



UNIVERSIDAD
NACIONAL
DE COLOMBIA

Índices radiométricos, multiespectrales y SAR, para la evaluación a gran escala de la calidad de hábitat en bosque húmedo tropical en zonas del Magdalena Medio, Colombia

Luis Hernando Romero Jiménez

Universidad Nacional de Colombia

Facultad de Ciencias Agrarias

Bogotá D.C., Colombia

2023

Índices radiométricos, multiespectrales y SAR, para la evaluación a gran escala de la calidad de hábitat en bosque húmedo tropical en zonas del Magdalena Medio, Colombia

Luis Hernando Romero Jiménez

Tesis como requisito para optar el título de:

Magister en Geomática

Director:

Ph.D. José Camilo Fagua González

Codirectora:

Ph.D. Susana Rodríguez Buriticá

Línea de Investigación:

Geoinformación para el uso sostenible de los recursos naturales

Grupo de Investigación:

Biodiversidad, Biotecnología y Conservación de Ecosistemas

Universidad Nacional de Colombia

Facultad de Ciencias Agrarias

Bogotá D.C., Colombia

2023

“No todo el que se esfuerza logra alcanzar el éxito, pero aquellos que han alcanzado el verdadero éxito ha sido gracias a su esfuerzo”
- Genji Kamogawa

“Disfruta del viaje y los pequeños desvíos. Las cosas que son más importantes que tu objetivo final, pueden estar ahí, a un lado del camino”
- Ging Freecs

A mi amada esposa, mis padres y mis hermanos que estuvieron brindándome su apoyo desde el inicio de este gran proceso... y a Kitty también.

Agradecimientos

Este trabajo no se podría haber concretado sin la ayuda del Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, cuyo equipo de trabajo del antiguo Programa de Evaluación y Monitoreo, me brindó el apoyo, las herramientas y el acompañamiento para sacar adelante este proyecto de tesis.

Agradezco mucho al profesor José Camilo Fagua, quien demostró un profundo interés en sacar adelante este proyecto, incluso cuando la pandemia retrasó y afectó el desempeño de la investigación. Aún con las diferentes problemáticas que se identificaron en el camino, siempre pude contar con su apoyo y su guía en la elaboración de esta investigación.

Un agradecimiento muy especial a la investigadora y mi mentora la doctora Susana Rodríguez Buriticá, quien desde siempre ha sido un ejemplo a seguir y un gran exponente para la ciencia y las mujeres científicas.

Resumen

Índices radiométricos, multiespectrales y SAR, para la evaluación a gran escala de la calidad de hábitat en bosque húmedo tropical en zonas del Magdalena Medio, Colombia

El uso de información espacial para el estudio de fenómenos ecológicos que involucran múltiples especies, es fundamental para el planteamiento de estrategias para la conservación, planeación y monitoreo de la biodiversidad. El hábitat, definido en este trabajo como las condiciones biofísicas que permiten la persistencia de la población de una especie en el espacio y en el tiempo, ha sido analizado tradicionalmente por medio de diferentes productos derivados de sensores remotos (SR), con la desventaja de que muchos de estos carecen de validación en campo, o tienen resoluciones espaciales o temáticas insuficientes para la planeación y el monitoreo de acciones de conservación. Para superar estas limitaciones, integramos información espectral de datos SAR (Radar de Apertura Sintética como Sentinel-1 y PALSAR), datos multiespectrales (Sentinel-2 y MODIS), y registros de especies tomados en campo y consultados en bases de datos nacionales e internacionales. Utilizando cinco algoritmos de aprendizaje (Machine Learning), proponemos un índice de calidad de hábitat específico para ecosistemas de bosque húmedo tropical, en una zona de estudio ubicada en el Magdalena Medio, entre los municipios de Puerto Berrío, Yondó, Cantagallo y Puerto Wilches. Como aproximación a la medida de calidad de bosques se utilizó el Índice de Condición Estructural de Hansen (SCI), que evalúa la estructura de los bosques, aunque no considera otros elementos relacionados con la integridad de este ecosistema, como su composición y función. Por esta razón, se modificó el SCI con datos de registros biológicos, ajustados para representar la dependencia que las especies tienen del bosque, y se construyó un índice de calidad de hábitat integral con datos de sensores remotos e información de campo. El algoritmo de Random Forest fue el que mostró el error más bajo de estimación del SCI (Accuracy = 0.675 y Kappa = 0.532); mientras que para la estimación de la dependencia de las especies hacia el bosque, fue el algoritmo de Máquinas de Soporte Vectorial (Accuracy = 0.643 y Kappa = 0.397). Se encontró que las variables que aportan mayor información para la estimación del SCI son Red edge 1, Red, SWIR_2, Green, el índice PSRI de Sentinel-2A y los índices de Radar HV, VHdivVV y VHdivHH. Se comprobó que los valores de exactitud temática son más bajos al utilizar las 18 categorías de SCI, por lo que se simplificó a cinco categorías. De manera similar, para la estimación de la calidad del bosque se encontró que las variables que aportan mayor información son HVdivHH y HV de Sentinel-1A y el Tasseled cap wetness, el índice MNDWI y la banda SWIR_1. Finalmente, el modelo que integra el SCI con la calidad de los bosques resultó con la mayor exactitud

temática, desarrollado con el algoritmo de Potenciación del Gradiente (Accuracy= 0.724 y Kappa= 0.493), permitiendo identificar áreas de incongruencia entre estos dos componentes. La exactitud de los modelos evidencia que las variables predictoras derivadas de SR, presentan relaciones que no son capturadas por las variables originales del SCI y que pueden contribuir a su mejoramiento, mientras que la estimación de dependencia de las especies al bosque refleja un sesgo en el muestreo. No obstante, el modelo final incorpora la incertidumbre de los dos primeros modelos, lo que fortalece los resultados encontrados en los modelos 1 y 2, pero así mismo con la capacidad de retroalimentarse con una mayor disponibilidad de registros biológicos curados.

Palabras clave: Sentinel-1A, Sentinel-2A, Aprendizaje supervisado, Machine Learning, Integridad ecológica, Calidad de Hábitat.

Abstract

Radiometric, multispectral and SAR indices, for the large-scale evaluation of habitat quality in tropical humid forest in areas of Magdalena Medio, Colombia

The use of spatial information for the study of ecological phenomena that involve multiple species, is fundamental for the approach of strategies for the conservation, planning and monitoring of biodiversity. The habitat, defined in this work as the biophysical conditions that allow the persistence of the population of a species in space and time, has traditionally been analyzed by means of different products derived from remote sensing (RS), with the disadvantage of that many of these lack validation in the field, or have insufficient spatial or thematic resolution for planning and monitoring conservation actions. To overcome these limitations, we integrate spectral information from SAR data (Synthetic Aperture Radar such as Sentinel-1 and PALSAR), multispectral data (Sentinel-2 and MODIS), and records of species taken in the field and consulted in national and international databases. Using five learning algorithms (Machine Learning), we propose a specific habitat quality index for tropical humid forest ecosystems, in a study area located in Magdalena Medio, between the municipalities of Puerto Berrío, Yondó, Cantagallo and Puerto Wilches. As an approximation to measure forest quality, we use the Hansen Structural Condition Index (SCI), which evaluates the structure of forests, although it does not consider other elements related to the integrity of this ecosystem, such as its composition and function. For this reason, the SCI was modified with biological records, adjusted to represent the dependence that species have on the forest, and a comprehensive habitat quality index was constructed with data from remote sensors and field information. The Random Forest algorithm was the one that showed the lowest SCI estimation error (Accuracy = 0.675 and Kappa = 0.532); while for the estimation of the dependence of the species on the forest, it was the Support Vector Machines algorithm (Accuracy = 0.643 and Kappa = 0.397). It was found that the variables that provide the most information for estimating the SCI are Red edge 1, Red, SWIR_2, Green, the PSRI index of Sentinel-2A and the Radar HV, VHdivVV and VHdivHH indices. It was found that the thematic accuracy values are lower when using the 18 SCI categories, so it was simplified to five categories. Similarly, for the estimation of forest quality it was found that the variables that provide the most information are HVdivHH and HV of Sentinel-1A and the Tasseled cap wetness, the MNDWI index and the SWIR_1 band. Finally, the model that integrates the SCI with the quality of the forests resulted with the greatest thematic accuracy, developed with the Gradient Boosting algorithm (Accuracy= 0.724 and Kappa= 0.493), allowing the identification of areas of incongruence between these two components. The accuracy of the models shows that the predictor variables derived from SR present relationships that are not captured by the original variables of the SCI and that can contribute to their improvement, while the estimate of dependence of the species on the forest reflects a bias in the sampling.

However, the final model incorporates the uncertainty of the first two models, which strengthens the results found in models 1 and 2, but also with the ability to feed back with a greater availability of curated biological records.

Keywords: Sentinel-1A, Sentinel-2A, Supervised Learning, Machine Learning, Ecological Integrity, Habitat Quality.

Contenido

Agradecimientos	7
Resumen	9
Lista de figuras	14
Lista de tablas	16
Introducción	17
1. Marco de Referencia	22
1.1 Marco Teórico	22
1.1.1 Sensores Remotos Multiespectrales.....	22
1.1.2 Sensores Remotos RADAR.....	28
1.1.3 Índices espectrales.....	29
1.1.4 Clasificación de Imágenes de Satélite	33
1.1.5 Clasificación Supervisada.....	34
1.1.6 Coberturas de la Tierra - Metodología CORINE Land Cover adaptada para Colombia	34
1.1.7 Registros Biológicos	35
1.1.8 Integridad Ecológica	36
1.2 Estado del Arte.....	37
1.2.1 Clasificación Supervisada y Técnicas de Aprendizaje de Máquinas	37
1.2.2 Integración de datos de múltiples productos satelitales y datos de campo	41
2. Objetivos	42
2.1 Objetivo General	42
2.2 Objetivos Específicos	42
3. Datos y Métodos	43
3.1 Zona de Estudio	43
3.2 Registros Biológicos.....	47
3.3 Metodología	48
3.3.1 Modelo 1: Estimación del Índice de Condición Estructural (SCI) con variables derivadas de Sensores Remotos.....	48
3.3.2 Modelo 2: Estimación de Calidad de Hábitat basada en registros, a partir de información de SR.....	57
3.3.3 Modelo 3: Propuesta de índice de calidad de hábitat, integrando SCI y registros biológicos	64
4. Resultados	65
4.1 Modelo 1	65
4.2 Modelo 2	72
4.3 Modelo 3	77
5. Discusión	84
6. Conclusiones y recomendaciones	99
Bibliografía	101

Lista de figuras

Figura 1-1. Regiones del espectro electromagnético desde rayos X hasta VNIR. Tomada de Zwinkles (2015).....	23
Figura 1-2. Dominios espectrales que responden a cambios de variables específicas en el estudio de la vegetación. Se compara la respuesta espectral obtenida por las bandas de Sentinel-2 (línea negra discontinua), por el sensor hiperespectral EnMap (cruces rojas) y por un espectrorradiómetro de campo cuasi continuo ASD FieldSpec4. Figura tomada de Hank et al. (2019).....	25
Figura 1-3. Ejemplo de marco de jerarquía de integridad ecológica para evaluar la condición en los ecosistemas naturales en función de las características del paisaje que sustentan las interacciones depredador-presa. Tomado de Mora (2017b).....	37
Figura 1-4. Tipos de técnicas de aprendizaje de máquina. Figura tomada de Sarker (2021)	38
Figura 1-5. Diagrama de flujo de un árbol de decisión (izquierda) y de un Random Forest (derecha). Figuras tomadas y adaptadas de (Abdulkareem & Abdulazeez, 2021).	39
Figura 1-6. Diferencias entre los métodos de ensamble más utilizados en modelos de aprendizaje de máquina. Figura tomada de Sheykhmousa et al. (2020).	40
Figura 1-7. Plano de separación por un hiperplano generado a partir de un Kernel lineal. Figura tomada de Sheykhmousa et al., 2020.	41
Figura 3-1. Área de estudio. A) Ubicación en Sudamérica. B) Ubicación del área en la ecorregión Magdalena-Urabá, situada al norte de Colombia y conectando el norte de las ecorregiones de Mesoamérica y el Chocó con la ecorregión andina. C) Polígono de estudio ubicado entre los municipios de Puerto Badel, Yondó, Cantagallo y Puerto Wilches, Antioquia. Fuente: Autor	45
Figura 3-2. Mosaico de imágenes PlanetScope con resolución espacial de 30 m.	46
Figura 3-3. Polígonos identificados en el área de estudio, donde se implementan estrategias de conservación y restauración de bosque húmedo tropical mediante diferentes estrategias.	47
Figura 3-4. Capa de SCI desarrollado por Hansen et al. (2019). El recuadro blanco de la imagen corresponderá a un acercamiento para mostrar en las siguientes figuras la distribución de los puntos de muestreo.	49
Figura 3-5. Muestreo por grillas equidistantes. A) Grilla de 1 km, B) Grilla de 500 metros. Al ser el SCI una capa continua para el área de estudio, estas grillas permiten realizar un muestreo equilibrado en todas las categorías de SCI disponibles.....	50
Figura 3-6. Registros biológicos disponibles en el área de estudio, a partir de los cuales se calcula la dependencia de las especies hacia los ecosistemas de bosque.	51
Figura 3-7. Polígonos de calidad de hábitat calculada con registros biológicos y ponderada con coberturas CORINE Land Cover adaptadas para Colombia.....	61
Figura 3-8. Determinación de las congruencias e incongruencias al integrar SCI con calidad de hábitat. Las categorías de SCI y calidad de hábitat intermedias fueron fusionadas con la categoría alta de cada uno de los componentes.	62
Figura 3-9. Áreas de congruencia e incongruencia entre SCI y Calidad de Bosque. Los niveles de congruencia en la leyenda dependen de los valores coincidentes entre los dos componentes. Baja calidad incongruente (SCI bajo, pero con alta dependencia de bosque), Alta calidad incongruente (SCI alto, pero con dependencia baja de bosque). Baja calidad congruente (SCI bajo y baja dependencia de bosque). Media calidad congruente (SCI medio y dependencia de bosque intermedia). Alta calidad congruente (SCI alto y alta dependencia de bosque). Es importante tener en cuenta que esta figura sólo es ilustrativa, ya que se	

construyó uniendo la figura 7 con el SCI de Hansen reclasificado. La variable respuesta para el modelo 3 se construirá con la predicción generada con el mejor algoritmo del modelo 2 y el SCI reclasificado..... 65

Figura 4-1. Importancia de variables medida previamente a la modelación del Modelo 1. Las 31 variables predictoras (Boxplots color verde) demostraron tener una importancia mayor que la generada únicamente por aleatoriedad al compararlas con las variables sombra (Boxplots de color azul). Entre las variables con mayor importancia encontramos la banda RED_EDGE_1, el índice PSRI, el porcentaje de bosque y el coeficiente de retrodispersión HV..... 67

Figura 4-2. Caracterización de diferencias entre modelos mediante el remuestreo de sus distribuciones para el Modelo 1..... 69

Figura 4-3. Resultado de la modelación del SCI reclasificado a partir de datos de SR, utilizando el algoritmo RF..... 70

Figura 4-4. Importancia de variables calculadas en R para el Modelo 1..... 71

Figura 4-5. Perfil espectral promedio por clase de SCI, utilizando las variables más importantes identificadas por el Modelo 1..... 72

Figura 4-6. Importancia de variables medida previamente a la modelación del Modelo 2. Las 31 variables predictoras demostraron tener una importancia mayor que la generada únicamente por aleatoriedad al compararlas con las variables sombra. Entre las variables con mayor importancia encontramos los índices de Radar HVdivHH y VHdivVV, el coeficiente de retrodispersión HV y los índices RNDVI y Fpar..... 73

Figura 4-7. Caracterización de diferencias entre modelos mediante el remuestreo de sus distribuciones para el Modelo 2..... 75

Figura 4-8. Resultado de la modelación de la calidad de bosque a partir de datos de SR, utilizando el algoritmo RF..... 75

Figura 4-9. Importancia de variables calculadas en R para el Modelo 2..... 76

Figura 4-10. Perfil espectral promedio por clase de calidad de hábitat, utilizando las variables más importantes identificadas por el Modelo 2..... 77

Figura 4-11. Áreas de congruencia e incongruencia entre SCI y Calidad de Bosque. A diferencia de la figura 3-9, este resultado si corresponde a la variable respuesta que se usará para el Modelo 3, creado con el mejor resultado del Modelo 2 y el SCI reclasificado de Hansen. Las áreas diferentes a los cuerpos de agua que no tienen valores, corresponden a valores que no pudieron clasificarse..... 78

Figura 4-12. Importancia de variables medida previamente a la modelación del Modelo 3. Las 8 variables predictoras (Boxplots color verde) demostraron tener una importancia mayor que la generada únicamente por aleatoriedad al compararlas con las variables sombra (Boxplots azules). Entre las variables con mayor importancia encontramos el coeficiente de retrodispersión HV y las bandas RED_EDGE_1 y SWIR_2..... 79

Figura 4-13. Caracterización de diferencias entre modelos mediante el remuestreo de sus distribuciones para el Modelo 3..... 80

Figura 4-14. Resultado de la modelación de la calidad de hábitat integral a partir de datos de SR, utilizando el algoritmo GBM. Adicionalmente se muestra la distribución de los registros biológicos para inferir las posibles explicaciones de las incongruencias..... 81

Figura 4-15. Importancia de variables calculadas en R para el Modelo 3. El paquete de R caret ofrece una perspectiva diferente al momento de medir la importancia de las variables cuando se utiliza el algoritmo de GBM..... 81

Figura 4-16. Perfil espectral promedio por clase del índice de calidad de hábitat integral, utilizando las variables más importantes identificadas para el Modelo 1 y el Modelo 2..... 82

Figura 5-1. Distribución de los valores de SCI reclasificado en los datos utilizados para el entrenamiento del Modelo 1..... 85

- Figura 5-2.** Variación espectral de las variables identificadas como más relevantes en el Modelo 1 de estimación de SCI reclasificado a cinco categorías. 86
- Figura 5-3.** Categorías de SCI reclasificado proyectados los datos sobre los dos primeros discriminantes que reúnen cerca del 90% del radio de discriminabilidad. 87
- Figura 5-4.** Matriz de correlación entre las variables más importantes del modelo 1. Nótese como muchas de las variables presentan una alta correlación, siendo las bandas RED y GREEN las de mayor valor (0.94), seguidas de GREEN y RED_EDGE_1 (0.92). 88
- Figura 5-5.** Matriz de correlación entre las variables más importantes del modelo 2. En este caso no se evidencia tanta correlación como con el Modelo 1, siendo las correlaciones entre las bandas MNDWI y wetness las de mayor valor (0.98) y entre wetness y SWIR_1 (-0.86). 89
- Figura 5-6.** Matriz de correlación entre las variables más importantes del modelo 3. Las variables con mayor correlación son las mismas que las presentadas en los Modelos 1 y 2, y al igual que los modelos anteriores, la variable respuesta no tiene correlaciones elevadas con las demás variables..... 93

Lista de tablas

Tabla 1-1. Sensores multiespectrales disponibles en la plataforma Google Earth Engine ..	25
Tabla 1-2. Sensores de radar disponibles en la plataforma Google Earth Engine.....	29
Tabla 3-1. Número total de registros encontrados en el área de estudio y agrupación por grupos.....	52
Tabla 3-2. Puntos de entrenamiento y validación para estimación del SCI	53
Tabla 3-3. Sensores multiespectrales disponibles en la plataforma GEE y utilizados en la modelación	54
Tabla 3-4. Hiperparámetros evaluados con cada algoritmo de aprendizaje, utilizando el paquete caret de R.	56
Tabla 3-5. Clasificación de dependencia de bosque para las especies identificadas en el área de estudio. Los valores entre paréntesis corresponden a las categorías de hábitat definidas por la IUCN.	59
Tabla 3-6. Ponderación de calidad de hábitat con categorías de coberturas de la tierra.....	60
Tabla 3-7. Puntos de entrenamiento y validación para estimación de la calidad de bosque ajustada por coberturas	62
Tabla 4-1. Resultados del ajuste de hiperparámetros para el Modelo 1 de estimación de SCI con variables de SR.	67
Tabla 4-2. Resultados del ajuste de hiperparámetros para el Modelo 2 de estimación de calidad de bosque con variables de SR.	74
Tabla 4-3. Resultados del ajuste de hiperparámetros para el Modelo 3 de estimación de calidad de bosque integrado con las variables más importantes de SR de los Modelos 1 y 2.....	79
Tabla 4-4. Decrecimiento promedio de Gini para las variables explicativas en los 3 modelos propuestos	82

Introducción

El desarrollo acelerado en la tecnología empleada para la observación de la tierra, ha permitido la generación de un gran volumen de información satelital, en gran parte disponible de manera gratuita, facilitando la evaluación y monitoreo espaciotemporal de los procesos que ocurren en la superficie terrestre como el cambio de uso y coberturas (Aniah et al., 2023; Dubertret et al., 2022; Nasir et al., 2022), la dinámica de incendios e inundaciones (Cheng et al., 2022; Gibson et al., 2020), la epidemiología (Carella et al., 2022; Louw et al., 2022), la degradación y pérdida de hábitats naturales (C. He et al., 2017; Moumane et al., 2022; Skowno et al., 2021), los fenómenos de expansión agrícola y urbana (Elbeih, 2021; Ul Din & Mak, 2021), entre otros atributos de la superficie terrestre. De igual manera, los Sensores Remotos (SR) son útiles para evaluar la estructura y función de los ecosistemas de bosque, como por ejemplo la estructura del dosel (Fagua et al., 2019; Reich, 2012), propiedades fotosintéticas (Alton, 2018; Kohzuma et al., 2021), heterogeneidad fenológica (Dronova et al., 2021), humedad (AghaKouchak et al., 2015; Ige et al., 2017), entre otras propiedades (Joshi et al., 2016; Zhang et al., 2021).

En últimas y más allá de entender propiedades ecosistémicas, los sensores remotos son determinantes en generar productos de análisis espacial para la gestión de los ecosistemas (Andrew et al., 2014; De Araujo Barbosa et al., 2015; Pettorelli et al., 2018). Uno de estos productos es la calidad del hábitat que se define como la capacidad del ambiente para proveer condiciones apropiadas para la persistencia de individuos y poblaciones (H. Wang et al., 2020). Se espera que un producto que integre información sobre calidad de hábitat para múltiples especies permitirá refinar estrategias de conservación y uso sostenible de la biodiversidad.

El uso de SR es una de las maneras más costo-efectivas de evaluar el nivel de transformación de los ecosistemas asociados con su calidad de hábitat, al permitir explorar cambios en la cobertura de bosque. No obstante, la mayoría de los indicadores derivados directamente de SR están enfocados en la estructura de la vegetación (estructura y cobertura de dosel, estructura de comunidades, abundancia de coberturas naturales, índices de reflectancia superficial, medidas de textura de superficies, altura de árboles, entre otros), así como de la estructura del paisaje (Índices de huella espacial humana, dinámicas de cambio y uso de la tierra, heterogeneidad del paisaje, conectividad del paisaje y fragmentación) (Rosenfield et al., 2023), los cuales usualmente no incorporan elementos ni procesos que analicen los componentes de composición y función de los ecosistemas de bosque (Pettorelli et al., 2018). Así mismo, la calidad de hábitat es especie-específica y dado que conceptualmente representa el enlace entre los atributos de un área y la persistencia de una especie, es un reto metodológico capturarla desde sensores remotos.

A pesar de que hay propuestas para la espacialización de atributos estructurales y funcionales de ecosistemas específicos con información de sensores remotos, aún es incierto determinar qué tanto de la variabilidad en esos atributos representa la variabilidad en calidad de hábitat para los organismos presentes en este ecosistema (Liao et al., 2013). Por otro lado, aproximaciones como los modelos de distribución, que intentan estimar el hábitat viable para las especies, enfatizan principalmente atributos bioclimáticos y tienden a hacer modelaciones de una especie a la vez (Lembrechts et al., 2019; Zurell et al., 2020). Más allá de los retos desde lo conceptual o lo metodológico, está la gran limitante de la información necesaria, ya sea para entender la información de sensores y la calidad de hábitat, o para calibrar modelos de hábitat, especialmente en hábitats difíciles de caracterizar como los sistemas complejos horizontales y verticales de los bosques (Vogeler & Cohen, 2016). Esto limita el uso de productos espaciales en la toma de decisiones, dado que en áreas específicas no está disponible la información requerida para una o varias especies. Más aún, puede que los tomadores de decisiones no requieran información muy sofisticada para realizar gestión territorial, siempre y cuando la información disponible sea utilizada para cumplir con los objetivos planteados y que se reconozca la incertidumbre asociada con el ejercicio de modelación. Así, aproximar la calidad de hábitat para múltiples especies representa un reto metodológico, especialmente para descripciones a escalas de análisis amplias (regional, nacional o global). Aprovechando los avances en tecnologías de sensores remotos y las capacidades mejoradas de cómputo, es posible generar una mejor aproximación a la calidad de hábitat, contribuyendo al fortalecimiento de la identificación, planeación y monitoreo de las dinámicas de cambio en la integridad de los hábitats, como las producidas en procesos de deforestación/restauración de bosques húmedos tropicales.

Un ejemplo de la problemática en el uso exclusivo de SR para la evaluación de la calidad de hábitat puede evidenciarse en el síndrome de los bosques vacíos, en donde bosques aparentemente bien conservados podrían estar en gran parte defaunados (Pires & Galetti, 2023). Por esta razón, contar con una descripción precisa de la estructura del paisaje no es suficiente para formular estrategias efectivas para la conservación, no obstante significan un punto de partida importante para la formulación de un índice que mida la calidad de hábitat en términos de los componentes de integridad ecológica (estructura, composición y función) (Tierney et al., 2009). Un ejemplo de índice global diseñado para evaluar la estructura de los ecosistemas de bosque es el Índice de Condición Estructural de Bosques (SCI) propuesto por Hansen (A. Hansen et al., 2019). Este índice integra información derivada de sensores remotos para clasificar los bosques húmedos tropicales de acuerdo con la altura de dosel, la cobertura de dosel y el tiempo de intervención para diferenciar entre bosques pequeños, abiertos y recientemente deforestados, de aquellos altos, densos y no perturbados.

Los SR utilizados para el cálculo del SCI corresponden a una combinación de datos ópticos pasivos para determinar la cobertura de árboles y la pérdida de bosque,

mediante interpretación visual y clasificación de imágenes. Los datos de entrenamiento utilizados se extrajeron de imágenes Quickbird de muy alta resolución espacial y datos PALSAR para calcular la altura de dosel (Hansen et al., 2019). A partir de esta metodología, se genera una cartografía temática altamente consistente; sin embargo, se desconoce cómo la definición de estas clases de condición estructural de bosque se relaciona con la calidad de hábitat para múltiples especies. Además, muchos de los insumos utilizados son de difícil acceso y al ser una interpretación visual resulta difícil su replicación, especialmente en zonas con presencia de nubes en la mayor parte del año debido al uso principal de sensores pasivos. Finalmente, como se mencionó anteriormente, el SCI tiene un enfoque exclusivo en la estructura de bosque, lo que corresponde a un solo componente del índice de calidad de hábitat que se propone.

Este trabajo corresponde a la evaluación de una propuesta metodológica, en la cual la información de sensores remotos se combina con información de campo para afinar los productos que puedan aproximar la calidad de hábitat para varias especies, aprovechando el desempeño de una de las herramientas más utilizadas en la clasificación y predicción de imágenes como lo son las técnicas de aprendizaje supervisado con algoritmos de aprendizaje de máquina (Allam et al., 2019) (Objetivo General). Al combinar la información de sensores remotos con información de campo, se propone una aproximación para entender la biodiversidad de manera más integral, al permitir su caracterización, análisis y modelamiento en múltiples escalas, entendiendo principalmente uno de los componentes que forman parte de nuestro índice de calidad de hábitat propuesto, como lo es la estructura de los bosques (Objetivo específico 1). Específicamente, el uso de imágenes multiespectrales y SAR es útil para estimar indicadores de calidad de hábitats como tolerancia de manejo, disponibilidad de luz, contenido de nutrientes, humedad, heterogeneidad espacial, y otras propiedades asociadas a la extensión y condición de los hábitats (Nagendra et al., 2013). En este caso particular, desde los SR se utilizará el SCI como proxy de la estructura de los bosques, el cual ofrece una solidez conceptual y un desempeño aceptable en el área de estudio, pero del que se desconoce su relación con otro tipo de información espacial como la derivada de sensores Sentinel y de Radar.

La información de campo asociada a la diversidad biológica nos brinda una perspectiva ecológica con la que es posible analizar componentes como el funcionamiento de los ecosistemas, la habilidad de estos para recuperarse luego de un disturbio y la composición de especies que soportan o perturban su funcionamiento (Rosenfield et al., 2023). Este tipo de análisis es imposible de detectar desde SR usados regularmente en la evaluación de integridad ecológica como las imágenes de satélite, por lo que se propone el apoyo con información tomada directamente desde campo, la cual será un complemento de la información derivada de SR y el núcleo de nuestra metodología de análisis para el cálculo del índice de calidad de hábitat (Objetivo específico 2). Por esta razón, se tomará como información de apoyo la presencia de grupos funcionales de flora y fauna dependiente de hábitat de bosque y

las coberturas de la tierra siguiendo la metodología CORINE Land Cover adaptada para Colombia. Siguiendo esta metodología, con este proyecto de tesis se podrán identificar zonas de congruencia e incongruencia entre la estructura y la composición de los bosques, mediante el desarrollo de un producto congruente conceptualmente entre la información captada por los SR y la información tomada en campo (Objetivo específico 3).

Una manera en la que podemos aprovechar esta aproximación integral, es mediante la identificación de áreas prioritarias para procesos de conservación y/o restauración, por medio de una herramienta costo-efectiva y replicable para el monitoreo de procesos de degradación de la biodiversidad que informen sobre el mantenimiento de servicios ecosistémicos. Estas herramientas son de especial importancia en áreas con ecosistemas estratégicos como los bosques húmedos tropicales, ya que albergan una alta biodiversidad y endemismo, además de ser ecosistemas muy desafiantes para ejecutar estrategias de restauración (Bechara et al., 2016). El desarrollo de esta metodología contribuirá a la efectividad de evaluaciones de proyectos de conservación y restauración locales, de forma refinada y estandarizada, combinando datos de SR y datos ecológicos tomados en campo, fomentando además la adquisición de información en diferentes escalas de otras áreas de la región, o incluso en otras regiones con incidencia de bosque húmedo tropical, para la generación de productos de calidad, utilizados para la investigación y el uso sostenible de recursos naturales.

1. Marco de Referencia

1.1 Marco Teórico

En esta sección se tratarán diversas temáticas necesarias para el entendimiento del proyecto. Inicialmente se revisó información asociada a los sensores remotos multiespectrales y de RADAR, los cuales formaron parte de nuestras variables explicativas. Adicional al valor de las bandas y los coeficientes de retrodispersión, para cada grupo de SR se calcularon índices que resaltan características del paisaje, principalmente vegetación y humedad. Posteriormente se hablará brevemente de la clasificación de imágenes, resaltando la estrategia de clasificación supervisada de imágenes que representa el paso más importante de nuestra metodología. Finalmente, se describen los datos auxiliares utilizados para calcular el índice de calidad de hábitat propuesto como lo son las coberturas de la tierra CORINE Land Cover adaptadas para Colombia y los registros biológicos.

1.1.1 Sensores Remotos Multiespectrales

La tecnología de sensores remotos ha avanzado considerablemente en las últimas cinco décadas, facilitando a los usuarios la captura e identificación de elementos geo-biofísicos del terreno (Temitope Yekeen & Balogun, 2020), y convirtiéndose en una de las herramientas más eficientes para estudiar la tierra en escalas locales, regionales y globales dado que permiten la recolección de grandes cantidades de información en el espacio y tiempo (Dubovik et al., 2021). Estos avances no solo refinaron la calidad y diversidad de los datos, al aprovechar la interacción de la radiación electromagnética con los objetos de la superficie (Roy et al., 2017). La energía que interactúa con los elementos del terreno y luego es recibida por los sensores, puede ser pasiva si corresponde a energía irradiada (por ejemplo, luz solar recibida por sensores multiespectrales e hiperespectrales, o el sonido recibido por grabadoras y otra tecnología acústica) o activa si es el retorno de una fuente de energía emitida durante la medición (microondas de los sensores SAR y pulsos láser de la tecnología LiDAR y de Escaneo Láser Aerotransportado). Esta diversidad de fuentes de datos, junto con el desarrollo de las capacidades de cómputo y algoritmos

cada vez más complejos, permite la aplicación de sensores remotos satelitales contribuyan al desarrollo de la ciencia, investigación, monitoreo e innovación en temas como la biodiversidad (Minghelli et al., 2021; Mutanga & Skidmore, 2007; Reddy, 2021; Thamaga et al., 2022; Vihervaara et al., 2017), agricultura (Jindo et al., 2021; Martos et al., 2021; Shahzaman et al., 2021; Ullo & Sinha, 2021; Zhang et al., 2022), planeación urbana (Baba et al., 2019; Halder et al., 2021; Shao et al., 2021; Zhang et al., 2021), entre otros campos.

Dependiendo el tipo de sensor que se emplee en la investigación, los datos capturados analizarán la naturaleza de una región del espectro electromagnético que viaja desde la fuente, interactúa con la atmósfera y los objetos muestreados y que finalmente es recibida por el sensor, siguiendo las leyes de la electricidad y el magnetismo. Diferentes regiones del espectro electromagnético son conocidos por distintos nombres como la luz visible, ondas de radio, radiación ultravioleta, microondas, entre otros, refiriéndose a la manera en la cual la radiación es generada o detectada (Chuvienco, 2020) (Figura 1-1). Los sensores multispectrales son sistemas que responden principalmente a la región visible y de infrarrojo cercano del espectro (VNIR, variando entre 400 a 900 nm), aunque algunos también generan respuestas en la región del infrarrojo térmico (TIR, variando entre los 10000 y 12500 nm) (Chuvienco 2020). En estos sistemas la radiación es recibida por el detector, el cual genera una señal eléctrica dependiente de la intensidad de radiación, generando una escena de radiación bidimensional representada en la imagen satelital. Esta clase de datos se denomina multispectrales, por su capacidad de recoger datos en diferentes regiones del espectro electromagnético, siendo los datos de Landsat MSS uno de los primeros productos libres utilizados por los investigadores (R. Pandey et al., 2021).

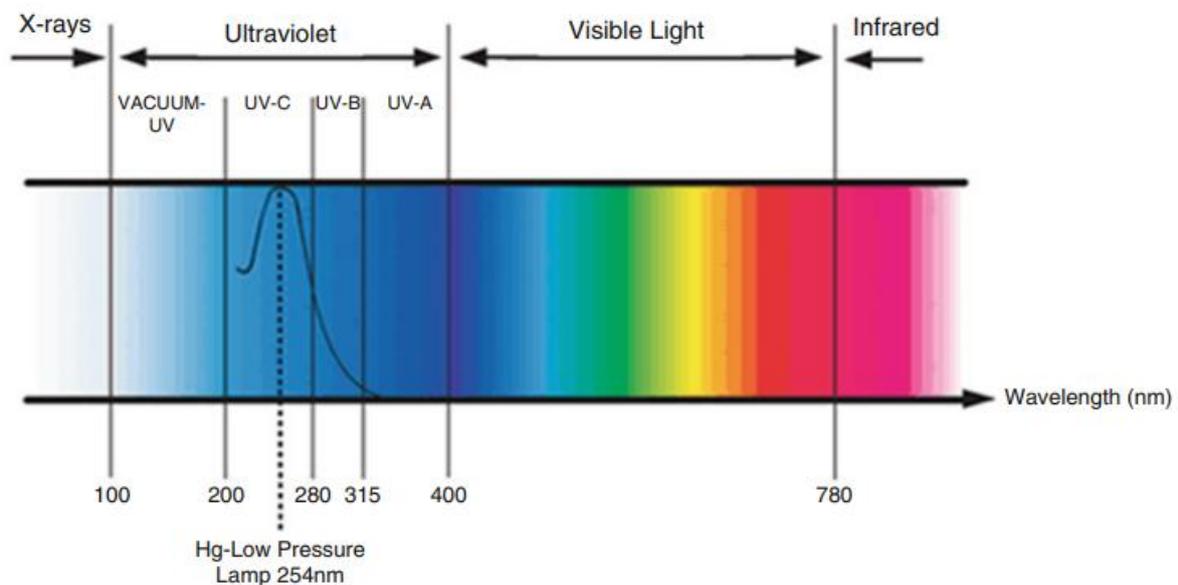


Figura 1-1. Regiones del espectro electromagnético desde rayos X hasta VNIR. Tomada de (Zwinkels, 2015).

La resolución espacial de los datos Landsat MSS comenzaron con un tamaño de pixel de 80 metros, el cual fue continuamente mejorando hasta el más reciente lanzamiento del mismo programa, correspondiente a Landsat 9, con una resolución espacial de 30 metros en VNIR y 15 metros en la banda pancromática (USGS 2022). Dependiendo de la estructura tridimensional de los objetos muestreados con SR, la reflectancia generará una señal espectral relacionada con sus propiedades biofísicas, permitiendo identificar el objeto y sus diferencias con otros elementos en el paisaje (Stenberg et al., 2008); Figura 1-2.

Los productos de Landsat son la mayor fuente de datos multiespectrales, siendo utilizados para análisis tanto a nivel detallado (escala mayor a 1:10.000) como exploratorio (escala menor a 1:100.000); no obstante, otros programas de observación de la tierra como SPOT, LISS III, LISS IV, Sentinel, WorldView-1 y WorldView-2 también desarrollaron sensores multiespectrales capaces de almacenar información con muy alta resolución espacial, llegando incluso a alcanzar valores centimétricos. Una mayor resolución espacial permite el desarrollo de productos con mejor resolución temática y la posibilidad de generar una mayor cantidad de datos para la elaboración de técnicas analíticas más sofisticadas, integradas con otro tipo de información geoespacial y de campo. Esta mejora se logra a partir de la incorporación de múltiples enfoques como las técnicas de segmentación y la facilitación para combinar otros tipos de datos como ópticos con diferente resolución, datos hiperespectrales, datos LiDAR, datos Radar, fusión con información derivada de Sistemas de Información Geográfica, entre muchas otras combinaciones (P. C. Pandey et al., 2021). No obstante, no todos los productos de SR con una detallada resolución espacial ofrecen información precisa sobre todos los atributos de la tierra, pues el sensor debe ser capaz de identificar diferentes características dentro de la escena de la imagen, determinado por su resolución espectral y los objetos que se desea muestrear. Por ejemplo, una fotografía aérea multibanda con resolución centimétrica, no es la más adecuada para evaluar la estructura vertical de la vegetación, mientras que otras opciones como Radar puede ser más acertado para este objetivo. Esta identificación de los objetivos del muestreo es un aspecto clave que debe realizarse previo a la adquisición de datos geoespaciales, para así identificar los productos o combinación de productos que mejor se ajustan a nuestra investigación.

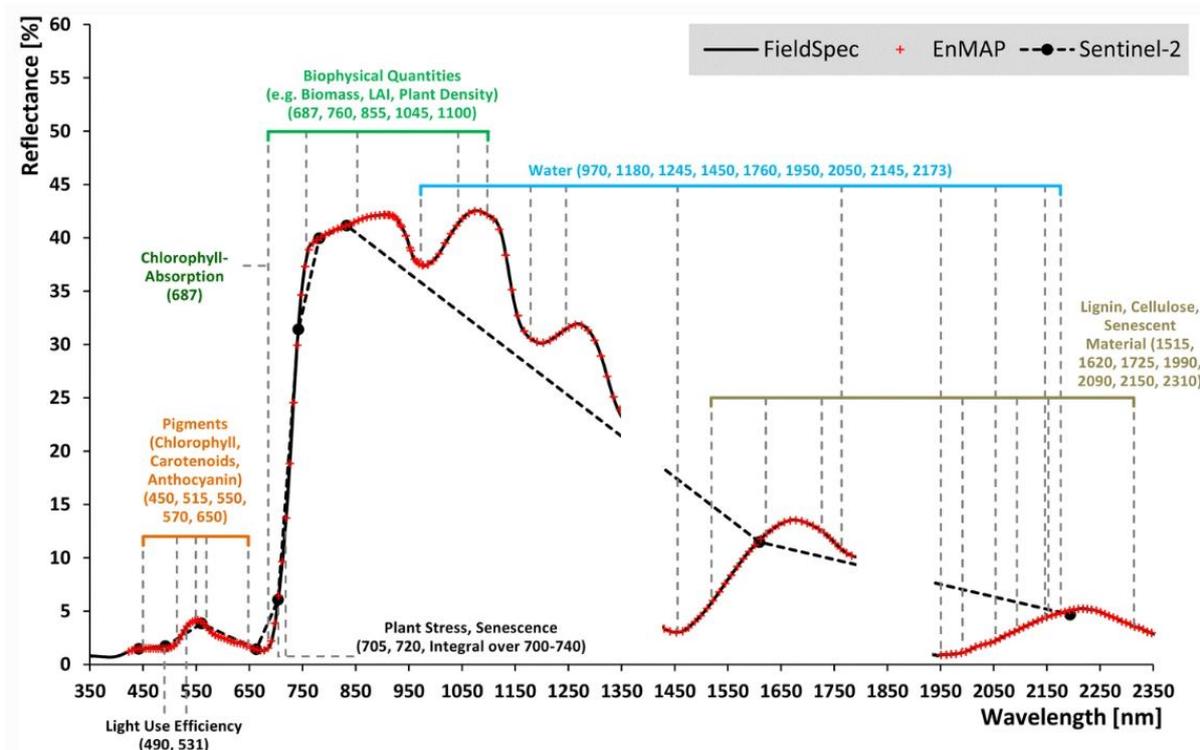


Figura 1-2. Dominios espectrales que responden a cambios de variables específicas en el estudio de la vegetación. Se compara la respuesta espectral obtenida por las bandas de Sentinel-2 (línea negra discontinua), por el sensor hiperespectral EnMap (cruces rojas) y por un espectrorradiómetro de campo cuasi continuo ASD FieldSpec4. Figura tomada de (Hank et al., 2019).

Otro elemento a tener en cuenta es la resolución temporal, que indica la periodicidad en que el sensor captura información de un sitio determinado, esto es especialmente importante cuando el objetivo es entender las dinámicas de cambio de un área. El programa Landsat con su extensa trayectoria, es uno de los insumos multispectrales con mayor resolución temporal y registro histórico de datos espaciales. En la Tabla 1-1 se presentan las características principales de sensores multispectrales potenciales para el desarrollo de esta tesis que se encuentran disponibles en la plataforma de Google Earth Engine (GEE).

Tabla 1-1. Sensores multispectrales disponibles en la plataforma Google Earth Engine

Sensor	Primer lanzamiento	Res. espacial	Bandas y Longitud de Onda (nm)	Res. temporal	Res. radiométrica	Res. espectral	Tamaño de la Imagen
Sentinel-2	2015	10m (B2, B3, B4, B8)	B1: Aerosol 430 - 450	5 días	4096 ND	13 bandas (4 VIS, 6	25 km transversal y 23 km

		<p>20m (B5, B6, B7, B8A, B11, B12)</p> <p>60m (B1, B9, B10)</p>	<p>B2: Blue 450 - 520</p> <p>B3: Green 540 - 570</p> <p>B4: Red 650 - 680</p> <p>B5: Red Edge 1 690 - 710</p> <p>B6: Red Edge 2 730 - 740</p> <p>B7: Red Edge 3 770 - 790</p> <p>B8: NIR 1 780 - 900</p> <p>B8A: NIR 2 850 - 870</p> <p>B9: Water vapour 930 - 950</p> <p>B10: Cirrus 1360 - 1390</p> <p>B11: SWIR 1 1560 - 1650</p> <p>B12: SWIR 2 2100 - 2280</p>			NIR, 3 SWIR	longitudinal para Nivel 0, 1A y 1B. 100 km x 100 km para Nivel 1C y 2A
Landsat 8	2013 (Landsat 1 - 1972)	<p>15m (OLI B8 Pancromática)</p> <p>30m (OLI B1- B7, B9)</p> <p>100m (TIRS B10, B11)</p>	<p>B1: Coastal Aerosol 430 – 450</p> <p>B2: Blue 450 – 510</p> <p>B3: Green 530 – 590</p> <p>B4: Red 640 – 670</p> <p>B5: NIR 850 – 880</p> <p>B6: SWIR 1 1570 – 1650</p>	16 días	4096 ND	11 bandas (5 VIS, 1 NIR, 3 SWIR, 2 It)	185 km x 185 km

			<p>B7: SWIR 2 2110 – 2290</p> <p>B8: Panchromatic 500 – 680</p> <p>B9: Cirrus 1360 – 1380</p> <p>B10: Thermal Infrared 1 10600 – 11190</p> <p>B11: Thermal Infrared 2 11500 - 12510</p>				
MODIS	1999	<p>250 m (B1, B2)</p> <p>500 m (B3-B7)</p> <p>1000 m (B8-B16)</p>	<p>B1: 620 – 670</p> <p>B2: 841 – 876</p> <p>B3: 459 – 479</p> <p>B4: 545 – 565</p> <p>B5: 1230 – 1250</p> <p>B6: 1628 – 1652</p> <p>B7: 2105 – 2155</p> <p>B8: 405 – 420</p> <p>B9: 438 – 448</p> <p>B10: 483 – 493</p> <p>B11: 526 – 536</p> <p>B12: 546 – 556</p> <p>B13: 662 – 672</p>	1 día	1024 ND	36 bandas (roja, azul, IR, case IR, IR media)	2330 km x 2330 km

			B14: 673 – 683				
			B15: 743 – 753				
			B16: 862 - 877				

1.1.2 Sensores Remotos RADAR

Los sistemas VNIR descritos anteriormente corresponden a sistemas pasivos, ya que la fuente de energía que es captada por el sensor proviene de la luz solar. Este aspecto permite conocer de los objetos de estudio propiedades como la intensidad de color, tono, textura y forma. Existen otra clase de sistemas de sensores remotos denominados activos, los cuales emiten su propia energía mediante pulsos activos de energía. Un ejemplo de estos sistemas activos corresponde a la tecnología de Radar (Radio detection and ranging), que opera en la frecuencia de las microondas. Los instrumentos con sistema de Radar posee una antena que transmite radiación de microondas hacia la superficie, parte de la radiación incide sobre los objetos muestreados, produciendo una retrodispersión en un amplio rango de direcciones, algunas de estas dirigiéndose nuevamente a la fuente de emisión (Chuvienco 2020). La longitud de la antena determinará el área de incidencia y su resolución espacial; pero, para evitar el uso de una antena de gran longitud que afecte la integridad física del satélite, se aprovecha el movimiento de este para simular una antena de mayor tamaño, denominada Apertura Sintética (Moreira et al., 2013).

Los sensores de radar de apertura sintética (SAR), que transmiten y reciben energía en frecuencia de microondas (McNairn & Shang, 2016), llevan más de 40 años en uso para la observación de la tierra. El sensor radar detecta las variaciones entre la energía transmitida y captada nuevamente, producidas principalmente por la geometría de las microondas (la cual puede ser polarizada horizontal o vertical en las radiaciones incidentes y/o dispersadas) y las propiedades de retrodispersión de la superficie. Dependiendo del ángulo de incidencia de las microondas y la estructura de los objetos, la retrodispersión capturada por el sensor puede variar, afectando la cantidad de energía recibida por el sensor. En la Tabla 1-2 se presentan las características principales de sensores de Radar potenciales para el desarrollo de esta tesis que se encuentran disponibles en la plataforma de Google Earth Engine.

Tabla 1-2. Sensores de radar disponibles en la plataforma Google Earth Engine

Sensor	Banda	Modos de adquisición y res. espacial	Res. temporal	Opciones de polarización	Tamaño de la escena
Sentinel-1	C ($\lambda = 3.75 - 7.5$ cm)	StripMap (SM, 5m x 5m) Interferometric Wide Swath (IW, 5m x 20m) Extra-Wide swath (EW, 20m x 40m) Wave (WV, 5m x 5m)	6 días	Single HH, VV Dual HH+HV, VV+VH (Excepto para WV)	Hasta 400 km
PALSAR	L ($\lambda = 15 - 30$ cm)	6,25 m 12,5 m	45 días	Single HH, HV	50 km x 70 km

Los sistemas de radar presentan algunas ventajas sobre los sistemas pasivos dada su capacidad de tomar datos de alta calidad tanto en el día como en la noche, además de que no es restringido en áreas de alta nubosidad gracias a la naturaleza físicas de las microondas. Sumado a estas ventajas de uso, los datos de radar muestran sensibilidad a las características estructurales del terreno, ofreciendo información importante respecto a los bordes de los objetos muestreados, forma, composición, entre otros, por lo que, al combinarse con información derivada de sistemas multiespectrales, puede facilitar la caracterización de los objetos analizados de una manera más integral (Pandey et al. 2021). Por supuesto, estos sistemas de microondas también presentan algunas dificultades, siendo una de las más conocidas un tipo característico de granularidad o ruido de la imagen denominada “Speckle”, producida por la interferencia entre todas las ondas de radar que ocurren de forma arbitraria en toda la imagen por efectos de dispersión y reflexión (Jaybhay & Shastri, 2015). Otras características asociadas a la geometría y al ángulo de incidencia del haz de microondas deben ser tomados en cuenta para evitar efectos geométricos y radiométricos, especialmente sobre áreas montañosas, como inversión del relieve, desplazamiento de estructuras y sombras, además de su limitada resolución temporal comparada en comparación con los satélites multiespectrales (Joshi et al., 2016).

1.1.3 Índices espectrales

Gracias a los avances recientes en analítica de sensores remotos, computación en la nube y técnicas de Machine y Deep Learning, los datos capturados por múltiples clases de sensores remotos pueden utilizarse para la construcción de diferentes índices que permitan monitorear procesos y dinámicas sobre la superficie de la tierra, manejo de recursos y para la toma de decisiones en materia ambiental (Maimaitijiang et al., 2020). A partir de esta información se pueden realizar estimaciones de la producción primaria (biomasa de la cobertura vegetal) y sus cambios (especialmente en zonas estacionales) y se evalúa su relación entre funcionamiento y calidad del ecosistema (Wiegand et al., 2008). Contar con un indicador de calidad de hábitat, puede ser estratégico para el monitoreo de áreas protegidas y sus alrededores a escala local, identificar la extensión y condición de hábitats en particular, identificar diversidad y amenaza de especies, identificar el efecto de presiones antrópicas tomando en cuenta tanto las dinámicas multitemporales a la actualidad, así como escenarios de cambio climático (Requena-Mullor et al., 2017). Este conocimiento es necesario a nivel nacional para promover el manejo adecuado de recursos para el cumplimiento de directivas y acuerdos nacionales e internacionales (Nagendra et al. 2013). A continuación, se presentan brevemente algunos índices espaciales que podrían contribuir en la construcción de un modelo espacial multivariado para la evaluación de la calidad de hábitat asociada a ecosistemas de bosque húmedo tropical.

NDVI (Normalized Difference Vegetation Index, (Gao et al., 2022; Huang et al., 2021; P. Li et al., 2021; S. Li et al., 2021): este índice de vegetación fue uno de los primeros en ser formulados para medir el estado de la salud de la vegetación, logrando una alta representatividad en las investigaciones asociadas a la medición de productividad, evapotranspiración, fenología y coberturas de la tierra en los últimos años, llegando a identificarse más de 17.000 publicaciones entre los años 1985 y 2021 (Y. Xu et al., 2022). A pesar de ser uno de los índices más utilizados, en áreas como los bosques tropicales, en donde hay una alta acumulación de biomasa, este índice tiende a saturarse en áreas con alta biomasa, lo que dificulta la discriminación de la vegetación con base en la estructura de dosel (Goswami et al., 2015; Huete et al., 2002; Santin-Janin et al., 2009).

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

Dado el extenso uso de este índice, ya ha sido probado con anterioridad su uso como indicador de calidad de hábitat, evaluando sus variaciones estacionales y la respuesta de especies con estrategias de hibernación (Wiegand et al., 2008), variaciones espaciotemporales en áreas de uso estratégico por especies terrestres y respuesta

de uso ante las características del paisaje (Lu et al., 2022; Mallegowda et al., 2015; Yohannes et al., 2021) y siendo usado como aproximación a la estructura de hábitat (Mairota et al., 2015; Messier, 2023).

EVI (Enhanced Vegetation Index), (Hasanah et al., 2020; Jarchow et al., 2018; Misra et al., 2020): este índice de vegetación es un parámetro clave para las ciencias de la tierra y útil para áreas tropicales, utilizado para monitoreo ecológico, medición de biomasa y vigor de vegetación, evaluación de cambios de coberturas y clasificación de la vegetación (Hasanah & Indrawan 2020). Originalmente fue desarrollado y calibrado para ser empleado por los satélites Terra y Aqua de MODIS, pero ya se estima en plataformas alternativas, demostrando una alta correspondencia con satélites Landsat, VIIRS y Sentinel (Misra et al., 2020; Jarchow et al. 2018). El cálculo multitemporal de este índice de vegetación ha sido empleado para evaluar cambios en la vegetación ante factores climáticos como estrategia para la evaluación de adaptaciones al cambio climático global (Zhong et al., 2021), estimaciones de respuestas ecológicas ante variabilidad climática (Fan et al., 2021), monitoreo de condiciones de inundación (Sangpradid et al., 2021), estimación de coberturas naturales (Sun et al., 2021), entre otros (X. Wang et al., 2022).

$$EVI = G * \frac{(NIR - Red)}{(NIR + C1 * Red - C2 * Blue + L)}$$

Siendo G un factor de ganancia, L un parámetro de ajuste al suelo, C1 y C2 parámetros que junto con la banda Blue ayudan a corregir los aerosoles atmosféricos. Para Sentinel-2, estos parámetros son iguales a 2.5, 1, 6 y 7.5 respectivamente (Fraga et al., 2014). Este índice ha sido empleado como predictor de heterogeneidad de hábitat y como índice correlacionado con la Productividad Primaria Neta, siendo importante como herramienta de monitoreo para medir los impactos producidos sobre el hábitat como la pérdida de biodiversidad y la degradación (Di Febbraro et al., 2018; Sims et al., 2008). También ha demostrado un buen desempeño como variable para medir el funcionamiento ecosistémico, especialmente en la modelación de la distribución potencial de las especies (Requena-Mullor et al., 2014), como medida directa de calidad de parches dentro de la configuración del paisaje para especies nativas e invasoras (Callaghan et al., 2018; Fachinetti & Grilli, 2023) y bajo escenarios de cambio climático (Requena-Mullor et al., 2017).

MNDWI (Modified Normalized Difference Water Index), (Acharya et al., 2018; H. Xu, 2006): esta clase de índices frecuentemente separan la estimación de cuerpos de agua a partir de un umbral óptimo definido, que debe tomarse en cuenta para no confundir los cuerpos de agua con otros elementos de la imagen como las sombras,

bosques, áreas artificiales, nieve y nubes (Acharya et al. 2018). El agua, considerada como parte vital de los ecosistemas y elemento de supervivencia de los seres vivos, se convierte en un indicador apropiado para la evaluación de la calidad de hábitat, no solo por los servicios ecosistémicos que brinda, sino también por sus deservicios como las inundaciones, en el que los sensores remotos se convierten en herramientas para el monitoreo de este proceso (Vishnu et al., 2019). Entre los inconvenientes al momento de realizar monitoreo hidrométrico, se encuentran el crecimiento de la vegetación sobre los cuerpos de agua y las dinámicas rápidas de inundaciones que dificultan el monitoreo a largo plazo, no obstante, el uso de datos de campo o información espacial de muy alta resolución, puede contribuir a resolver esta problemática (Ogilvie et al., 2018).

$$MNDWI = \frac{(Green - SWIR1)}{(Green + SWIR1)}$$

Este índice ha sido empleado principalmente para la evaluación de ecosistemas acuáticos y su calidad de hábitat, como por ejemplo los corredores riparios, delineando la alteración hidrológica producto de las actividades humanas (Pal et al., 2020), así como por los cambios de coberturas y usos de la tierra (Rahimi et al., 2020). Así mismo, se han analizado sus dinámicas espaciotemporales para determinar los efectos de las dinámicas de inundación sobre los servicios ecosistémicos y el uso de hábitat de especies asociadas a cuerpos de agua (Mu et al., 2022).

Vegetation continuous Fields - VCF, (DiMiceli et al., 2021; Sexton et al., 2013): índice para analizar la cobertura de árboles de manera continua, en lugar de manera categórica inferida a través de categorías de coberturas como bosques o sabanas. Esta información puede encontrarse como un set de datos de MODIS, con una resolución espacial de 250 m y una extensión global desde el año 2000 al 2020 (MODIS 2022). Este producto está basado en un algoritmo de árboles de regresión flexible, el cual es más capaz de incorporar información empírica para mejorar la correlación de estimaciones de la cobertura de árboles medida, generando un valor de porcentaje de bosques que puede calcularse de manera continua en el área de estudio (Sexton et al. 2013). A diferencia de los índices anteriores, el VCF también ha sido utilizado como variable auxiliar para caracterizar la conectividad y los efectos de borde sobre las especies a múltiples escalas, además de la degradación y fragmentación de los bosques, siendo esta variable un complemento importante de variables categóricas como las asociadas a coberturas de la tierra (Brennan et al., 2020; Dantas De Paula et al., 2016; Kupfer, 2006; Vergara et al., 2021). Por supuesto, también ha sido utilizado como proxy para medir la estructura y dinámica de la calidad de hábitat para algunas especies (Alaniz et al., 2021).

Plant Senescence Reflectance Index - PSRI, (Vergara et al., 2019). Las especies animales tienen múltiples requerimientos estructurales de hábitat, por lo que no es posible identificar un solo tipo de área que pueda ser considerada útil para todas las especies. Por ejemplo, existen algunas especies que tienen preferencia por la vegetación senescente, como es el caso del carpintero negro de Magallanes (*Campephilus magellanicus*) llegando incluso a cambiar su modo de comportamiento de acuerdo con la calidad de hábitat disponible en su área de distribución (Alaniz et al., 2021; Soto et al., 2017; Vergara et al., 2021). No obstante, para este proyecto se considerará que si la vegetación se encuentra intensamente transformada, muchas especies dependientes de bosques maduros no encontrarán las condiciones necesarias para subsistir en esta área. Por esta razón, el PSRI será un indicador adecuado para cuantificar la senescencia de la vegetación, por medio de la relación carotenoide/clorofila de la reflectancia espectral de las hojas (Vergara et al. 2019).

$$PSRI = \frac{(B04 - B02)}{B06}$$

En los ejercicios de monitoreo de recuperación de parches degradados, el PSRI ha sido empleado como indicador de la tasa de recuperación y resiliencia de parches de vegetación (Z. Song et al., 2023), al ser un indicador eficiente para medir el estrés de dosel (Arshad et al., 2020). Al igual que el NDVI, el PSRI también ha sido evaluado junto con otros índices de vegetación como calidad de hábitat, medido como las áreas con disponibilidad de sustratos de alimentación, particularmente para especies con requerimientos especializados de hábitat (Ansaldó et al., 2021; Soto et al., 2017). Entre los grupos de fauna que mayor se ha evaluado su diversidad y selección de hábitat asociada a este índice de vegetación son las aves (Moudry et al., 2021; Vergara et al., 2021).

Leaf Area Index (LAI, Zheng & Moskal 2009; Song 2013; Kganyago et al. 2020; Li et al. 2018): el fenómeno de intercambio de gases producido por la vegetación puede ser afectado por diferentes procesos como la fotosíntesis, evaporación, transpiración, interceptación de lluvias o flujo de carbono, entender este fenómeno contribuye a un entendimiento de la dinámica de cambios en productividad e impactos climáticos sobre ecosistemas de bosque (G. Zheng & Moskal, 2009). El cálculo de este índice comenzó con enfoques empíricos, localizados en un área determinada y fue evolucionando a enfoques biofísicos, los cuales pueden ser aplicados globalmente (C. Song, 2013). De esta manera, se desarrollaron diferentes métodos para la determinación de esta característica biofísica, incluyendo el uso de diferentes tipos de índices de vegetación calculados a partir de combinación de bandas espectrales, y empleando diferentes métodos de análisis como redes neuronales, inversión de modelos de transferencia radiativa, múltiples enfoques estadísticos multiespectrales, entre otros (Viña et al., 2011). Actualmente, el programa de Sentinel Application

Platform (SNAP) genera este índice y evalúa su consistencia junto con otros productos LAI globales existentes (Kganyago et al., 2020; Q. Li et al., 2018).

$$LAI = \frac{Total\ canopy\ (m^2)}{Ground\ Area\ (m^2)}$$

El índice LAI es otro ejemplo de medición relacionada con importantes procesos, especialmente en la vegetación de dosel como la caracterización de las hojas, la predicción de regímenes de luz, el cálculo de biomasa superficial y la estimación de producción primaria en múltiples escalas, entre otras características biofísicas (Kalacska, 2004; Kanniah et al., 2021; Nemani et al., 1993; Parker, 2020). Además, ya se ha propuesto un índice de calidad de hábitat, usando a LAI como variable de entrada para la expresión del cambio de estructura de los ecosistemas, monitoreando principalmente el cambio de la vegetación, contribuyendo al análisis de cambio climático y al análisis del proceso de fotosíntesis y precipitación (Ma & Liang, 2022).

Fraction of Photosynthetically Active Radiation (FAPAR, Zheng et al. 2022; Yu et al. 2021): Al igual que LAI, el índice FAPAR se correlaciona con la actividad fotosintética de las plantas (Y. Zheng et al., 2022), pero se diferencia en que se refiere solo a la parte verde del dosel, mientras que LAI está definida como la mitad del área con capacidad fotosintética. Este índice se usa al igual que otros como el NDVI, EVI, el Ratio Vegetation Index (RVI) y LAI, como indicador para el monitoreo de degradación de bosques, evaluando el decrecimiento de la calidad de bosques y dando un mejor entendimiento de las emisiones de carbono regionales o globales relacionadas con el cambio climático (Yu et al., 2021). Entre los índices de vegetación más empleados para estimar el índice FAPAR son el NDVI y el Simple Ratio Vegetation Index (SR), no obstante, ya se probaron otros índices de vegetación más apropiados para su estimación que toman en cuenta tanto el área foliar como la geometría de observación de los datos de sensores remotos (Peng et al., 2018).

1.1.4 Clasificación de Imágenes de Satélite

Una de las estrategias utilizadas para caracterizar un paisaje es mediante la clasificación de imágenes de satélite, a partir de las cuales es posible recolectar información relacionada con procesos ecológicos en varios periodos y a escalas desde muy detallada (por encima de la escala 1:10.000) hasta escalas globales. Uno de los análisis más desarrollados que incorporan la clasificación de imágenes de satélite es el monitoreo de cambio de coberturas de la tierra, en donde es posible identificar coberturas de la tierra en múltiples períodos y analizar sus dinámicas de cambio, gracias a la creciente disponibilidad de datos geoespaciales (Sheykhoussa

et al., 2020). Estas dinámicas de transformación ofrecen una perspectiva general del impacto que tienen las actividades humanas sobre las coberturas naturales y sus relaciones para la conservación del ambiente, el manejo de recursos y la planeación de uso del suelo en ecosistemas terrestres y acuáticos (Alam et al., 2020; Allam et al., 2019; Hussain & Karuppanan, 2023; Ruelland et al., 2008; Souza et al., 2020). Una de las constelaciones de satélites artificiales que más imágenes ha generado a lo largo de la historia es la constelación Landsat, la cual captura imágenes satelitales alrededor del mundo desde 1972, siendo de acceso libre desde 2008 (Wulder et al., 2019).

Además de estudiar en conjunto las coberturas de la tierra, muchas investigaciones se centran en coberturas específicas como bosques (Bodart et al., 2013; M. C. Hansen et al., 2013), zonas urbanas (Liu & Yang, 2015), manglares (Sasmito et al., 2019), cultivos (Taiwo et al., 2023), sabanas (Aniah et al., 2023), cuerpos de agua (Leta et al., 2021), entre otros. El poder de clasificación de estos modelos ha aumentado en los últimos años, gracias al aumento de la capacidad de cómputo que permite la elaboración de modelos más complejos, y el acelerado desarrollo de plataformas de múltiples plataformas para la visualización, análisis y descarga de imágenes satelitales, como por ejemplo Google Earth Engine (Perilla & Mas, 2020).

1.1.5 Clasificación Supervisada

Uno de los tipos de aprendizaje automático dentro de los métodos de aprendizaje, es el aprendizaje supervisado, cuyo objetivo principal es el de producir un modelo a partir de datos de entrenamiento etiquetados, que nos permita formular pronósticos sobre datos ocultos o futuros (Rasckka, Liu & Mirjalili, 2023). Básicamente, a los datos usados como entrenamiento se les provee las clases de la variable respuesta que se está modelando, además de los valores de reflectancia, retrodispersión o de la naturaleza que mejor convenga y esté disponible, para que sean usados como clasificadores de entrenamiento y clasificar los datos espectrales en el mapa temático (Dhingra & Kumar, 2019). Se sugiere además, que para efectuar una clasificación supervisada adecuada, se debe tener un conocimiento previo de la zona y los datos (Jog & Dixit, 2016). El resultado de la clasificación puede servir como un producto intermedio, como en el caso de los estudios de degradación y transformación de las coberturas mencionados anteriormente, modelamiento del paisaje, manejo de recursos y de servicios ecosistémicos, entre otras aplicaciones, o ser el producto final como método estandarizado y replicable para la clasificación (Perumal & Bhaskaran, 2010).

1.1.6 Coberturas de la Tierra - Metodología CORINE Land Cover adaptada para Colombia

Esta metodología fue originalmente construida para los países europeos en el año 1985, pero luego en el año 2004 comenzó su implementación en Colombia sobre la cuenca hidrológica del Río Magdalena, abarcando una superficie de más de 274.000 km cuadrados (Feranec, 2016). A pesar de que diversos autores generaron algoritmos para la determinación de las coberturas de la tierra, identificando las categorías definidas por CORINE (Coordination of Information on the Environment) (Ruggeri et al., 2021), el mapa temático oficial para Colombia es desarrollado mediante fotointerpretación de imágenes pre-procesadas, con apoyo de datos auxiliares (Mancera, 2019).

En términos generales, la metodología CORINE Land Cover identifica cinco tipos principales de cobertura (denominados Nivel 1) que posteriormente se van subdividiendo en coberturas más específicas. De acuerdo con el Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM), las coberturas Nivel 1 para Colombia se definen como sigue:

1. **Territorios artificializados:** Comprende las áreas de las ciudades y las poblaciones y, aquellas áreas periféricas que están siendo incorporadas a las zonas urbanas...
2. **Territorios agrícolas:** Terrenos dedicados principalmente a la producción de alimentos, fibras y otras materias primas industriales, ya sea que se encuentren con cultivos, pastos, en rotación y en descanso, etc...
3. **Bosques y áreas seminaturales:** Grupos de coberturas boscosos, arbustivos y herbáceos, desarrolladas sobre diferentes sustratos y pisos altitudinales que son el resultado de procesos climáticos...
4. **Áreas húmedas:** Coberturas constituidas por terrenos anegadizos, que pueden ser totalmente inundados y estar parcialmente cubiertos por vegetación acuática, ubicados en los bordes marinos e interior del continente...
5. **Superficies de agua:** Cuerpos y cauces de aguas permanentes, intermitentes y estacionales, localizados al interior del continente y adyacentes a la línea de costa continental...

Esta información auxiliar permitirá generalizar la información de los registros biológicos (detallados a continuación), la cual pasa de ser una información puntual (registro de ocurrencia) a la de un parche (polígono) en la que se asume que el individuo puede aprovechar todo el parche delimitado, y que la cobertura captura su requerimiento de hábitat, por lo que tendrá condiciones ambientales similares. De esta manera, capturamos la presencia de las especies desde una perspectiva ecológica, y adicionalmente ampliamos el área con información de presencia de las especies evaluadas en este proyecto.

1.1.7 Registros Biológicos

Las bases de datos en la nube para el almacenamiento de registros biológicos, han adquirido gran popularidad debido a su capacidad de almacenar una gran cantidad de registros biológicos, describiendo no solo las distribuciones actuales de especies de interés, sino las tendencias en los sesgos temporales y espaciales obtenidas al analizar datos históricos (Bowler et al., 2022). Con esta gran cantidad de información histórica y presente, es posible abordar diversas preguntas de investigación como las estimaciones de diversidad, riqueza y biomasa de especies clave (Zattara & Aizen, 2021), abordar problemáticas del propio almacenamiento y curaduría de esta vasta cantidad de información (Zizka et al., 2020), así como uno de sus principales análisis que es la identificación de herramientas y metodologías para estimar de manera precisa la distribución potencial y nicho ecológico de las especies (Aiello-Lammens et al., 2015; Curd et al., 2023; Melo-Merino et al., 2020; Mondanaro et al., 2023).

Únicamente en la plataforma de Global Biodiversity Information Facility (GBIF) hay disponibles más de dos mil millones de registros biológicos, incluyendo datos de colecciones paleontológicas (Zizka et al., 2019). A pesar de almacenar esta gran cantidad de información, un porcentaje considerable de los registros son imprecisos en su georreferenciación, en su fecha o en su resolución taxonómica (Zizka et al., 2020), por lo que el porcentaje de registros útiles para el modelamiento es mucho menor. Entre menor sea el número de registros disponibles para la modelación, mayor será su efecto en el desempeño de los modelos de distribución de especies, evidenciado en la disminución de la predicción del modelo e incremento en la variabilidad entre especies y modelos (Beck et al., 2014; Wisz et al., 2008). A pesar de estas limitaciones, hay múltiples investigaciones enfocadas en el desarrollo de filtros para la eliminación de registros o de métodos que puedan sobrellevar estas problemáticas, facilitado además por las múltiples interfases desarrolladas para conectar la plataforma de GBIF con herramientas para la búsqueda, preprocesamiento y análisis de datos como R, Python, Ruby, y para el caso de Colombia, el Sistema de Información sobre Biodiversidad de Colombia (Chamberlain & Boettiger, 2017).

Estos registros brindarán la información de presencia de especies, desde las que son altamente dependientes de los bosques, hasta las especies introducidas, invasoras o que no dependen primordialmente del bosque. De esta manera, se identificarán los sitios que poseen una alta concentración de especies y se contrastarán con la información derivada de sensores remotos para identificar sitios de congruencia (valores de índices de vegetación asociadas a alta calidad de bosque con especies de alta dependencia al bosque) e incongruencia (valores de índices de vegetación asociados a alta calidad de bosque con especies que no dependen del bosque o que

son indicativas de degradación del bosque, o valores de índices de baja calidad de bosque con especies altamente dependiente de bosque). Con la intersección de los datos de SR y los registros biológicos, se podrán identificar las fortalezas, aspectos a mejorar y estrategias de investigación que se podrán realizar en la zona de estudio, a través de nuestro índice de calidad de hábitat propuesto.

1.1.8 Integridad Ecológica

La integridad ecológica puede definirse como las características que mantienen un ecosistema estable, sostenible y activo, manteniendo su organización y autonomía a través del tiempo, además de brindarle la capacidad de retomar a las condiciones anteriores a una perturbación (Mora, 2017). Generalmente, este término es asociado con tres componentes que deben tenerse en cuenta para una evaluación real de la integridad ecológica: Estructura, Composición y Función (Tierney et al., 2009). No obstante, los indicadores utilizados para medir cada componente, deben ser cuidadosamente elegidos, dependiendo el tipo de proceso o interacción que se desea evaluar (Figura 1-3).

Los indicadores de integridad ecológica que se consideren relevantes para un estudio en particular se pueden utilizar como estrategias para la creación de herramientas para la toma de decisiones en la escala predeterminada (Mora, 2019; Pan et al., 2021). Los indicadores de integridad ecológica permiten el desarrollo de métodos efectivos para la generación de conocimiento, especialmente ante la pérdida de integridad de los sistemas naturales producto de las actividades antrópicas y que incorporan herramientas multidisciplinarias para la integración de factores bióticos y abiótico como estrategia para la medición, evaluación y monitoreo de los ecosistemas estudiados (Borja et al., 2008; Karr et al., 2022). Varios índices de SR satelitales han demostrado ser útiles como indicadores de integridad ecológica. Por ejemplo, (Zelený et al., 2021), cuantifican la integridad ecológica regional a partir de índices como el NDVI, temperatura y heterogeneidad de la vegetación de superficie, al considerarlos como proxys de la captura de exergía y de la heterogeneidad biótica y abiótica del paisaje. Este es un ejemplo de cómo la integración de datos de SR pueden ser usados para la identificación y potencial monitoreo de la integridad ecológica y sus servicios ecosistémicos derivados, al igual que estos otros trabajos (Sánchez-Giraldo et al., 2021; Soubry et al., 2021).

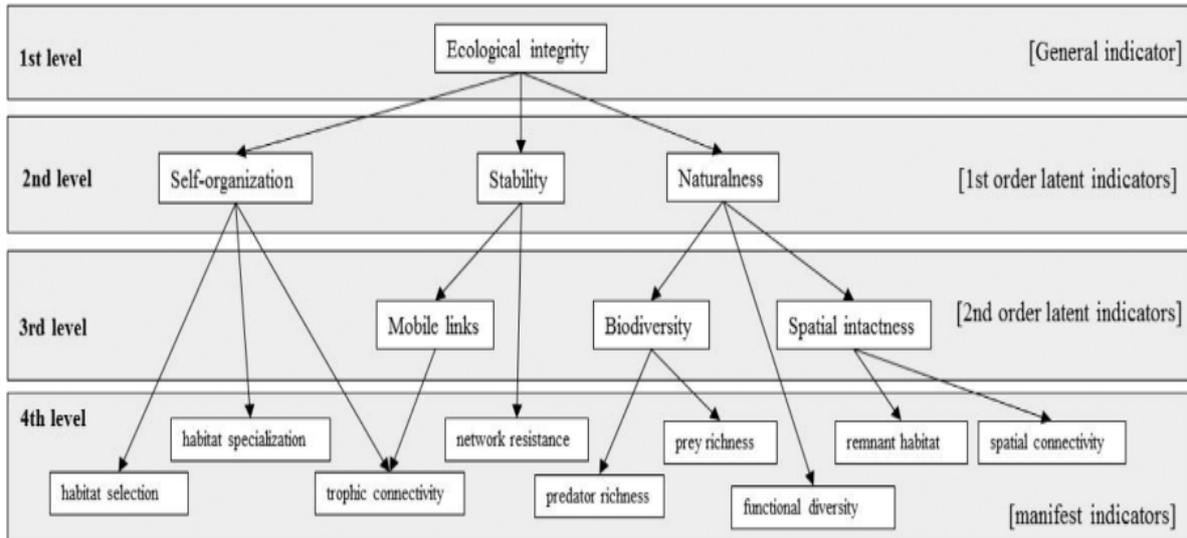


Figura 1-3. Ejemplo de marco de jerarquía de integridad ecológica para evaluar la condición en los ecosistemas naturales en función de las características del paisaje que sustentan la interacción depredador-presa. Tomado de (Mora, 2017).

1.2 Estado del Arte

1.2.1 Clasificación Supervisada y Técnicas de Aprendizaje de Máquinas

El catálogo de algoritmos disponibles para la clasificación de imágenes satelitales se ha ampliado en las últimas décadas, gracias a la evolución de las capacidades de cómputo y la amplia variedad de información geoespacial generada por SR, que buscan responder a las necesidades de mejorar en la toma de decisiones de múltiples industrias (Sajjad & Kumar, 2019). Esta disponibilidad de datos y herramientas de cómputo ha resultado en su amplio uso, no solo por parte de las instituciones de investigación y gubernamentales, sino también por sectores privados y usuarios independientes que buscan generar soluciones a problemas complejos. Dado que un creciente volumen de información está directamente ligado con el desarrollo de técnicas de almacenamiento y análisis más complejos, los algoritmos de aprendizaje también se han diversificado.

Se denomina aprendizaje de máquina dada la capacidad que tiene un equipo de cómputo para tomar decisiones y hacer predicciones, con base en la experiencia asignada mediante los datos de entrenamiento (Ray, 2019). Los algoritmos de aprendizaje pueden dividirse en cuatro tipos diferentes: 1) Supervisado, 2) Semi-Supervisado, 3) No Supervisado y 4) Aprendizaje de refuerzo (Figura 1-4). Este

proyecto se centrará en los algoritmos de aprendizaje supervisado, que utiliza datos etiquetados (labeled) y una colección de datos de entrenamiento para inferir una función.

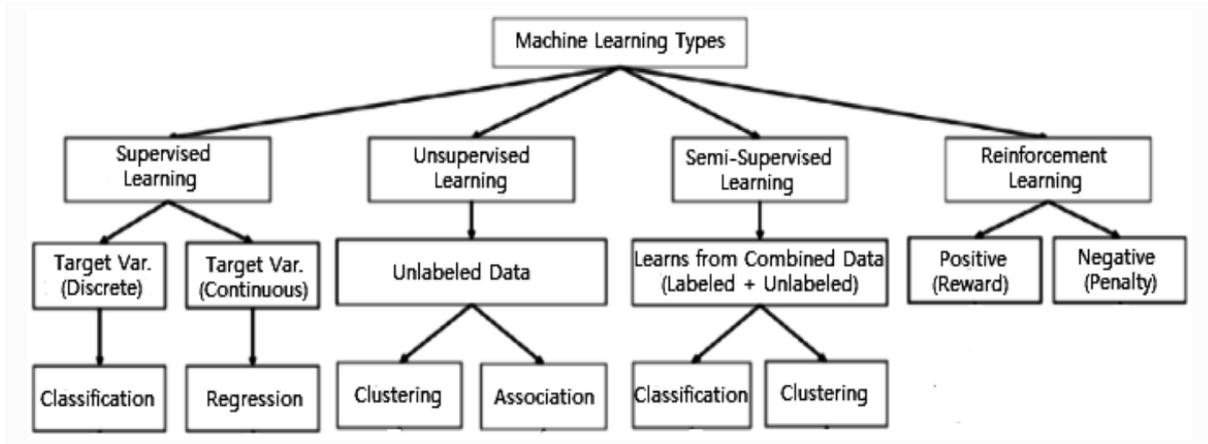


Figura 1-4. Tipos de técnicas de aprendizaje de máquina. Figura tomada de (Sarker, 2021)

La efectividad y eficiencia de un ejercicio de clasificación supervisada, dependerá en gran medida de la naturaleza de los datos, el algoritmo seleccionado y del ajuste de los hiperparámetros (Sarker, 2021). Entre los algoritmos más empleados para realizar ejercicios de clasificación supervisada se encuentran los árboles de decisión, el cual puede ser utilizado tanto para ejercicios de clasificación como de regresión. En un árbol de decisión usado para clasificación, la etiqueta se encuentra en las hojas, mientras que las variables predictivas se dividen en los nodos (Ray, 2019). Este algoritmo tiene las ventajas de ser fácilmente interpretable, tener una alta eficiencia computacional y, cuando se realiza un ensamble de múltiples árboles, puede reducir el sobreajuste del resultado final, al ser un ensamble en donde participan múltiples clasificadores, disminuyendo la varianza especialmente en el caso de clasificadores inestables (Figura 1-5) (Sheykhmousa et al., 2020).

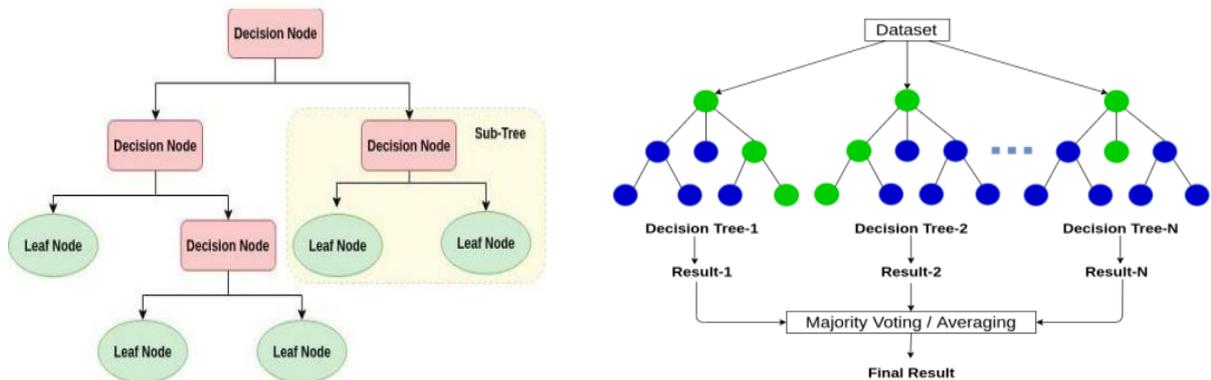


Figura 1-5. Diagrama de flujo de un árbol de decisión (izquierda) y de un Random Forest (derecha). Figuras tomadas y adaptadas de (Abdulkareem & Abdulazeez, 2021).

El ensamble de RF, que permite disminuir la varianza de un modelo complejo, es una modificación de la técnica de Bagging, el cual está diseñado para mejorar la estabilidad y exactitud, considerando los árboles individuales como modelos independientes, y ajustando el error mediante decisión mayoritaria (Strobl et al., 2009). La diferencia principal entre RF y BAG, es que el primero tiene mayor capacidad para lidiar con la correlación entre árboles, al añadir un componente de aleatoriedad en la construcción de estos. Este nuevo componente se obtiene de elegir de manera aleatoria los predictores candidatos para los cortes binarios que formarán los nodos, reduciendo así la colinealidad entre los árboles que conformarán el bosque aleatorio (Casal et al., 2021).

Otra técnica de ensamble de modelos para aumentar la exactitud integrando múltiples clasificadores es el método de Boosting, que a diferencia de Bagging no genera múltiples clasificadores independientes, sino que crea una secuencia de iteraciones donde cada modelo busca ajustar el error del modelo anterior (Figura 1-6) (Sheykhmousa et al., 2020). Uno de los algoritmos que utiliza este método de ensamble es el conocido como Adaptive Boosting (AdaBoost), el cual está estrechamente relacionado con otro algoritmo denominado LogitBoost, ya que ambos desarrollan una regresión logística adaptativa. La diferencia entre estos dos algoritmos, es que LogitBoost está basado en probabilidades condicionales de Bernoulli, minimizando la pérdida logística y dándole un mayor peso a las clases menos representadas, las cuales afectan el desempeño de AdaBoost (Maalouf, 2011; J. Song et al., 2011).

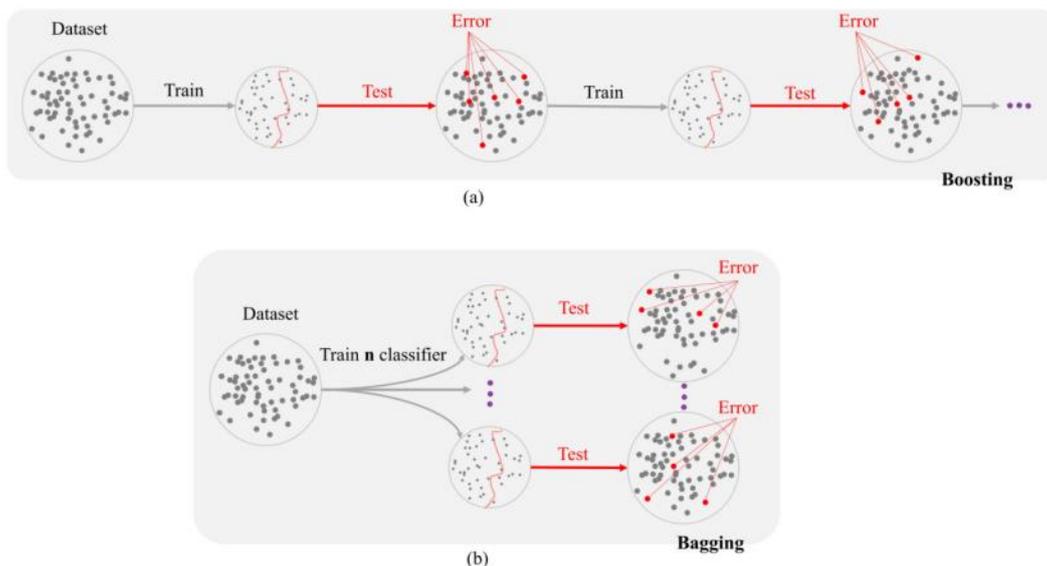


Figura 1-6. Diferencias entre los métodos de ensamble más utilizados en modelos de aprendizaje de máquina. Figura tomada de Sheykhmousa et al. (2020).

Gracias a las mejoras identificadas con los métodos de ensamble, se diseñó un método híbrido entre boosting y bagging denominado Stochastic Gradient Boosting, en el que se generan muestras aleatorias, a partir de las cuales se desarrolla Boosting. Cada iteración se basa en un algoritmo de gradiente pronunciado, formando árboles pequeños que posteriormente son sumados, seleccionando el valor de clasificación más común (Lawrence, 2004). La combinación de estos efectos reduce la sensibilidad del resultado final a los datos de entrenamiento inadecuados, outliers, y datos desbalanceados.

Otro de los algoritmos de clasificación más ampliamente utilizados es Support Vector Machine (SVM), el cual es un algoritmo de aprendizaje basado en Kernels y con especial uso en clasificación de imágenes. SVM pertenece a las técnicas supervisadas no paramétricas, los cuales son insensibles a la distribución de los datos, y cuyo objetivo es resolver un problema de optimización cuadrática convexa para obtener un hiperplano o grupo de hiperplanos, para separar el set de datos en un número discreto de clases predefinidas, usando los datos de entrenamiento (Figura 1-7) (Sarker, 2021; Sheykhmousa et al., 2020). Entre las dificultades identificadas con esta técnica, es que el desempeño disminuye con grandes cantidades de datos, ya que dificulta encontrar el kernel apropiado para la clasificación.

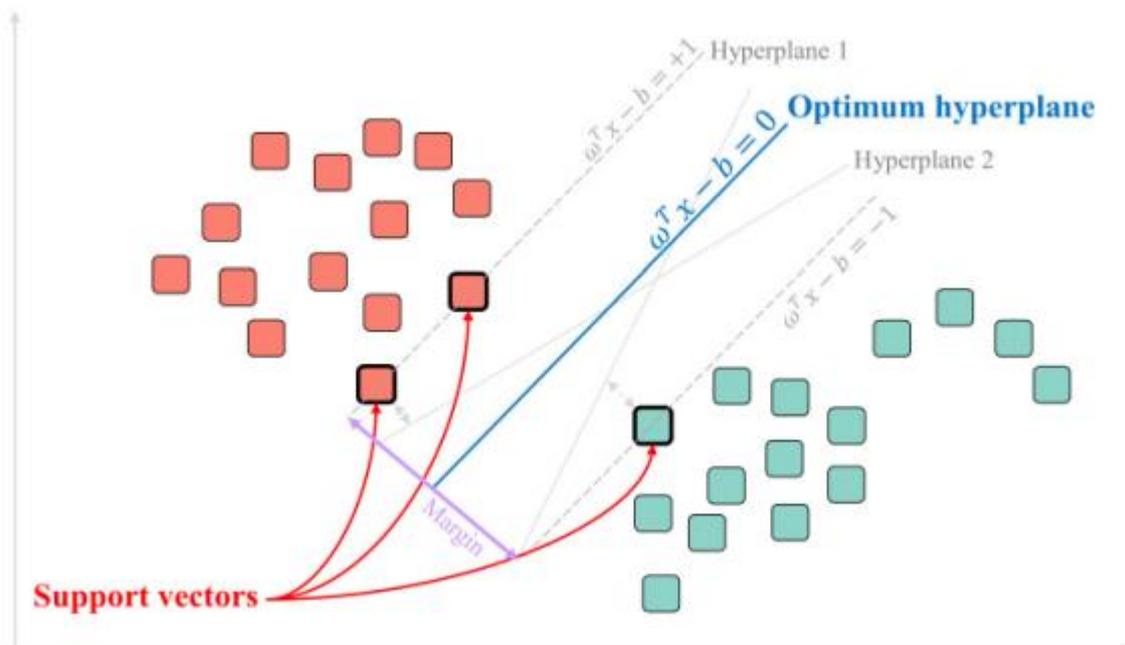


Figura 1-7. Plano de separación por un hiperplano generado a partir de un Kernel lineal. Figura tomada de Sheykhmousa et al., 2020.

1.2.2 Integración de datos de múltiples productos satelitales y datos de campo

La integración de datos consiste en un modelo estadístico que incorpora múltiples fuentes de datos dentro de un marco analítico unificado (Zipkin et al., 2021). Entre los objetivos de esta integración de datos, se encuentra expandir el alcance espacio-temporal de la investigación para la realización de inferencias, aumentar la precisión de los parámetros estimados, y analizar múltiples fuentes de incertidumbre en procesos de estimación multiescala.

Frente a la integración de diferentes tipos de datos capturados por SR, la combinación de datos multiespectrales y de Radar es relevante para la investigación de la ecología de especies, estructura horizontal y vertical, composición y función de los ecosistemas en múltiples escalas (Schulte To Bühne & Pettorelli, 2018). (X. Wang et al., 2022) integraron la información disponible en tres fuentes diferentes de SR (Lidar, MODIS y ALOS PALSAR) para la identificación de biomasa superficial en ecosistemas de bosque, alcanzando niveles de exactitud temática de cerca del 82%. También se evidencia que las estrategias de fusión de datos pueden brindar información más precisa sobre la geología y geomorfología de la tierra (Abdelkareem et al., 2020).

La integración de los datos no necesariamente debe hacerse exclusivamente entre SR satelitales, también la integración con datos de campo y otro tipo de información espacial diferente a la de los satélites, pueden generar productos de alta calidad. Como se presentó a lo largo del texto, desde los SR se pueden analizar variables del paisaje como características espectrales, de textura, temporales y espaciales, mientras que de otras fuentes como las bases de datos geográficas oficiales, es posible identificar características relacionadas a censos sociales, datos de medios sociales y estrategias comunitarias para la generación de información geográfica local, todo esto utilizado para la planeación y el mapeo del uso de la tierra (Yin et al., 2021).

En la sección 1.1.7 se mencionó la importancia de los registros para calibrar y validar SDM, así como la robustez que tienen estos modelos. Muchos de los SDM dependen de las estimaciones de variables bioclimáticas, que no tienen la resolución espacial adecuada para análisis locales. No obstante, con la integración de información adicional a la de registros, es posible realizar un monitoreo más detallado al incluir aspectos relacionados con presiones humanas directas e indirectas, disturbios físicos

como interacciones bióticas, regímenes de fuego, y regulaciones climáticas, micro climáticas y topográficas (Leitão & Santos, 2019; Randin et al., 2020; Szpakowski & Jensen, 2019; R. Wang & Gamon, 2019).

2. Objetivos

2.1 Objetivo General

Estudiar la concordancia entre información de sensores remotos y la calidad de hábitat en ecosistemas de bosque húmedo tropical en áreas del Magdalena Medio colombiano, obtenida a partir de la implementación de técnicas de clasificación con algoritmos de aprendizaje de máquina supervisada en imágenes multiespectrales y SAR y ajustado con datos de biodiversidad tomados en campo.

2.2 Objetivos Específicos

- Evaluar el uso de bandas e índices radiométricos derivados de sensores multiespectrales y SAR, mediante la medición de importancia de las variables explicativas en cada uno de los modelos basados en algoritmos de aprendizaje de máquina, para la identificación de la estructura de ecosistemas de bosque húmedo tropical en áreas del Magdalena Medio colombiano.
- Generar una metodología de análisis con bandas e índices radiométricos para aproximar la calidad de hábitat de bosques con datos de biodiversidad in situ, a partir de los modelos de aprendizaje de máquina con mejor desempeño, disponibles en áreas de bosque húmedo tropical en el Magdalena Medio colombiano.
- Identificar zonas de congruencia e incongruencia entre la estructura y la composición de bosques húmedos tropicales, para evaluar y mejorar el modelamiento de calidad de hábitat, incorporando la identificación de vacíos y potencial de investigación en estos ecosistemas.

3. Datos y Métodos

3.1 Zona de Estudio

Este análisis fue realizado en un área de aproximadamente 290.000 hectáreas de los humedales y bosques húmedos del Valle del Magdalena Medio (VMM), que pertenecen al hotspot de biodiversidad Tumbes-Choco-Magdalena, altamente amenazado del cual solo queda el 24% de su distribución potencial (Achury & Suarez, 2018). El área también forma parte de la ecorregión de bosques húmedos Magdalena-Urabá, un ecosistema único con un alto grado de endemismo y diversidad de especies, y el lugar de llegada de muchas aves migratorias. En el área se han registrado ~18% de las aves colombianas (345 especies), 20% de roedores (27 especies), 16% de primates (6 especies), 29% de los murciélagos (58 especies), 3% de las plantas vasculares (723 especies) y 5% de los anfibios (42 especies).

Las especies clave en el área incluyen monos araña (*Ateles hybridus*, CR, casi endémico y una de las diez especies de primates más amenazadas del mundo), paujiles de pico azul (*Crax alberti*, CR, endémico), titíes de patas blancas (*Saguinus leucopus*, EN, endémica), titíes cabeciblancos (*Saguinus oedipus*, CR, endémica), tapires sudamericanos (*Tapirus terrestris*, VU), entre otros. Los impulsores de la pérdida y degradación de los ecosistemas naturales en esta área son la concentración de la propiedad de la tierra, la rápida expansión de la agricultura extensiva, la ganadería, la minería, la explotación petrolera, la extracción de madera, el tráfico ilegal de especies, y la contaminación de sus ríos más importantes (Magdalena y Cauca) principalmente (Gómez-Lora et al., 2021; Schipper, s. f.). En 2015, las autoridades regionales declararon un área protegida en la región, y desde entonces se ejecutan proyectos de conservación y restauración de los bosques húmedos de la zona, convirtiendo a este ecosistema en una estrategia de conocimiento, divulgación y apropiación del territorio (Figura 3-1).

El límite del área de estudio fue definido con el fin de representar un gradiente de impactos sobre los ecosistemas naturales y el grado de respuesta que presentan los habitantes frente a la problemática de la degradación y transformación de las coberturas y usos de la tierra. En la parte norte del área de estudio, se ubica un área denominada Guane-Kalypso, en donde se están adelantando proyectos piloto de investigación integral sobre yacimientos no convencionales de hidrocarburos, utilizando la técnica de fracturamiento hidráulico multietapa con perforación horizontal

(Instituto Humboldt, 2022). Se levantó una Línea Base general en la zona, incluyendo entre otros aspectos, información sobre ecosistemas y biodiversidad (Figura 3-2). Esta área representa entonces una sección del gradiente de transformación, resaltando áreas con un alto grado de impacto sobre sus ecosistemas naturales y donde las actividades minero energéticas prevalecen sobre las estrategias de conservación y restauración.

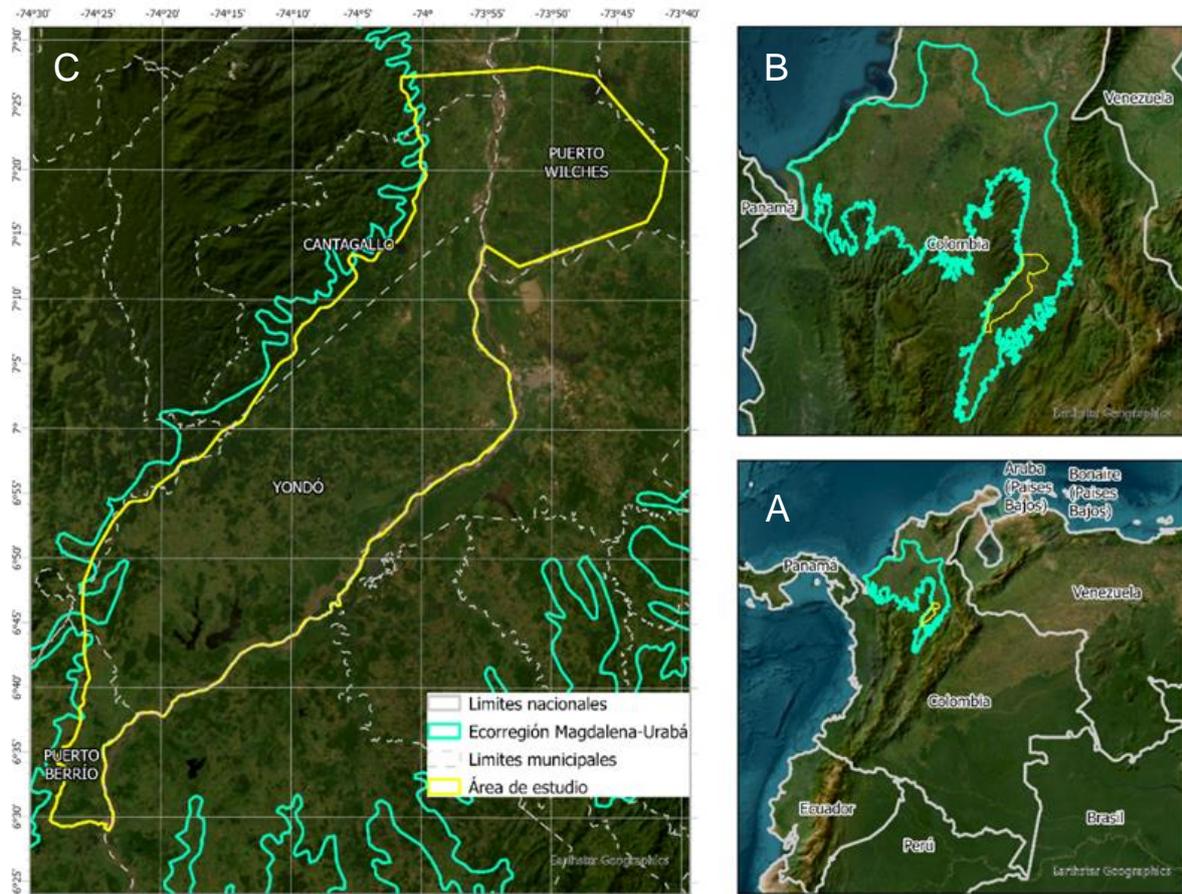


Figura 3-1. Área de estudio. A) Ubicación en Sudamérica. B) Ubicación del área en la ecorregión Magdalena-Urabá, situada al norte de Colombia y conectando el norte de las ecorregiones de Mesoamérica y el Chocó con la ecorregión andina. C) Polígono de estudio ubicado entre los municipios de Puerto Badel, Yondó, Cantagallo y Puerto Wilches, Antioquia. Fuente: Autor

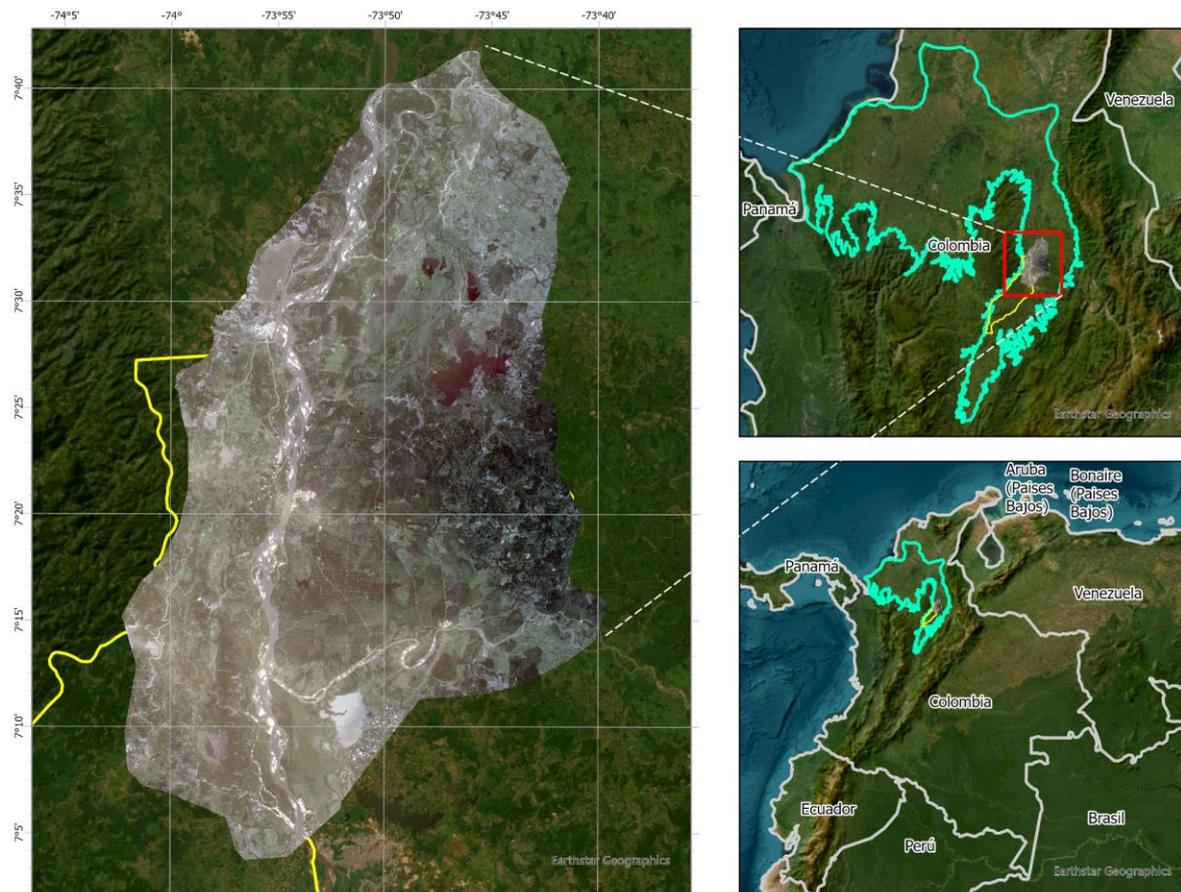


Figura 3-2. Mosaico de imágenes PlanetScope con resolución espacial de 30 m.

Por otro lado, en la región sur del área de estudio, algunos propietarios han mostrado interés en actividades de reconversión de ganadería extensiva a estrategias silvopastoriles que fomenten la conservación de los últimos relictos de bosque, además de incluir la delimitación del DRMI Ciénaga de Barbacoas, declarada como área protegida en 2017 (Fundación Biodiversa Colombia, 2014; Figura 3-3). En la región sur se encuentran entonces áreas históricamente degradadas, entre las cuales algunas muestran recuperación de sus coberturas naturales, además de presentar los parches de bosques mejor conservados en el Magdalena Medio, como lo son los parches ubicados en las inmediaciones del Río San Bartolo. Para incluir todo este gradiente de transformación y estrategias para combatirlo, se construyó el polígono de la Figura 3-1, teniendo como guía no solo los límites político-administrativos, sino la configuración del paisaje para responder las preguntas de investigación en un área definida desde una perspectiva ecológica.

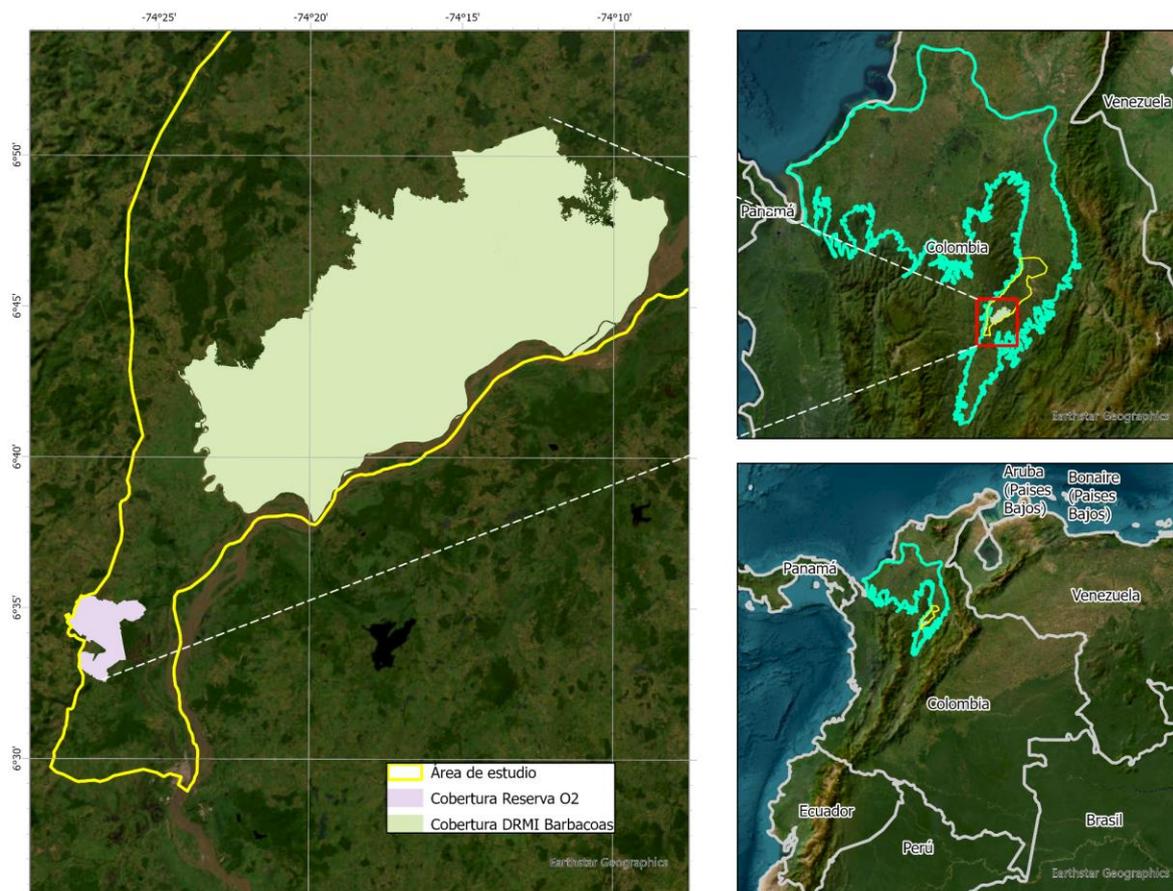


Figura 3-3. Polígonos identificados en el área de estudio, donde se implementan estrategias de conservación y restauración de bosque húmedo tropical mediante diferentes estrategias.

3.2 Registros Biológicos

Entre los registros utilizados para este proyecto, se encuentran los datos generados en el Convenio Interadministrativo No. 21-095 (264 de 2021 ANH) entre la Agencia Nacional de Hidrocarburos y el Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, en donde se realizó un levantamiento de: “*flora, mamíferos, anfibios, reptiles, aves, colémbolos, himenópteros terrestres, lepidópteros diurnos, coleópteros escarabeidos y melolóntidos, e insectos estridulantes, microorganismos, ictiofauna, macrófitas, macroinvertebrados, perifiton, fitoplancton y zooplancton*”. En este ejercicio de levantamiento de línea base, las áreas de muestreo se definieron para evaluar la variabilidad espacial y temporal de diferentes grupos biológicos y niveles de organización, con relación a factores bióticos y abióticos del territorio. Las unidades de muestreo finales se enfocaron en el margen derecho del Río Magdalena, en donde se encuentra el Heliobioma e Hidrobioma Magdalena Medio y Depresión Momposina, identificando las áreas con mayor biodiversidad de biomas, además de capturar un gradiente de transformación medida con el Índice de Huella Espacial Humana.

Para cada grupo biológico se seleccionaron entre 20 y 100 unidades de muestreo, en la que cada área reúne características específicas de cobertura vegetal, tipo de suelo y orden de drenaje. Las técnicas de muestreo y su esfuerzo por unidades de muestreo se ajustaron de acuerdo con el grupo biológico a estudiar, utilizando estrategias de puntos de conteo y redes de niebla (aves y mamíferos pequeños voladores), búsquedas por encuentro visual (anfibios y reptiles), trampas Sherman (mamíferos pequeños terrestres), parcelas (plantas), trampas de caída, trampas Winkler y cebos de atún (hormigas y coleópteros), trampas van Someren Rydon (Lepidópteros diurnos), puntos de muestreo (Colémbolos), cámaras trampa y grabadoras Audiomoth.

Además del levantamiento de registros del Convenio 21-095, se realizó una búsqueda exhaustiva en bases de datos nacionales e internacionales, principalmente GBIF y los registros curados utilizados por el Instituto Alexander von Humboldt, usados para la elaboración de Biomodelos, la cual es una plataforma elaborada para desarrollar modelación de distribución de especies con ayuda de una red de más de 500 expertos en diferentes taxones. También revisamos registros provenientes de proyectos recientes de caracterización en varios grupos, y que no se encuentran todavía en el SIB Colombia.

3.3 Metodología

Este proyecto se desarrolló en tres fases de modelamiento y en cada una se explora de forma diferente qué información de SR es útil para describir la calidad del hábitat para múltiples especies. Estos modelos se describen a continuación:

3.3.1 Modelo 1: Estimación del Índice de Condición Estructural (SCI) con variables derivadas de Sensores Remotos

En este modelo exploramos qué información de SR se correlaciona mejor con el índice de condición estructural forestal (SCI).

Variable respuesta: El SCI es una variable de tipo cuantitativa y discreta, que distingue los bosques pequeños, de dosel abierto, o recientemente deforestados (SCI = 1), de los bosques altos, de dosel cerrado e intacto en su estructura (SCI= 18) (Hansen et al., 2019; Figura 3-4). El SCI está compuesto por tres variables que se enfocan en la estructura de los bosques: 1) la altura, 2) la cobertura del dosel por pixel, y 3) el tiempo de intervención, por lo que consideramos esta variable como un primer proxy para evaluar la estructura de los bosques y como primer componente del índice de calidad de hábitat.

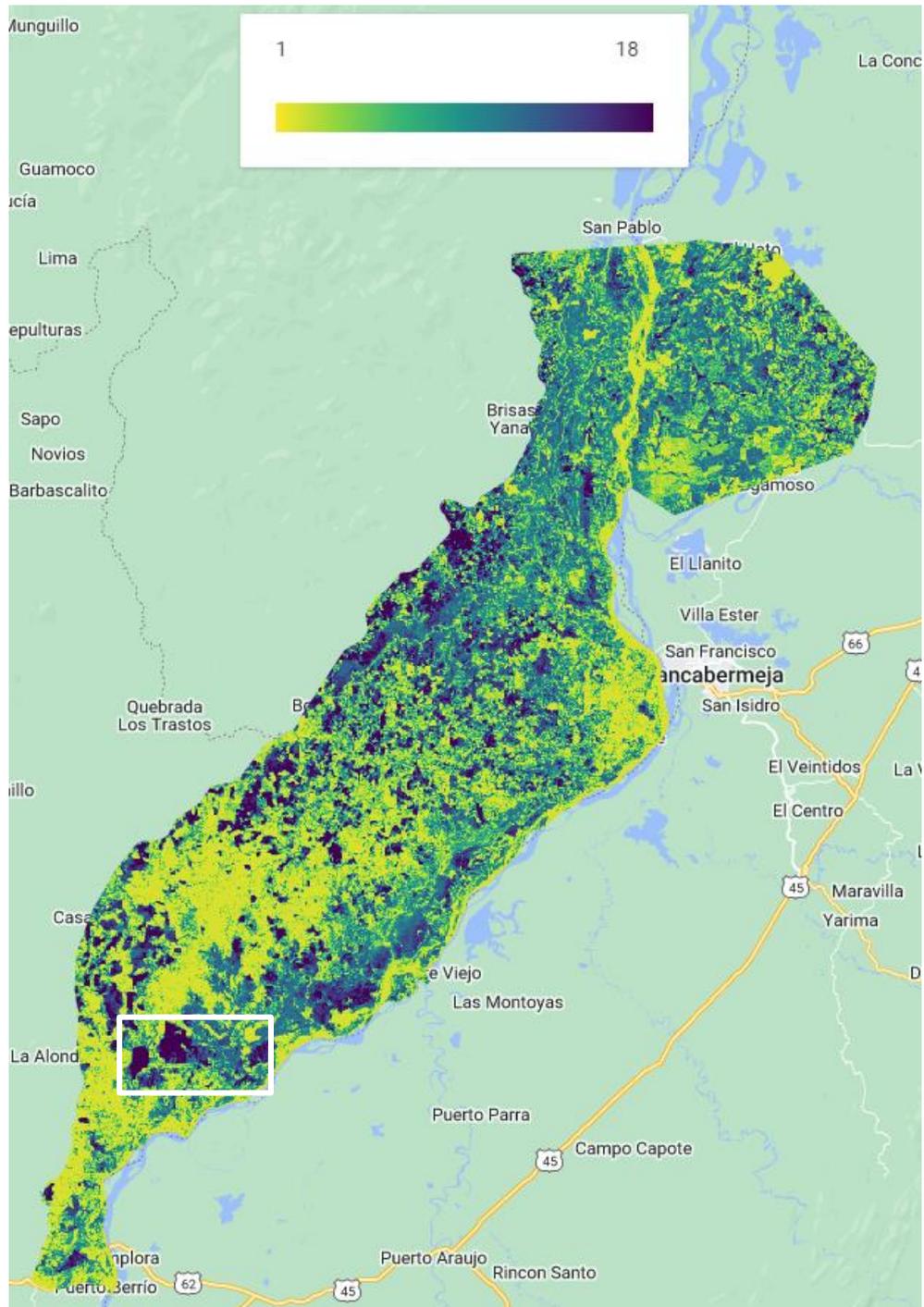


Figura 3-4. Capa de SCI desarrollado por Hansen et al. (2019). El recuadro blanco de la imagen corresponderá a un acercamiento para mostrar en las siguientes figuras la distribución de los puntos de muestreo.

Muestreo: se realizaron tres ejercicios de muestreo para generar los datos de entrenamiento y prueba de los modelos de aprendizaje: 1) se desplegó una grilla de puntos equidistantes con una separación de 500 metros, 2) una grilla de 1 km, y 3)

los registros biológicos filtrados para la zona de estudio. Las distancias de separación entre las grillas se eligieron debido a que la resolución espacial más gruesa dentro de las variables predictoras fue de 500 metros, además de que se quería evaluar el efecto del tamaño de la muestra en los resultados de la modelación. Estas estrategias de muestreo contrastan una estrategia que garantiza la representatividad de todo el rango de valores definidos en el SCI para el área de estudio (mediante el uso de las grillas), y otra restringida a áreas con registros biológicos, donde la representatividad de categorías de SCI puede verse restringida. Se obtuvo un total de 11.588 y 2.867 puntos para las grillas de 500 m y 1km respectivamente (Figura 3-5), los cuales se dividieron como datos de entrenamiento (70%) y de prueba (30%) para la evaluación de los modelos, de manera totalmente aleatoria, verificando que en cada uno de los ejercicios de modelación se utilizara el mismo conjunto de puntos de entrenamiento y validación, además de que los puntos de validación no fueran utilizados en ninguno de los ejercicios de entrenamiento.

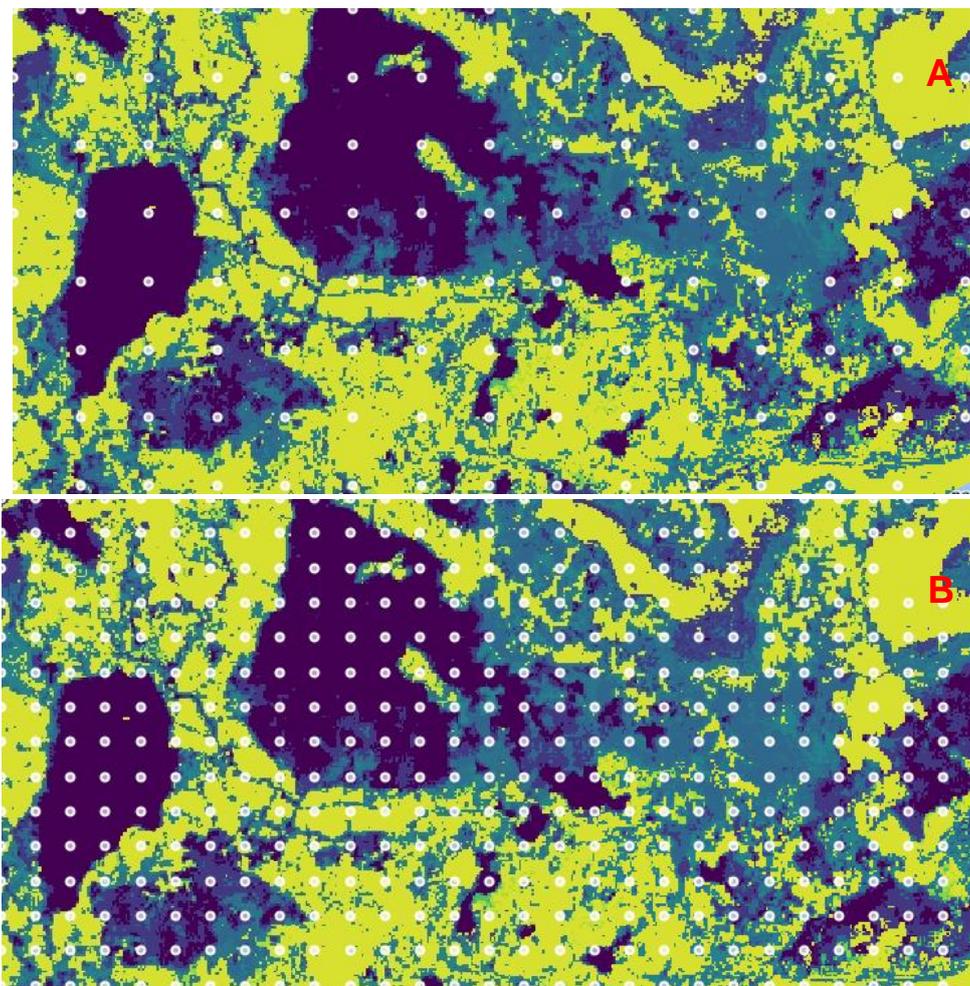


Figura 3-5. Muestreo por grillas equidistantes. A) Grilla de 1 km, B) Grilla de 500 metros. Al ser el SCI una capa continua para el área de estudio, estas grillas permiten realizar un muestreo equilibrado en todas las categorías de SCI disponibles.

En cuanto a los registros biológicos, solo se seleccionaron aquellos con información disponible desde 2018, y se eliminaron todos los duplicados, tomando en cuenta el nombre de la especie, el año y las coordenadas geográficas. En total se identificaron 49.959 registros de 3107 especies que fueron agrupadas como anfibios, aves, plantas, colémbolos, escarabajos, hormigas, mamíferos, mariposas, reptiles y otros (Figura 3-6; Tabla 3-1).

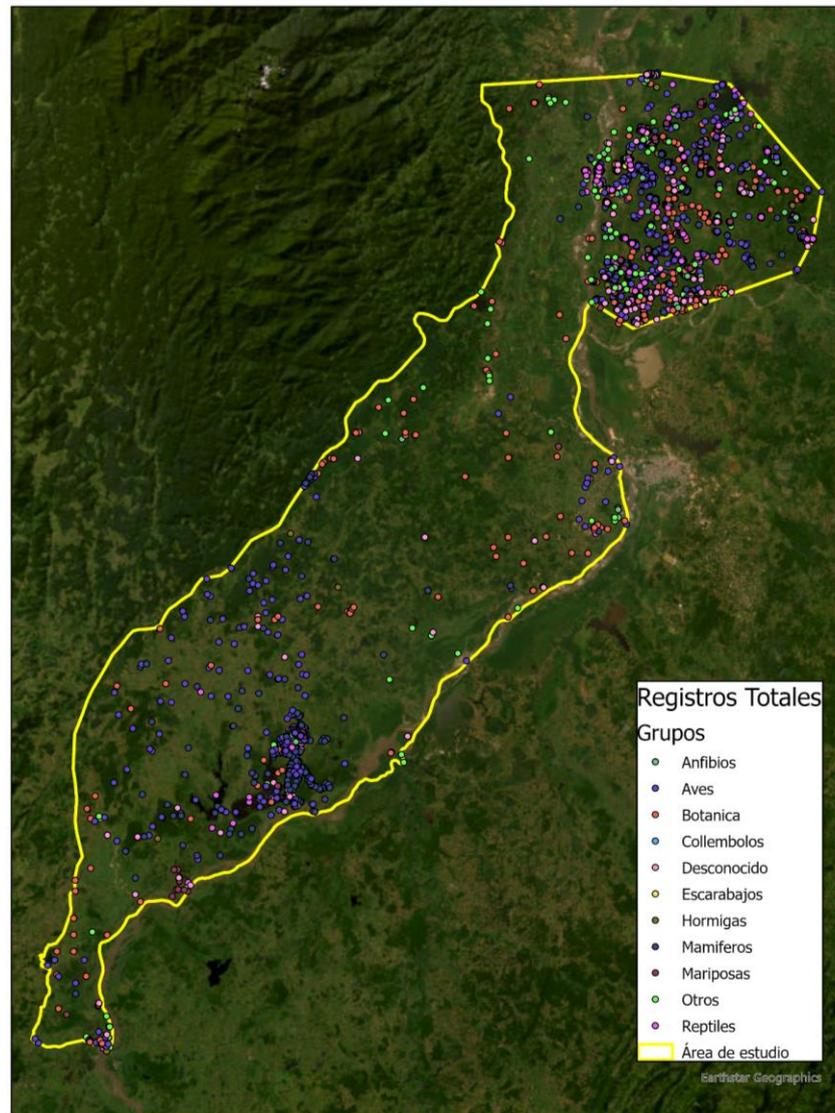


Figura 3-6. Registros biológicos disponibles en el área de estudio, a partir de los cuales se calcula la dependencia de las especies hacia los ecosistemas de bosque.

Tabla 3-1. Número total de registros encontrados en el área de estudio y agrupación por grupos.

Grupos de especies	Número de registros
Anfibios	1590
Aves	11619
Botánica	5727
Colémbolos	4054
Escarabajos	3464
Hormigas	10777
Mamíferos	4290
Mariposas	4194
Otros	3420
Reptiles	824
Total	49959

Los filtros utilizados para el muestreo con los registros tomaron en cuenta dos aspectos: 1) eliminación de registros redundantes con base en los valores de las variables predictoras, y 2) eliminación de registros con vacíos debido a las limitaciones de RF. Se optó por la eliminación de los registros con valores faltantes, ya que estos datos no superan el 5% del total y el desempeño de la modelación no se vio afectado al incluirlos (Cismondi et al., 2013). El primer filtro se realizó, para evidenciar redundancia en variables por registros muy cercanos geográficamente (distancia inferior a 10 m, resolución espacial más detallada dentro de las variables predictoras). Para evitar la generación de pseudoreplicas en los datos, se eliminaron los registros con atributos y salidas idénticas, utilizando únicamente las combinaciones de valores únicos de la matriz de datos (Lazic et al., 2020). De los 49.959 registros, luego de aplicar los filtros anteriores, se seleccionaron 5.787 registros. Cada uno de los muestreos fue dividido en puntos de entrenamiento (70%) y validación (30%) como se muestra en la Tabla 3-2.

Tabla 3-2. Puntos de entrenamiento y validación para estimación del SCI

Categoría SCI	Train 500 m	Test 500 m	Train km 1	Test 1 km	Train registros	Test registros
1	409	202	36	12	1135	468
2	693	273	138	54	88	45
3	562	265	117	60	58	24
4	565	241	146	40	44	22
5	631	256	147	70	20	12
6	670	293	184	73	35	11
7	693	305	225	87	35	9
8	712	307	238	113	4	0
9	666	293	210	98	0	0
10	634	278	192	73	324	128
11	520	231	128	61	332	142
12	376	168	92	39	325	124
13	241	112	41	38	95	46
14	219	82	50	12	243	81
15	160	87	27	11	750	326
16	138	53	20	10	0	0
17	112	32	13	7	49	19
18	79	30	4	1	554	239
Total	8080	3508	2008	859	4091	1696

Variables predictoras: En GEE, los puntos de muestreo se usaron para extraer la información espectral de los productos satelitales utilizados como variables predictoras (Tabla 3-3). Los productos Sentinel (COPERNICUS/S1_GRD_FLOAT y COPERNICUS/S2_SR) se filtraron para el año 2022, mientras que los productos PALSAR (JAXA/ALOS/PALSAR/YEARLY/SAR_EPOCH) y MODIS (MODIS/061/MCD15A3H y MODIS/006/MOD44B) se filtraron para los años 2020 y 2021 debido a que, a la fecha de la búsqueda de las imágenes, no existían productos

para el año 2022. Estos productos fueron elegidos, por su alta resolución espacial (particularmente Sentinel-2 para cada una de sus bandas), su alta calidad de los datos y su uso común en algunos de los índices que se utilizaron en la modelación (particularmente MODIS con los índices de Fpar, LAI y Vegetation Continuous Fields). Sobre las colecciones de imágenes, se corrigieron errores geométricos y atmosféricos, y luego sus valores se promediaron. De esta manera se generó una imagen para cada producto, las cuales conforman el stack final del que se extrajo la información espectral y los índices para los puntos de muestreo. Todas las variables predictivas fueron obtenidas utilizando el catálogo de la plataforma GEE. De la misma forma, las funciones utilizadas para la elaboración de los conjuntos de datos de entrenamiento adecuados, el preprocesamiento de los datos, y la elaboración de los modelos de clasificación generados con algoritmos de Machine Learning, se pueden ejecutar en la plataforma GEE.

Tabla 3-3. Sensores multiespectrales disponibles en la plataforma GEE y utilizados en la modelación

Satélite	Banda, Índice, Coeficiente de retrodispersión	Longitud de onda o definición
Sentinel-2 (Multiespectral)	B2: Blue B3: Green B4: Red B5: Red Edge 1 B6: Red Edge 2 B7: Red Edge 3 B8: NIR 1 B8a: Red Edge 4 B11: SWIR 1 B12: SWIR 2 NDVI: Normalized Difference Vegetation Index EVI: Enhanced Vegetation Index SAVI: Soil Adjusted Vegetation Index SVVI: Spectral Variability Vegetation Index RNDVI: Red Edge Vegetation Index MNDWI: Modified Normalized Difference Water Index PSRI: Plant Senescence Reflectance Index Brightness Greenness Wetness	450 – 520 nm 540 – 570 nm 650 – 680 nm 690 – 710 nm 730 – 740 nm 770 – 790 nm 780 – 900 nm 850 – 870 nm 1560 – 1650 nm 2100 – 2280 nm $(B8-B4)/(B8+B4)$ $G*(B8-B4)/(B8+C1*B4-C2*B2+L)$ $(1+L)*(B8-B4)/(B8+B4+L)$ $Stdv(B2,B3,B4,B8,B11,B12) - Stdv(B8,B11,B12)$ $(B8-B6)/(B8+B6)$ $(B3-B11)/(B3+B11)$ $(B4-B3)/B6$ Tasseled Cap 1 Tasseled Cap 2 Tasseled Cap 3
Sentinel-1 (SAR)	VV de Banda C	3.75 – 7.5 cm

	VH de Banda C VH/VV de Banda C VV – VH de Banda C	3.75 – 7.5 cm 3.75 – 7.5 cm 3.75 – 7.5 cm
PALSAR (SAR)	HH de Banda L HV de Banda L HV/HH de Banda L HH – HV de Banda L	15 – 30 cm 15 – 30 cm 15 – 30 cm 15 – 30 cm
MODIS (Multispectral)	FPAR: Fraction of photosynthetically active radiation LAI: Leaf Area Index VCF: Vegetation Continuous Field	500 m 500 m 250 m

La creación del mosaico de Sentinel-2 se realizó a partir de dos colecciones de imágenes: MSI-2A (Multispectral Instrument Level 2A) y Sentinel-2 –Cloud Probability. La colección MSI-2A provee imágenes con reflectancia de superficie en geometría cartográfica, mientras que Sentinel-2 Cloud Probability provee información para enmascarar píxeles con una alta probabilidad de nubes usando la librería LightGBM (Fagua et al., 2023). Estas colecciones fueron combinadas siguiendo la metodología de Braaten (2022) para enmascarar nubes identificadas con el producto Sentinel-2 Cloud Probability, y sombras de nubes definidas por una intersección entre las proyecciones de nubes con los píxeles de baja reflectancia en la banda de infrarrojo cercano (NIR). Luego de corregir y estandarizar las bandas multispectrales, estas fueron utilizadas para calcular los índices y Tasseled Cap que resumieron información relevante del stack de imágenes, principalmente el brillo, la vegetación y la humedad del suelo. Por otro lado, el mosaico de Sentinel-1 se construyó con el producto Sentinel-1 SAR GRD (C-Band Synthetic Aperture Radar Ground Range Detected) que provee datos con orientación y distancia correctas con respecto al plano terrestre. Se aplicó además una corrección de pendiente radiométrica basada en ángulos (Vollrath et al., 2020), utilizando el modelo de corrección usado para coberturas vegetales y el Modelo Digital de Superficie Global (ALOS, AW3D30) para estimar la pendiente.

Modelación: Antes de la elaboración de los modelos, se desarrolló un análisis de Boruta (Kursa & Rudnicki, 2010), el cual es un algoritmo utilizado para identificar la relevancia de los predictores previo a la modelación. Este algoritmo se basa en la misma idea de los clasificadores como RF, donde se agrega aleatoriedad al sistema y se colectan los resultados de conjuntos de muestras aleatorias de cada uno de los predictores, denominadas predictores sombra. Los predictores originales cuya importancia sea mayor que los predictores sombra se declararán como variables importantes para la modelación (Fagua et al., 2023; Kursa & Rudnicki, 2010; Masrur Ahmed et al., 2021).

Posteriormente se evaluaron y compararon cinco algoritmos de aprendizaje que aplican diferentes métodos estadísticos para predecir la variable respuesta: 1) Random Forest (RF), 2) Gradient Boosting Machine (GBM), 3) Supported Vector Machine (SVM), 4) Bootstrap aggregating trees o Bagging (BAG), y 5) Boosted Logistic Regression. Dado que se van a evaluar diferentes estrategias de muestreo, cada algoritmo presenta variaciones tanto en sus puntos de muestreo, como en la configuración de sus hiperparámetros. Como se mencionó en la sección 1.2.1, estos algoritmos fueron seleccionados por el gran desempeño que han demostrado en ejercicios de clasificación de imágenes de satélite, además de ser ensambles que potencian la capacidad de clasificación por medio de la combinación de múltiples clasificadores débiles (con excepción del algoritmo SVM), reduciendo así el problema de sobreajuste y lidiar con otras problemáticas evidenciadas con algoritmos de aprendizaje individuales (Zhang et al., 2022).

A cada uno de los modelos se le realizó un control de entrenamiento mediante validación cruzada, utilizando 10 iteraciones y 2 repeticiones; es decir, que el conjunto de datos es dividido aleatoriamente en subconjuntos de entrenamiento en 10 iteraciones sin sustitución. Nueve de las iteraciones se utilizan para entrenar el modelo, mientras que la décima iteración se usa para evaluar el rendimiento, generando así para cada uno de los algoritmos 10 modelos y estimaciones de rendimiento que son promediadas, para obtener la mejor estimación del modelo. Este proceso se repite dos veces y los dos resultados finales son nuevamente promediados (Raschka, Liu & Mirjalili, 2023). Adicionalmente, en cada modelo se probó una optimización a partir de múltiples parámetros, para así identificar la combinación de parámetros que ofrecen el menor error posible (Tabla 3-4). Este ajuste de parámetros se realizó usando el paquete caret de R, y se seleccionó el modelo identificando el de mayor tasa de acuerdo general promediada sobre las iteraciones de validación cruzada (Accuracy) y el estadístico Kappa de Cohen promediada entre los resultados de remuestreo.

Tabla 3-4. Hiperparámetros evaluados con cada algoritmo de aprendizaje, utilizando el paquete caret de R.

Algoritmo	Hiperparámetros	Valores
Random Forest	Predictores seleccionados aleatoriamente (mtry) Regla de división (splitrule) Tamaño mínimo de nodo (min.node.size) Número de árboles (ntrees)	[2,3,4,sqrt(31),6,7,8,9,10] ["Variance", "Extratrees"] [1,3,5,10] 1000
Gradient Boosting Machine	Profundidad de Interacción (interaction depth) Contracción (shrinkage) Número mínimo de observaciones en nodo (n.minobsinnode) Número de árboles (ntrees)	[3,5,9] 0.01 [10,20,30] [500,1000,1500,2000]

Algoritmo	Hiperparámetros	Valores
Bootstrap Aggregating Trees	nbagg minsplit cp	[10,20,30] [2,4,6] [0,0.01]
Supported Vector Machine	Costo (cost) Sigma Kernel	[0.1,1,10,100] [0.01,0.1,1] Radial
Boosted Logistic Regression	Número de Iteraciones (nIter)	[500,1000,5000,10000]

En cada uno de los muestreos se calculó la importancia de los predictores para los algoritmos de aprendizaje, utilizando el paquete de R caret. Este paquete aplica un método de estimación de importancia de variables para cada algoritmo y set de muestra, permitiendo la comparación de importancia entre las diferentes combinaciones de modelos. Los algoritmos de BAG, BRT y RF utilizan una medida similar de importancia basada en permutaciones, la cual adoptan un enfoque basado en la predicción mediante el uso del error de predicción atribuible a la variable (Ishwaran & Lu, 2019). Cada predictor tiene una puntuación de mejora que corresponde a la suma de las mejoras al cuadrado sobre los nodos internos de cada árbol para el cual el predictor fue elegido como variable de partición, los cuales son promediados para generar un valor estabilizado más confiable en grandes conjuntos de árboles (Greenwell et al., 2018).

Luego de generar y comparar cada uno de los modelos, se eligieron los datos e hiperparámetros que generan el menor error, para ser utilizados como insumos en la generación de la salida gráfica final, utilizando las funciones disponibles en GEE.

Supuestos:

- El SCI funciona como un indicador fiable de la estructura del bosque. Los tres componentes del SCI, son variables identificadas como estratégicas para el mapeo de los bosques, describiendo la cobertura (porcentaje de dosel), la estructura de la vegetación (altura de dosel) y los disturbios (tiempo de intervención) (Lechner et al., 2020). Además de esto, estas características también son objeto de estudio por medio de diferentes productos de sensores remotos, por lo que se sustenta la elección tanto del índice como de la información de SR con la que se busca encontrar su relación.

3.3.2 Modelo 2: Estimación de Calidad de Hábitat basada en registros, a partir de información de SR

El objetivo con este modelo es explorar la correlación entre información de SR y la calidad de hábitat inferida desde la detección de múltiples especies con requerimientos conocidos de calidad de coberturas.

Variable respuesta: para este segundo modelo se estima la dependencia de Bosque, la cual es una propuesta de categorización, utilizando los registros biológicos disponibles en el área de estudio para determinar el hábitat más frecuente para cada especie, y ajustarlo con los polígonos de cobertura de la tierra de CORINE Land Cover.

Muestreo: Para cada una de las especies se determinó el hábitat mediante dos herramientas: 1) el paquete de R rredlist que es un API (Application Programming Interface) conectada a la base de datos de la IUCN y donde es posible consultar el hábitat de ciertas especies, y 2) directamente del mapa de hábitat de la IUCN desarrollado por (Jung et al., 2020). Para la identificación de los hábitats utilizando el mapa de hábitat, se realizó una búsqueda a nivel nacional de los registros disponibles para estas especies. De esta manera, se reconocieron los hábitats utilizados por estas especies, no solo dentro del área de estudio, sino con toda la información nacional disponible. Luego de esta revisión nacional de registros, a cada especie se le asignó el o los hábitats más frecuentes, y gracias a la disponibilidad global del mapa de hábitats, fue posible utilizar todos los 49.959 registros para estimar la dependencia de las especies con los bosques y ajustar este componente con las coberturas de la tierra. A diferencia del modelo 1, la cercanía de los registros biológicos no afecta la determinación de la dependencia de bosque, sino al contrario, resalta la importancia del polígono al registrar un número elevado de individuos. Por otro lado, de las 3107 especies registradas, solo 649 tenían información en la plataforma de la lista roja de la IUCN, reduciendo los registros a 2.064.

Los registros fueron agrupados en los polígonos de coberturas de la tierra descritos con la metodología CORINE Land Cover. En cada uno de los polígonos de coberturas se determinó el o los valores más frecuentes de hábitats a partir de los registros, asignando una clasificación de dependencia de bosque siguiendo las reglas mostradas en la Tabla 3-5. Finalmente, dependiendo de la categoría de cobertura, el valor de dependencia de bosque es ponderado para maximizar las congruencias entre hábitat y coberturas, y penalizar las incongruencias (Tabla 3-6) siendo esta última clasificación el segundo componente del índice de calidad de hábitat para ecosistemas de bosque.

Tabla 3-5. Clasificación de dependencia de bosque para las especies identificadas en el área de estudio. Los valores entre paréntesis corresponden a las categorías de hábitat definidas por la IUCN.

Tipo de Hábitat	Hábitat más frecuente	Clasificación de dependencia de Bosque por registros
Hábitats de bosque	Bosque tropical/subtropical seco (1.5) Bosque tropical/subtropical húmedo de tierras bajas (1.6) Bosque tropical/subtropical pantanos (1.8) Bosque tropical/subtropical húmedo de montaña(1.9)	5
Hábitats de bosque y otros hábitats naturales	Hábitats de bosque Sabana seca (2.1) Arbustal tropical/subtropical seco (3.5) Arbustal tropical/subtropical húmedo (3.6) Arbustal tropical/subtropical de montaña (3.7) Herbazal tropical/subtropical seco (4.5) Humedales (5) Humedales, Ríos y arroyos permanentes (5.1) Humedales, Ríos y arroyos intermitentes/estacionales (5.2) Humedales, Lagos permanentes de agua dulce (>8 ha) (5.5)	4
Hábitats naturales exceptuando bosques	Otros hábitats naturales	3
Hábitats de bosque, otros bosques naturales y hábitats artificiales	Hábitats de bosque Otros hábitats naturales Tierra cultivable (14.1) Pastizal (14.2) Plantaciones (14.3) Áreas urbanas (14.5)	2
Hábitats artificiales y especies invasoras	Hábitats artificiales Especies invasoras	1

Tabla 3-6. Ponderación de calidad de hábitat con categorías de coberturas de la tierra

Categoría CORINE (Nivel 2) en área de estudio	Valor de ponderación
1.1 Zonas urbanizadas 1.2 Zonas industriales o comerciales y redes de comunicación 1.3 Zonas de extracción mineras y escombreras	1
2.2 Cultivos permanentes 2.3 Pastos 3.3 Áreas abiertas sin, o con poca vegetación	2
2.4 Áreas agrícolas heterogéneas	3
3.2 Áreas con vegetación herbácea y/o arbustiva 4.1 Áreas húmedas continentales 5.1 Aguas continentales	4
3.1 Bosques	5

Con la información descrita en las tablas 3-5 y 3-6, lo que se busca es encontrar áreas de congruencia entre información de hábitat y de coberturas. Así, un polígono con un valor de dependencia de bosque de 5 y cobertura de bosques, tendrá un valor mayor que un polígono con cobertura de pastizales, a pesar de tener el mismo valor promedio de dependencia de bosque (alta calidad congruente). De igual manera, se busca identificar los polígonos que tienen un valor promedio de hábitat bajo con polígonos de baja ponderación de cobertura (baja calidad congruente), y valores promedios intermedios de dependencia y ponderación media de cobertura (calidad intermedia congruente). La ponderación se realiza simplemente multiplicando los valores promedio de dependencia de bosque de la tabla 3-6, con los valores de ponderación de la tabla 3-7, dividiendo entre 25 para mantener una escala similar en cada uno de los procesos de modelación (Figura 3-7). Al definir estos polígonos con la variable de calidad de hábitat que integra los registros y las coberturas, fue posible ampliar las zonas con información para la generación de los modelos. Gracias a esta generalización, fue posible realizar un muestreo similar a la del modelo 1 dentro de los polígonos de la figura 3-7, desplegando una grilla de 500 m y de 1 km (Tabla 3-7). Las áreas sin información son zonas en las que no se encontraron registros biológicos.

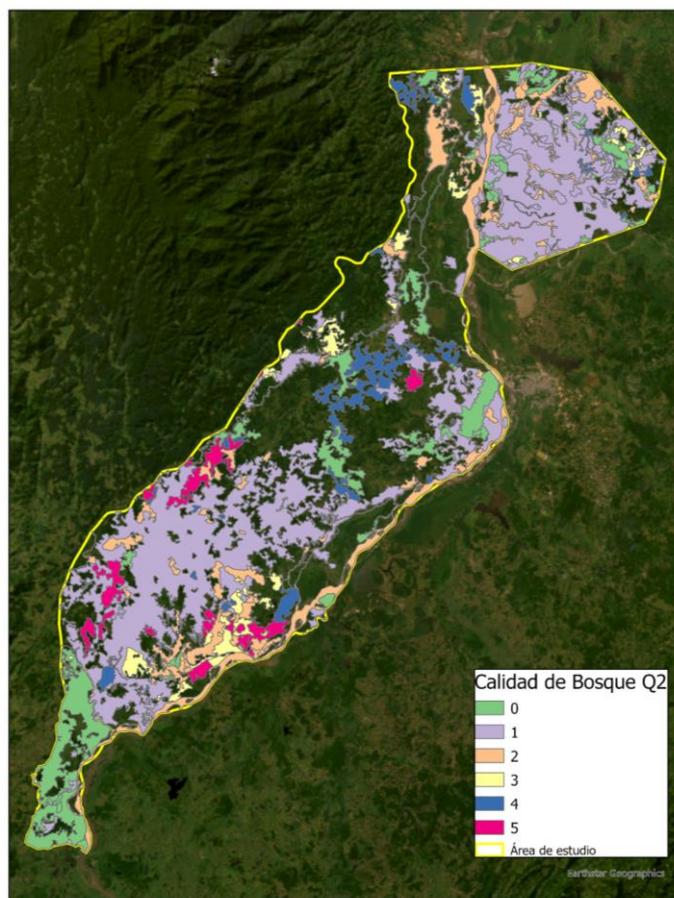


Figura 3-7. Polígonos de calidad de hábitat calculada con registros biológicos y ponderada con coberturas CORINE Land Cover adaptadas para Colombia.

Como puede apreciarse en la Figura 3-7, las áreas con calidad intermedia son muy bajas, y tanto para el SCI como para la calidad de hábitat, la categoría intermedia es poco intuitiva. Para el SCI por ejemplo, la categoría intermedia ($6 \leq SCI \leq 9$, [3] incluye árboles con alturas de más de 15 metros y cobertura de dosel variable (superior a 25%), correspondiendo a árboles maduros que pueden albergar una mayor diversidad de especies (Chen et al., 2020; Klimes et al., 2012). Por otro lado, una calidad de hábitat intermedia significa que en el área conviven un alto número de especies, con una variada dependencia de bosque (aunque en su mayoría no son altamente dependientes como las de calidad 4 o 5, tampoco son únicamente especies no dependientes de bosques, introducidas o invasoras). Por estas razones, se decidió que para las áreas con SCI y calidad de bosque intermedia, se evaluará su congruencia o incongruencia como si pertenecieran al nivel alto (Figura 3-8).

Tabla 3-7. Puntos de entrenamiento y validación para estimación de la calidad de bosque ajustada por coberturas

Categoría Calidad de Bosque	Puntos de entrenamiento 500 m	Puntos de validación 500 m	Puntos de entrenamiento 1 km	Puntos de validación 1 km
0	744	317	199	75
1	3056	1270	723	337
2	1158	484	249	125
3	387	135	104	33
4	381	188	90	42
5	192	83	53	23
Total	5918	2477	1418	635

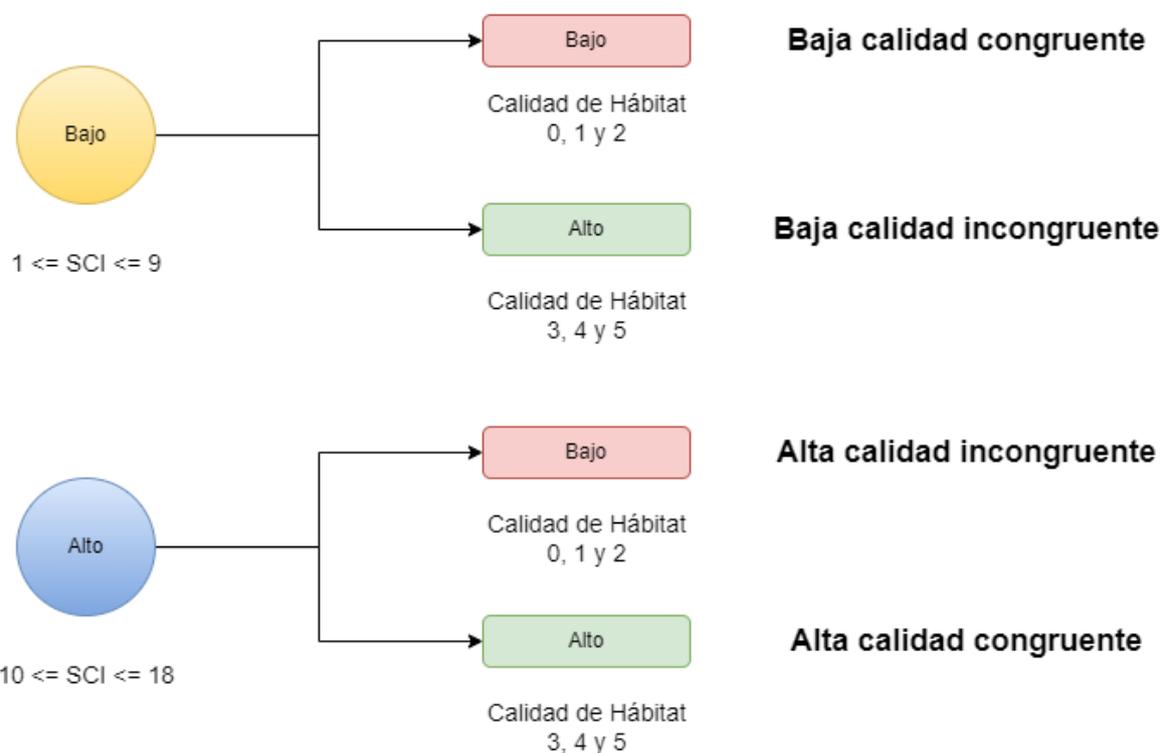


Figura 3-8. Determinación de las congruencias e incongruencias al integrar SCI con calidad de hábitat. Las categorías de SCI y calidad de hábitat intermedias fueron fusionadas con la categoría alta de cada uno de los componentes.

VARIABLES PREDICTORAS: El preprocesamiento de imágenes y cálculo de índices para el modelo 2 son idénticos que en el modelo 1.

MODELACIÓN: La evaluación de las variables predictivas y los algoritmos de aprendizaje son los mismos que en el modelo 1.

SUPUESTOS:

- Los registros biológicos reflejan fielmente la preferencia de las especies en cuanto a calidad de coberturas. Es decir, el registro se encontró en un polígono particular porque la especie hace uso del mismo y lo necesita para su subsistencia. Si los registros están en lugares que no corresponden a lo que se sabe por sus requerimientos de calidad de hábitat, es posible que no tengamos un conocimiento muy refinado de dichos requerimientos, que nuestros descriptores no estén capturando atributos claves que determinan calidad de hábitat, o que los puntos registrados tengan errores de localización. Los resultados de este trabajo darán luces sobre las dos primeras opciones y nuestra generalización de calidad de hábitat por polígonos y cobertura, reduce el impacto de la tercera opción.
- La generalización realizada por parches de coberturas de la tierra asume que el individuo puede aprovechar todo el parche delimitado y que la cobertura captura la calidad de hábitat, por lo que cada área del polígono tendrá condiciones ambientales similares.
- Las especies dentro de los polígonos se evaluaron con el mismo peso. La agrupación de las clases de los registros biológicos se mostró principalmente para evidenciar el esfuerzo de muestreo que se realizó en campo, pero no se generó ningún tipo de ponderación a los grupos.
- La identificación de las especies dentro de un polígono, el valor promedio de sus dependencias al bosque y el tipo de cobertura de la tierra es suficiente para estimar la calidad de bosque del polígono generalizado, esto significa, que un polígono dentro de una categoría de cobertura dada, tendrá la misma calidad de hábitat en toda la extensión del polígono.
- Dado que el área de estudio tiene una distribución potencial de ecosistemas de bosques, los hábitats de mejor calidad corresponden a las áreas con mejor cobertura boscosa. Esto influye en el puntaje otorgado a las coberturas de la tierra y el hábitat predominante de las especies, dando un mayor puntaje a las especies que necesitan el bosque.

3.3.3 Modelo 3: Propuesta de índice de calidad de hábitat, integrando SCI y registros biológicos

El objetivo de este modelo es combinar el conocimiento de los dos modelos anteriores para determinar tanto las áreas de congruencias entre el SCI de Hansen y la calidad de bosque, como sus áreas de incongruencias. Estos dos atributos brindan información muy importante para tener un enfoque integral de la calidad de hábitat, puesto que las áreas congruentes corresponden a aquellas que brindan información acertada de la calidad de bosque, sea alta o baja desde ambas metodologías, y las áreas incongruentes corresponden a aquellas que deben revisarse minuciosamente para encontrar la fuente de incongruencia que debe ser corregida. Ambos escenarios de la integración de estas dos metodologías permiten reforzar con datos de campo la información que puede capturarse desde SR. Así, el análisis de las áreas donde se presenta o no esa congruencia y las variables de SR responsables por predecirla, nos permite evaluar nuestra capacidad para predecir la calidad de hábitat de forma integral y no solo desde productos exclusivamente derivados desde SR e inferir qué características capturables desde SR pueden mejorar esta capacidad predictiva.

Variable respuesta: Este índice de calidad de hábitat integrado, descrito en el párrafo anterior, puede definirse como un SCI ajustado con registros de múltiples especies, creado con la finalidad de detectar áreas de congruencia entre SCI y la calidad de bosque propuesta en este proyecto. De esta manera, se fortalece el alcance del SCI para detectar no solo la estructura de los bosques sino también la composición (Figura 3-9).

Muestreo: Los puntos de entrenamiento y validación son los mismos que los del Modelo 2, puesto que corresponden a las áreas en donde se puede medir simultáneamente el SCI y la calidad de bosque.

Variables predictoras: El preprocesamiento de imágenes y cálculo de índices para el modelo 3 son idénticos que en el modelo 1.

Modelación: La evaluación de las variables predictivas, los algoritmos de aprendizaje y el entrenamiento de control son los mismos que en el modelo 1.

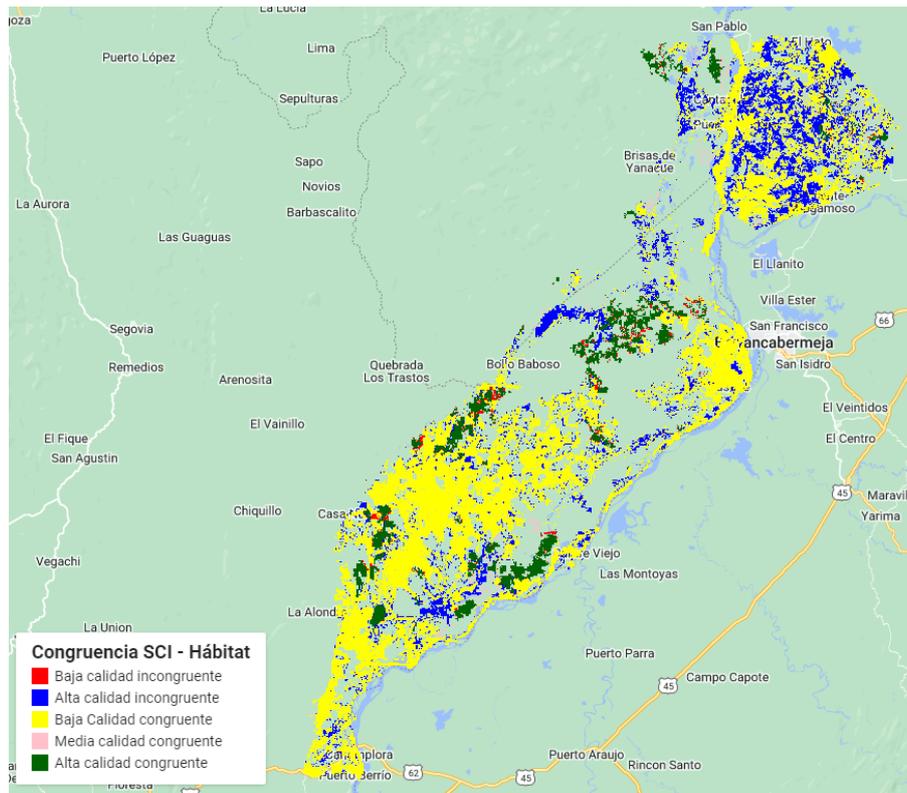


Figura 3-9. Áreas de congruencia e incongruencia entre SCI y Calidad de Bosque. Los niveles de congruencia en la leyenda dependen de los valores coincidentes entre los dos componentes. Baja calidad incongruente (SCI bajo, pero con alta dependencia de bosque), Alta calidad incongruente (SCI alto, pero con dependencia baja de bosque). Baja calidad congruente (SCI bajo y baja dependencia de bosque). Media calidad congruente (SCI medio y dependencia de bosque intermedia). Alta calidad congruente (SCI alto y alta dependencia de bosque). Es importante tener en cuenta que esta figura sólo es ilustrativa, ya que se construyó uniendo la figura 7 con el SCI de Hansen reclasificado. La variable respuesta para el modelo 3 se construirá con la predicción generada con el mejor algoritmo del modelo 2 y el SCI reclasificado.

4. Resultados

4.1 Modelo 1

Los resultados del análisis de Boruta mostraron que todos los predictores extraídos de SR son significativos en la modelación del SCI, es decir, que su aporte en la modelación es superior que el debido al azar, por lo que ninguna variable predictora fue excluida. Un resultado importante que se evidenció en un primer ejercicio de modelación, es que el valor de exactitud del modelo era muy bajo si se usaban las 18 categorías del SCI (Accuracy = 0.184 y Kappa = 0.101), esto posiblemente a la baja variabilidad espectral entre algunas categorías. Para resolver esta problemática, el SCI fue reclasificado en 5 categorías. La categoría 1 permaneció igual, ya que en la metodología de Hansen, esta categoría tiene un amplio número de combinaciones de sus variables constituyentes, mientras que las demás categorías se agruparon en clases iguales. Teniendo esto en cuenta, las categorías se definieron de la siguiente manera: $SCI = 1, [1]$; $2 \leq SCI \leq 5, [2]$; $6 \leq SCI \leq 9, [3]$; $10 \leq SCI \leq 14, [4]$; $15 \leq SCI \leq 18, [5]$. La prueba de Boruta arrojó los mismos resultados de variables significativas con las cinco clases que con las categorías originales (Figura 4-1).

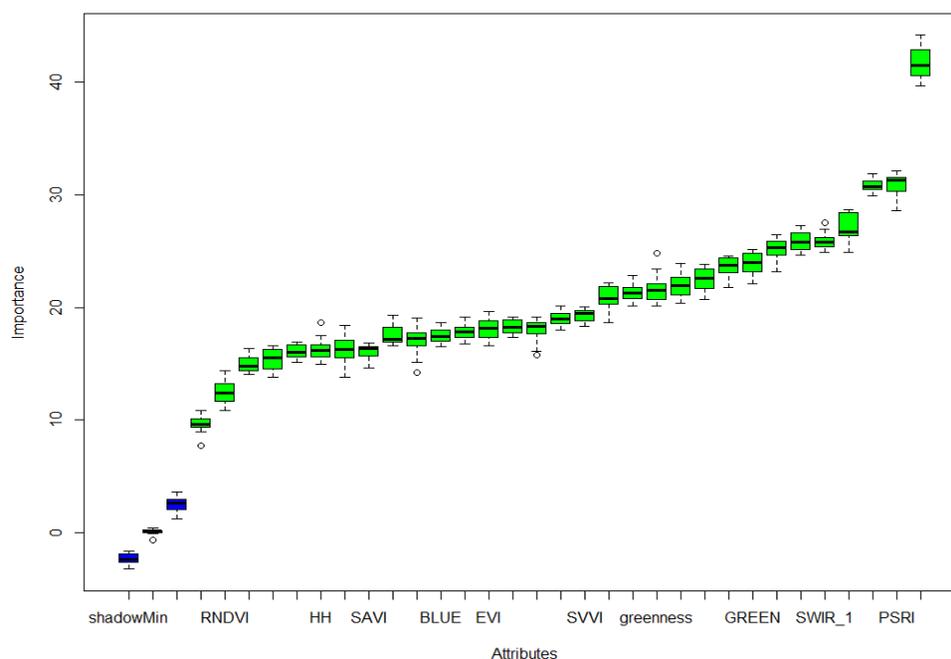


Figura 4-1. Importancia de variables medida previamente a la modelación del Modelo 1. Las 31 variables predictoras (Boxplots color verde) demostraron tener una importancia mayor que la generada únicamente por aleatoriedad al compararlas con las variables sombra (Boxplots de color azul). Entre las variables con mayor importancia encontramos la banda RED_EDGE_1, el índice PSRI, el porcentaje de bosque y el coeficiente de retrodispersión HV.

Los puntos de muestreo fueron desplegados en la plataforma GEE, donde se realizó la partición utilizando un subconjunto del 70% de los datos elegidos de manera aleatoria para el entrenamiento de los modelos, y el 30% restante como datos de prueba. Para el muestreo realizado con los registros biológicos, luego de aplicar los filtros descritos en la sección 3.2.2, de los 49.959 registros, quedaron 5787 registros, divididos en 4091 registros para el entrenamiento de los modelos y 1696 registros de validación. Estos registros corresponden a los valores únicos del conjunto de 31 variables predictoras para evitar el sobreajuste con pseudoréplicas, y que no presentan valores faltantes. Con esta partición se realizaron ejercicios de clasificación con los cinco algoritmos de aprendizaje de máquina y los 3 muestreos diferentes (Grillas de 500 m y 1 km, y los registros biológicos totales). Los resultados para el modelo 1 se muestran a continuación en la Tabla 4-1, en donde se puede identificar que el modelo Random Forest, calculado a partir de los registros biológicos, es el que presenta una mayor exactitud temática; no obstante, el desempeño de los modelos internos generados en el entrenamiento de control y ajuste de hiperparámetros fue muy similar entre los algoritmos utilizados, con un rendimiento ligeramente menor con los generados por Boosted Logistic Regression y Bagging (Figura 4-2). Los valores de hiperparámetros con mejor desempeño de la tabla 4-1 fueron utilizados para realizar la predicción del SCI reclasificado con el raster stack construido en GEE (Figura 4-3).

Tabla 4-1. Resultados del ajuste de hiperparámetros para el Modelo 1 de estimación de SCI con variables de SR.

Muestreo	Algoritmo	Parámetros	Valor de Parámetros	Exactitud	Valor de Exactitud
Registros Totales	Random Forest	mtry splitrule min.node.size ntrees	10 gini 1 1000	Accuracy Kappa	0.6753543 0.532126
	Gradient Boosting Machine	n.trees interaction.depth shrinkage n.minobsinnode	1000 9 0.01 10	Accuracy Kappa	0.6667899 0.5201481
	Bagging	nbagg minsplit cp	30 4 0	Accuracy Kappa	0.6561745 0.4963392

	Supported Vector Machine	cost sigma kernel	1 0.1 radial	Accuracy Kappa	0.6474446 0.490844824
	Boosted Logistic Regression	niter	500	Accuracy Kappa	0.5705293 0.3745908
Grilla 500 m	Random Forest	mtry splitrule min.node.size ntrees	4 gini 1 1000	Accuracy Kappa	0.5823948 0.4117366
	Gradient Boosting Machine	n.trees interaction.depth shrinkage n.minobsinnode	500 9 0.01 10	Accuracy Kappa	0.5794863 0.409919
	Bagging	nbagg minsplit cp	30 6 0	Accuracy Kappa	0.562419 0.3860447
	Supported Vector Machine	cost sigma kernel	10 0.01 radial	Accuracy Kappa	0.5797942 0.402144707
	Boosted Logistic Regression	niter	5000	Accuracy Kappa	0.4901043 0.2887168
Grilla 1 km	Random Forest	mtry splitrule min.node.size ntrees	8 extratrees 10 1000	Accuracy Kappa	0.6378749 0.4437009
	Gradient Boosting Machine	n.trees interaction.depth shrinkage n.minobsinnode	500 9 0.01 5	Accuracy Kappa	0.6222559 0.4276597
	Bagging	nbagg minsplit cp		Accuracy Kappa	0.6157118 0.417191
	Supported Vector Machine	cost sigma kernel	1 0.1 radial	Accuracy Kappa	0.63363 0.4344153
	Boosted Logistic Regression	niter	10000	Accuracy Kappa	0.4604037 0.2125741

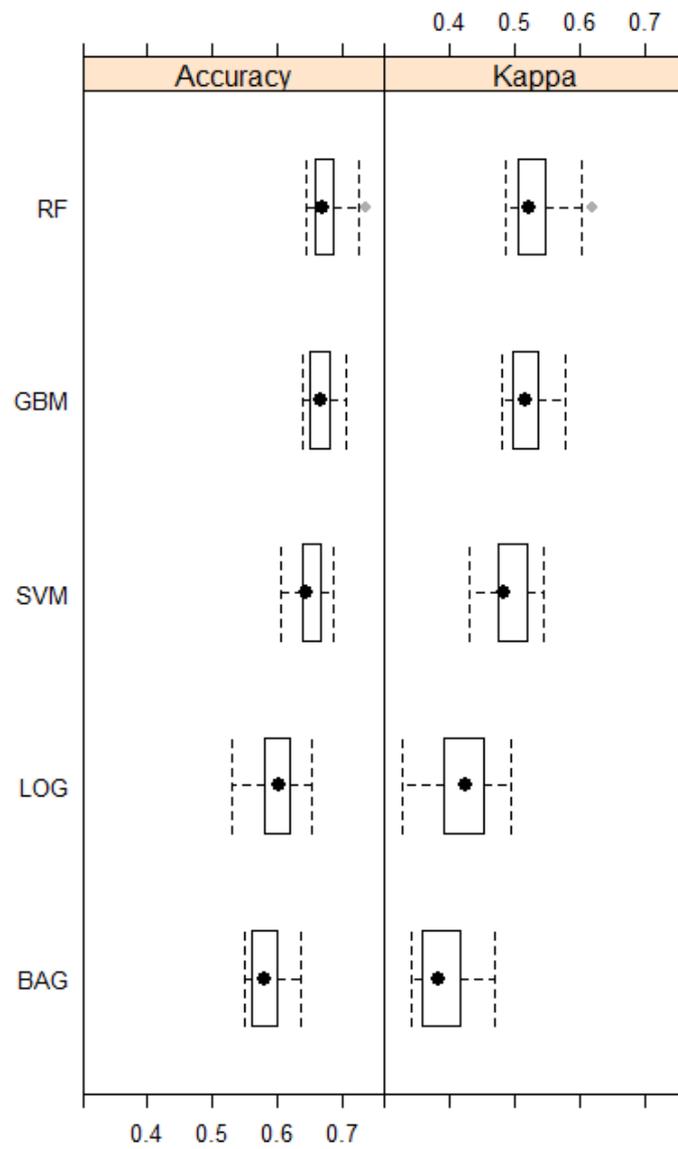


Figura 4-2. Caracterización de diferencias entre modelos mediante el remuestreo de sus distribuciones para el Modelo 1.

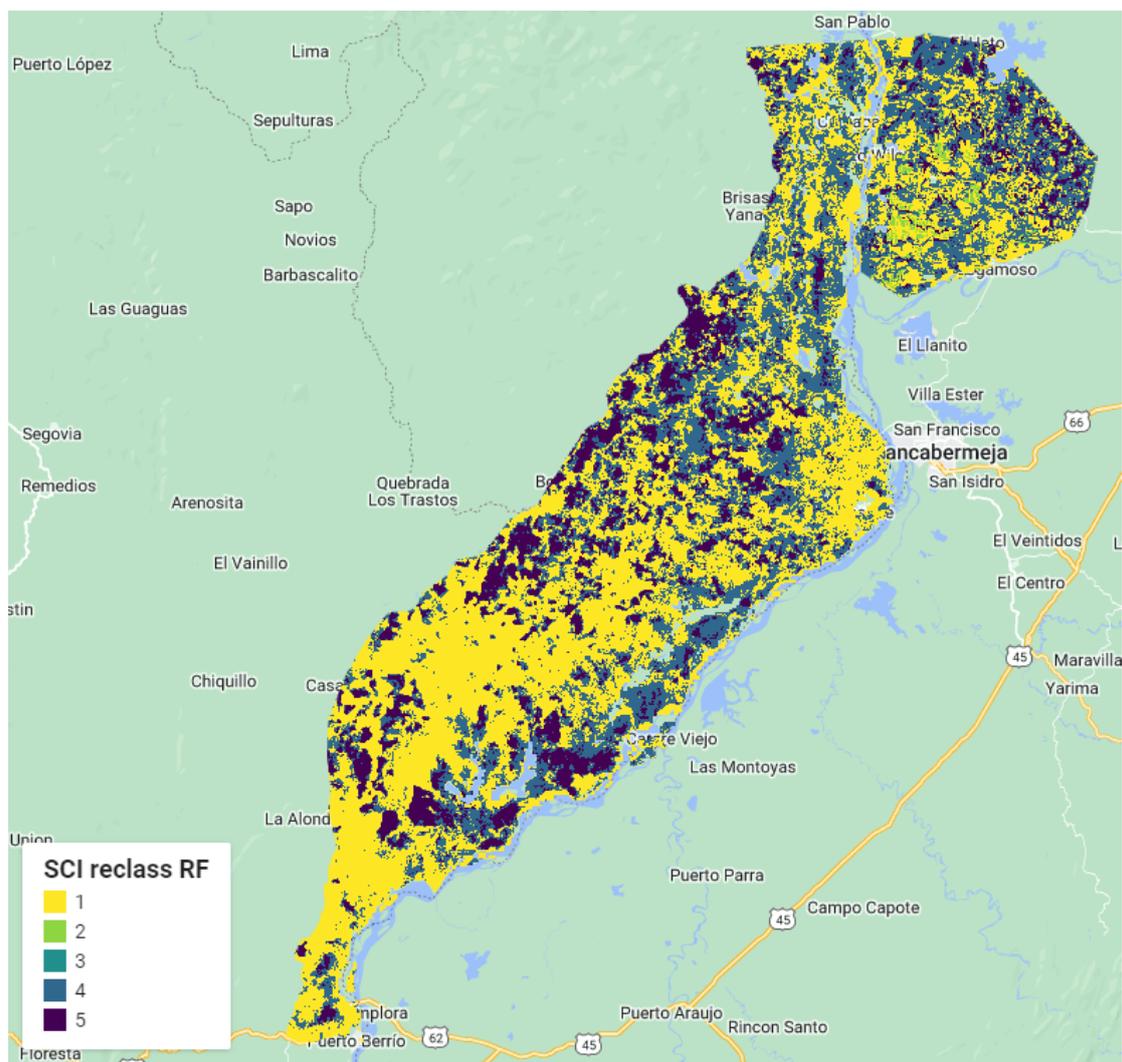


Figura 4-3. Resultado de la modelación del SCI reclasificado a partir de datos de SR, utilizando el algoritmo RF.

Como se mencionó en la sección 3.2.1, el objetivo de la elaboración de este modelo es identificar la información de SR que más se relaciona con SCI, encontrando esas regiones del espectro que alcanzan capturar la naturaleza de este indicador. Por esta razón, luego de determinar el mejor modelo para la estimación del SCI reclasificado, se identificaron las variables más importantes para cumplir con este objetivo. En el caso de RF, adicional a la medición de exactitud temática presentada en la Tabla 4-1, el algoritmo permite además generar dos indicadores de revisión de la clasificación y evaluación de las variables predictoras, los cuales son el índice de disminución media de la precisión (Mean Decrease Accuracy, MDA) y el índice de disminución media de Gini (Mean Decrease Gini, MDG) (Figura 4-4). El MDA permite evaluar las variables que mejoran la precisión de la clasificación al comparar el error promedio antes y después de la permutación de las variables en cada uno de los árboles del ensamble, mientras que MDG permite identificar qué variables presentan mayor homogeneidad en la clasificación, disminuyendo la impureza de Gini que se traduce

como la probabilidad de que un nuevo registro sea incorrectamente clasificado (Hong Han et al., 2016); Mancera, 2019).

Importancia de variables

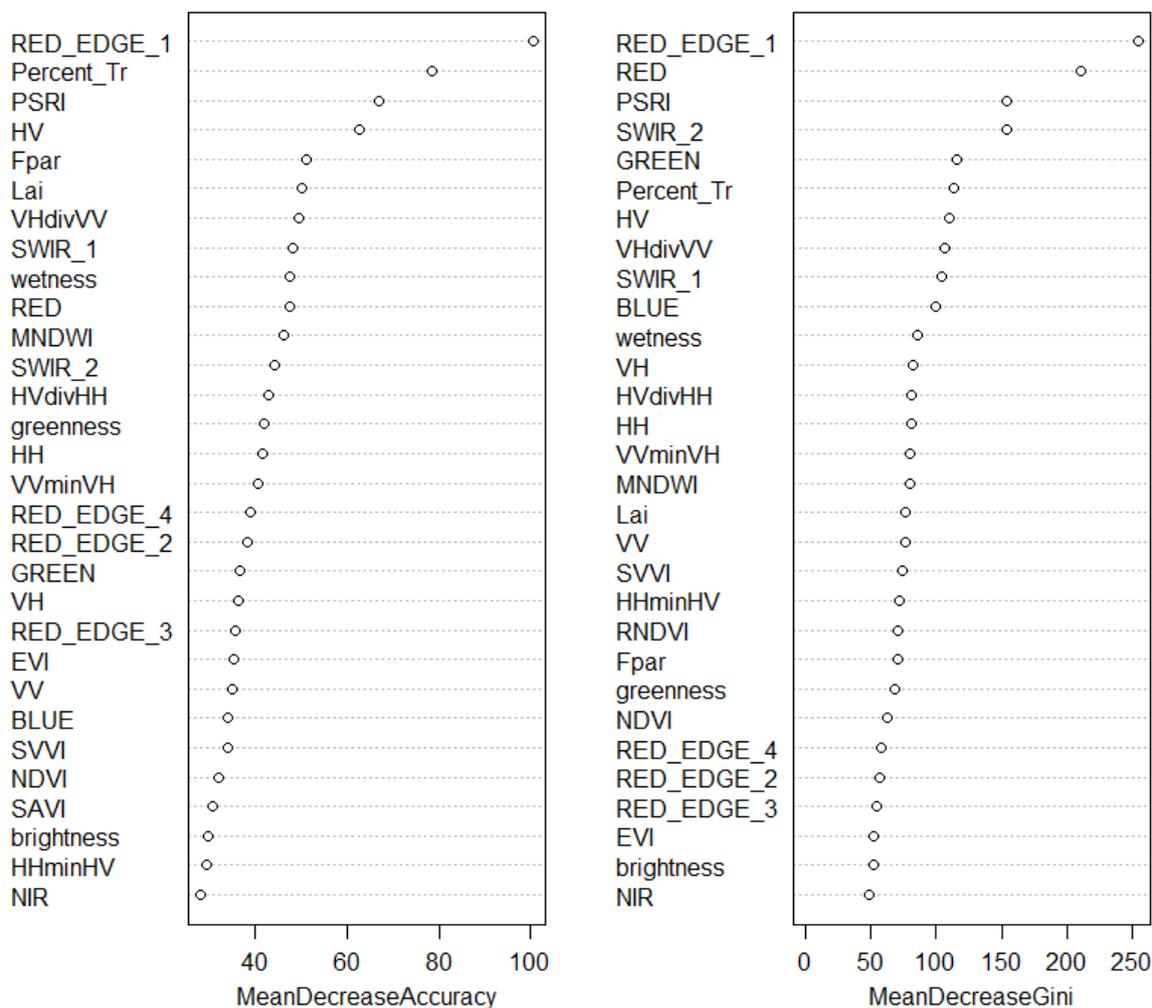


Figura 4-4. Importancia de variables calculadas en R para el Modelo 1.

Dado que el objetivo del Modelo 1 es determinar la relación existente entre datos de SR y el SCI, se tomarán las primeras cinco variables de mayor importancia detectadas por el decrecimiento promedio del índice de Gini, y se ilustrará su comportamiento en relación con las categorías de SCI (Figura 4-5). Dado que MDG permite identificar la variable que más aportan a la disminución de la impureza de los nodos, dando árboles más homogéneos localmente, en lugar de brindar un resultado global de la modelación como lo hace MDA, se optó por elegir el orden de las variables de acuerdo con el indicador MDG.

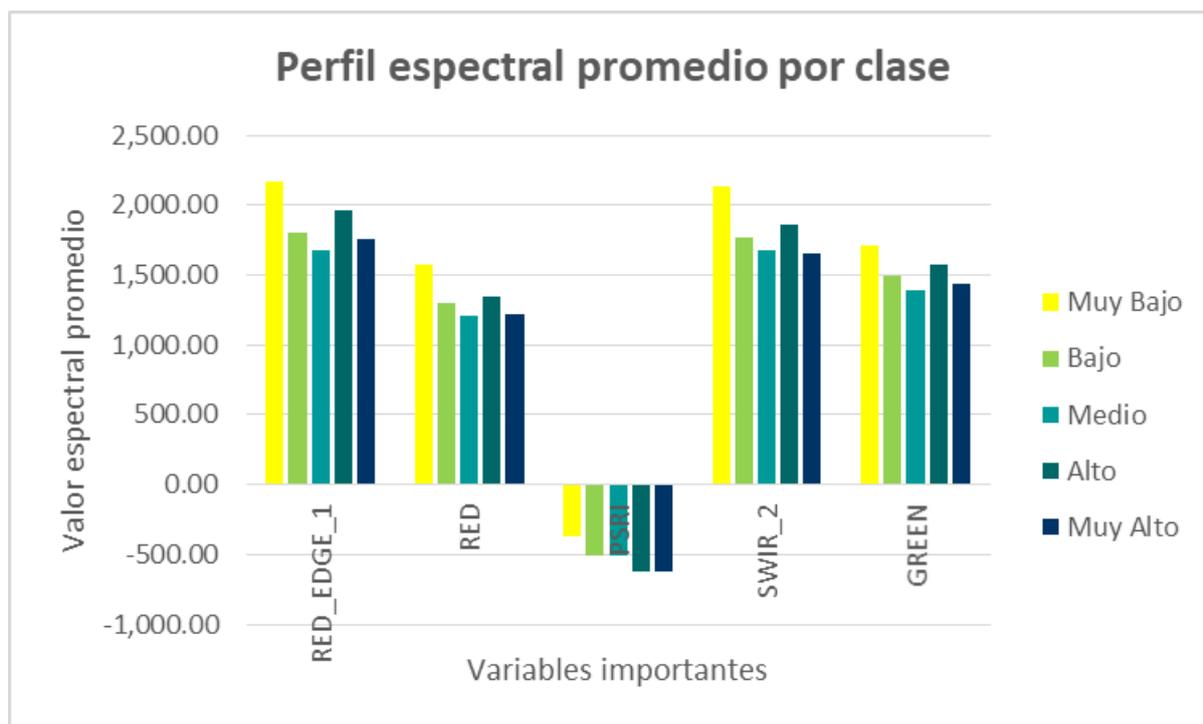


Figura 4-5. Perfil espectral promedio por clase de SCI, utilizando las variables más importantes identificadas por el Modelo 1.

Con este primer ejercicio de modelación identifiqué el algoritmo que mejor se ajusta a las categorías de SCI, así como las variables de sensores remotos de mayor importancia para la modelación. Esta información será relevante para la elaboración del Modelo 3, en el que se tomarán las variables más importantes de los modelos 1 y 2 para estimar el índice de calidad de hábitat propuesto.

4.2 Modelo 2

Para el segundo componente de calidad de bosque, se siguió la misma metodología y se evaluaron los mismos hiperparámetros con los cinco algoritmos. Al igual que con el Modelo 1, el análisis de Boruta consideró a las 31 variables como relevantes para el modelamiento, por encima de la aleatoriedad, por lo que ninguna variable fue excluida del análisis (Figura 4-6). En la Figura 3-6 se presentaron unos polígonos en los que fue posible determinar la calidad de bosque con base en la información de registros biológicos, su categorización de hábitat y su ponderación con coberturas. En estos polígonos se desplegó una grilla de 500 m y 1 km para extraer estos valores y modelarlos con las variables de sensores remotos para identificar las variables más importantes como se hizo con el Modelo 1 (Tabla 4-2). En este caso, también se

identificó que el modelo Random Forest, calculado a partir de la grilla de 500 m, es el que presenta una mayor exactitud temática; no obstante, el desempeño de los modelos internos generados en el entrenamiento de control y ajuste de hiperparámetros fue muy similar entre los algoritmos utilizados, exceptuando los generados por Boosted Logistic Regression, el cual presentó una amplia variabilidad en la exactitud temática dentro del ejercicio de entrenamiento de control por validación cruzada (Figura 4-7). Los valores de hiperparámetros con mejor desempeño de la tabla 11 fueron utilizados para realizar la predicción de la calidad de bosque con el raster stack construido en GEE (Figura 4-8).

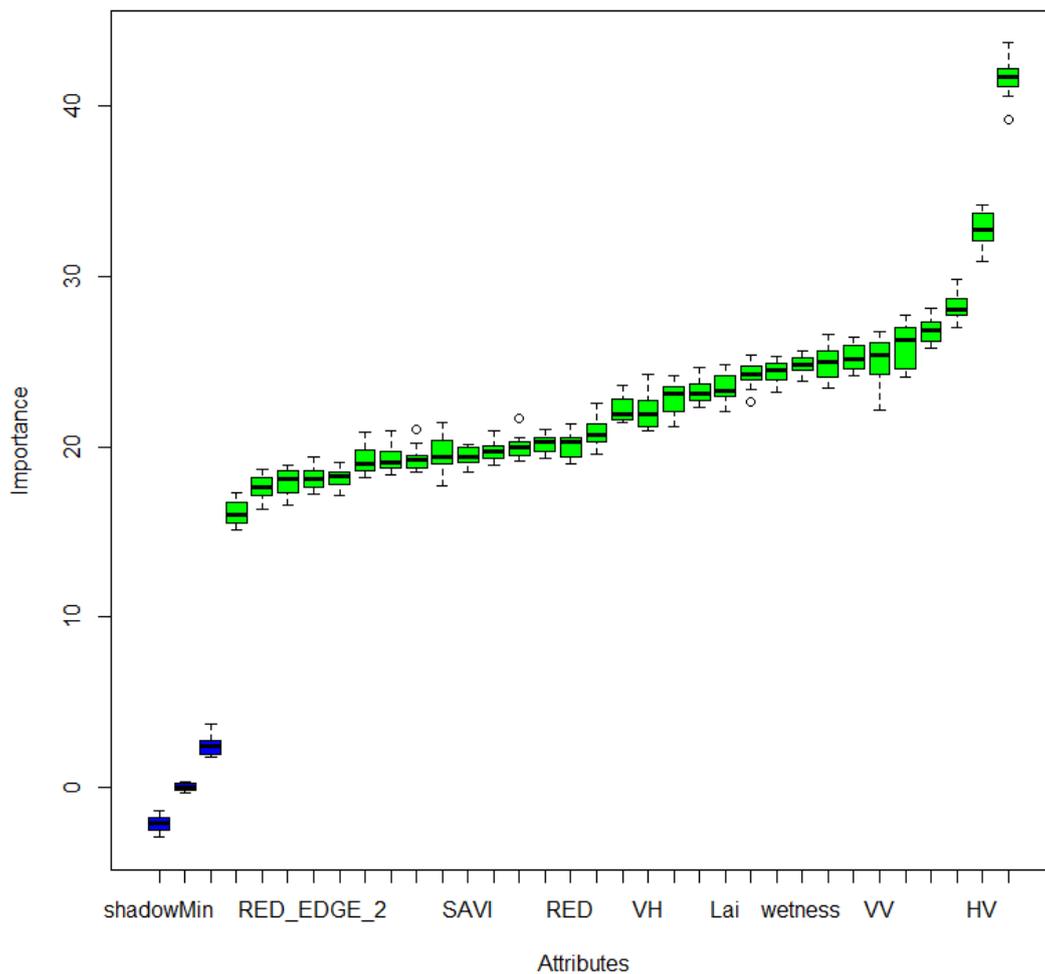


Figura 4-6. Importancia de variables medida previamente a la modelación del Modelo 2. Las 31 variables predictoras demostraron tener una importancia mayor que la generada únicamente por aleatoriedad al compararlas con las variables sombra. Entre las variables con mayor importancia encontramos los índices de Radar HVdivHH y VHdivVV, el coeficiente de retrodispersión HV y los índices RNDVI y Fpar.

Tabla 4-2. Resultados del ajuste de hiperparámetros para el Modelo 2 de estimación de calidad de bosque con variables de SR.

Muestreo	Algoritmo	Parámetros	Valor de Parámetros	Exactitud	Valor de Exactitud
Grilla 500 m	Random Forest	mtry splitrule min.node.size ntrees	10 gini 3 1000	Accuracy Kappa	0.6407577 0.3977078
	Gradient Boosting Machine	n.trees interaction.depth shrinkage n.minobsinnode	1000 9 0.01 10	Accuracy Kappa	0.6401547 0.4026838
	Bagging	nbagg minsplit cp	30 4 0	Accuracy Kappa	0.6287607 0.3897406
	Supported Vector Machine	cost sigma kernel	100 0.01 radial	Accuracy Kappa	0.6437961 0.3974233
	Boosted Logistic Regression	niter	500	Accuracy Kappa	0.4508657 0.2102942
Grilla 1 km	Random Forest	mtry splitrule min.node.size ntrees	8 extratrees 3 1000	Accuracy Kappa	0.5976182 0.3141892
	Gradient Boosting Machine	n.trees interaction.depth shrinkage n.minobsinnode	500 9 0.01 1	Accuracy Kappa	0.5955705 0.3294731
	Bagging	nbagg minsplit cp	30 6 0.01	Accuracy Kappa	0.5800314 0.2633887
	Supported Vector Machine	cost sigma kernel	100 0.01 radial	Accuracy Kappa	0.5797942 0.402144707
	Boosted Logistic	niter	5000	Accuracy Kappa	0.5995034 0.3325034

	Regression				
--	------------	--	--	--	--

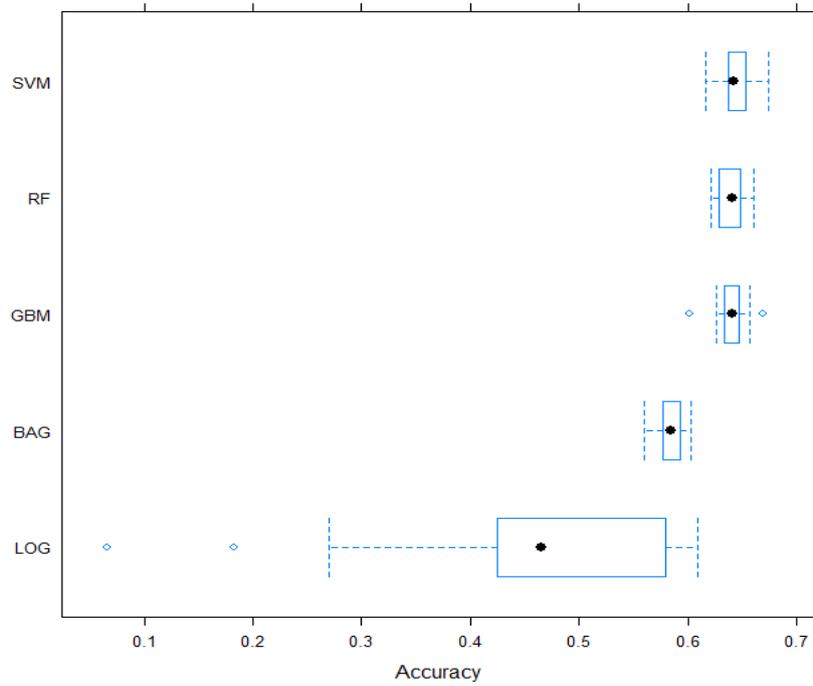


Figura 4-7. Caracterización de diferencias entre modelos mediante el remuestreo de sus distribuciones para el Modelo 2.

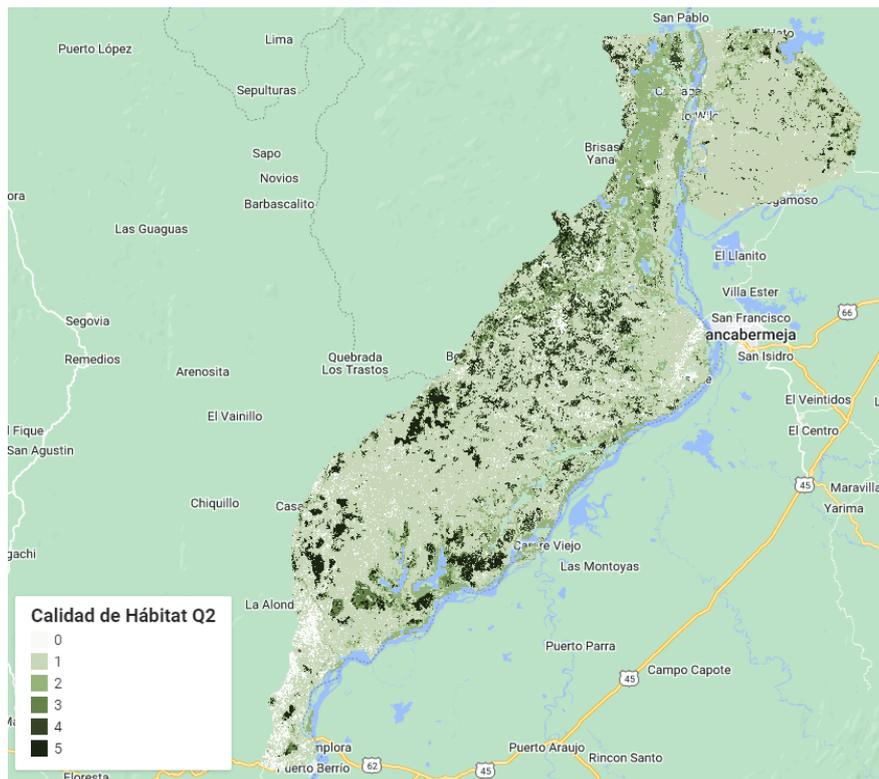


Figura 4-8. Resultado de la modelación de la calidad de bosque a partir de datos de SR, utilizando el algoritmo RF.

Luego de determinar el mejor modelo para la estimación de la calidad de hábitat, se identificaron las variables más importantes para encontrar la relación de este índice propuesto con los SR (Figura 4-9).

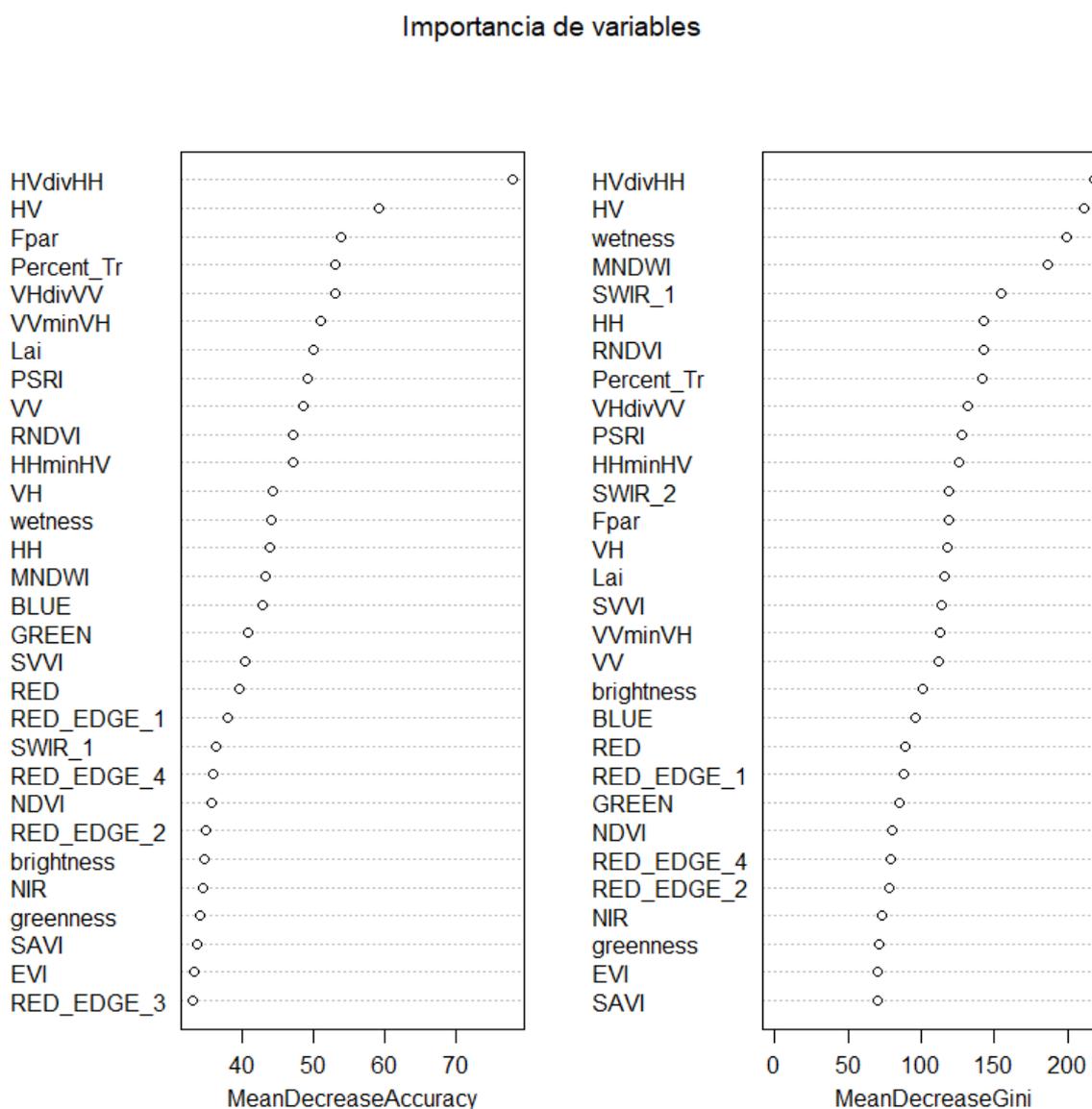


Figura 4-9. Importancia de variables calculadas en R para el Modelo 2.

Al igual que con el Modelo 1, nos interesa determinar la relación existente entre datos de SR y la calidad de bosque propuesta a partir de registros y coberturas de la tierra. Por esta razón, se tomarán las cinco primeras variables de mayor importancia

detectadas por la media de decrecimiento del índice de Gini y se ilustrará su comportamiento en relación con las categorías de calidad de bosque (Figura 4-10).

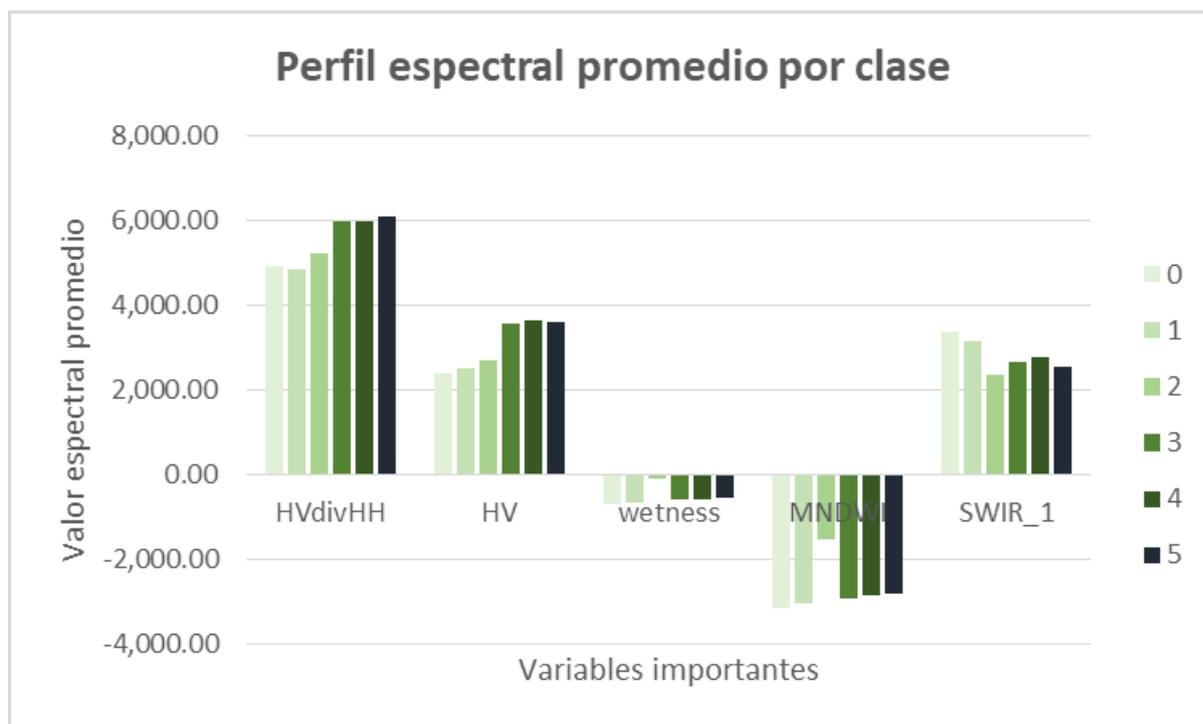


Figura 4-10. Perfil espectral promedio por clase de calidad de hábitat, utilizando las variables más importantes identificadas por el Modelo 2.

Con este segundo ejercicio de modelación identifiqué el algoritmo que mejor se ajusta a las categorías de calidad bosque, así como las variables de sensores remotos de mayor importancia para la modelación. Este resultado se convertirá en la variable respuesta del Modelo 3, cuyas variables predictoras corresponden a las variables más importantes de los modelos 1 y 2.

4.3 Modelo 3

Para el modelo 3 de calidad de hábitat integral, se siguió la misma metodología y evaluación de hiperparámetros de los modelos 1 y 2. Los puntos de entrenamiento y validación son los mismos que se emplearon para el modelo 2, ya que con estos puntos puede medirse simultáneamente el SCI y la calidad de bosque para identificar las áreas de congruencia e incongruencia entre estos dos componentes (Figura 4-11).

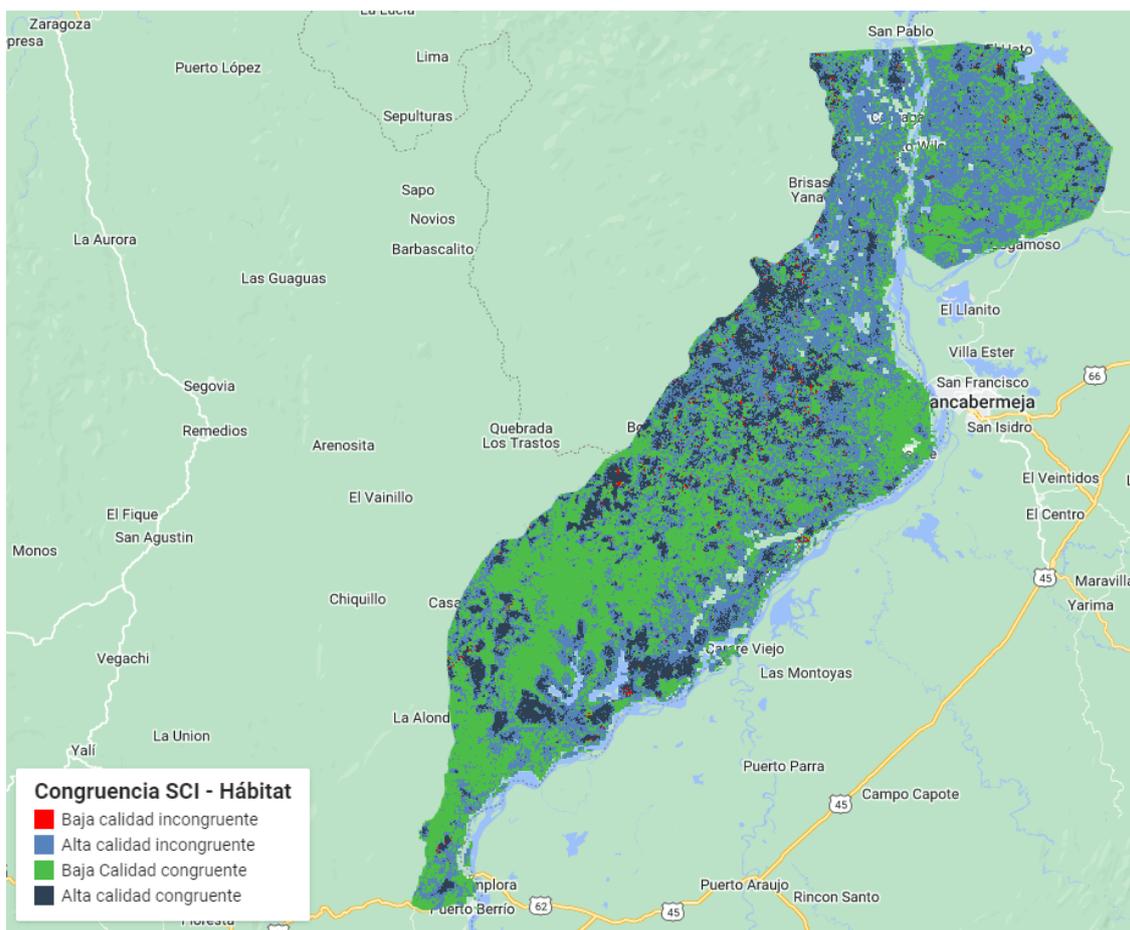


Figura 4-11. Áreas de congruencia e incongruencia entre SCI y Calidad de Bosque. A diferencia de la figura 3-9, este resultado si corresponde a la variable respuesta que se usará para el Modelo 3, creado con el mejor resultado del Modelo 2 y el SCI reclasificado de Hansen. Las áreas diferentes a los cuerpos de agua que no tienen valores, corresponden a valores que no pudieron clasificarse

Este modelo solo tuvo en cuenta 8 variables predictoras (las 4 variables más importantes de los modelos 1 y 2), y al realizar el análisis de Boruta, como era de esperarse, las consideró todas como relevantes para el modelamiento, por encima de la aleatoriedad, por lo que ninguna variable fue excluida del análisis (Figura 4-12). El muestreo que se utilizó para este modelo es el mismo que se usó en el modelo 2 para obtener el mejor modelo, es decir, la grilla de 500 m mostrada en la Tabla 4-2. En este caso, el modelo con la mejor exactitud temática fue el generado por GBM, guardando una muy estrecha diferencia con SVM y RF (Tabla 4-3; Figura 4-13). Los valores de hiperparámetros con mejor desempeño de la tabla 12 fueron utilizados para realizar la predicción de la calidad de bosque con el raster stack construido en GEE (Figura 4-14).

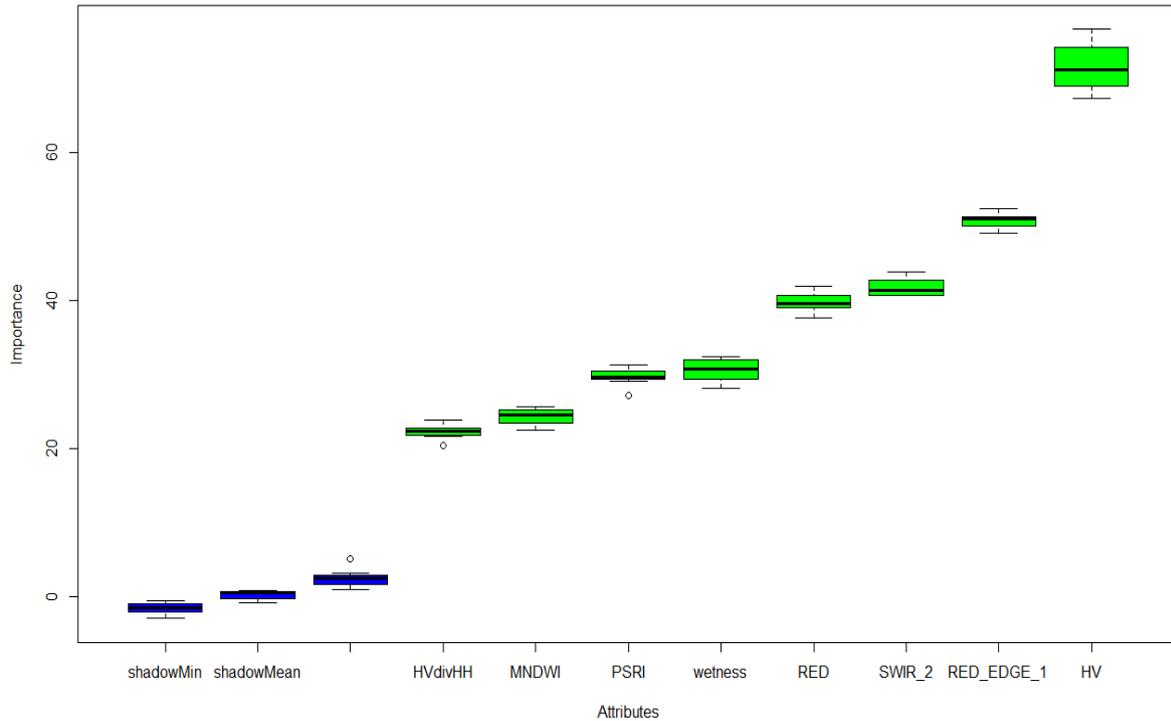
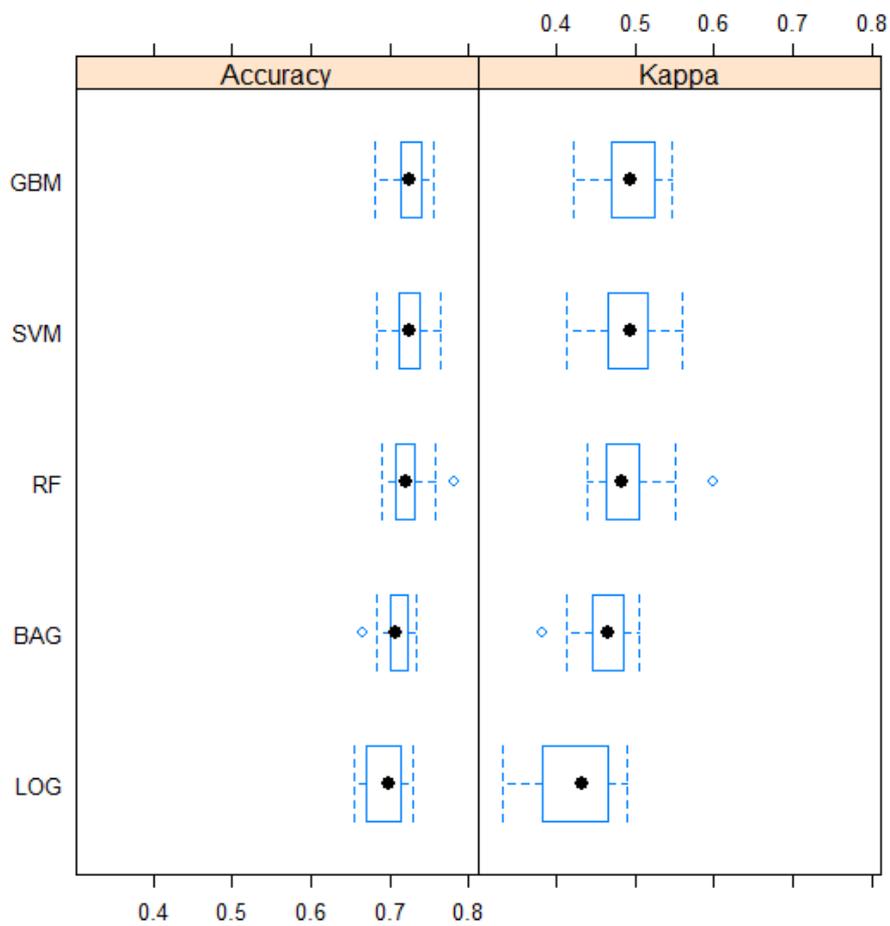


Figura 4-12. Importancia de variables medida previamente a la modelación del Modelo 3. Las 8 variables predictoras (Boxplots color verde) demostraron tener una importancia mayor que la generada únicamente por aleatoriedad al compararlas con las variables sombra (Boxplots azules). Entre las variables con mayor importancia encontramos el coeficiente de retrodispersión HV y las bandas RED_EDGE_1 y SWIR_2.

Tabla 4-3. Resultados del ajuste de hiperparámetros para el Modelo 3 de estimación de calidad de bosque integrado con las variables más importantes de SR de los Modelos 1 y 2.

Muestreo	Algoritmo	Parámetros	Valor de Parámetros	Exactitud	Valor de Exactitud
Grilla 500 m	Random Forest	mtry splitrule min.node.size ntrees	4 extratrees 5 1000	Accuracy Kappa	0.7221914 0.4883358
	Gradient Boosting Machine	n.trees interaction.depth shrinkage n.minobsinnode	500 9 0.01 1	Accuracy Kappa	0.7244411 0.4932218

	Bagging	nbagg minsplit cp	10 2 0.01	Accuracy Kappa	0.7115002 0.4668948
	Supported Vector Machine	cost sigma kernel	10 0.1 radial	Accuracy Kappa	0.7236835 0.4893216
	Boosted Logistic Regression	niter	500	Accuracy Kappa	0.6955947 0.4257818



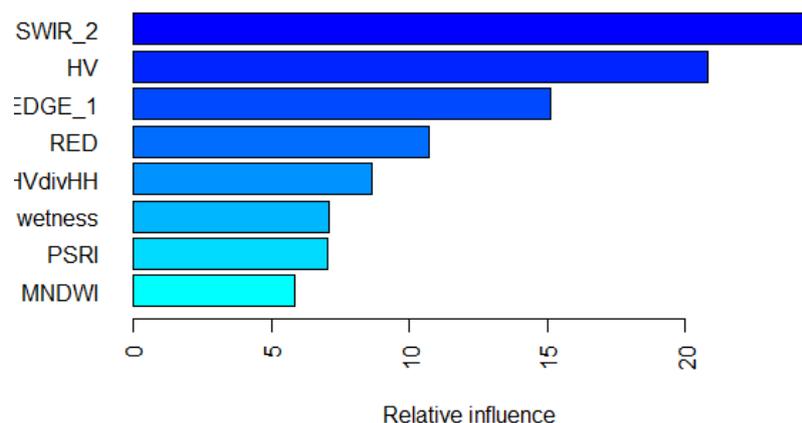


Figura 4-15. Importancia de variables calculadas en R para el Modelo 3. El paquete de R caret ofrece una perspectiva diferente al momento de medir la importancia de las variables cuando se utiliza el algoritmo de GBM.

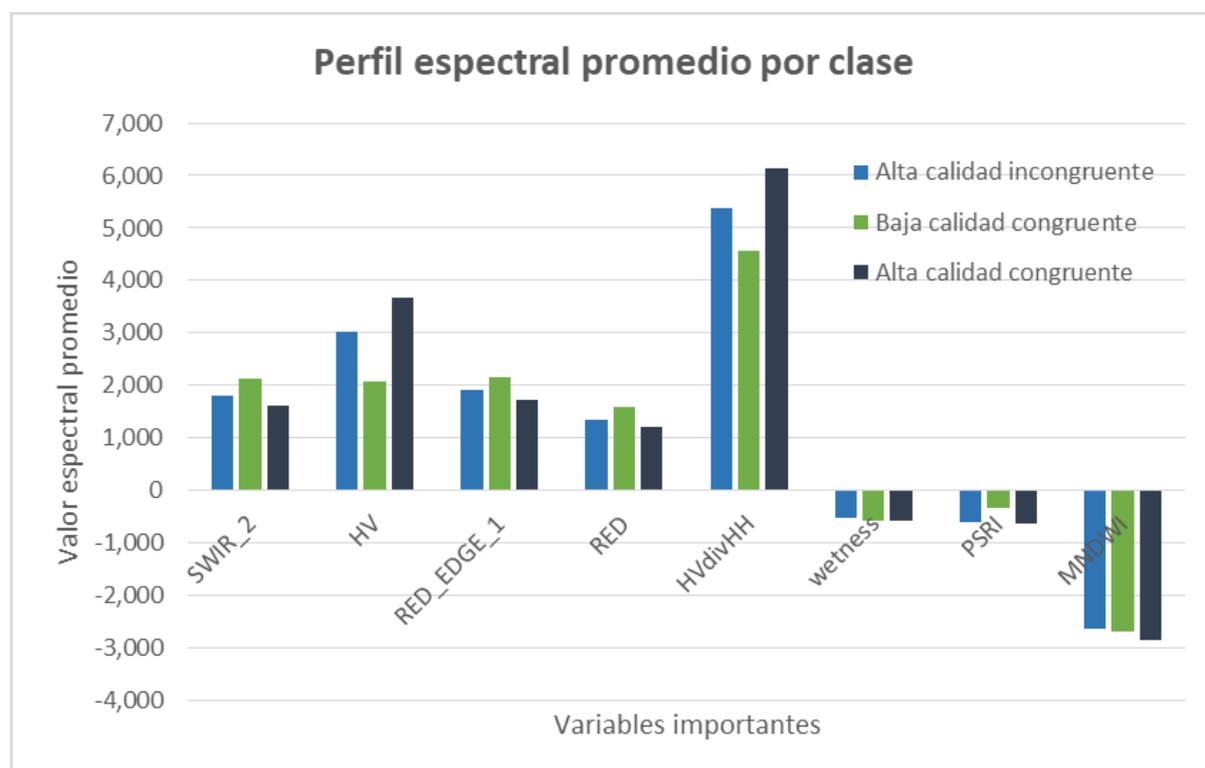


Figura 4-16. Perfil espectral promedio por clase del índice de calidad de hábitat integral, utilizando las variables más importantes identificadas para el Modelo 1 y el Modelo 2.

Llegados a este punto, se identificaron las variables más importantes para los modelos 1 y 2, las cuales fueron utilizadas para la construcción del modelo 3. En la Tabla 4-4 se presentan la totalidad de variables explicativas utilizadas en este proyecto y el valor de importancia utilizando el promedio de decrecimiento de Gini calculado para el mejor algoritmo de cada modelo.

Tabla 4-4. Decrecimiento promedio de Gini para las variables explicativas en los 3 modelos propuestos

Variable	Rango de importancia medida con decrecimiento promedio de Gini (1= variable más importante)		
	Modelo 1	Modelo 2	Modelo 3
SWIR_2	4	12	1
HV	7	2	2
RED_EDGE_1	1	22	3
RED	2	21	4
HVdivHH	13	1	5
wetness	11	3	6
PSRI	3	10	7
MNDWI	16	4	8
GREEN	5	23	--
Percent_tree	6	8	--
VHdivVV	8	9	--
SWIR_1	9	5	--
BLUE	10	20	--
VH	12	14	--
HH	14	6	--
VVminVH	15	17	--
LAI	17	15	--
VV	18	18	--
SVVI	19	16	--
HHminHV	20	11	--
greenness	21	28	--
RNDVI	22	7	--

FPAR	23	13	--
NDVI	24	24	--
RED_EDGE_4	25	25	--
RED_EDGE_2	26	26	--
RED_EDGE_3	27	31	--
EVI	28	29	--
Brightness	29	19	--
NIR	30	27	--
SAVI	31	30	--

5. Discusión

Con el desarrollo de este proyecto identificamos una metodología para medir la calidad de hábitat en términos de la condición estructural del bosque y la presencia de especies con alta dependencia a este ecosistema. Con un enfoque principalmente desde la geomática, determinamos las bandas e índices espectrales con mayor incidencia sobre el SCI, la calidad de bosque desde los registros biológicos y la congruencia de estos dos componentes. Este es el primer paso para generar un indicador más robusto de calidad de hábitat. El desempeño de diferentes algoritmos fue muy similar, no obstante, se observaron diferencias al cambiar los datos utilizados para entrenamiento y prueba de los modelos. A pesar de que los modelos no obtuvieron una exactitud temática por encima del 70%, entendemos que la evaluación de una propiedad compleja como la calidad de hábitat asociada al uso de hábitat, es una característica que no puede medirse únicamente desde los SR satelitales. Por ejemplo, para el modelo 1 entre SCI y SR, no es posible capturar el componente de tiempo de intervención, pues la metodología se planteó para analizar la información de las bandas del presente y no las dinámicas de cambios que podrían aproximarse a este componente del SCI. El modelo 2 por otro lado, solo se basa en la presencia de las especies, la cual es extendida con ayuda de los polígonos de coberturas de la tierra, y dentro de los cuales puede existir una gran variabilidad en la respuesta espectral. Esta variabilidad dificulta el ejercicio de modelación al encontrarse con un amplio rango de valores de radiancia en áreas que pueden clasificarse con valores cercanos de nuestro índice de calidad de hábitat.

A lo largo de los ejercicios de modelación identificamos el comportamiento de las variables de SR en respuesta al índice sintético del SCI y a la calidad de hábitat de bosque percibida a partir de los registros de las especies. En el primer modelo, el muestreo con la mejor exactitud temática fue el producido con los registros biológicos, lo cual puede deberse a un efecto del sesgo muestral de los registros, en donde se evidencia una mayor representatividad en ciertos valores del SCI, favoreciendo la predicción de estas clases. Este problema del sesgo muestral fue evidenciado desde el principio del ejercicio de modelación con las 18 categorías de SCI. Como se mencionó en la sección 4.1 de los resultados, las categorías del SCI tuvieron que ser reducidas a 5 clases, pues el desempeño de los modelos era muy bajo, debido al poco balance de muestreo en algunas categorías, afectando así el correcto entrenamiento de los modelos (Luque et al., 2019). Aún con esta consideración, no fue posible reducir por completo el desequilibrio entre las clases, aunque sí se aumentó considerablemente el desempeño final de los modelos. En la Figura 5-1 se muestra la distribución de los valores de SCI reclasificado, utilizados en los puntos de entrenamiento del modelo 1, en donde las clases más representadas son las extremas (1, 3 y 4); esto quiere decir que el modelo se ajusta e identifica las clases que son mayormente separables. El problema del sesgo muestral de los registros tiende a suavizarse en el modelo 2 con ayuda de las coberturas de la tierra, no obstante, es una característica que también afecta los resultados de los modelos 2 y 3.

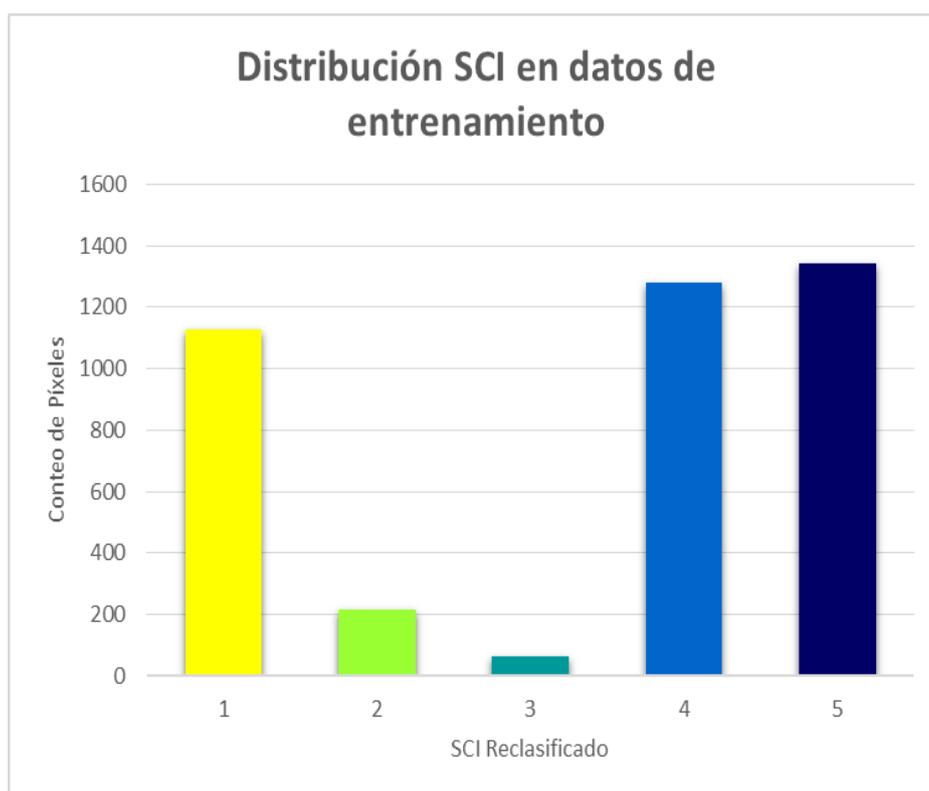


Figura 5-1. Distribución de los valores de SCI reclasificado en los datos utilizados para el entrenamiento del Modelo 1.

Además del efecto identificado por el desequilibrio de clases, la manera en cómo están definidas las clases produce una alta variabilidad en la respuesta espectral de las variables explicativas, lo que dificulta el ejercicio de clasificación (Figura 5-2). Entre los componentes del SCI, la penalización que recibe un píxel producto del año de pérdida es lo que dificulta la identificación de relaciones con los SR y la separabilidad de las clases predichas. Para comprobar la dificultad de la separabilidad de clases, se redujo el grado de sobreajuste producto de la dimensionalidad mediante una transformación lineal utilizando la técnica de análisis discriminante lineal (Figura 5-3), evidenciando que, efectivamente, las categorías de SCI son difícilmente separables.

Estos dos efectos de desequilibrio de clases y alta variabilidad espectral dentro de cada clase se evidencian con unos valores muy bajos de sensibilidad, especialmente para las clases 2 y 3 (0.284 y 0.129 respectivamente), mientras que las clases mayormente representadas tuvieron valores por encima de 0.6, siendo la clase 5 la de mayor valor (0.86). Por otro lado, la especificidad de todas las clases estuvieron por encima de 0.8, demostrando la capacidad del modelo de identificar correctamente los patrones negativos (Hossin & Sulaiman, 2019).

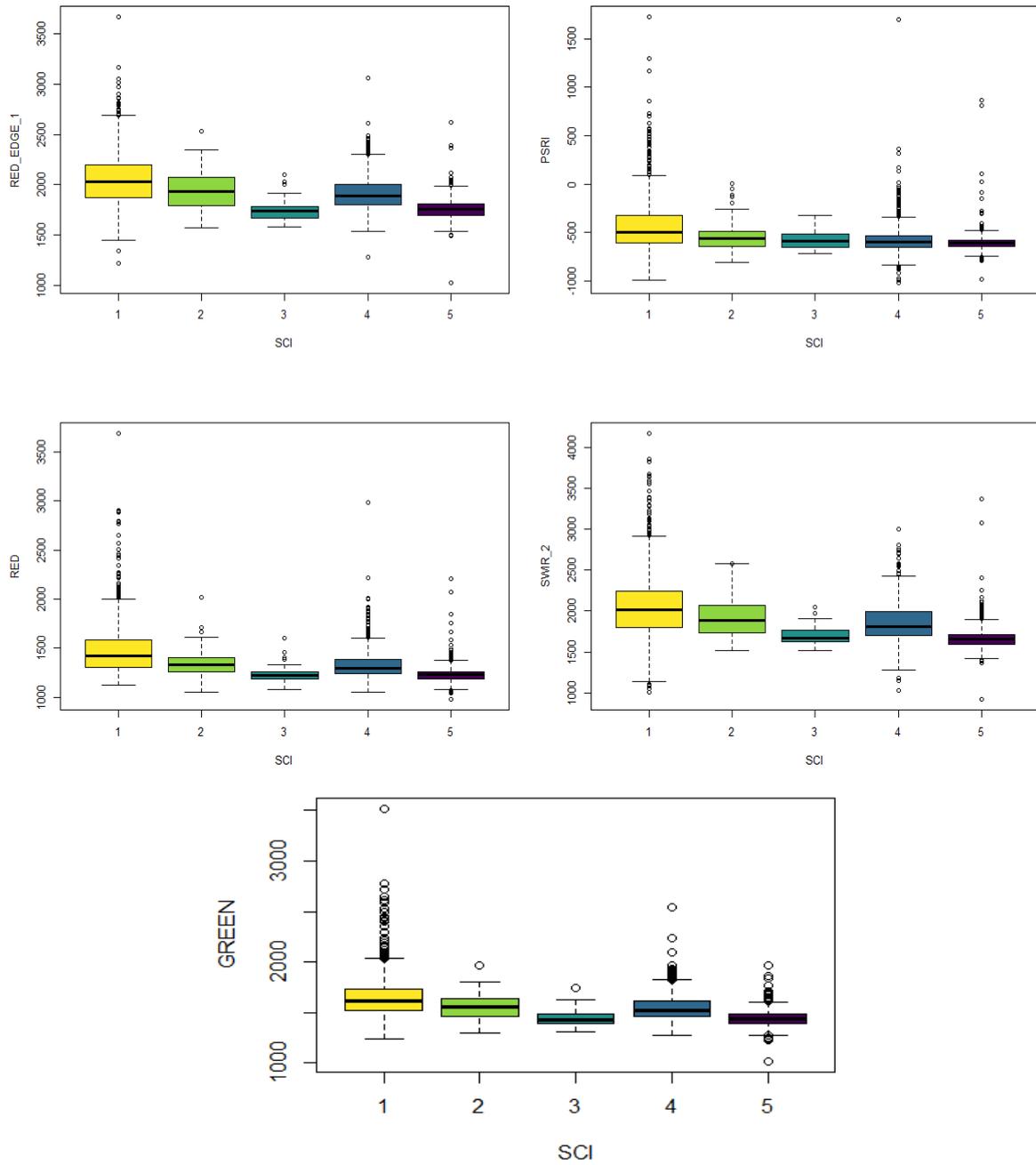


Figura 5-2. Variación espectral de las variables identificadas como más relevantes en el Modelo 1 de estimación de SCI reclasificado a cinco categorías.

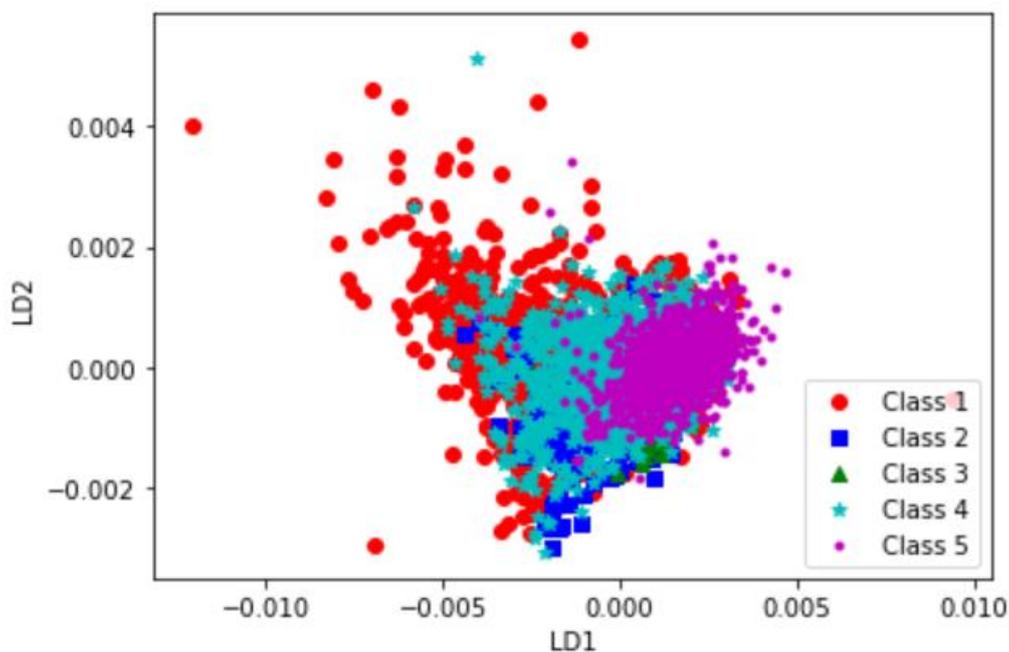


Figura 5-3. Categorías de SCI reclasificado proyectados los datos sobre los dos primeros discriminantes que reúnen cerca del 90% del radio de discriminabilidad.

Es importante recordar, que el objetivo de la pregunta 1 es identificar las variables de SR que mejor se relacionan con el SCI. Teniendo esto en cuenta, las bandas de mayor importancia identificadas fueron RED_EDGE_1, RED, SWIR_2, PSRI y GREEN. En el caso de RED_EDGE_1, estudios demuestran que es una banda importante para el monitoreo de parámetros asociados al índice de área foliar (LAI) y al contenido de clorofila utilizados como bioindicadores directos del estatus de salud actual de las plantas y su productividad primaria bruta (Delegido et al., 2011; Mutanga & Skidmore, 2007), lo que podría estar obedeciendo al componente de porcentaje de dosel del SCI de Hansen. De igual manera, la banda RED también es sensible al LAI y a la orientación de las hojas en el dosel, pues se encarga de absorber la fluorescencia de clorofila inducida por el sol (SIF), que es emitida por las hojas, dispersada y reabsorbida dentro del dosel, y que posteriormente es captado por los SRs para ser utilizado como estimativo de la fotosíntesis y de la Productividad Primaria Bruta (Qiu et al., 2019).

La banda SWIR_2 junto con las bandas RED y NIR son componentes relevantes para la capacidad de separar entre bosques saludables y perturbados, por lo que existen índices que combinan estas bandas para evaluar la estructura del dosel, humedad y estrés (Bolgen et al., 2018; Gibson et al., 2020; Huo et al., 2021; Lastovicka et al., 2020; Morcillo-Pallarés et al., 2019). La banda verde es un factor importante para la predicción de variables de bosque asociadas a su estructura como el volumen y diámetro de tallos, altura de árboles, área basal, entre otros (Astola et al., 2019; Grabska et al., 2019). Finalmente, el índice PSRI que se constituye fundamentalmente de las regiones Red edge, red y green, también está entre las

variables más importantes debido a la complementariedad de este índice con las bandas anteriormente nombradas, indicando que hay aspectos en esta región espectral, que son detectables con el uso del PSRI y no se observaron con las bandas individuales (Figura 5-4). Por otro lado, la variable SCI no muestra ninguna correlación fuerte con ninguna variable, siendo la más significativa con RED_EDGE_1 y la menos significativa con PSRI.

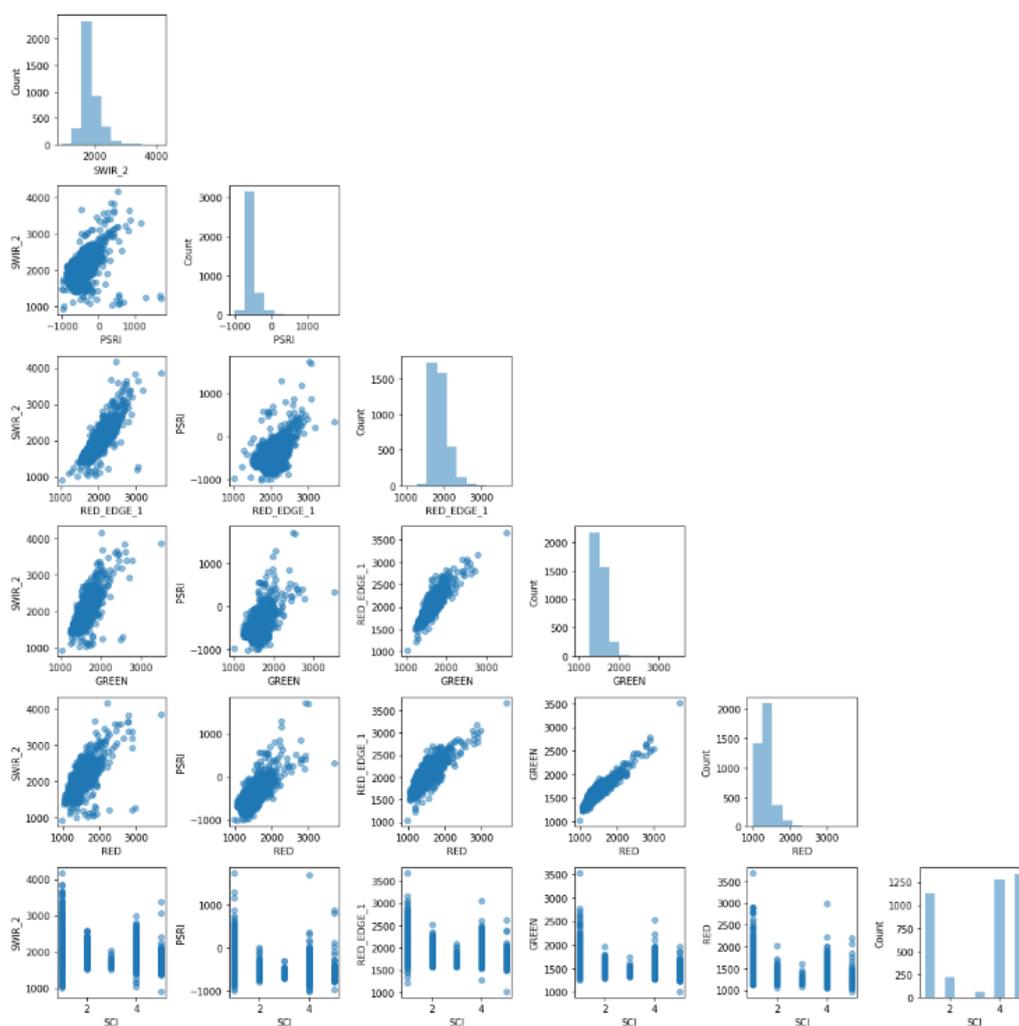


Figura 5-4. Matriz de correlación entre las variables más importantes del modelo 1. Nótese como muchas de las variables presentan una alta correlación, siendo las bandas RED y GREEN las de mayor valor (0.94), seguidas de GREEN y RED_EDGE_1 (0.92).

Analizando los resultados del modelo 2, su desempeño fue muy similar al del Modelo 1, con una exactitud temática de 0.641 e identificando como variables más importantes los coeficientes de retrodispersión HVdivHH, HV, el tasseled cap de humedad, el índice MNDWI y la banda SWIR_1 (Figura 5-5). De manera similar que con el Modelo 1, se identificó una alta correlación entre MNDWI y Tasseled Cap

wetness, ya que ambas variables se utilizan para la interpretación de cuerpos de agua, mientras que SWIR tiene una mejor respuesta ante la estructura densa de los bosques. De igual manera, la variable calidad de bosque no muestra ninguna correlación fuerte con ninguna variable, siendo la más significativa con HVdivHH (0.43) y la menos significativa con MNDWI (0.17). Se esperaba que los datos de Radar fueran relevantes para el modelo 1, ya que podrían asociarse con el componente de la altura de dosel, sin embargo, aparecieron en el modelo 2.

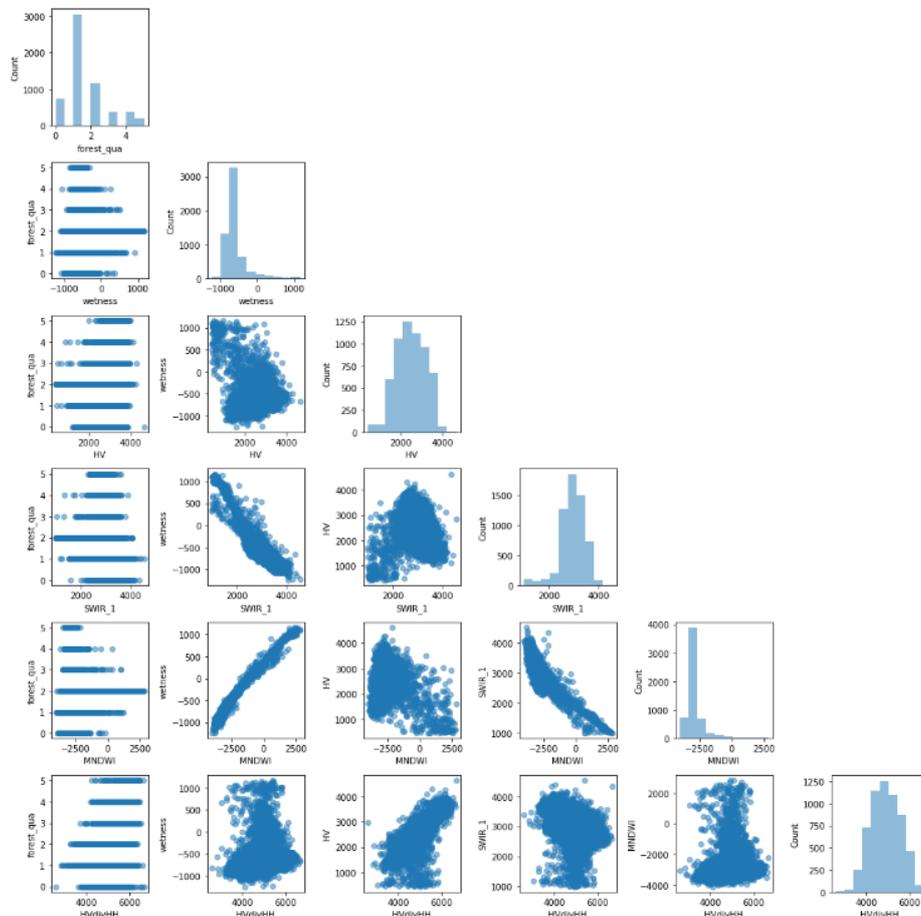


Figura 5-5. Matriz de correlación entre las variables más importantes del modelo 2. En este caso no se evidencia tanta correlación como con el Modelo 1, siendo las correlaciones entre las bandas MNDWI y wetness las de mayor valor (0.98) y entre wetness y SWIR_1 (-0.86).

Gracias a la generalización de los registros planteada por polígonos y coberturas de la tierra, fue posible disminuir el error generado por el sesgo en el muestreo. Esto aplica especialmente en polígonos con registros de especies de plantas, donde la generalización puede sustentarse con ayuda de la fenología de las especies que genera una respuesta espectral relativamente homogénea que puede detectarse a través de sensores remotos. Por otro lado, los polígonos con registros mayoritariamente de especies animales, siguen los supuestos mencionados en la sección 3.2.2, siendo el supuesto de la persistencia de la especie en un polígono uno de los que tiene mayor efecto en los resultados. La persistencia de una especie en un

área determinada no depende únicamente de las condiciones físicas del ambiente, sino que también interfieren aspectos de la ecología poblacional que excede los alcances de este proyecto, pero que valdría la pena considerarlos en un escenario donde la identidad de las especies sea tomada en cuenta.

Una variable de particular interés utilizada en los modelos 1 y 2 es la de Vegetation Continuous Field, medida como el porcentaje de árboles por píxel. Esta variable, aunque no es la misma utilizada por Hansen, puede considerarse como una de las variables constituyentes del SCI por su relación con el porcentaje de cobertura de dosel. A pesar de esta cualidad, VCF no aparece entre las variables más importantes cuando se utiliza el indicador de importancia MDG, aunque con el indicador MDA si se posiciona como la segunda variable más importante después de RED_EDGE_1. Esto se debe, a que si la variable VCF es retirada durante el entrenamiento, el rendimiento promedio de los árboles que conforman el ensamble se verá afectado, lo que es lógico puesto que el SCI considera el porcentaje de cobertura de dosel para crear sus categorías (Han, Guo & Yu, 2016). Por otro lado, cada componente del SCI tiene una relación uno a varios, lo que aumenta la variabilidad de las categorías para un mismo criterio, reduciendo la capacidad predictiva de una variable predictiva y aumentando la complejidad de los nodos. Por citar un ejemplo, un píxel con una altura de dosel de entre 5 y 15 metros, con porcentaje de cobertura de dosel de más del 95% puede categorizarse como un SCI de 1, 4 o 12, repitiéndose este comportamiento a mayores alturas de dosel. De esta manera, cuando se usa VCF como variable para la creación de nodos, la pureza de estos se ve afectada debido a su respuesta multiclase con el SCI. Desafortunadamente, este efecto no pudo solucionarse con la clasificación de las cinco categorías finales, pues cada categoría incluye también múltiples valores de porcentaje de dosel y altura, especialmente la categoría 1 que se mantuvo como en la metodología original de Hansen y que muestra una alta variabilidad en la Figura 5-2. Un análisis similar se puede realizar con los coeficientes de retrodispersión de RADAR que pueden relacionarse con la altura de dosel.

El objetivo del desarrollo de esta metodología fue el de construir un índice apoyado en SR y datos de campo que evalúe la calidad de hábitat en términos no solo de la estructura de los bosques, sino su composición. Este objetivo se cumplió al identificar que las variables más importantes en la modelación 1 y 2 son diferentes, con la excepción de que ambos ejercicios consideraron la región SWIR del espectro electromagnético (Tabla 4-4). Esta similitud se debe a que la región SWIR es sensible a la humedad del suelo y la vegetación, y tanto la estructura como la composición medida con el uso de hábitat de especies dependientes de bosque (por tratarse de un área con ecosistemas potenciales de bosque húmedo tropical), están relacionadas con la humedad determinación de sus categorías (Fatehi et al., 2015). Gracias a este resultado, nos damos cuenta que se capturan procesos diferentes, materializándose en los patrones identificados por los algoritmos de aprendizaje. En caso de que las variables más importantes hubiesen sido las mismas o muy similares, hubiera sido un

indicador de que las aproximaciones a la estructura y composición de los bosques, no serían las más adecuadas para ser integradas en nuestro nuevo índice de calidad del Modelo 3, por lo cual no habría espacio para mejorar los indicadores con información de SR.

Uno de los resultados más importantes de este proyecto, es la identificación de las congruencias e incongruencias entre el SCI de Hansen, reclasificado a cinco categorías, y la calidad de bosque que se utilizaron como variable respuesta para el modelo 3. Como se evidenció en las figuras 3.9 y 4.11, fueron pocas las áreas identificadas como baja calidad incongruente en la variable respuesta utilizada en el modelo 3; por esta razón, el modelo GBM final no logró discriminar esta clase. Esto quiere decir que no fue posible identificar los atributos de SR que definen las áreas de alta calidad de hábitat desde los registros, pero de baja calidad desde el SCI. Estas áreas son de gran importancia, pues son indicativas de la plasticidad individual de especies objeto de conservación (que usualmente requieren de alta calidad de bosques), aprovechando los remanentes de parches de bosque dentro de un paisaje transformado, y que por ende es un área que requiere de medidas urgentes de conservación o restauración.

La poca identificación de las áreas de baja calidad incongruente es otro efecto producido por el sesgo muestral de los registros, debido principalmente a que los esfuerzos de muestreo tienden a concentrarse en áreas de baja intervención, reduciendo nuestra capacidad de detectar especies clave en estas áreas. Sin embargo, es una limitante que puede superarse al incrementar la representatividad de los muestreos biológicos en áreas de SCI bajo y verificar la presencia o ausencia de especies clave. Este tipo de adiciones de registros es una prueba de la continua adaptabilidad y actualización que ofrece esta metodología, permitiendo no solo ajustar la clasificación, sino aumentar la capacidad de predecir incongruencias perjudiciales para la conservación de las especies. Otra posible explicación sobre esta clase de incongruencia, puede ser un error en la delimitación e identidad de los polígonos de coberturas, pues la mayoría de estas incongruencias se encuentran al borde de parches de bosque con alta calidad congruente. Esto puede resolverse con la elaboración de cartografía a escalas muy finas (mayor que 1:5000), para garantizar que los polígonos representan fielmente la cobertura actual del área de estudio, así como ir actualizando esta información con la periodicidad necesaria para el cumplimiento de los objetivos de gestión planteados.

A diferencia de las áreas de baja calidad incongruente, sí se identificaron muchas áreas de alta calidad incongruente, donde el SCI reporta bosques densos, altos y poco disturbados, pero donde no se evidencian especies dependientes de bosques o hay una predominancia de especies invasoras, introducidas o de baja dependencia al bosque. Esta última incongruencia puede explicarse desde tres puntos principalmente: 1) analizado a nivel local, el SCI en estas zonas realiza una sobrestimación de la condición estructural de los bosques húmedos, lo que puede

comprobarse al compararlos con la calidad de hábitat de bosque medida con registros; 2) el muestreo de los registros biológicos es insuficiente para estas zonas y la medida de calidad de bosque no refleja el verdadero valor del polígono; o 3) tanto el SCI como la calidad de bosque medida por registros son correctos y cuentan con un nivel adecuado de muestreo, por lo que sería evidencia del fenómeno de bosques vacíos, mencionado en la introducción del documento. A pesar de que los tres puntos de vista son plausibles al momento de explicar esta clase de incongruencia, como se mencionó anteriormente, los esfuerzos de muestreo tienden a concentrarse en los parches de bosque mayormente conservados, además de que en el área de Guane-Kalypso se encuentra un elevado número de registros pertenecientes a todos los grupos estudiados en este proyecto. Por estas razones, la presencia de estas áreas de alta calidad incongruente, parece apoyarse en el fenómeno de bosques vacíos, donde valdría la pena recopilar mayor información de campo para estudiar este fenómeno la opción de que puede existir una sobrestimación del SCI en la zona.

Aún con todas las consideraciones mencionadas previamente, el desempeño de los algoritmos para el Modelo 3 fue muy similar entre los diferentes algoritmos, mostrando una exactitud temática de poco más de 0.7, siendo este el valor más alto entre los ejercicios de modelación. Sin embargo, también es cierto que el modelo solo estimó tres de las cuatro clases de la variable respuesta, lo que pudo contribuir también a que aumentara su exactitud temática (Figura 5-6).

Una de las consideraciones que deben tenerse en cuenta al momento de realizar ejercicios de clasificación es la reducción de la dimensionalidad. De acuerdo con (Abramovich & Pensky, 2015), incluso para una clasificación con dos clases, la clasificación con múltiples variables sin una reducción de la dimensionalidad adecuada, afectará negativamente el desempeño de los modelos, debido a la generación de ruido producido especialmente por los atributos que no ofrecen una adecuada separabilidad de las clases, también denominados atributos débiles. En este caso particular, quisimos conocer el comportamiento de todas las bandas, coeficientes de retrodispersión e índices con detalle, ya que se desconocían los patrones de radiancia de las variables respuesta. Esto es evidente especialmente con la clase 1 del SCI, que abarca todo el espectro de valores posibles en cobertura de dosel y altura, distinguiéndose únicamente por su valor del último año de intervención (Figura 5-2). Es por esto, que se decidió evaluar la totalidad de las variables, con el fin de identificar relaciones poco intuitivas y generar una mayor aproximación a los objetivos del proyecto.

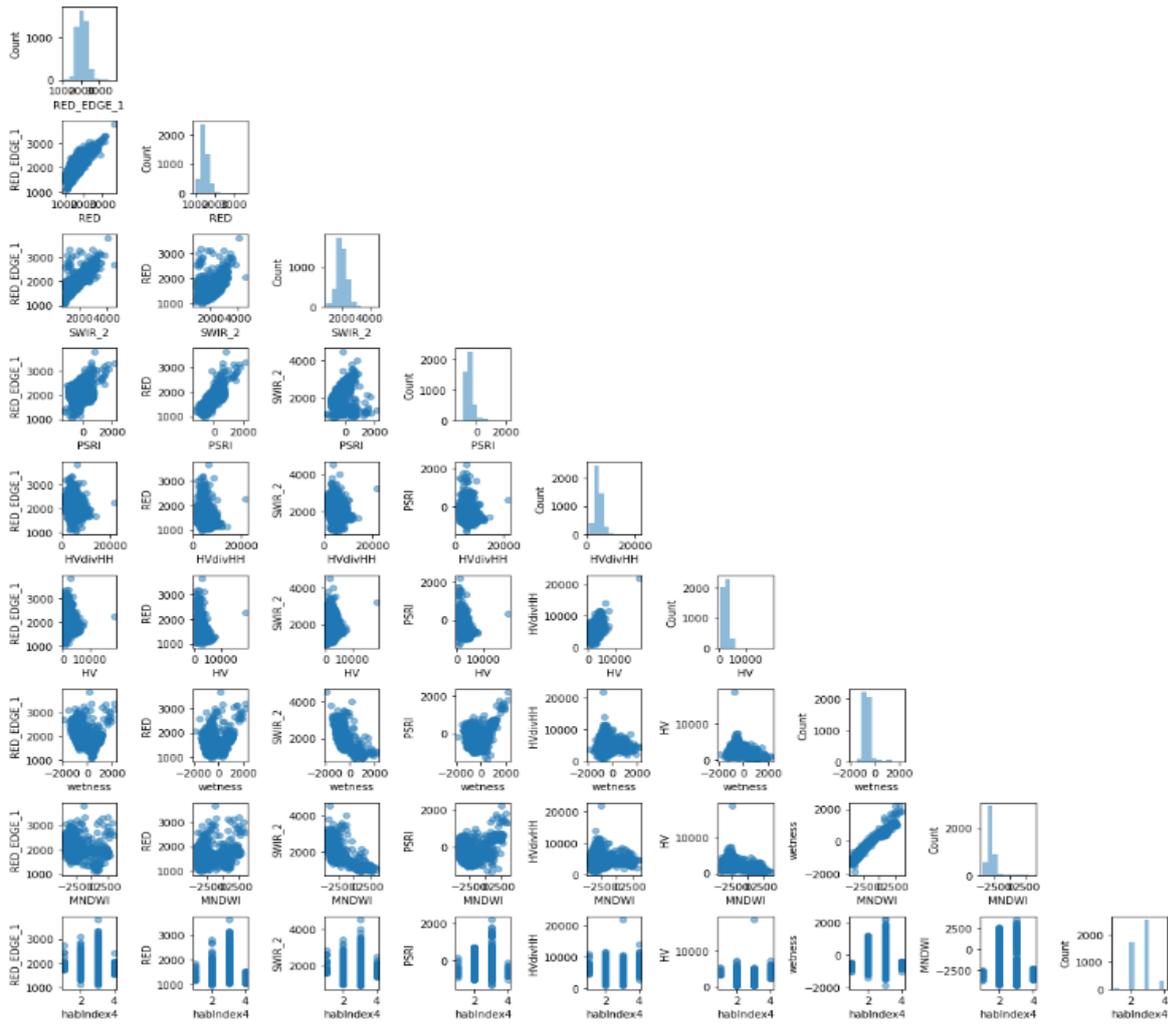


Figura 5-6. Matriz de correlación entre las variables más importantes del modelo 3. Las variables con mayor correlación son las mismas que las presentadas en los Modelos 1 y 2, y al igual que los modelos anteriores, la variable respuesta no tiene correlaciones elevadas con las demás variables.

Para comprobar el efecto del número de variables, se repitió la modelación con mejor exactitud temática del modelo 1, haciendo previamente una reducción de dimensionalidad mediante un ANOVA y seleccionando las cinco variables que generan un mayor impacto en la variable respuesta. A diferencia del análisis de Boruta que se ejecutó antes de todos los modelos, este análisis no compara el desempeño de las variables comparadas con la aleatoriedad, sino que compara su variabilidad con respecto a las etiquetas, identificando las diferencias significativas, y por ende el impacto que generan sobre la variable respuesta. El resultado de esta reducción de dimensionalidad fue muy similar a lo mostrado en la Figura 4.1, seleccionando como más relevantes las bandas SWIR_2, RED_EDGE_1, GREEN y RED como las más relevantes (junto con SVVI, siendo esta la única que aparece con una importancia relativamente menor en la figura 4.1). No obstante, al realizar la modelación con los

mismos hiperparámetros del Modelo 1, la exactitud temática arrojó un valor inferior a la de este modelo (Accuracy: 0.606), al igual que el modelo con nuevos hiperparámetros calculados únicamente con las variables seleccionadas por el ANOVA (Accuracy: 0.600). En esta ocasión, a pesar de que con la reducción de la dimensionalidad se busca eliminar las variables menos relevantes para aumentar el desempeño del modelo, en términos estadísticos fue mejor incluir la totalidad de las variables. Este comportamiento puede deberse a que el ejercicio con menos variables utilizó el algoritmo RF, el cual puede lidiar con múltiples variables (Chen et al., 2020), por lo que el desempeño inicialmente observado no disminuyó con la reducción de las variables.

Otra fuente de error que pudo afectar nuestros análisis fue el sesgo muestral producido por los registros biológicos en el modelo 2, así como por otras inexactitudes como en la georreferenciación de los registros, su resolución taxonómica y los vacíos de información en zonas del país (K. S. He et al., 2015). La generalización de la dependencia de bosque con polígonos de cobertura permite ampliar el área muestral espectral para los modelos y puede contribuir (en casos particulares) el impacto de imprecisiones espaciales cuando los registros están cerca del centroide de los polígonos. Gracias a esto, se logró muestrear un área considerable de toda el área de estudio, a pesar de que el sesgo producido por la distribución de los registros, cerca del 90% de los registros se encontraban ubicados en el área de Guane-Kalypso, gracias al ejercicio del levantamiento de Línea Base que se hizo en la zona. Este es un escenario muy común al momento de priorizar áreas a escala local con fines de restauración/conservación, y es la falta de datos asociados a la biodiversidad de la zona. A pesar de que esta estrategia brindó una mayor área de muestreo espectral, tiene la limitante de asumir una homogeneidad para cada polígono que puede que espectralmente no exista, aumentando así la variabilidad dentro de cada categoría. No obstante, esta generalización por polígonos se llevó a cabo desde un enfoque biológico, al asumir que si una especie está en un área y su rango de movilidad es mayor o cercana a esta área, aunque se produce un incremento en la variabilidad espectral, ofrece una precisión desde el punto de vista ecológico.

Generalmente, un paso previo para la priorización de áreas es realizar un análisis con insumos regionales, nacionales o incluso globales de indicadores para la identificación de estas áreas (por ejemplo SCI, el índice de huella espacial humana, la probabilidad de deforestación, el índice de integridad de bosque a escala de paisaje, índices de conectividad del paisaje, y la integración de información espacial asociada al uso de la tierra como áreas de minería, áreas protegidas, coberturas de la tierra, entre otras) (Correa Ayram et al., 2020; Grantham et al., 2020; A. Hansen et al., 2019; Rosa et al., 2013), pero no siempre estos insumos cuentan con la resolución temática o espacial suficiente para resolver esta problemática. A pesar de estas consideraciones, el objetivo 2 sobre generar una metodología que integre datos de SR con datos de biodiversidad, se cumplió con esta identificación de la

dependencia de las especies al ecosistema de bosque y su ajuste con coberturas de la tierra.

Una de las estrategias utilizadas para la determinación de hábitats es mediante los modelos de distribución de especies (SDM). Los SDM pueden integrarse con SR para identificar de manera precisa los requerimientos de hábitat de una especie en particular (Zielewska-Büttner et al., 2018). No obstante, el resultado deseado con este índice de calidad de hábitat, no es la medición de un hábitat específico para una sola especie, sino generar una herramienta para la toma de decisiones que beneficien a la mayoría de las especies y la mayoría de los ecosistemas, evitando mayor degradación ambiental. La gran ventaja de este índice es que toma la información de múltiples especies en un tiempo razonable para la toma de decisiones y con una incertidumbre por lo menos reconocida. De esta manera se detectan áreas con incongruencia asociada a su integridad en múltiples niveles, favoreciendo una toma de decisiones informada e identificando áreas para enfocar estudios futuros que beneficiarán a múltiples especies al mismo tiempo.

Si dentro de los objetivos en las estrategias que quieran implementarse en un área de interés, se priorizan una o múltiples especies, con información suficiente, es posible ajustar la metodología propuesta en este índice de calidad de hábitat e incluir la modelación de los SDM de las especies priorizadas. Sin embargo, para que la estimación de la distribución potencial de cada especie sea correcta, se requiere de un número mínimo de registros y de una estrategia de muestreo para disminuir el error, además de depender de las estrategias de parametrización, la selección del modelo y la contribución de los predictores para identificar los patrones de ocurrencia entre los registros de la especie y una serie de variables bioclimáticas (Araújo & Guisan, 2006). Cada una de estas consideraciones debe tenerse en cuenta por especie, lo que significa que la incertidumbre final para una comunidad se verá afectado por especies raras, con información insuficiente o sesgo espacial, o por especies relacionadas con procesos ecológicos particulares como las especies invasoras (Caradima et al., 2019; Srivastava et al., 2019). Adicionalmente, las variables bioclimáticas utilizadas en la modelación, suelen ser de una resolución muy gruesa para los fines que busca este proyecto de servir como herramienta de apoyo para los tomadores de decisión al momento de implementar estrategias de conservación/restauración (Booth, 2018; Waltari et al., 2014). Por estas razones, la agrupación de los registros en grupos taxonómicos ayuda a generalizar la respuesta espectral medida por estos, permitiendo su evaluación con datos de SR en zonas más extensas dentro del área de estudio. Esta decisión genera un trade-off en el que se pierde resolución taxonómica, impidiendo la identificación de especies específicas en el área de estudio, pero a cambio se aumenta el número de datos útiles para evaluar el hábitat.

Tomando todo esto en cuenta, la estrategia utilizada para la determinación de la calidad de hábitat de bosque en este proyecto, funciona como una medición práctica,

fácilmente replicable, de fácil interpretación y que tiende a disminuir los errores acumulados que pueden generarse siguiendo otras metodologías. Además, el uso de registros biológicos permite aprovechar una base de información cada vez más creciente y que puede encontrarse casi en cualquier parte del país. Por supuesto, esta es una metodología que puede mejorarse dependiendo de la información disponible de la zona, de los objetivos de gestión que deseen alcanzarse y de los recursos disponibles. Frente a la metodología, un aspecto importante a mejorar es el ajuste de las variables predictoras para capturar fenómenos asociados a la sucesión, restauración o recuperación de la cobertura boscosa. Incorporar estos componentes podría ayudar a entender la dinámica temporal de los bosques, siendo además un proxy para la evaluación del componente de tiempo de intervención propuesto por Hansen.

Por ejemplo, si en los objetivos de gestión no se requiere el conocimiento específico de un grupo de especies en particular, esta metodología se convierte en una opción adecuada para la medición de calidad de bosque, reconociendo principalmente los registros y las preferencias de hábitats. Por otro lado, si se tiene algún interés en la identidad taxonómica de las especies, esta metodología ya no permite reconocer de manera particular la respuesta de una sola especie. Una opción para resolver este requerimiento de gestión, es trabajar únicamente con los registros de las especies de interés, corriendo el riesgo de que la información no sea suficiente para generar puntos de entrenamiento adecuados para la generación del modelo e incurrir en errores de subestimación o modificar el supuesto utilizado en este proyecto de asignar el mismo peso a todos los registros, y en su lugar realizar una ponderación de los grupos taxonómicos para encontrar un óptimo entre resolución funcional (dado por la identidad taxonómica) y resolución espectral capturada desde los registros en sí.

En el desarrollo de este proyecto procuramos utilizar herramientas de uso libre para los usuarios, con el fin de que cualquier persona pudiera replicar los resultados. La herramienta Google Earth Engine ofreció una interfaz muy útil para la detección y manipulación de las imágenes multispectrales y de radar utilizadas, así como de los productos que contenían algunos de los índices utilizados. Una de las mayores ventajas de GEE es que nos evita la descarga de las imágenes, cuyo desempeño dependerá de las características del equipo de cómputo y la conexión de internet. De igual manera, los servidores de GEE permiten el análisis de múltiples imágenes que fueron esenciales para la construcción del mosaico final para la adquisición de datos en nuestra área de estudio. La principal limitante de GEE es la limitada gama de herramientas para el ajuste (Tuning) de los modelos, así como el reducido número de algoritmos de aprendizaje que se pueden implementar en este entorno. Por esta razón, complementamos las herramientas de GEE con paquetes estadísticos de R que suplieron esta necesidad del fortalecimiento de los modelos de aprendizaje.

A pesar de que esta estrategia permitió evaluar la calidad de hábitat de bosque a partir de los registros y su información de hábitat, el desempeño del modelo no es lo

suficientemente alto para determinar una relación directa entre los SR y la presencia de las especies. Este resultado también era de esperarse, puesto que el uso del hábitat por parte de las especies no puede simplificarse únicamente a los valores de SR satelitales, especialmente cuando se integran múltiples especies. Más allá de las características del paisaje que se pueden identificar con SR, la persistencia de las especies en un hábitat como los bosques húmedos tropicales, también responden a otros procesos como la fragmentación, la pérdida de hábitat per se, la configuración del paisaje, interacciones entre especies, además de presiones de ámbito social y cultural (Roques & Stoica, 2007; Villard & Metzger, 2014). No obstante, la respuesta del modelo 3 permite agregar las incertidumbres de los primeros dos modelos, al definir categorías ajustadas que incorporen la congruencia e incongruencia entre los dos niveles de información (SCI y registros), aunque no elimina del todo la incertidumbre de los atributos que definen la calidad de hábitat, pero que no pudieron ser capturados por SR. En este sentido, el modelo 3 es una mejoría sobre los modelos 1 y 2 tomados por separado, aunque aún pueden realizarse ajustes a su metodología para que pueda considerarse como una solución ideal. La utilidad principal de este resultado, es la de determinar a la escala de las predicciones, dónde hay que enfocar la investigación biológica y ecológica futura, y en qué áreas hay un valor más certero de los bosques. Dado esto, se puede afirmar que se cumplieron con los objetivos propuestos en este proyecto de maestría.

6. Conclusiones y recomendaciones

La evaluación de los resultados de las clasificaciones generadas al integrar datos multiespectrales, radar e índices espectrales por medio de algoritmos de aprendizaje de máquina, demostraron tener un elevado potencial para identificar las relaciones entre estas variables de SR y las variables asociadas a la estructura y composición de los bosques húmedos tropicales. Las exactitudes temáticas obtenidas, a pesar de ser un ejercicio de clasificación con variables que no tienen una relación directa con la definición de sus clases, son suficientes para identificar la información de las bandas e índices que más están relacionadas con la variable respuesta. El desempeño del modelo 3, el cual integraba los elementos más relevantes de los modelos 1 y 2, resultó ser el de mayor exactitud temática.

Este estudio demuestra que parte de la esencia del índice de condición estructural de bosques de Hansen, puede traducirse en la respuesta espectral de una combinación de sensores remotos principalmente ópticos, lo cual ofrece una alternativa para evaluar la estructura de los bosques a partir de esta clase de SR, los cuales están cada vez más disponibles para los usuarios, y algunos con una historia de captura de imágenes muy extensa para evaluar escenarios pasados, presentes y futuros. Además de esto, es posible complementar el componente estructural del SCI con datos de registros biológicos para incluir una aproximación del componente de composición y acercarnos a un índice de calidad de hábitat que incluya los tres componentes de la integridad ecológica como son estructura, composición y función.

En este proyecto no se incluyó el componente funcional de la integridad ecológica para ser incluido en la construcción del índice de calidad de hábitat final o "SCI Ajustado". Una alternativa para evaluar la funcionalidad del ecosistema es mediante el índice de huella espacial humana, puesto que, entre mayor sea el valor de la huella espacial, mayor es la presión ejercida sobre los ecosistemas naturales, lo que puede afectar la funcionalidad del ecosistema. No se realizaron pruebas con huella espacial humana en este proyecto, puesto que el objetivo principal es el desarrollo de una metodología para la identificación de áreas prioritarias para la gestión y elaboración de estrategias de manejo sostenible de los bosques tropicales, a partir de información de imágenes satelitales multiespectrales y de radar. Aún no se ha evaluado a fondo la relación entre sensores remotos y huella espacial humana desarrollada para Colombia, pero resultaría interesante incorporarlo en los análisis. Otras aproximaciones es incorporar productos de sensores remotos como proxy a funciones ecosistémicas como captura de carbono, ciclaje de nutrientes, transferencia de energía, entre otros elementos funcionales.

Es importante tener en cuenta que este método obtuvo resultados aceptables gracias a la amplia disponibilidad de registros que habían en la zona, pero pudo haber presentado valores más elevados si se reducen las áreas de vacíos en donde no se encontraron registros biológicos de ninguna especie. Afortunadamente, la adición de nuevos registros, así como el desarrollo de nuevas tecnologías aeroespaciales seguirán presentándose al ser herramientas cada vez más relevantes en la resolución de problemas socio-ecológicos. Aún con las limitaciones anteriormente nombradas, se logró identificar ciertos parches en el área de estudio en donde, con los recursos suficientes, se podrá identificar la razón por la que se detecta la incongruencia entre el SCI y la calidad de hábitat.

Con este proyecto queremos hacer una invitación, para que previamente a la réplica de este método, se tengan en cuenta dos aspectos fundamentales: 1) reunir la mayor cantidad de información disponible, ya sea de registros como se propone en esta metodología o la información de campo que se considere relevante, y 2) definir el objetivo que se busca con la aplicación del método, para así darse cuenta previamente si esta metodología es la más apropiada, o es mejor optar por otras estrategias como los SDM.

En síntesis, este proyecto sugiere que, como medida exploratoria para identificar parches de bosque en donde se desee realizar estrategias de conservación/restauración, que integre tanto la estructura como la composición de los bosques, esta metodología resulta en una estrategia práctica, actualizable y replicable, que con la elección de hiperparámetros y entrenamiento de datos adecuados, responderá de manera adecuada al objetivo planteado.

Bibliografía

- Abdelkareem, M., Bamousa, A. O., Hamimi, Z., & Kamal El-Din, G. M. (2020). Multispectral and RADAR images integration for geologic, geomorphic, and structural investigation in southwestern Arabian Shield, Al Qunfudhah area, Saudi Arabia. *Journal of Taibah University for Science*, 14(1), 383-401.
<https://doi.org/10.1080/16583655.2020.1741957>
- Abdulkareem, N. M., & Abdulazeez, A. M. (2021). *Machine Learning Classification Based on Radom Forest Algorithm: A Review*. <https://doi.org/10.5281/ZENODO.4471118>
- Abramovich, F., & Pensky, M. (2015). *Classification with many classes: Challenges and pluses*. <https://doi.org/10.48550/ARXIV.1506.01567>
- Acharya, T., Subedi, A., & Lee, D. (2018). Evaluation of Water Indices for Surface Water Extraction in a Landsat 8 Scene of Nepal. *Sensors*, 18(8), 2580.
<https://doi.org/10.3390/s18082580>
- Achury, R., & Suarez, A. V. (2018). Richness and Composition of Ground-dwelling Ants in Tropical Rainforest and Surrounding Landscapes in the Colombian Inter-Andean Valley. *Neotropical Entomology*, 47(6), 731-741. <https://doi.org/10.1007/s13744-017-0565-4>
- AghaKouchak, A., Farahmand, A., Melton, F. S., Teixeira, J., Anderson, M. C., Wardlow, B. D., & Hain, C. R. (2015). Remote sensing of drought: Progress, challenges and opportunities: REMOTE SENSING OF DROUGHT. *Reviews of Geophysics*, 53(2), 452-480. <https://doi.org/10.1002/2014RG000456>
- Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B., & Anderson, R. P. (2015). spThin: An R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*, 38(5), 541-545.
<https://doi.org/10.1111/ecog.01132>

- Alam, A., Bhat, M. S., & Maheen, M. (2020). Using Landsat satellite data for assessing the land use and land cover change in Kashmir valley. *GeoJournal*, *85*(6), 1529-1543. <https://doi.org/10.1007/s10708-019-10037-x>
- Alaniz, A. J., Carvajal, M. A., Fierro, A., Vergara-Rodríguez, V., Toledo, G., Ansaldo, D., Moreira-Arce, D., Rojas-Osorio, A., & Vergara, P. M. (2021). Remote-sensing estimates of forest structure and dynamics as indicators of habitat quality for Magellanic woodpeckers. *Ecological Indicators*, *126*, 107634. <https://doi.org/10.1016/j.ecolind.2021.107634>
- Allam, M., Bakr, N., & Elbably, W. (2019). Multi-temporal assessment of land use/land cover change in arid region based on landsat satellite imagery: Case study in Fayoum Region, Egypt. *Remote Sensing Applications: Society and Environment*, *14*, 8-19. <https://doi.org/10.1016/j.rsase.2019.02.002>
- Alton, P. B. (2018). Decadal trends in photosynthetic capacity and leaf area index inferred from satellite remote sensing for global vegetation types. *Agricultural and Forest Meteorology*, *250-251*, 361-375. <https://doi.org/10.1016/j.agrformet.2017.11.020>
- Andrew, M. E., Wulder, M. A., & Nelson, T. A. (2014). Potential contributions of remote sensing to ecosystem service assessments. *Progress in Physical Geography: Earth and Environment*, *38*(3), 328-353. <https://doi.org/10.1177/0309133314528942>
- Aniah, P., Bawakyillenuo, S., Codjoe, S. N. A., & Dzanku, F. M. (2023). Land use and land cover change detection and prediction based on CA-Markov chain in the savannah ecological zone of Ghana. *Environmental Challenges*, *10*, 100664. <https://doi.org/10.1016/j.envc.2022.100664>
- Ansaldo, D., Vergara, P. M., Carvajal, M. A., Alaniz, A. J., Fierro, A., ReinaldoVargas-Castillo, Quiroz, M., Moreira-Arce, D., & Pizarro, J. (2021). Tree decay modulates the functional response of lichen communities in Patagonian temperate forests. *Science of The Total Environment*, *771*, 145360. <https://doi.org/10.1016/j.scitotenv.2021.145360>

- Araújo, M. B., & Guisan, A. (2006). Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, 33(10), 1677-1688. <https://doi.org/10.1111/j.1365-2699.2006.01584.x>
- Arshad, M., Eid, E. M., & Hasan, M. (2020). Mangrove health along the hyper-arid southern Red Sea coast of Saudi Arabia. *Environmental Monitoring and Assessment*, 192(3), 189. <https://doi.org/10.1007/s10661-020-8140-6>
- Astola, H., Häme, T., Sirro, L., Molinier, M., & Kilpi, J. (2019). Comparison of Sentinel-2 and Landsat 8 imagery for forest variable prediction in boreal region. *Remote Sensing of Environment*, 223, 257-273. <https://doi.org/10.1016/j.rse.2019.01.019>
- Baba, K. A., Lal, D., & Bello, A. (2019). *Application of Remote sensing and GIS Techniques in Urban Planning, Development and Management. (A case study of Allahabad District, India)*. 10(6).
- Bechara, F. C., Dickens, S. J., Farrer, E. C., Larios, L., Spotswood, E. N., Mariotte, P., & Suding, K. N. (2016). Neotropical rainforest restoration: Comparing passive, plantation and nucleation approaches. *Biodiversity and Conservation*, 25(11), 2021-2034. <https://doi.org/10.1007/s10531-016-1186-7>
- Beck, J., Böller, M., Erhardt, A., & Schwanghart, W. (2014). Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecological Informatics*, 19, 10-15. <https://doi.org/10.1016/j.ecoinf.2013.11.002>
- Bodart, C., Brink, A. B., Donnay, F., Lupi, A., Mayaux, P., & Achard, F. (2013). Continental estimates of forest cover and forest cover changes in the dry ecosystems of Africa between 1990 and 2000. *Journal of Biogeography*, 40(6), 1036-1047. <https://doi.org/10.1111/jbi.12084>
- Bolyn, C., Michez, A., Gaucher, P., Lejeune, P., & Bonnet, S. (2018). Forest mapping and species composition using supervised per pixel classification of Sentinel-2 imagery. *BASE*, 172-187. <https://doi.org/10.25518/1780-4507.16524>

- Booth, T. H. (2018). Why understanding the pioneering and continuing contributions of BIOCLIM to species distribution modelling is important. *Austral Ecology*, *43*(8), 852-860. <https://doi.org/10.1111/aec.12628>
- Borja, A., Bricker, S. B., Dauer, D. M., Demetriades, N. T., Ferreira, J. G., Forbes, A. T., Hutchings, P., Jia, X., Kenchington, R., Marques, J. C., & Zhu, C. (2008). Overview of integrative tools and methods in assessing ecological integrity in estuarine and coastal systems worldwide. *Marine Pollution Bulletin*, *56*(9), 1519-1537. <https://doi.org/10.1016/j.marpolbul.2008.07.005>
- Bowler, D. E., Callaghan, C. T., Bhandari, N., Henle, K., Benjamin Barth, M., Koppitz, C., Klenke, R., Winter, M., Jansen, F., Bruelheide, H., & Bonn, A. (2022). Temporal trends in the spatial bias of species occurrence records. *Ecography*, *2022*(8). <https://doi.org/10.1111/ecog.06219>
- Brennan, A., Beytell, P., Aschenborn, O., Du Preez, P., Funston, P. J., Hanssen, L., Kilian, J. W., Stuart-Hill, G., Taylor, R. D., & Naidoo, R. (2020). Characterizing multispecies connectivity across a transfrontier conservation landscape. *Journal of Applied Ecology*, *57*(9), 1700-1710. <https://doi.org/10.1111/1365-2664.13716>
- Callaghan, C. T., Major, R. E., Lyons, M. B., Martin, J. M., & Kingsford, R. T. (2018). The effects of local and landscape habitat attributes on bird diversity in urban greenspaces. *Ecosphere*, *9*(7), e02347. <https://doi.org/10.1002/ecs2.2347>
- Caradima, B., Schuwirth, N., & Reichert, P. (2019). From individual to joint species distribution models: A comparison of model complexity and predictive performance. *Journal of Biogeography*, *46*(10), 2260-2274. <https://doi.org/10.1111/jbi.13668>
- Carella, E., Orusa, T., Viani, A., Meloni, D., Borgogno-Mondino, E., & Orusa, R. (2022). An Integrated, Tentative Remote-Sensing Approach Based on NDVI Entropy to Model Canine Distemper Virus in Wildlife and to Prompt Science-Based Management Policies. *Animals*, *12*(8), 1049. <https://doi.org/10.3390/ani12081049>
- Casal, R., Costa, J., & Oviedo, M. (2021). *Aprendizaje Estadístico*. https://rubenfcasal.github.io/aprendizaje_estadistico/index.html

- Chamberlain, S. A., & Boettiger, C. (2017). *R Python, and Ruby clients for GBIF species occurrence data* [Preprint]. PeerJ Preprints.
<https://doi.org/10.7287/peerj.preprints.3304v1>
- Chen, R.-C., Dewi, C., Huang, S.-W., & Caraka, R. E. (2020). Selecting critical features for data classification based on machine learning methods. *Journal of Big Data*, 7(1), 52.
<https://doi.org/10.1186/s40537-020-00327-4>
- Cheng, J., Song, C., Liu, K., Fan, C., Ke, L., Chen, T., Zhan, P., & Yao, J. (2022). Satellite and UAV-based remote sensing for assessing the flooding risk from Tibetan lake expansion and optimizing the village relocation site. *Science of The Total Environment*, 802, 149928. <https://doi.org/10.1016/j.scitotenv.2021.149928>
- Chuvieco, E. (2020). *Fundamentals of satellite remote sensing: An environmental approach* (Third edition). CRC Press.
- Cismondi, F., Fialho, A. S., Vieira, S. M., Reti, S. R., Sousa, J. M. C., & Finkelstein, S. N. (2013). Missing data in medical databases: Impute, delete or classify? *Artificial Intelligence in Medicine*, 58(1), 63-72. <https://doi.org/10.1016/j.artmed.2013.01.003>
- Correa Ayram, C. A., Etter, A., Díaz-Timoté, J., Rodríguez Buriticá, S., Ramírez, W., & Corzo, G. (2020). Spatiotemporal evaluation of the human footprint in Colombia: Four decades of anthropic impact in highly biodiverse ecosystems. *Ecological Indicators*, 117, 106630. <https://doi.org/10.1016/j.ecolind.2020.106630>
- Curd, A., Chevalier, M., Vasquez, M., Boyé, A., Firth, L. B., Marzloff, M. P., Bricheno, L. M., Burrows, M. T., Bush, L. E., Cordier, C., Davies, A. J., Green, J. A. M., Hawkins, S. J., Lima, F. P., Meneghesso, C., Mieszkowska, N., Seabra, R., & Dubois, S. F. (2023). Applying landscape metrics to species distribution model predictions to characterize internal range structure and associated changes. *Global Change Biology*, 29(3), 631-647. <https://doi.org/10.1111/gcb.16496>
- Dantas De Paula, M., Groeneveld, J., & Huth, A. (2016). The extent of edge effects in fragmented landscapes: Insights from satellite measurements of tree cover. *Ecological Indicators*, 69, 196-204. <https://doi.org/10.1016/j.ecolind.2016.04.018>

- De Araujo Barbosa, C. C., Atkinson, P. M., & Dearing, J. A. (2015). Remote sensing of ecosystem services: A systematic review. *Ecological Indicators*, *52*, 430-443. <https://doi.org/10.1016/j.ecolind.2015.01.007>
- Delegido, J., Verrelst, J., Alonso, L., & Moreno, J. (2011). Evaluation of Sentinel-2 Red-Edge Bands for Empirical Estimation of Green LAI and Chlorophyll Content. *Sensors*, *11*(7), 7063-7081. <https://doi.org/10.3390/s110707063>
- Dhingra, S., & Kumar, D. (2019). A review of remotely sensed satellite image classification. *International Journal of Electrical and Computer Engineering (IJECE)*, *9*(3), 1720. <https://doi.org/10.11591/ijece.v9i3.pp1720-1731>
- Di Febraro, M., Sallustio, L., Vizzarri, M., De Rosa, D., De Lisio, L., Loy, A., Eichelberger, B. A., & Marchetti, M. (2018). Expert-based and correlative models to map habitat quality: Which gives better support to conservation planning? *Global Ecology and Conservation*, *16*, e00513. <https://doi.org/10.1016/j.gecco.2018.e00513>
- DiMiceli, C., Townshend, J., Carroll, M., & Sohlberg, R. (2021). Evolution of the representation of global vegetation by vegetation continuous fields. *Remote Sensing of Environment*, *254*, 112271. <https://doi.org/10.1016/j.rse.2020.112271>
- Dronova, I., Taddeo, S., Hemes, K. S., Knox, S. H., Valach, A., Oikawa, P. Y., Kasak, K., & Baldocchi, D. D. (2021). Remotely sensed phenological heterogeneity of restored wetlands: Linking vegetation structure and function. *Agricultural and Forest Meteorology*, *296*, 108215. <https://doi.org/10.1016/j.agrformet.2020.108215>
- Dubertret, F., Le Tourneau, F.-M., Villarreal, M. L., & Norman, L. M. (2022). Monitoring Annual Land Use/Land Cover Change in the Tucson Metropolitan Area with Google Earth Engine (1986–2020). *Remote Sensing*, *14*(9), 2127. <https://doi.org/10.3390/rs14092127>
- Dubovik, O., Schuster, G. L., Xu, F., Hu, Y., Bösch, H., Landgraf, J., & Li, Z. (2021). Grand Challenges in Satellite Remote Sensing. *Frontiers in Remote Sensing*, *2*, 619818. <https://doi.org/10.3389/frsen.2021.619818>

- Elbeih, S. F. (2021). Evaluation of agricultural expansion areas in the Egyptian deserts: A review using remote sensing and GIS. *The Egyptian Journal of Remote Sensing and Space Science*, 24(3), 889-906. <https://doi.org/10.1016/j.ejrs.2021.10.004>
- Fachinetti, R., & Grilli, M. P. (2023). The quality matters: Assessing the roles of patch quality and configuration on the abundance of an invading cerambycid species through a multi-model inference approach. *Ecological Entomology*, 48(5), 557-567. <https://doi.org/10.1111/een.13247>
- Fagua, J. C., Jantz, P., Rodriguez-Buritica, S., Duncanson, L., & Goetz, S. J. (2019). Integrating LiDAR, Multispectral and SAR Data to Estimate and Map Canopy Height in Tropical Forests. *Remote Sensing*, 11(22), 2697. <https://doi.org/10.3390/rs11222697>
- Fagua, J. C., Rodríguez-Buriticá, S., & Jantz, P. (2023). Advancing High-Resolution Land Cover Mapping in Colombia: The Importance of a Locally Appropriate Legend. *Remote Sensing*, 15(10), 2522. <https://doi.org/10.3390/rs15102522>
- Fan, X., Song, Y., Zhu, C., Balzter, H., & Bai, Z. (2021). Estimating Ecological Responses to Climatic Variability on Reclaimed and Unmined Lands Using Enhanced Vegetation Index. *Remote Sensing*, 13(6), 1100. <https://doi.org/10.3390/rs13061100>
- Fatehi, P., Damm, A., Schaepman, M., & Kneubühler, M. (2015). Estimation of Alpine Forest Structural Variables from Imaging Spectrometer Data. *Remote Sensing*, 7(12), 16315-16338. <https://doi.org/10.3390/rs71215830>
- Feranec, J. (Ed.). (2016). *European landscape dynamics: Corine land cover data*. CRC Press.
- Fraga, H., Amraoui, M., Malheiro, A. C., Moutinho-Pereira, J., Eiras-Dias, J., Silvestre, J., & Santos, J. A. (2014). Examining the relationship between the Enhanced Vegetation Index and grapevine phenology. *European Journal of Remote Sensing*, 47(1), 753-771. <https://doi.org/10.5721/EuJRS20144743>
- Gao, W., Zheng, C., Liu, X., Lu, Y., Chen, Y., Wei, Y., & Ma, Y. (2022). NDVI-based vegetation dynamics and their responses to climate change and human activities

- from 1982 to 2020: A case study in the Mu Us Sandy Land, China. *Ecological Indicators*, 137, 108745. <https://doi.org/10.1016/j.ecolind.2022.108745>
- Gibson, R., Danaher, T., Hehir, W., & Collins, L. (2020). A remote sensing approach to mapping fire severity in south-eastern Australia using sentinel 2 and random forest. *Remote Sensing of Environment*, 240, 111702. <https://doi.org/10.1016/j.rse.2020.111702>
- Gómez-Lora, J. W., Gallo-Ramos, V. H., & Camacho-Zorogastúa, K. D. C. (2021). Evaluación del bosque húmedo tropical mediante el análisis de la cobertura fraccional y técnicas SIG en la subcuenca del río Yuracyacu, Amazonía peruana. *Madera y Bosques*, 27(2). <https://doi.org/10.21829/myb.2021.2722109>
- Goswami, S., Gamon, J., Vargas, S., & Tweedie, C. (2015). *Relationships of NDVI, Biomass, and Leaf Area Index (LAI) for six key plant species in Barrow, Alaska* [Preprint]. PeerJ PrePrints. <https://doi.org/10.7287/peerj.preprints.913v1>
- Grabska, E., Hostert, P., Pflugmacher, D., & Ostapowicz, K. (2019). Forest Stand Species Mapping Using the Sentinel-2 Time Series. *Remote Sensing*, 11(10), 1197. <https://doi.org/10.3390/rs11101197>
- Grantham, H. S., Duncan, A., Evans, T. D., Jones, K. R., Beyer, H. L., Schuster, R., Walston, J., Ray, J. C., Robinson, J. G., Callow, M., Clements, T., Costa, H. M., DeGemmis, A., Elsen, P. R., Ervin, J., Franco, P., Goldman, E., Goetz, S., Hansen, A., ... Watson, J. E. M. (2020). Anthropogenic modification of forests means only 40% of remaining forests have high ecosystem integrity. *Nature Communications*, 11(1), 5978. <https://doi.org/10.1038/s41467-020-19493-3>
- Greenwell, B. M., Boehmke, B. C., & McCarthy, A. J. (2018). *A Simple and Effective Model-Based Variable Importance Measure*. <https://doi.org/10.48550/ARXIV.1805.04755>
- Halder, B., Bandyopadhyay, J., & Banik, P. (2021). Monitoring the effect of urban development on urban heat island based on remote sensing and geo-spatial approach in Kolkata and adjacent areas, India. *Sustainable Cities and Society*, 74, 103186. <https://doi.org/10.1016/j.scs.2021.103186>

- Hank, T. B., Berger, K., Bach, H., Clevers, J. G. P. W., Gitelson, A., Zarco-Tejada, P., & Mauser, W. (2019). Spaceborne Imaging Spectroscopy for Sustainable Agriculture: Contributions and Challenges. *Surveys in Geophysics*, *40*(3), 515-551.
<https://doi.org/10.1007/s10712-018-9492-0>
- Hansen, A., Barnett, K., Jantz, P., Phillips, L., Goetz, S. J., Hansen, M., Venter, O., Watson, J. E. M., Burns, P., Atkinson, S., Rodríguez-Buritica, S., Ervin, J., Virnig, A., Supples, C., & De Camargo, R. (2019). Global humid tropics forest structural condition and forest structural integrity maps. *Scientific Data*, *6*(1), 232.
<https://doi.org/10.1038/s41597-019-0214-3>
- Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., & Townshend, J. R. G. (2013). High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science*, *342*(6160), 850-853.
<https://doi.org/10.1126/science.1244693>
- Hasanah, A., Supriatna, & Indrawan, M. (2020). Assessment of tropical forest degradation on a small island using the enhanced vegetation index. *IOP Conference Series: Earth and Environmental Science*, *481*(1), 012061. <https://doi.org/10.1088/1755-1315/481/1/012061>
- He, C., Gao, B., Huang, Q., Ma, Q., & Dou, Y. (2017). Environmental degradation in the urban areas of China: Evidence from multi-source remote sensing data. *Remote Sensing of Environment*, *193*, 65-75. <https://doi.org/10.1016/j.rse.2017.02.027>
- He, K. S., Bradley, B. A., Cord, A. F., Rocchini, D., Tuanmu, M., Schmidtlein, S., Turner, W., Wegmann, M., & Pettorelli, N. (2015). Will remote sensing shape the next generation of species distribution models? *Remote Sensing in Ecology and Conservation*, *1*(1), 4-18. <https://doi.org/10.1002/rse2.7>
- Hong Han, Xiaoling Guo, & Hua Yu. (2016). Variable selection using Mean Decrease Accuracy and Mean Decrease Gini based on Random Forest. *2016 7th IEEE*

- International Conference on Software Engineering and Service Science (ICSESS)*, 219-224. <https://doi.org/10.1109/ICSESS.2016.7883053>
- Huang, S., Tang, L., Hupy, J. P., Wang, Y., & Shao, G. (2021). A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. *Journal of Forestry Research*, 32(1), 1-6. <https://doi.org/10.1007/s11676-020-01155-1>
- 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*, 83(1-2), 195-213. [https://doi.org/10.1016/S0034-4257\(02\)00096-2](https://doi.org/10.1016/S0034-4257(02)00096-2)
- Huo, L., Persson, H. J., & Lindberg, E. (2021). Early detection of forest stress from European spruce bark beetle attack, and a new vegetation index: Normalized distance red & SWIR (NDRS). *Remote Sensing of Environment*, 255, 112240. <https://doi.org/10.1016/j.rse.2020.112240>
- Hussain, S., & Karuppanan, S. (2023). Land use/land cover changes and their impact on land surface temperature using remote sensing technique in district Khanewal, Punjab Pakistan. *Geology, Ecology, and Landscapes*, 7(1), 46-58. <https://doi.org/10.1080/24749508.2021.1923272>
- Ige, S. O., Ajayi, V. O., Adeyeri, O. E., & Oyekan, K. S. A. (2017). Assessing remotely sensed temperature humidity index as human comfort indicator relative to landuse landcover change in Abuja, Nigeria. *Spatial Information Research*, 25(4), 523-533. <https://doi.org/10.1007/s41324-017-0118-2>
- Ishwaran, H., & Lu, M. (2019). Standard errors and confidence intervals for variable importance in random forest regression, classification, and survival. *Statistics in Medicine*, 38(4), 558-582. <https://doi.org/10.1002/sim.7803>
- Jarchow, C., Didan, K., Barreto-Muñoz, A., Nagler, P., & Glenn, E. (2018). Application and Comparison of the MODIS-Derived Enhanced Vegetation Index to VIIRS, Landsat 5

- TM and Landsat 8 OLI Platforms: A Case Study in the Arid Colorado River Delta, Mexico. *Sensors*, 18(5), 1546. <https://doi.org/10.3390/s18051546>
- Jaybhay, J., & Shastri, R. (2015). A Study of Speckle Noise Reduction Filters. *Signal & Image Processing : An International Journal*, 6(3), 71-80.
<https://doi.org/10.5121/sipij.2015.6306>
- Jindo, K., Kozan, O., Iseki, K., Maestrini, B., Van Evert, F. K., Wubengeda, Y., Arai, E., Shimabukuro, Y. E., Sawada, Y., & Kempenaar, C. (2021). Potential utilization of satellite remote sensing for field-based agricultural studies. *Chemical and Biological Technologies in Agriculture*, 8(1), 58. <https://doi.org/10.1186/s40538-021-00253-4>
- Jog, S., & Dixit, M. (2016). Supervised classification of satellite images. *2016 Conference on Advances in Signal Processing (CASP)*, 93-98.
<https://doi.org/10.1109/CASP.2016.7746144>
- Joshi, N., Baumann, M., Ehammer, A., Fensholt, R., Grogan, K., Hostert, P., Jepsen, M., Kuemmerle, T., Meyfroidt, P., Mitchard, E., Reiche, J., Ryan, C., & Waske, B. (2016). A Review of the Application of Optical and Radar Remote Sensing Data Fusion to Land Use Mapping and Monitoring. *Remote Sensing*, 8(1), 70.
<https://doi.org/10.3390/rs8010070>
- Jung, M., Dahal, P. R., Butchart, S. H. M., Donald, P. F., De Lamo, X., Lesiv, M., Kapos, V., Rondinini, C., & Visconti, P. (2020). A global map of terrestrial habitat types. *Scientific Data*, 7(1), 256. <https://doi.org/10.1038/s41597-020-00599-8>
- Kalacska, M. (2004). Leaf area index measurements in a tropical moist forest: A case study from Costa Rica. *Remote Sensing of Environment*, 91(2), 134-152.
<https://doi.org/10.1016/j.rse.2004.02.011>
- Kanniah, K. D., Kang, C. S., Sharma, S., & Amir, A. A. (2021). Remote Sensing to Study Mangrove Fragmentation and Its Impacts on Leaf Area Index and Gross Primary Productivity in the South of Peninsular Malaysia. *Remote Sensing*, 13(8), 1427.
<https://doi.org/10.3390/rs13081427>

- Karr, J. R., Larson, E. R., & Chu, E. W. (2022). Ecological integrity is both real and valuable. *Conservation Science and Practice*, 4(2). <https://doi.org/10.1111/csp2.583>
- Kganyago, M., Mhangara, P., Alexandridis, T., Laneve, G., Ovakoglou, G., & Mashiyi, N. (2020). Validation of sentinel-2 leaf area index (LAI) product derived from SNAP toolbox and its comparison with global LAI products in an African semi-arid agricultural landscape. *Remote Sensing Letters*, 11(10), 883-892. <https://doi.org/10.1080/2150704X.2020.1767823>
- Klimes, P., Idigel, C., Rimandai, M., Fayle, T. M., Janda, M., Weiblen, G. D., & Novotny, V. (2012). Why are there more arboreal ant species in primary than in secondary tropical forests?: Why are there more arboreal ants in primary forests? *Journal of Animal Ecology*, 81(5), 1103-1112. <https://doi.org/10.1111/j.1365-2656.2012.02002.x>
- Kohzuma, K., Sonoike, K., & Hikosaka, K. (2021). Imaging, screening and remote sensing of photosynthetic activity and stress responses. *Journal of Plant Research*, 134(4), 649-651. <https://doi.org/10.1007/s10265-021-01324-1>
- Kupfer, J. A. (2006). National assessments of forest fragmentation in the US. *Global Environmental Change*, 16(1), 73-82. <https://doi.org/10.1016/j.gloenvcha.2005.10.003>
- Kursa, M. B., & Rudnicki, W. R. (2010). Feature Selection with the **Boruta** Package. *Journal of Statistical Software*, 36(11). <https://doi.org/10.18637/jss.v036.i11>
- Lastovicka, J., Svec, P., Paluba, D., Kobliuk, N., Svoboda, J., Hladky, R., & Stych, P. (2020). Sentinel-2 Data in an Evaluation of the Impact of the Disturbances on Forest Vegetation. *Remote Sensing*, 12(12), 1914. <https://doi.org/10.3390/rs12121914>
- Lawrence, R. (2004). Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. *Remote Sensing of Environment*, 90(3), 331-336. <https://doi.org/10.1016/j.rse.2004.01.007>
- Lazic, S. E., Mellor, J. R., Ashby, M. C., & Munafa, M. R. (2020). A Bayesian predictive approach for dealing with pseudoreplication. *Scientific Reports*, 10(1), 2366. <https://doi.org/10.1038/s41598-020-59384-7>

- Lechner, A. M., Foody, G. M., & Boyd, D. S. (2020). Applications in Remote Sensing to Forest Ecology and Management. *One Earth*, 2(5), 405-412.
<https://doi.org/10.1016/j.oneear.2020.05.001>
- Leitão, P. J., & Santos, M. J. (2019). Improving Models of Species Ecological Niches: A Remote Sensing Overview. *Frontiers in Ecology and Evolution*, 7, 9.
<https://doi.org/10.3389/fevo.2019.00009>
- Lembrechts, J. J., Nijs, I., & Lenoir, J. (2019). Incorporating microclimate into species distribution models. *Ecography*, 42(7), 1267-1279.
<https://doi.org/10.1111/ecog.03947>
- Leta, M. K., Demissie, T. A., & Tränckner, J. (2021). Modeling and Prediction of Land Use Land Cover Change Dynamics Based on Land Change Modeler (LCM) in Nashe Watershed, Upper Blue Nile Basin, Ethiopia. *Sustainability*, 13(7), 3740.
<https://doi.org/10.3390/su13073740>
- Li, P., Wang, J., Liu, M., Xue, Z., Bagherzadeh, A., & Liu, M. (2021). Spatio-temporal variation characteristics of NDVI and its response to climate on the Loess Plateau from 1985 to 2015. *CATENA*, 203, 105331.
<https://doi.org/10.1016/j.catena.2021.105331>
- Li, Q., Lu, X., Wang, Y., Huang, X., Cox, P. M., & Luo, Y. (2018). Leaf area index identified as a major source of variability in modeled CO₂ fertilization. *Biogeosciences*, 15(22), 6909-6925. <https://doi.org/10.5194/bg-15-6909-2018>
- Li, S., Xu, L., Jing, Y., Yin, H., Li, X., & Guan, X. (2021). High-quality vegetation index product generation: A review of NDVI time series reconstruction techniques. *International Journal of Applied Earth Observation and Geoinformation*, 105, 102640.
<https://doi.org/10.1016/j.jag.2021.102640>
- Liao, J., Li, Z., Hiebeler, D. E., Iwasa, Y., Bogaert, J., & Nijs, I. (2013). Species persistence in landscapes with spatial variation in habitat quality: A pair approximation model. *Journal of Theoretical Biology*, 335, 22-30. <https://doi.org/10.1016/j.jtbi.2013.06.015>

- Liu, T., & Yang, X. (2015). Monitoring land changes in an urban area using satellite imagery, GIS and landscape metrics. *Applied Geography*, *56*, 42-54.
<https://doi.org/10.1016/j.apgeog.2014.10.002>
- Louw, A. S., Fu, J., Raut, A., Zulhilmi, A., Yao, S., McAlinn, M., Fujikawa, A., Siddique, M. T., Wang, X., Yu, X., Mandvikar, K., & Avtar, R. (2022). The role of remote sensing during a global disaster: COVID-19 pandemic as case study. *Remote Sensing Applications: Society and Environment*, *27*, 100789.
<https://doi.org/10.1016/j.rsase.2022.100789>
- Lu, Y., Zhao, J., Qi, J., Rong, T., Wang, Z., Yang, Z., & Han, F. (2022). Monitoring the Spatiotemporal Dynamics of Habitat Quality and Its Driving Factors Based on the Coupled NDVI-InVEST Model: A Case Study from the Tianshan Mountains in Xinjiang, China. *Land*, *11*(10), 1805. <https://doi.org/10.3390/land11101805>
- Luque, A., Carrasco, A., Martín, A., & De Las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, *91*, 216-231.
<https://doi.org/10.1016/j.patcog.2019.02.023>
- Ma, H., & Liang, S. (2022). Development of the GLASS 250-m leaf area index product (version 6) from MODIS data using the bidirectional LSTM deep learning model. *Remote Sensing of Environment*, *273*, 112985.
<https://doi.org/10.1016/j.rse.2022.112985>
- Maalouf, M. (2011). Logistic regression in data analysis: An overview. *International Journal of Data Analysis Techniques and Strategies*, *3*(3), 281.
<https://doi.org/10.1504/IJDATS.2011.041335>
- Maimaitijiang, M., Sagan, V., Sidike, P., Daloye, A. M., Erkbol, H., & Fritschi, F. B. (2020). Crop Monitoring Using Satellite/UAV Data Fusion and Machine Learning. *Remote Sensing*, *12*(9), 1357. <https://doi.org/10.3390/rs12091357>
- Mairota, P., Cafarelli, B., Labadessa, R., Lovergine, F. P., Tarantino, C., Nagendra, H., & Didham, R. K. (2015). Very high resolution Earth Observation features for testing the

direct and indirect effects of landscape structure on local habitat quality. *International Journal of Applied Earth Observation and Geoinformation*, 34, 96-102.

<https://doi.org/10.1016/j.jag.2014.07.003>

Mallegowda, P., Rengaiyan, G., Krishnan, J., & Niphadkar, M. (2015). Assessing Habitat Quality of Forest-Corridors through NDVI Analysis in Dry Tropical Forests of South India: Implications for Conservation. *Remote Sensing*, 7(2), 1619-1639.

<https://doi.org/10.3390/rs70201619>

Mancera, R. (2019). Evaluación de imágenes de radar Sentinel-1^a e imágenes multiespectrales Sentinel-2^a en la clasificación de cobertura del suelo en diferentes niveles de detalle. [Tesis de maestría no publicada]. Universidad Nacional de Colombia.

Martos, V., Ahmad, A., Cartujo, P., & Ordoñez, J. (2021). Ensuring Agricultural Sustainability through Remote Sensing in the Era of Agriculture 5.0. *Applied Sciences*, 11(13),

5911. <https://doi.org/10.3390/app11135911>

Masrur Ahmed, A. A., Deo, R. C., Feng, Q., Ghahramani, A., Raj, N., Yin, Z., & Yang, L. (2021). Deep learning hybrid model with Boruta-Random forest optimiser algorithm for streamflow forecasting with climate mode indices, rainfall, and periodicity. *Journal of Hydrology*, 599, 126350. <https://doi.org/10.1016/j.jhydrol.2021.126350>

McNairn, H., & Shang, J. (2016). A Review of Multitemporal Synthetic Aperture Radar (SAR) for Crop Monitoring. En Y. Ban (Ed.), *Multitemporal Remote Sensing* (Vol. 20, pp. 317-340). Springer International Publishing. https://doi.org/10.1007/978-3-319-47037-5_15

Melo-Merino, S. M., Reyes-Bonilla, H., & Lira-Noriega, A. (2020). Ecological niche models and species distribution models in marine environments: A literature review and spatial analysis of evidence. *Ecological Modelling*, 415, 108837.

<https://doi.org/10.1016/j.ecolmodel.2019.108837>

Messier, A. (2023). *Patterns of greenness (NDVI) in the Southern Great Plains and their influence on the habitat quality and reproduction of a declining prairie grouse*

[Thesis]. <https://krex.k-state.edu/handle/2097/42970>

- Minghelli, A., Vadakke-Chanat, S., Chami, M., Guillaume, M., & Peirache, M. (2021). Benefit of the Potential Future Hyperspectral Satellite Sensor (BIODIVERSITY) for Improving the Determination of Water Column and Seabed Features in Coastal Zones. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 1222-1232. <https://doi.org/10.1109/JSTARS.2020.3031729>
- Misra, G., Cawkwell, F., & Wingler, A. (2020). Status of Phenological Research Using Sentinel-2 Data: A Review. *Remote Sensing*, *12*(17), 2760. <https://doi.org/10.3390/rs12172760>
- Mondanaro, A., Di Febbraro, M., Castiglione, S., Melchionna, M., Serio, C., Girardi, G., Belfiore, A. M., & Raia, P. (2023). *ENPHYLO*: A new method to model the distribution of extremely rare species. *Methods in Ecology and Evolution*, *14*(3), 911-922. <https://doi.org/10.1111/2041-210X.14066>
- Mora, F. (2017). A structural equation modeling approach for formalizing and evaluating ecological integrity in terrestrial ecosystems. *Ecological Informatics*, *41*, 74-90. <https://doi.org/10.1016/j.ecoinf.2017.05.002>
- Mora, F. (2019). The use of ecological integrity indicators within the natural capital index framework: The ecological and economic value of the remnant natural capital of México. *Journal for Nature Conservation*, *47*, 77-92. <https://doi.org/10.1016/j.jnc.2018.11.007>
- Morcillo-Pallarés, P., Rivera-Caicedo, J. P., Belda, S., De Grave, C., Burriel, H., Moreno, J., & Verrelst, J. (2019). Quantifying the Robustness of Vegetation Indices through Global Sensitivity Analysis of Homogeneous and Forest Leaf-Canopy Radiative Transfer Models. *Remote Sensing*, *11*(20), 2418. <https://doi.org/10.3390/rs11202418>
- Moreira, A., Prats-Iraola, P., Younis, M., Krieger, G., Hajnsek, I., & Papathanassiou, K. P. (2013). A tutorial on synthetic aperture radar. *IEEE Geoscience and Remote Sensing Magazine*, *1*(1), 6-43. <https://doi.org/10.1109/MGRS.2013.2248301>
- Moudrý, V., Moudrá, L., Barták, V., Bejček, V., Gdulová, K., Hendrychová, M., Moravec, D., Musil, P., Rocchini, D., Šťastný, K., Volf, O., & Šálek, M. (2021). The role of the

vegetation structure, primary productivity and senescence derived from airborne LiDAR and hyperspectral data for birds diversity and rarity on a restored site.

Landscape and Urban Planning, 210, 104064.

<https://doi.org/10.1016/j.landurbplan.2021.104064>

Moumane, A., Al Karkouri, J., Benmansour, A., El Ghazali, F. E., Fico, J., Karmaoui, A., & Batchi, M. (2022). Monitoring long-term land use, land cover change, and desertification in the Ternata oasis, Middle Draa Valley, Morocco. *Remote Sensing Applications: Society and Environment*, 26, 100745.

<https://doi.org/10.1016/j.rsase.2022.100745>

Mu, S., Yang, G., Xu, X., Wan, R., & Li, B. (2022). Assessing the inundation dynamics and its impacts on habitat suitability in Poyang Lake based on integrating Landsat and MODIS observations. *Science of The Total Environment*, 834, 154936.

<https://doi.org/10.1016/j.scitotenv.2022.154936>

Mutanga, O., & Skidmore, A. K. (2007). Red edge shift and biochemical content in grass canopies. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62(1), 34-42.

<https://doi.org/10.1016/j.isprsjprs.2007.02.001>

Nagendra, H., Lucas, R., Honrado, J. P., Jongman, R. H. G., Tarantino, C., Adamo, M., & Mairota, P. (2013). Remote sensing for conservation monitoring: Assessing protected areas, habitat extent, habitat condition, species diversity, and threats. *Ecological Indicators*, 33, 45-59. <https://doi.org/10.1016/j.ecolind.2012.09.014>

Nasir, S. M., Kamran, K. V., Blaschke, T., & Karimzadeh, S. (2022). Change of land use / land cover in kurdistan region of Iraq: A semi-automated object-based approach. *Remote Sensing Applications: Society and Environment*, 26, 100713.

<https://doi.org/10.1016/j.rsase.2022.100713>

Nemani, R., Pierce, L., Running, S., & Band, L. (1993). Forest ecosystem processes at the watershed scale: Sensitivity to remotely-sensed Leaf Area Index estimates.

International Journal of Remote Sensing, 14(13), 2519-2534.

<https://doi.org/10.1080/01431169308904290>

- Ogilvie, A., Belaud, G., Massuel, S., Mulligan, M., Le Goulven, P., & Calvez, R. (2018). Surface water monitoring in small water bodies: Potential and limits of multi-sensor Landsat time series. *Hydrology and Earth System Sciences*, *22*(8), 4349-4380. <https://doi.org/10.5194/hess-22-4349-2018>
- Pal, S., Talukdar, S., & Ghosh, R. (2020). Damming effect on habitat quality of riparian corridor. *Ecological Indicators*, *114*, 106300. <https://doi.org/10.1016/j.ecolind.2020.106300>
- Pan, Z., He, J., Liu, D., Wang, J., & Guo, X. (2021). Ecosystem health assessment based on ecological integrity and ecosystem services demand in the Middle Reaches of the Yangtze River Economic Belt, China. *Science of The Total Environment*, *774*, 144837. <https://doi.org/10.1016/j.scitotenv.2020.144837>
- Pandey, P. C., Koutsias, N., Petropoulos, G. P., Srivastava, P. K., & Ben Dor, E. (2021). Land use/land cover in view of earth observation: Data sources, input dimensions, and classifiers—a review of the state of the art. *Geocarto International*, *36*(9), 957-988. <https://doi.org/10.1080/10106049.2019.1629647>
- Pandey, R., Goswami, S., Sarup, J., & Matin, S. (2021). The thermal–optical trapezoid model-based soil moisture estimation using Landsat-8 data. *Modeling Earth Systems and Environment*, *7*(2), 1029-1037. <https://doi.org/10.1007/s40808-020-00975-8>
- Parker, G. G. (2020). Tamm review: Leaf Area Index (LAI) is both a determinant and a consequence of important processes in vegetation canopies. *Forest Ecology and Management*, *477*, 118496. <https://doi.org/10.1016/j.foreco.2020.118496>
- Peng, D., Zhang, H., Yu, L., Wu, M., Wang, F., Huang, W., Liu, L., Sun, R., Li, C., Wang, D., & Xu, F. (2018). Assessing spectral indices to estimate the fraction of photosynthetically active radiation absorbed by the vegetation canopy. *International Journal of Remote Sensing*, *39*(22), 8022-8040. <https://doi.org/10.1080/01431161.2018.1479795>

- Perilla, G. A., & Mas, J.-F. (2020). Google Earth Engine (GEE): Una poderosa herramienta que vincula el potencial de los datos masivos y la eficacia del procesamiento en la nube. *Investigaciones Geográficas*, 101. <https://doi.org/10.14350/rig.59929>
- Perumal, K., & Bhaskaran, R. (2010). *Supervised Classification Performance of Multispectral Images* (arXiv:1002.4046). arXiv. <http://arxiv.org/abs/1002.4046>
- Pettorelli, N., Schulte To Bühne, H., Tulloch, A., Dubois, G., Macinnis-Ng, C., Queirós, A. M., Keith, D. A., Wegmann, M., Schrod, F., Stellmes, M., Sonnenschein, R., Geller, G. N., Roy, S., Somers, B., Murray, N., Bland, L., Geijzendorffer, I., Kerr, J. T., Broszeit, S., ... Nicholson, E. (2018). Satellite remote sensing of ecosystem functions: Opportunities, challenges and way forward. *Remote Sensing in Ecology and Conservation*, 4(2), 71-93. <https://doi.org/10.1002/rse2.59>
- Pires, M. M., & Galetti, M. (2023). Beyond the “empty forest”: The defaunation syndromes of Neotropical forests in the Anthropocene. *Global Ecology and Conservation*, 41, e02362. <https://doi.org/10.1016/j.gecco.2022.e02362>
- Qiu, B., Chen, J. M., Ju, W., Zhang, Q., & Zhang, Y. (2019). Simulating emission and scattering of solar-induced chlorophyll fluorescence at far-red band in global vegetation with different canopy structures. *Remote Sensing of Environment*, 233, 111373. <https://doi.org/10.1016/j.rse.2019.111373>
- Rahimi, L., Malekmohammadi, B., & Yavari, A. R. (2020). Assessing and Modeling the Impacts of Wetland Land Cover Changes on Water Provision and Habitat Quality Ecosystem Services. *Natural Resources Research*, 29(6), 3701-3718. <https://doi.org/10.1007/s11053-020-09667-7>
- Randin, C. F., Ashcroft, M. B., Bolliger, J., Cavender-Bares, J., Coops, N. C., Dullinger, S., Dirnböck, T., Eckert, S., Ellis, E., Fernández, N., Giuliani, G., Guisan, A., Jetz, W., Joost, S., Karger, D., Lembrechts, J., Lenoir, J., Luoto, M., Morin, X., ... Payne, D. (2020). Monitoring biodiversity in the Anthropocene using remote sensing in species distribution models. *Remote Sensing of Environment*, 239, 111626. <https://doi.org/10.1016/j.rse.2019.111626>

- Raschka, S., Liu, Y. H. & Mirjalili, V. (2023). *Machine Learning con PyTorch y Scikit-Learn: Cómo desarrollar modelos de Machine Learning y Deep Learning con Python*. Alpha Editorial.
- Ray, S. (2019). A Quick Review of Machine Learning Algorithms. *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, 35-39. <https://doi.org/10.1109/COMITCon.2019.8862451>
- Reddy, C. S. (2021). Remote sensing of biodiversity: What to measure and monitor from space to species? *Biodiversity and Conservation*, 30(10), 2617-2631. <https://doi.org/10.1007/s10531-021-02216-5>
- Reich, P. B. (2012). Key canopy traits drive forest productivity. *Proceedings of the Royal Society B: Biological Sciences*, 279(1736), 2128-2134. <https://doi.org/10.1098/rspb.2011.2270>
- Requena-Mullor, J. M., López, E., Castro, A. J., Alcaraz-Segura, D., Castro, H., Reyes, A., & Cabello, J. (2017). Remote-sensing based approach to forecast habitat quality under climate change scenarios. *PLOS ONE*, 12(3), e0172107. <https://doi.org/10.1371/journal.pone.0172107>
- Requena-Mullor, J. M., López, E., Castro, A. J., Cabello, J., Virgós, E., González-Miras, E., & Castro, H. (2014). Modeling spatial distribution of European badger in arid landscapes: An ecosystem functioning approach. *Landscape Ecology*, 29(5), 843-855. <https://doi.org/10.1007/s10980-014-0020-4>
- Roques, L., & Stoica, R. S. (2007). Species persistence decreases with habitat fragmentation: An analysis in periodic stochastic environments. *Journal of Mathematical Biology*, 55(2), 189-205. <https://doi.org/10.1007/s00285-007-0076-8>
- Rosa, I. M. D., Purves, D., Souza, C., & Ewers, R. M. (2013). Predictive Modelling of Contagious Deforestation in the Brazilian Amazon. *PLoS ONE*, 8(10), e77231. <https://doi.org/10.1371/journal.pone.0077231>
- Rosenfield, M. F., Jakovac, C. C., Vieira, D. L. M., Poorter, L., Brancalion, P. H. S., Vieira, I. C. G., De Almeida, D. R. A., Massoca, P., Schiatti, J., Albernaz, A. L. M., Ferreira, M.

- J., & Mesquita, R. C. G. (2023). Ecological integrity of tropical secondary forests: Concepts and indicators. *Biological Reviews*, *98*(2), 662-676.
<https://doi.org/10.1111/brv.12924>
- Roy, P. S., Behera, M. D., & Srivastav, S. K. (2017). Satellite Remote Sensing: Sensors, Applications and Techniques. *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences*, *87*(4), 465-472. <https://doi.org/10.1007/s40010-017-0428-8>
- Ruelland, D., Dezetter, A., Puech, C., & Ardoin-Bardin, S. (2008). Long-term monitoring of land cover changes based on Landsat imagery to improve hydrological modelling in West Africa. *International Journal of Remote Sensing*, *29*(12), 3533-3551.
<https://doi.org/10.1080/01431160701758699>
- Ruggeri, S., Henao-Cespedes, V., Garcés-Gómez, Y. A., & Parra Uzcátegui, A. (2021). Optimized unsupervised CORINE Land Cover mapping using linear spectral mixture analysis and object-based image analysis. *The Egyptian Journal of Remote Sensing and Space Science*, *24*(3), 1061-1069. <https://doi.org/10.1016/j.ejrs.2021.10.009>
- Sajjad, H., & Kumar, P. (2019). Future Challenges and Perspective of Remote Sensing Technology. En P. Kumar, M. Rani, P. Chandra Pandey, H. Sajjad, & B. S. Chaudhary (Eds.), *Applications and Challenges of Geospatial Technology* (pp. 275-277). Springer International Publishing. https://doi.org/10.1007/978-3-319-99882-4_16
- Sánchez-Giraldo, C., Correa Ayram, C., & Daza, J. M. (2021). Environmental sound as a mirror of landscape ecological integrity in monitoring programs. *Perspectives in Ecology and Conservation*, *19*(3), 319-328.
<https://doi.org/10.1016/j.pecon.2021.04.003>
- Sangpradid, S., Uttaruk, Y., Rotjanakusol, T., & Laosuwan, T. (2021). FORECASTING TIME SERIES CHANGE OF THE AVERAGE ENHANCED VEGETATION INDEX TO MONITORING DROUGHT CONDITION BY USING TERRA/MODIS DATA. *The*

Journal «Agriculture and Forestry», 67(4).

<https://doi.org/10.17707/AgricultForest.67.4.11>

Santin-Janin, H., Garel, M., Chapuis, J.-L., & Pontier, D. (2009). Assessing the performance of NDVI as a proxy for plant biomass using non-linear models: A case study on the Kerguelen archipelago. *Polar Biology*, 32(6), 861-871.

<https://doi.org/10.1007/s00300-009-0586-5>

Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, 2(3), 160. <https://doi.org/10.1007/s42979-021-00592-x>

Sasmito, S. D., Taillardat, P., Clendenning, J. N., Cameron, C., Friess, D. A., Murdiyarso, D., & Hutley, L. B. (2019). Effect of land-use and land-cover change on mangrove blue carbon: A systematic review. *Global Change Biology*, 25(12), 4291-4302.

<https://doi.org/10.1111/gcb.14774>

Schipper, J. (s. f.). *Magdalena-Urabá Moist Forests* [Divulgativa].

<https://www.oneearth.org/ecoregions/magdalena-uraba-moist-forests/>

Schulte To Bühne, H., & Pettorelli, N. (2018). Better together: Integrating and fusing multispectral and radar satellite imagery to inform biodiversity monitoring, ecological research and conservation science. *Methods in Ecology and Evolution*, 9(4), 849-865. <https://doi.org/10.1111/2041-210X.12942>

Sexton, J. O., Song, X.-P., Feng, M., Noojipady, P., Anand, A., Huang, C., Kim, D.-H., Collins, K. M., Channan, S., DiMiceli, C., & Townshend, J. R. (2013). Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error. *International Journal of Digital Earth*, 6(5), 427-448. <https://doi.org/10.1080/17538947.2013.786146>

Shahzaman, M., Zhu, W., Bilal, M., Habtemicheal, B. A., Mustafa, F., Arshad, M., Ullah, I., Ishfaq, S., & Iqbal, R. (2021). Remote Sensing Indices for Spatial Monitoring of Agricultural Drought in South Asian Countries. *Remote Sensing*, 13(11), 2059.

<https://doi.org/10.3390/rs13112059>

- Shao, Z., Sumari, N. S., Portnov, A., Ujoh, F., Musakwa, W., & Mandela, P. J. (2021). Urban sprawl and its impact on sustainable urban development: A combination of remote sensing and social media data. *Geo-Spatial Information Science*, *24*(2), 241-255. <https://doi.org/10.1080/10095020.2020.1787800>
- Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., & Homayouni, S. (2020). Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 6308-6325. <https://doi.org/10.1109/JSTARS.2020.3026724>
- Sims, D., Rahman, A., Cordova, V., Elmasri, B., Baldocchi, D., Bolstad, P., Flanagan, L., Goldstein, A., Hollinger, D., & Misson, L. (2008). A new model of gross primary productivity for North American ecosystems based solely on the enhanced vegetation index and land surface temperature from MODIS. *Remote Sensing of Environment*, *112*(4), 1633-1646. <https://doi.org/10.1016/j.rse.2007.08.004>
- Skowno, A. L., Jewitt, D., & Slingsby, J. A. (2021). Rates and patterns of habitat loss across South Africa's vegetation biomes. *South African Journal of Science*, *117*(1/2). <https://doi.org/10.17159/sajs.2021/8182>
- Song, C. (2013). Optical remote sensing of forest leaf area index and biomass. *Progress in Physical Geography: Earth and Environment*, *37*(1), 98-113. <https://doi.org/10.1177/0309133312471367>
- Song, J., Lu, X., Liu, M., & Wu, X. (2011). A new LogitBoost algorithm for multiclass unbalanced data classification. *2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, 974-977. <https://doi.org/10.1109/FSKD.2011.6019654>
- Song, Z., Li, X., Su, X., & Li, C. (2023). Analyzing the recovery mechanisms of patchy degradation and its response to mowing and plateau pika disturbances in alpine meadow. *Ecological Indicators*, *154*, 110565. <https://doi.org/10.1016/j.ecolind.2023.110565>

- Soto, G. E., Pérez-Hernández, C. G., Hahn, I. J., Rodewald, A. D., & Vergara, P. M. (2017). Tree senescence as a direct measure of habitat quality: Linking red-edge Vegetation Indices to space use by Magellanic woodpeckers. *Remote Sensing of Environment*, 193, 1-10. <https://doi.org/10.1016/j.rse.2017.02.018>
- Soubry, I., Doan, T., Chu, T., & Guo, X. (2021). A Systematic Review on the Integration of Remote Sensing and GIS to Forest and Grassland Ecosystem Health Attributes, Indicators, and Measures. *Remote Sensing*, 13(16), 3262. <https://doi.org/10.3390/rs13163262>
- Souza, C. M., Z. Shimbo, J., Rosa, M. R., Parente, L. L., A. Alencar, A., Rudorff, B. F. T., Hasenack, H., Matsumoto, M., G. Ferreira, L., Souza-Filho, P. W. M., De Oliveira, S. W., Rocha, W. F., Fonseca, A. V., Marques, C. B., Diniz, C. G., Costa, D., Monteiro, D., Rosa, E. R., Vélez-Martin, E., ... Azevedo, T. (2020). Reconstructing Three Decades of Land Use and Land Cover Changes in Brazilian Biomes with Landsat Archive and Earth Engine. *Remote Sensing*, 12(17), 2735. <https://doi.org/10.3390/rs12172735>
- Srivastava, V., Lafond, V., & Griess, V. C. (2019). Species distribution models (SDM): Applications, benefits and challenges in invasive species management. *CABI Reviews*, 2019, 1-13. <https://doi.org/10.1079/PAVSNNR201914020>
- Stenberg, P., Möttus, M., & Rautiainen, M. (2008). Modeling the Spectral Signature of Forests: Application of Remote Sensing Models to Coniferous Canopies. En S. Liang (Ed.), *Advances in Land Remote Sensing* (pp. 147-171). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6450-0_6
- Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*, 14(4), 323-348. <https://doi.org/10.1037/a0016973>

- Sun, Y., Liu, H., & Guo, Z. (2021). Capsule network-based approach for estimating grassland coverage using time series data from enhanced vegetation index. *Artificial Intelligence in Geosciences*, 2, 26-34. <https://doi.org/10.1016/j.aiig.2021.08.001>
- Szpakowski, D., & Jensen, J. (2019). A Review of the Applications of Remote Sensing in Fire Ecology. *Remote Sensing*, 11(22), 2638. <https://doi.org/10.3390/rs11222638>
- Taiwo, B. E., Kafy, A.-A., Samuel, A. A., Rahaman, Z. A., Ayowole, O. E., Shahrier, M., Duti, B. M., Rahman, M. T., Peter, O. T., & Abosedo, O. O. (2023). Monitoring and predicting the influences of land use/land cover change on cropland characteristics and drought severity using remote sensing techniques. *Environmental and Sustainability Indicators*, 18, 100248. <https://doi.org/10.1016/j.indic.2023.100248>
- Temitope Yekeen, S., & Balogun, A.-L. (2020). Advances in Remote Sensing Technology, Machine Learning and Deep Learning for Marine Oil Spill Detection, Prediction and Vulnerability Assessment. *Remote Sensing*, 12(20), 3416. <https://doi.org/10.3390/rs12203416>
- Thamaga, K. H., Dube, T., & Shoko, C. (2022). Advances in satellite remote sensing of the wetland ecosystems in Sub-Saharan Africa. *Geocarto International*, 37(20), 5891-5913. <https://doi.org/10.1080/10106049.2021.1926552>
- Tierney, G. L., Faber-Langendoen, D., Mitchell, B. R., Shriver, W. G., & Gibbs, J. P. (2009). Monitoring and evaluating the ecological integrity of forest ecosystems. *Frontiers in Ecology and the Environment*, 7(6), 308-316. <https://doi.org/10.1890/070176>
- UI Din, S., & Mak, H. W. L. (2021). Retrieval of Land-Use/Land Cover Change (LUCC) Maps and Urban Expansion Dynamics of Hyderabad, Pakistan via Landsat Datasets and Support Vector Machine Framework. *Remote Sensing*, 13(16), 3337. <https://doi.org/10.3390/rs13163337>
- Ullo, S. L., & Sinha, G. R. (2021). Advances in IoT and Smart Sensors for Remote Sensing and Agriculture Applications. *Remote Sensing*, 13(13), 2585. <https://doi.org/10.3390/rs13132585>

- Vergara, P. M., Fierro, A., Alaniz, A. J., Carvajal, M. A., Lizama, M., & Llanos, J. L. (2021). Landscape-scale effects of forest degradation on insectivorous birds and invertebrates in austral temperate forests. *Landscape Ecology*, *36*(1), 191-208. <https://doi.org/10.1007/s10980-020-01133-2>
- Vergara, P. M., Soto, G. E., Rodewald, A. D., & Quiroz, M. (2019). Behavioral switching in Magellanic woodpeckers reveals perception of habitat quality at different spatial scales. *Landscape Ecology*, *34*(1), 79-92. <https://doi.org/10.1007/s10980-018-0746-5>
- Vihervaara, P., Auvinen, A.-P., Mononen, L., Törmä, M., Ahlroth, P., Anttila, S., Böttcher, K., Forsius, M., Heino, J., Heliölä, J., Koskelainen, M., Kuussaari, M., Meissner, K., Ojala, O., Tuominen, S., Viitasalo, M., & Virkkala, R. (2017). How Essential Biodiversity Variables and remote sensing can help national biodiversity monitoring. *Global Ecology and Conservation*, *10*, 43-59. <https://doi.org/10.1016/j.gecco.2017.01.007>
- Villard, M.-A., & Metzger, J. P. (2014). REVIEW: Beyond the fragmentation debate: a conceptual model to predict when habitat configuration really matters. *Journal of Applied Ecology*, *51*(2), 309-318. <https://doi.org/10.1111/1365-2664.12190>
- Viña, A., Gitelson, A. A., Nguy-Robertson, A. L., & Peng, Y. (2011). Comparison of different vegetation indices for the remote assessment of green leaf area index of crops. *Remote Sensing of Environment*, *115*(12), 3468-3478. <https://doi.org/10.1016/j.rse.2011.08.010>
- Vishnu, C. L., Sajinkumar, K. S., Oommen, T., Coffman, R. A., Thri vikramji, K. P., Rani, V. R., & Keerthy, S. (2019). Satellite-based assessment of the August 2018 flood in parts of Kerala, India. *Geomatics, Natural Hazards and Risk*, *10*(1), 758-767. <https://doi.org/10.1080/19475705.2018.1543212>
- Vogeler, J. C., & Cohen, W. B. (2016). A review of the role of active remote sensing and data fusion for characterizing forest in wildlife habitat models. *Revista de Teledetección*, *45*, 1. <https://doi.org/10.4995/raet.2016.3981>

- Vollrath, A., Mullissa, A., & Reiche, J. (2020). Angular-Based Radiometric Slope Correction for Sentinel-1 on Google Earth Engine. *Remote Sensing*, 12(11), 1867.
<https://doi.org/10.3390/rs12111867>
- Waltari, E., Schroeder, R., McDonald, K., Anderson, R. P., & Carnaval, A. (2014). Bioclimatic variables derived from remote sensing: Assessment and application for species distribution modelling. *Methods in Ecology and Evolution*, 5(10), 1033-1042.
<https://doi.org/10.1111/2041-210X.12264>
- Wang, H., Tang, L., Qiu, Q., & Chen, H. (2020). Assessing the Impacts of Urban Expansion on Habitat Quality by Combining the Concepts of Land Use, Landscape, and Habitat in Two Urban Agglomerations in China. *Sustainability*, 12(11), 4346.
<https://doi.org/10.3390/su12114346>
- Wang, R., & Gamon, J. A. (2019). Remote sensing of terrestrial plant biodiversity. *Remote Sensing of Environment*, 231, 111218. <https://doi.org/10.1016/j.rse.2019.111218>
- Wang, X., Liu, C., Lv, G., Xu, J., & Cui, G. (2022). Integrating Multi-Source Remote Sensing to Assess Forest Aboveground Biomass in the Khingan Mountains of North-Eastern China Using Machine-Learning Algorithms. *Remote Sensing*, 14(4), 1039.
<https://doi.org/10.3390/rs14041039>
- Wiegand, T., Naves, J., Garbulsky, M. F., & Fernández, N. (2008). ANIMAL HABITAT QUALITY AND ECOSYSTEM FUNCTIONING: EXPLORING SEASONAL PATTERNS USING NDVI. *Ecological Monographs*, 78(1), 87-103.
<https://doi.org/10.1890/06-1870.1>
- Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H., Guisan, A., & NCEAS Predicting Species Distributions Working Group†. (2008). Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, 14(5), 763-773. <https://doi.org/10.1111/j.1472-4642.2008.00482.x>
- Wulder, M. A., Loveland, T. R., Roy, D. P., Crawford, C. J., Masek, J. G., Woodcock, C. E., Allen, R. G., Anderson, M. C., Belward, A. S., Cohen, W. B., Dwyer, J., Erb, A., Gao, F., Griffiths, P., Helder, D., Hermosilla, T., Hipple, J. D., Hostert, P., Hughes, M. J., ...

- Zhu, Z. (2019). Current status of Landsat program, science, and applications. *Remote Sensing of Environment*, 225, 127-147.
<https://doi.org/10.1016/j.rse.2019.02.015>
- Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, 27(14), 3025-3033. <https://doi.org/10.1080/01431160600589179>
- Xu, Y., Yang, Y., Chen, X., & Liu, Y. (2022). Bibliometric Analysis of Global NDVI Research Trends from 1985 to 2021. *Remote Sensing*, 14(16), 3967.
<https://doi.org/10.3390/rs14163967>
- Yin, J., Dong, J., Hamm, N. A. S., Li, Z., Wang, J., Xing, H., & Fu, P. (2021). Integrating remote sensing and geospatial big data for urban land use mapping: A review. *International Journal of Applied Earth Observation and Geoinformation*, 103, 102514.
<https://doi.org/10.1016/j.jag.2021.102514>
- Yohannes, H., Soromessa, T., Argaw, M., & Dewan, A. (2021). Spatio-temporal changes in habitat quality and linkage with landscape characteristics in the Beressa watershed, Blue Nile basin of Ethiopian highlands. *Journal of Environmental Management*, 281, 111885. <https://doi.org/10.1016/j.jenvman.2020.111885>
- Yu, T., Liu, P., Zhang, Q., Ren, Y., & Yao, J. (2021). Detecting Forest Degradation in the Three-North Forest Shelterbelt in China from Multi-Scale Satellite Images. *Remote Sensing*, 13(6), 1131. <https://doi.org/10.3390/rs13061131>
- Zattara, E. E., & Aizen, M. A. (2021). Worldwide occurrence records suggest a global decline in bee species richness. *One Earth*, 4(1), 114-123.
<https://doi.org/10.1016/j.oneear.2020.12.005>
- Zelený, J., Mercado-Bettín, D., & Müller, F. (2021). Towards the evaluation of regional ecosystem integrity using NDVI, brightness temperature and surface heterogeneity. *Science of The Total Environment*, 796, 148994.
<https://doi.org/10.1016/j.scitotenv.2021.148994>

- Zhang, Y., Liu, J., & Shen, W. (2022). A Review of Ensemble Learning Algorithms Used in Remote Sensing Applications. *Applied Sciences*, 12(17), 8654.
<https://doi.org/10.3390/app12178654>
- Zhang, Y., Zhao, L., Zhao, H., & Gao, X. (2021). Urban development trend analysis and spatial simulation based on time series remote sensing data: A case study of Jinan, China. *PLOS ONE*, 16(10), e0257776. <https://doi.org/10.1371/journal.pone.0257776>
- Zheng, G., & Moskal, L. M. (2009). Retrieving Leaf Area Index (LAI) Using Remote Sensing: Theories, Methods and Sensors. *Sensors*, 9(4), 2719-2745.
<https://doi.org/10.3390/s90402719>
- Zheng, Y., Xiao, Z., Li, J., Yang, H., & Song, J. (2022). Evaluation of Global Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) Products at 500 m Spatial Resolution. *Remote Sensing*, 14(14), 3304. <https://doi.org/10.3390/rs14143304>
- Zhong, R., Wang, P., Mao, G., Chen, A., & Liu, J. (2021). Spatiotemporal variation of enhanced vegetation index in the Amazon Basin and its response to climate change. *Physics and Chemistry of the Earth, Parts A/B/C*, 123, 103024.
<https://doi.org/10.1016/j.pce.2021.103024>
- Zielewska-Büttner, K., Heurich, M., Müller, J., & Braunisch, V. (2018). Remotely Sensed Single Tree Data Enable the Determination of Habitat Thresholds for the Three-Toed Woodpecker (*Picoides tridactylus*). *Remote Sensing*, 10(12), 1972.
<https://doi.org/10.3390/rs10121972>
- Zipkin, E. F., Zylstra, E. R., Wright, A. D., Saunders, S. P., Finley, A. O., Dietze, M. C., Itter, M. S., & Tingley, M. W. (2021). Addressing data integration challenges to link ecological processes across scales. *Frontiers in Ecology and the Environment*, 19(1), 30-38. <https://doi.org/10.1002/fee.2290>
- Zizka, A., Antunes Carvalho, F., Calvente, A., Rocio Baez-Lizarazo, M., Cabral, A., Coelho, J. F. R., Colli-Silva, M., Fantinati, M. R., Fernandes, M. F., Ferreira-Araújo, T., Gondim Lambert Moreira, F., Santos, N. M. C., Santos, T. A. B., Dos Santos-Costa, R. C., Serrano, F. C., Alves Da Silva, A. P., De Souza Soares, A., Cavalcante De

- Souza, P. G., Calisto Tomaz, E., ... Antonelli, A. (2020). No one-size-fits-all solution to clean GBIF. *PeerJ*, 8, e9916. <https://doi.org/10.7717/peerj.9916>
- Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Duarte Ritter, C., Edler, D., Farooq, H., Herdean, A., Ariza, M., Scharn, R., Svantesson, S., Wengström, N., Zizka, V., & Antonelli, A. (2019). COORDINATECLEANER: Standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution*, 10(5), 744-751. <https://doi.org/10.1111/2041-210X.13152>
- Zurell, D., Franklin, J., König, C., Bouchet, P. J., Dormann, C. F., Elith, J., Fandos, G., Feng, X., Guillera-Aroita, G., Guisan, A., Lahoz-Monfort, J. J., Leitão, P. J., Park, D. S., Peterson, A. T., Rapacciuolo, G., Schmatz, D. R., Schröder, B., Serra-Diaz, J. M., Thuiller, W., ... Merow, C. (2020). A standard protocol for reporting species distribution models. *Ecography*, 43(9), 1261-1277. <https://doi.org/10.1111/ecog.04960>
- Zwinkels, J. (2015). Light, Electromagnetic Spectrum. En R. Luo (Ed.), *Encyclopedia of Color Science and Technology* (pp. 1-8). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-27851-8_204-1