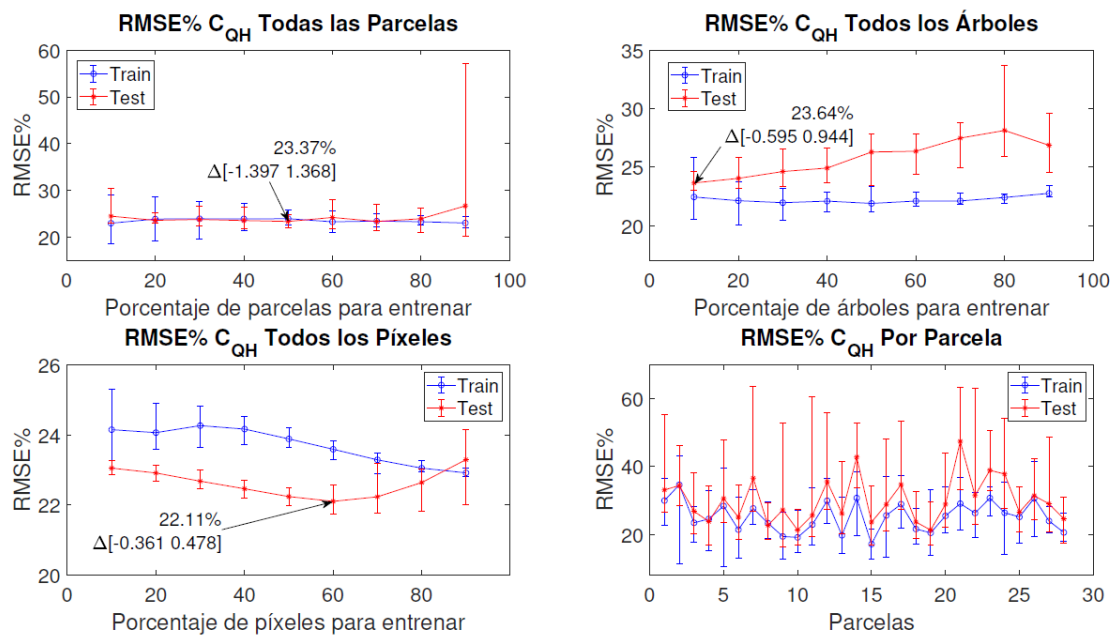
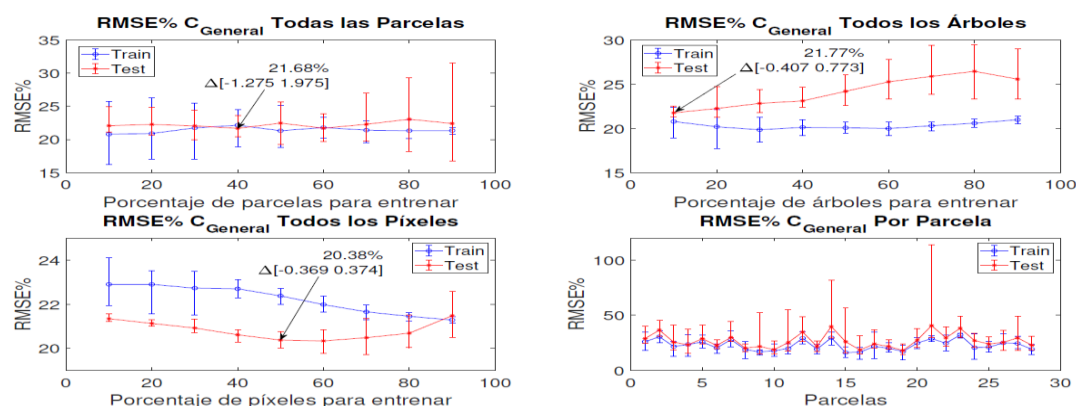


Figura 25.

Gráficas del RMSE% obtenido al estimar el $C_{General}$ utilizando una regresión lineal con el índice de vegetación NDVI.

**Tabla 3.**

Resumen de resultados con el método de regresión lineal con NDVI para las imágenes de carbono

C_{QH} y $C_{General}$.

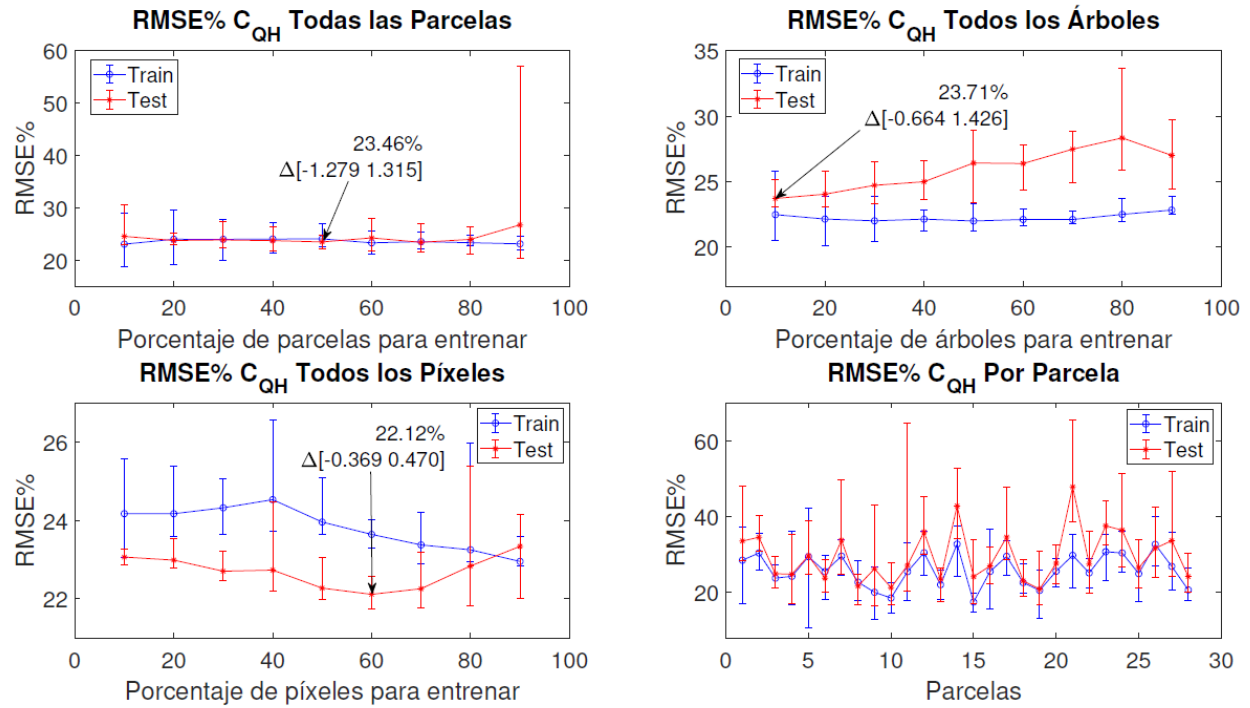
Organización de datos	C_{QH}		$C_{General}$	
	RMSE%	Variabilidad	RMSE%	Variabilidad
	Promedio	Promedio	Promedio	Promedio
Todas las parcelas	24,1157	34,7218%	22,2240	29,6556%
Todos los árboles	25,8104	15,0543%	24,1590	16,0007%
Todos los píxeles	<u>22,6272</u>	<u>3,9460%</u>	<u>20,8192</u>	<u>4,7203%</u>
Por parcela	30,0260	75,2801%	26,8103	81,0707%

5.4.3 Entrenamiento con la red neuronal propuesta usando el índice de vegetación NDVI

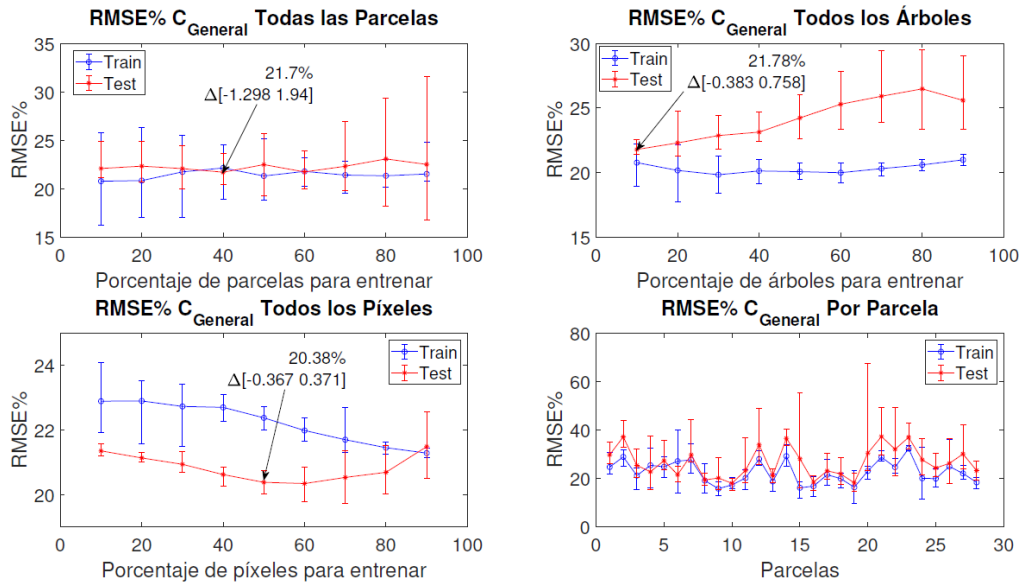
En las Figura 26 y Figura 27 se presentan los resultados de RMSE% con las imágenes de carbono C_{QH} y $C_{General}$ utilizando la red neuronal propuesta, en la Tabla 4 se presenta el valor de RMSE% promedio junto con la variabilidad promedio en los resultados obtenidos en los entrenamientos. Estos resultados son similares a los obtenidos por medio de la regresión lineal usando el índice de vegetación NDVI, sin embargo, la diferencia se encuentra en el método de organización de datos por parcela, ya que disminuye en la variabilidad de los resultados obtenidos. Al igual que en la regresión lineal con NDVI, se observa que en el método de todas las parcelas en la imagen de C_{QH} al usar el 90% de los datos hay un aumento considerable de la variabilidad, al remover este valor para estudiar los otros que son más cercanos se tiene que el RMSE% promedio es de 23,8572% y la variabilidad promedio es de 20,4311%. De estos resultados se puede concluir que la imagen de $C_{General}$ es la que mejor puede ser modelada con la estimación de carbono usando una red neuronal con el índice de vegetación NDVI, esto debido a que sus valores de RMSE% son más bajos que los de la imagen de C_{QH} , y pese a que las variabilidades son mayores en $C_{General}$ al ser cercanas a las de C_{QH} , el valor del RMSE% se convierte en el factor decisivo.

Figura 26.

Gráficas del RMSE% obtenido al estimar el C_{QH} utilizando la red neuronal propuesta con el índice de vegetación NDVI utilizando las diferentes organizaciones de datos de entrenamiento propuesta.

**Figura 27.**

Gráficas del RMSE% obtenido al estimar el $C_{General}$ utilizando la red neuronal propuesta con el índice de vegetación NDVI utilizando las diferentes organizaciones de datos de entrenamiento propuesta.



5.4.4 Entrenamiento con la red neuronal propuesta usando imágenes satelitales

La Figura 28 presenta los resultados de RMSE% con las imágenes de carbono C_{QH} y $C_{General}$ utilizando la red neuronal propuesta sobre un cubo de datos generados con una imagen satelital. En la Tabla 5 se presenta el valor medio de RMSE% junto con el porcentaje de variabilidad en los resultados obtenidos en los entrenamientos. De estos resultados, lo primero que resalta es que los porcentajes de error en esta metodología fueron elevados, entregando errores medios cercanos al 100%, esto se debe a que para generar estos modelos se entrenó una red neuronal con un conjunto de 28 datos, lo cual es poco para entrenar una red neuronal como la propuesta. Otra característica de este tipo de estimación de carbono fue que al usar el 10 y el 90% de los píxeles para entrenar la variabilidad de los resultados aumentaba en gran medida, por esta razón se decidió no presentarlos en la Figura 28, debido a que llegaban a escalas mayores al 1000% de error. En este entrenamiento la imagen que mejor se acomoda a las estimaciones es la

de C_{QH} , sin embargo, hay que tener en cuenta que errores cercanos al 100% con variabilidades de más de 60% no son confiables.

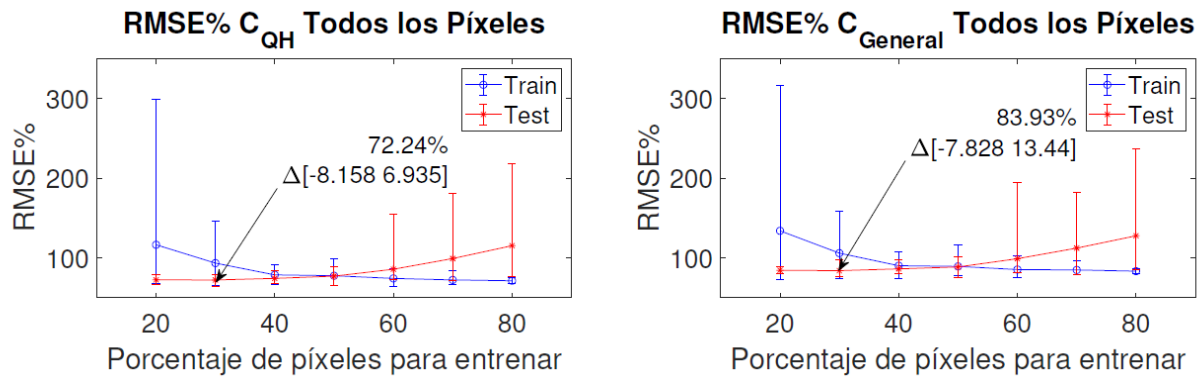
Tabla 4.

Resumen de resultados con el método propuesto de redes neuronales para las imágenes de carbono C_{QH} y $C_{General}$ usando la red neuronal propuesta con el índice de vegetación NDVI.

Organización de datos	C_{QH}		$C_{General}$	
	RMSE%	Variabilidad	RMSE%	Variabilidad
	Promedio	Promedio	Promedio	Promedio
Todas las parcelas	24,1762	34,8375%	22,2538	29,4922%
Todos los árboles	25,8362	15,8617%	24,1719	15,9919%
Todos los píxeles	<u>22,7038</u>	<u>6,4867%</u>	<u>20,8287</u>	<u>4,7181%</u>
Por parcela	29,5476	56,0611%	26,5643	55,5312%

Figura 28.

Gráficas del RMSE% obtenido al estimar el C_{QH} y $C_{General}$ utilizando la red neuronal propuesta con un cubo de datos generado con una imagen satelital.

**Tabla 5.**

Resumen de resultados con el método propuesto de redes neuronales para las imágenes de carbono C_{QH} y $C_{General}$ usando un cubo de datos adquirido con una imagen satelital.

Organización de datos	C_{QH}		$C_{General}$	
	RMSE%	Variabilidad	RMSE%	Variabilidad
	Promedio	Promedio	Promedio	Promedio
Todos los píxeles	<u>85,1311</u>	<u>67,3411%</u>	97,4020	64,8487%

5.4.5 Entrenamiento usando regresión lineal con el índice de vegetación NDVI de las imágenes satelitales

La **Figura 29** presenta los resultados de RMSE% con las imágenes de carbono C_{QH} y $C_{General}$ utilizando una regresión lineal sobre el índice de vegetación NDVI de los datos satelitales. En el cuadro Tabla 6 se presenta el valor medio de RMSE% junto con el porcentaje de variabilidad en los resultados obtenidos en los entrenamientos. De estos resultados, lo primero que resalta es que los valores de RMSE% medio son mucho más pequeños en comparación a los obtenidos usando redes neuronales con datos satelitales, sin embargo, estos tienen variabilidades elevadas, siendo estas superiores al 100%. Otra característica de este tipo de estimación de carbono fue que al usar el 10 y el 90% de los píxeles para entrenar la variabilidad de los resultados aumentaba en gran medida, por esta razón se decidió no presentarlos en la Figura 29, debido a que llegaban a escalas mayores al 800% de error. En este entrenamiento ambas imágenes de carbono tienen resultados similares, ya que tienen una media cercana y variabilidades muy grandes, lo cual hace que este tipo de estimación no sea confiable.

Figura 29.

Gráficas del RMSE% obtenido al estimar el C_{QH} y $C_{General}$ utilizando una regresión lineal del índice de vegetación NDVI con un cubo de datos generado con una imagen satelital.

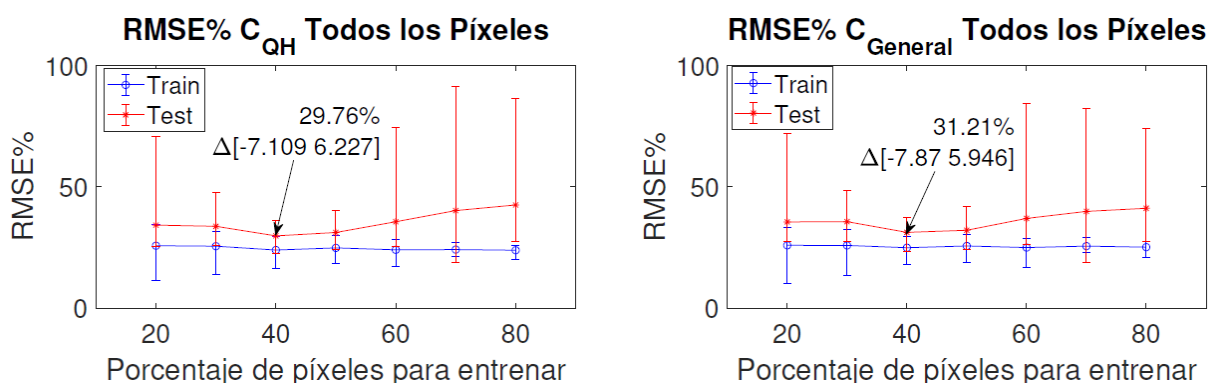


Tabla 6.

Resumen del promedio de resultados con el método de regresión lineal con NDVI para las imágenes de carbono C_{QH} y $C_{General}$ usando un cubo de datos adquirido con una imagen satelital.

Organización de datos	C_{QH}		$C_{General}$	
	RMSE%	Variabilidad	RMSE%	Variabilidad
	Promedio	Promedio	Promedio	Promedio
Todos los píxeles	35,3057	112,1112%	<u>36,0577</u>	<u>105,3272%</u>

5.5 Resumen de resultados

La idea principal de esta sección es presentar los resultados de RMSE% obtenidos con cada prueba para estimar carbono en función de las metodologías de organización de datos, con el fin de comparar que tipo de estimación se comporta mejor en cada tipo de organización de resultados.

5.5.1 Todas las parcelas

En la Tabla 7 se presenta el promedio de los resultados de RMSE% y variabilidad de todos los porcentajes de datos a las estimaciones de carbono sobre las cuales se pudo aplicar la metodología de organización de datos de todas las parcelas, y en el cuadro Tabla 8 se presentan los mejores resultados en cada una de las estimaciones de carbono. De estos resultados se pueden observar varios comportamientos, el primero fue que la imagen de carbono C_{QH} fue la que mejor se acoplo a las estimaciones realizadas con este tipo de organización de datos de entrenamiento. Por otro lado, las tres formas de estimación de carbono presentaron resultados muy cercanos, tanto en el mejor caso de estimación como en los resultados promedios, sin embargo, debido a que la

regresión lineal con NDVI es la que en promedio entrega las variabilidades más bajas se selecciona esta estimación de carbono como la mejor para este tipo de organización de datos usando la imagen de carbono C_{QH} .

Tabla 7.

Resumen del promedio de todos los resultados usando la organización de datos de todas las parcelas.

Tipo de Estimación	C_{QH}		$C_{General}$	
	RMSE%	Variabilidad	RMSE%	Variabilidad
	Promedio	Promedio	Promedio	Promedio
Red neuronal UAV	24,0561	21,0212%	23,5712	37,5449%
Regresión con NDVI UAV	<u>23,7941</u>	<u>20,2530%</u>	22,2240	29,6556%
Red neuronal Con NDVI UAV	23,8572	20,4311%	22,2538	29,4922%

5.5.2 Todos los árboles

En la Tabla 9 se presentan los promedios de los resultados de las estimaciones de carbono sobre las cuales se pudo aplicar la metodología de organización de datos de todos los árboles, y en la

Tabla **10** están los mejores resultados en cada una de las estimaciones de carbono. De estos resultados se observa que la imagen de $C_{General}$ fue la que entregó mejores resultados en esta metodología de organización de datos. Esta metodología de organización de datos entrega los

mejores resultados usando bajos porcentajes de datos, ya que en la mayoría de las pruebas entregó los mejores resultados usando porcentajes entre el 10 y el 30% de los datos de entrenamiento. Finalmente, utilizando la estimación de carbono por medio de una red neuronal usando el índice de vegetación NDVI es posible obtener los mejores resultados, sin embargo, los resultados son cercanos a los de la regresión lineal usando el índice de vegetación NDVI.

Tabla 8.

Resumen de los mejores resultados de RMSE% usando la organización de datos de todas las parcelas.

Tipo de Estimación	C_{QH}			$C_{General}$		
	Mejor RMSE%	Var	Datos usados %	Mejor RMSE%	Var	Datos usados %
Red neuronal UAV	23,71	10,5989%	50	22,36	17,093%	60
Regresión con NDVI UAV	23,37	11,8314%	50	21,68	14,9908%	40
Red neuronal Con NDVI UAV	23,46	11,0571%	50	21,7	14,9217%	40

Tabla 9.

Resumen del promedio de todos los resultados usando la organización de datos de todos los árboles.

Tipo de Estimación	C_{QH}		$C_{General}$	
	RMSE%	Variabilidad	RMSE%	Variabilidad
	Promedio	Promedio	Promedio	Promedio
Red neuronal UAV	25,0878	16,5602%	24,7526	26,4564%
Regresión con NDVI UAV	25,8104	15,0543%	24,1590	16,0007%
Red neuronal Con NDVI UAV	25,8362	15,8617%	<u>24,1719</u>	<u>15,9919%</u>

Tabla 10.

Resumen de los mejores resultados de RMSE% usando la organización de datos de todos los árboles.

Tipo de Estimación	C_{QH}			$C_{General}$		
	Mejor	Var	Datos	Mejor	Var	Datos
	RMSE%		usados %	RMSE%		usados %
Red neuronal UAV	24,37	12,5154%	20	23,06	19,6683%	30
Regresión con	23,64	6,5102%	10	21,77	5,4203%	10

NDVI UAV

Red neuronal	23,71	8,8148%	10	<u>21,78</u>	<u>5,2388%</u>	<u>10</u>
Con NDVI UAV						

5.2.3 Todos los píxeles

Este método de organización de datos permite comparar todas las estimaciones de carbono realizadas, en la Tabla 11 se presentan los promedios de los resultados de todas las estimaciones de carbono usando la organización de datos de todos los píxeles, y en la Tabla 12 se presentan los mejores resultados en cada una de las estimaciones de carbono. De estos resultados sobresalen los resultados de las estimaciones usando imágenes satelitales, ya que entregan valores de RMSE% y variabilidades más altas en comparación a las estimaciones de carbono usando imágenes capturadas desde UAV, lo cual era de esperarse, debido a que la resolución espacial que tienen las imágenes satelitales Landsat 8, hace que la zona de estudio sea representada únicamente con 28 píxeles, haciendo que la estimación de carbono propuesta que se basa en redes neuronales no tenga una cantidad de datos apropiada para construir el modelo, y en el caso de la regresión lineal usando NDVI los resultados fueron mejores que los del método de estimación propuesto sobre el cubo de datos satelital, sin embargo, sus altos valores de variabilidad hacen que estos resultados no sean confiables. En esta metodología de organización de datos, las estimaciones de carbono que mejor entregan resultados son la regresión lineal con NDVI y la red neuronal con NDVI utilizando la imagen de carbono $C_{General}$, los resultados de estas dos estimaciones son muy cercanos, y entregan un mejor resultado usando el 50% de los datos de entrenamiento. En general, la metodología de organización de datos de todos los píxeles fue la que entregó los valores de RMSE% y variabilidades más bajas de todas las pruebas.

Tabla 11.

Resumen del promedio de todos los resultados usando la organización de datos de todos los píxeles.

Tipo de Estimación	C_{QH}		$C_{General}$	
	RMSE%	Variabilidad	RMSE%	Variabilidad
	Promedio	Promedio	Promedio	Promedio
Red neuronal UAV	22,7848	10,75%	21,6397	19,7435%
Regresión con NDVI UAV	22,6272	3,9460%	<u>20,8192</u>	<u>4,7203%</u>
Red neuronal Con NDVI UAV	22,7038	6,4867%	<u>20,8287</u>	<u>4,7181%</u>
Red neuronal satelital	85,1311	67,3411%	97,4020	64,8487%
Regresión con NDVI satelital	35,3057	112,1112%	36,0577	105,3272%

Tabla 12.

Resumen de los mejores resultados de RMSE% usando la organización de datos de todos los píxeles.

Tipo de Estimación	C_{QH}			$C_{General}$		
	Mejor RMSE%	Var	Datos usados %	Mejor RMSE%	Var	Datos usados %
Red neuronal UAV	21,95	5,8907%	60	20,43	9,8434%	70
Regresión con NDVI UAV	22,11	3,7947%	60	<u>20,38</u>	<u>3,6457%</u>	<u>50</u>
Red neuronal Con NDVI UAV	22,12	3,7929%	60	<u>20,38</u>	<u>3,6212%</u>	<u>50</u>
Red neuronal satelital	72,24	20,8929%	30	83,93	25,3402%	30
Regresión con NDVI satelital	29,76	44,8118%	30	31,21	44,2679%	30

5.5.4 Por parcela

En este método de organización de datos solamente se presenta la Tabla 13 con los promedios de los resultados de las estimaciones de carbono sobre las cuales se pudo aplicar la metodología por parcela, no se presenta un cuadro con los mejores resultados, ya que los resultados con este método de organización buscan mostrar el comportamiento de las diferentes estimaciones

en promedio sobre todas las parcelas, en vez de entrar en detalle en cada una de las parcelas. Esta metodología de organización de datos, en general, fue la que entregó los errores más altos en comparación a las otras, la razón por la cual esto sucede es debido a la cantidad de datos que existen en las parcelas, ya que al generar un modelo por parcela, si dentro de la parcela no existe información suficiente, un modelo de redes neuronales como el planteado seguramente no será entrenado correctamente y dará lugar a variabilidades altas en las estimaciones, algo similar a lo que sucedía con estimaciones que utilizaban imágenes satelitales, sin embargo, se observa que la estimación de carbono usando la red neuronal con NDVI sobre la imagen de carbono $C_{General}$, permite reducir la variabilidad de los resultados, y dada la situación con la que se cuenta con una parcela que tenga la cantidad de datos necesaria, esta podría realzar estimaciones con porcentajes de variabilidad más bajos.

Tabla 13.

Resumen del promedio de todos los resultados usando la organización de datos de todos los píxeles.

Tipo de Estimación	C_{QH}		$C_{General}$	
	RMSE%	Variabilidad	RMSE%	Variabilidad
	Promedio	Promedio	Promedio	Promedio
Red neuronal UAV	35,6659	66,9347%	37,4936	89,2409%
Regresión con NDVI UAV	30,0260	75,2801%	26,8103	81,0707%
Red neuronal	29,5476	56,0611%	<u>26,5643</u>	<u>55,5312%</u>

Con NDVI UAV

5.6 Comparación de resultados

En esta sección se hace una comparación de los mejores resultados obtenidos en cada una de las organizaciones de datos, la Tabla 14 presenta los mejores resultados en cada método de organización de datos, en este cuadro se observa que el método de organización de datos de todos los píxeles entrega el porcentaje de error y variabilidad promedio más bajos, siempre y cuando se utilice la imagen de carbono $C_{General}$ con el método de estimación de carbono de red neuronal con NDVI.

Tabla 14.

Comparación de los mejores resultados de cada uno de los métodos de organización de datos.

Organización de datos	Imagen de carbono	Tipo de estimación	RMESE% Promedio	Variabilidad promedio
Todas las parcelas	C_{QH}	Regresión con NDVI	23,7941	20,2530%
Todos los árboles	$C_{General}$	Red neuronal con NDVI	24,1917	15,9919%
<u>Todos los píxeles</u>	$C_{General}$	Red neuronal con NDVI	<u>20,8287</u>	<u>4,7181%</u>
Por parcela	$C_{General}$	Red neuronal con NDVI	26,5643	55,4312%

6. Conclusiones

- Se desarrolló una metodología para estimar bancos de carbono a partir de imágenes RGB-NIR de bosques capturadas a partir de drones.
- Se caracterizó la capacidad de absorción de carbono en una zona de estudio ubicada en la república de Colombia en el departamento de Santander, dentro del municipio de Molagavita en la vereda de Potrero de Rodríguez sobre la cual se voló un UAV para obtener las imágenes R-G-B-NIR utilizadas en el proyecto.
- En el proyecto se estudiaron cuatro formas de organizar la información de carbono utilizada para entrenar el modelo de estimación de bancos de carbono. Las formas fueron: todas las parcelas, todos los árboles, todos los píxeles y modelo por parcela. De las cuales, se concluye que la forma de todos los píxeles fue la que entregó los errores más bajos, con un valor de RMSE% de 20,38% y una variabilidad de 3,62% usando el 50% de los datos de entrenamiento.
- Dentro de la metodología planteada un paso importante es el entrenamiento del modelo matemático, donde se busca que el modelo relacione los datos de la expedición de campo y las imágenes capturadas con el UAV con el menor error posible. En total se probaron 5 modelos matemáticos, los cuales tres usan imágenes de UAV y dos usan imágenes satelitales. Las metodologías implementadas fueron: una red neuronal con 15 índices de vegetación y bandas R-G-B-NIR de la zona de estudio, una con datos de UAV y otra con satelitales; una regresión lineal usando el índice de vegetación NDVI de la zona de estudio para entrenar el modelo, una con datos de UAV y otra con satelitales; finalmente, una red neuronal que usa el índice de vegetación NDVI de la zona de estudio para entrenar el modelo. A partir

de estos modelos, se encontró que los mejores resultados fueron entregados por la red neuronal propuesta entrenada con el índice de vegetación NDVI obtenido con imágenes capturadas desde el UAV, de la imagen de carbono $C_{General}$, entregando un error de 20,82% en promedio con variabilidad promedio entre estimaciones de 4,71%, la cual mejora en un 0,01% la variabilidad del resultado obtenido usando regresión lineal con NDVI usando imágenes de UAV, también mejora el RMSE% en un 11,21% y la variabilidad en un 40,64% en comparación a las estimaciones usando una regresión NDVI usando imágenes satelitales.

- En base a los resultados obtenidos, se tiene que la característica más importante utilizada en el cubo de datos fue el índice de vegetación NDVI. Esto debido a que los entrenamientos realizados únicamente con este índice de vegetación fueron los que entregaron los resultados con errores más bajos.

Trabajo Futuro

En este trabajo de investigación se diseñó una metodología con el fin de obtener información del estado de los bosques, sin embargo, como se describió en el capítulo 2, la metodología desarrollada es una de las dos fases presentes en un modelo de monitoreo forestal permanente, donde en este trabajo se implementó únicamente la fase de entrenamiento. Debido a lo anterior, un trabajo futuro a desarrollar es aplicar la metodología teniendo en cuenta ambas fases del modelo planteado.

Otro punto interesante a tener en cuenta son los métodos de organización de datos, los cuales se plantearon con la intención de reducir la cantidad de información a medir al realizar una expedición de campo, sin embargo, dada la extensión del área de estudio, no fue factible realizar más de una expedición de campo para validar si efectivamente la reducción de medición de datos es factible. Por esta razón, un trabajo futuro a realizar sería verificar si se pueden llegar a los mismos resultados de RMSE, usando únicamente el porcentaje de datos que entrega los mejores resultados con las formas de organización de datos planteadas.

Pese a que los porcentajes de error encontrados son buenos en comparación a otras metodologías, un porcentaje de error en el orden del 20% no es muy atractivo. Así que un trabajo futuro a realizar sería probar configuraciones de redes neuronales que permitan reducir este porcentaje de error.

Por último, otro trabajo futuro de esta investigación es la integración de más información en el modelo de estimación de carbono para disminuir el error, información como la nube de puntos obtenida durante la construcción del modelo digital de la zona de estudio con las imágenes de UAV, la cual puede aportar información referente a las alturas de los árboles que se encuentra

fuertemente relacionada con el carbono, incluso el acoplamiento de otra clase de sensores al UAV para capturar información diferente a imágenes espectrales, como láseres o una nariz electrónica.

Bibliografía

- Alpaydin, E. (2020). *Introduction to machine learning*. MIT press.
- Berie, H. T., & Burud, I. (2018). Application of unmanned aerial vehicles in earth resources monitoring: focus on evaluating potentials for forest monitoring in Ethiopia. *European journal of remote sensing*.
- Byrd, K. B., Ballanti, L., Thomas, N., Nguyen, D., Holmquist, J. R., Simard, M., & Windham-Myers, L. (2018). A remote sensing-based model of tidal marsh aboveground carbon stocks for the conterminous United States. *ISPRS journal of photogrammetry and remote sensing*, 139, 255–271.
- Cabrera E., G. Galindo, D.M. Vargas. (2011). *Protocolo de Procesamiento Digital de Imágenes para la Cuantificación de la Deforestación en Colombia, Nivel Nacional Escala Gruesa y Fina*. Tech. rep., Instituto de Hidrología, Meteorología y Estudios Ambientales-IDEAM.
- Carlson, T. N., & Ripley, D. A. (1997). On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote sensing of Environment*, 62, 241–252.
- Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M. S., Delitti, W. B., . . . others. (2014). Improved allometric models to estimate the aboveground biomass of tropical trees. *Global change biology*.
- Chen, G., Ozelkan, E., Singh, K. K., Zhou, J., Brown, M. R., & Meentemeyer, R. K. (2017). Uncertainties in mapping forest carbon in urban ecosystems. *Journal of environmental management*, 187, 229–238.
- de Almeida, C. T., Galvão, L. S., Ometto, J. P., Jacon, A. D., de Souza Pereira, F. R., Sato, L. Y., . . . others. (2019). Combining LiDAR and hyperspectral data for aboveground biomass

- modeling in the Brazilian Amazon using different regression algorithms. *Remote Sensing of Environment*, 232, 111323.
- Domingo, D., Ørka, H. O., Næsset, E., Kachamba, D., & Gobakken, T. (2019). Effects of UAV Image Resolution, Camera Type, and Image Overlap on Accuracy of Biomass Predictions in a Tropical Woodland. *Remote Sensing*, 11, 948.
- Eisenbeiß, H. (2009). *UAV photogrammetry*.
- FAO. (2018). *Measuring and modelling soil carbon stocks and stock changes in livestock production systems – Guidelines for assessment (Draft for public review)*. FAO Publications.
- Farquhar, G. D., & Sharkey, T. D. (1982). Stomatal conductance and photosynthesis. *Annual review of plant physiology*, 33, 317–345.
- Ferreira, R. O. (2005). Herramientas para el Manejo de Bosques Inventario Forestal. *Siguatepeque, Honduras IV*.
- Galidaki, G., Zianis, D., Gitas, I., Radoglou, K., Karathanassi, V., Tsakiri–Strati, M., . . . Mallinis, G. (2017). Vegetation biomass estimation with remote sensing: focus on forest and other wooded land over the Mediterranean ecosystem. *International Journal of Remote Sensing*, 38, 1940–1966.
- Gao, L., Wang, X., Johnson, B. A., Tian, Q., Wang, Y., Verrelst, J., . . . Gu, X. (2020). Remote sensing algorithms for estimation of fractional vegetation cover using pure vegetation index values: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 159, 364–377.
- Giri, C. P. (2016). *Remote sensing of land use and land cover: principles and applications*. CRC press.

- Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote sensing of Environment*.
- Gitelson, A. A., Kaufman, Y. J., Stark, R., & Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. *Remote sensing of Environment*.
- Gomez, C., White, J. C., & Wulder, M. A. (2016). Optical remotely sensed time series data for land cover classification: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 116, 55–72.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.
- Goslee, K., Walker, S. M., Grais, A., Murray, L., Casarim, F., & Brown, S. (2015). Module C-CS: calculations for estimating carbon stocks. *LEAF REDD+ Technical Guidance Series*.
- Gutiérrez, N., Yepes, A. P., Cabrera, E., González, J. J., Galindo, G., Barbosa, A. P., . . . Pulido, A. D. (2018). *Sistema de Medición/Monitoreo, Reporte y Verificación (M/MRV) en Colombia: Avances y hoja de ruta para su consolidación*. Tech. rep., Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM), Ministerio de Ambiente y Desarrollo Sostenible (MADS).
- Hall, D. (2012). *Remote sensing of ice and snow*. Springer Science & Business Media.
- Haykin, S. (2010). *Neural Networks and Learning Machines*, 3/E. Pearson Education India.
- 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.
- Howard, J., Hoyt, S., Isensee, K., Telszewski, M., Pidgeon, E., & others. (2014). *Coastal Blue Carbon: Methods for assessing carbon stocks and emissions factors in mangroves, tidal*

- salt marshes, and seagrasses*. Conservation International, Intergovernmental Oceanographic Commission of
- Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote sensing of environment*.
- ITU. (2018). ITU-T, Environment, Climate Change and Circular Economy at a Glance. *ITU-T, Environment, Climate Change and Circular Economy at a Glance*. Obtenido de <https://www.itu.int/en/ITU-T/climatechange/Pages/ictccenv.aspx>
- Jayathunga, S., Owari, T., & Tsuyuki, S. (2018). The use of fixed-wing UAV photogrammetry with LiDAR DTM to estimate merchantable volume and carbon stock in living biomass over a mixed conifer-broadleaf forest. *International journal of applied earth observation and geoinformation*, 73, 767–777.
- Kraus, K. (2011). *Photogrammetry: geometry from images and laser scans*. Walter de Gruyter.
- Lu, G., & Fei, B. (2014). Medical hyperspectral imaging: a review. *Journal of biomedical optics*.
- MADS, & IDEAM. (2018). *Manual de Campo Inventario Forestal Nacional de Colombia*. IDEAM.
- Navarro, J., Algeet, N., Fernández-Landa, A., Esteban, J., Rodríguez-Noriega, P., & Guillén-Climent, M. (2019). Integration of uav, sentinel-1, and sentinel-2 data for mangrove plantation aboveground biomass monitoring in senegal. *Remote Sensing*, 11, 77.
- Norwegian Agency for Development Cooperation, A. N., & FECOFUN, I. C. (2010). *Forest Carbon Stock Measurement Guidelines for measuring carbon stocks in community-managed forests*. ANSAB.

- Ota, T., Ahmed, O. S., Minn, S. T., Khai, T. C., Mizoue, N., & Yoshida, S. (2019). Estimating selective logging impacts on aboveground biomass in tropical forests using digital aerial photography obtained before and after a logging event from an unmanned aerial vehicle. *Forest ecology and management*, 433, 162–169.
- Pandey, P. C., Srivastava, P. K., Chetri, T., Choudhary, B. K., & Kumar, P. (2019). Forest biomass estimation using remote sensing and field inventory: a case study of Tripura, India. *Environmental monitoring and assessment*, 191, 593.
- Pearson, R. L., & Miller, L. D. (1972). Remote mapping of standing crop biomass for estimation of the productivity of the shortgrass prairie. *Remote Sensing of Environment*, VIII.
- Ponce-Corona, E., Sánchez, M. G., Fajardo-Delgado, D., Castro, W., De-la-Torre, M., & Avila-George, H. (2019). Detection of Vegetation Using Unmanned Aerial Vehicles Images: A Systematic Review. *2019 8th International Conference On Software Process Improvement (CIMPS)*, (págs. 1–7).
- Remondino, F., Barazzetti, L., Nex, F., Scaioni, M., & Sarazzi, D. (2011). UAV photogrammetry for mapping and 3d modeling—current status and future perspectives. *International archives of the photogrammetry, remote sensing and spatial information sciences*.
- Rhyma, P. P., Norizah, K., Hamdan, O., Faridah-Hanum, I., & Zulfa, A. W. (2020). Integration of normalised different vegetation index and Soil-Adjusted Vegetation Index for mangrove vegetation delineation. *Remote Sensing Applications: Society and Environment*, 17, 100280.
- Rouse, J. W., Hass, R. H., Schell, J. A., & Deering, D. W. (1973). Monitoring vegetation systems in the great plains with ERTS. *Third ERTS Symposium, NASA (1973)*.

- Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited,.
- Saarela, S., Holm, S., Healey, S., Andersen, H.-E., Petersson, H., Prentius, W., . . . Ståhl, G. (2018). Generalized Hierarchical Model-Based Estimation for Aboveground Biomass Assessment Using GEDI and Landsat Data. *Remote Sensing*, 10, 1832.
- Schowengerdt, R. A. (2006). *Remote sensing: models and methods for image processing*. Elsevier.
- Shaw, G., & Burke, H.-h. (2003). Spectral Imaging for Remote sensing. *Lincoln Laboratory Journal*.
- Shi-Long, P. I., Fangjing-Yun, H. J.-s., & Yu, X. (2004). Spatial distribution of grassland biomass in China.
- Stocker, T. (2014). Climate change 2013: the physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change. *Climate change 2013: the physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge university press.
- Sun, X., Li, B., Du, Z., Li, G., Fan, Z., Wang, M., & Yue, T. (2019). Surface modelling of forest aboveground biomass based on remote sensing and forest inventory data. *Geocarto International*, 1–16.
- Swinfield, T., Lindsell, J. A., Williams, J. V., Harrison, R. D., Gemita, E., Schönlieb, C. B., . . . others. (2019). Accurate Measurement of Tropical Forest Canopy Heights and Aboveground Carbon Using Structure From Motion. *Remote Sensing*, 11, 928.

- Tejada, G., Görgens, E. B., Espírito-Santo, F. D., Cantinho, R. Z., & Ometto, J. P. (2019). Evaluating spatial coverage of data on the aboveground biomass in undisturbed forests in the Brazilian Amazon. *Carbon balance and management*, 14, 11.
- UNFCCC. (2015). *Measurements for Estimation of Carbon Stocks in Afforestation and Reforestation Project Activities under the Clean Development Mechanism: A Field Manual*. United Nations Framework Convention on Climate Change (UNFCCC).
- Urbazaev, M., Thiel, C., Cremer, F., Dubayah, R., Migliavacca, M., Reichstein, M., & Schmullius, C. (2018). Estimation of forest aboveground biomass and uncertainties by integration of field measurements, airborne LiDAR, and SAR and optical satellite data in Mexico. *Carbon balance and management*, 13, 5.
- Wall, M. E., Rechtsteiner, A., & Rocha, L. M. (2003). Singular value decomposition and principal component analysis. En *A practical approach to microarray data analysis* (págs. 91–109). Springer.
- Warfield, A. D., & Leon, J. X. (2019). Estimating Mangrove Forest Volume Using Terrestrial Laser Scanning and UAV-Derived Structure-from-Motion. *Drones*, 3, 32.
- Yepes, A., Sierra, A., Niño, L. M., López, M., Garay, C., Vargas, D., . . . Barbosa, A. (2016). Biomasa y carbono total almacenado en robledales del sur de los Andes Colombianos: aportes para el enfoque REDD+ a escala de proyectos. *Revista de Biología Tropical*, 64, 399–412.
- Zhang, R., Zhou, X., Ouyang, Z., Avitabile, V., Qi, J., Chen, J., & Giannico, V. (2019). Estimating aboveground biomass in subtropical forests of China by integrating multisource remote sensing and ground data. *Remote Sensing of Environment*, 232, 111341.

- Zhao, K., Suarez, J. C., Garcia, M., Hu, T., Wang, C., & Londo, A. (2018). Utility of multitemporal lidar for forest and carbon monitoring: Tree growth, biomass dynamics, and carbon flux. *Remote Sensing of Environment*.
- Zimmermann, T., Rietdorf, J., & Pepperkok, R. (2003). Spectral imaging and its applications in live cell microscopy. *FEBS letters*.
- Zitova, B., & Flusser, J. (2003). Image registration methods: a survey. *Image and vision computing*, 21, 977–1000.