

# **Reuniendo conocimiento experto con datos de cobertura terrestre para la evaluación de cambios en los ecosistemas en las áreas protegidas de Colombia**

## **Informe de estado de avance. Diciembre 11, 2023**

**Victor H. Gutierrez Velez, PhD ([victorhugo@temple.edu](mailto:victorhugo@temple.edu))**

### **Introducción**

El trabajo se enmarca como un componente principal del proyecto financiado por NASA, titulado “avance de herramientas para un sistema integrado de monitoreo para las áreas protegidas de Colombia. El objetivo del proyecto es facilitar la integración de las observaciones de la Tierra y herramientas de información en un sistema de monitoreo de biodiversidad para la toma de decisiones de conservación en Colombia. El proyecto tiene como investigadora principal a Mary Blair del Museo Americano de Historia Natural y a Victor Gutiérrez de la Universidad de Temple como co-investigador. El proyecto se está llevando a cabo en colaboración con El Sistema de Parques Nacionales Naturales de Colombia (PNN) y el Instituto Alexander von Humboldt.

Las actividades descritas aquí se refieren al componente de clasificación de coberturas, coordinado por Victor Gutierrez. Este componente surgió a través de un proceso de consulta extendido con representantes del Sistema nacional de Parques y otros colaboradores del proyecto. Dicho proceso consistió en reuniones periódicas y la realización de talleres en los que se identificaron necesidades y prioridades de trabajo. Una de las prioridades identificadas consistió en la necesidad de mejorar la frecuencia y consistencia en la producción y acceso de mapas de cobertura de la tierra con el fin de informar apropiadamente la producción de indicadores de conservación y su uso para la toma de decisiones. Estos mapas son claves para la toma de decisiones ya que son insumos para:

- La producción de indicadores usados para evaluar la contribución de las áreas de conservación a la biodiversidad: (por ejemplo, representatividad y conectividad de ecosistemas y especies)
- La identificación de áreas prioritarias para intervenciones y para la declaración de nuevas áreas protegidas.
- Informar el éxito de las intervenciones de conservación para reducir la presión sobre la biodiversidad.

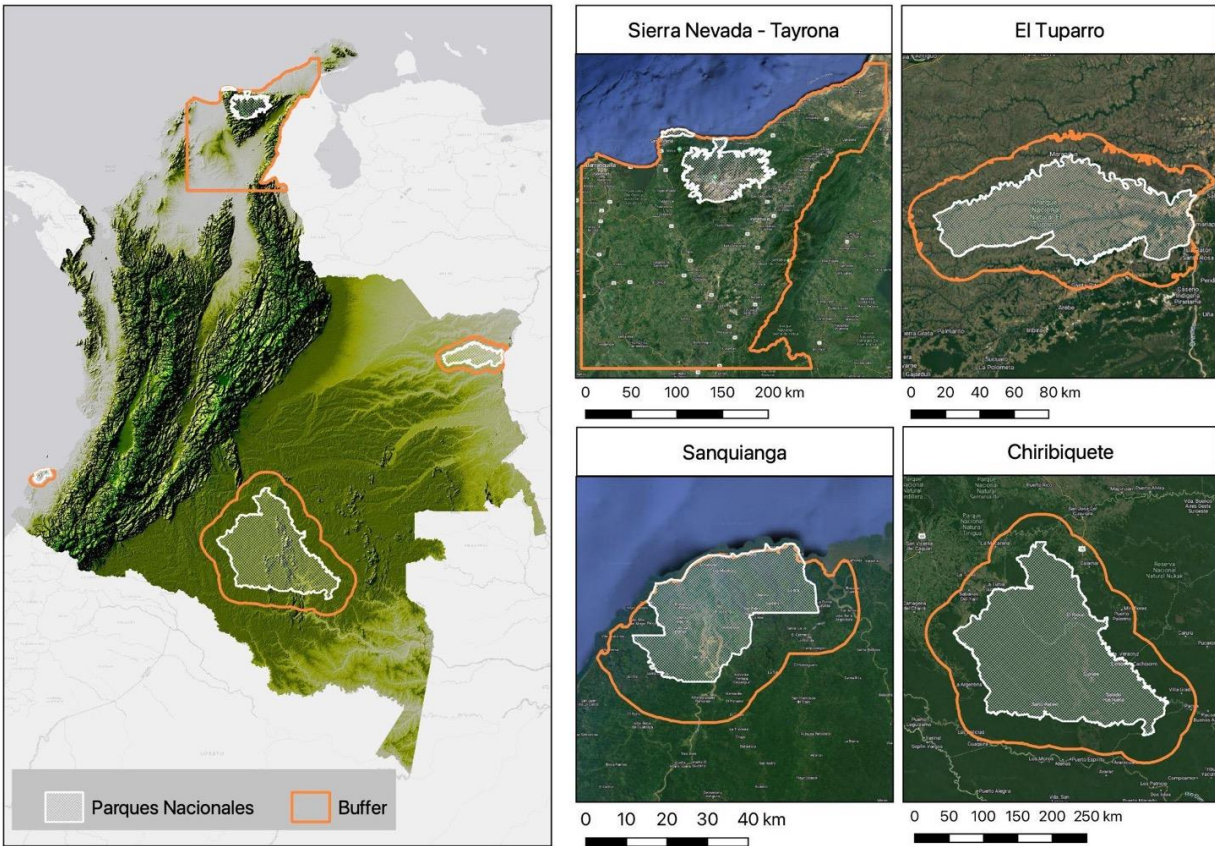
El mapa de referencia usado para la construcción de indicadores, es el desarrollado con base en la adaptación del Sistema de CORINE a Colombia y coordinado por el Instituto de Estudios Ambientales (IDEAM). La metodología CORINE consiste en la participación de un número extenso de expertos de diferentes instituciones quienes delinear polígonos generalizados que

representan coberturas vegetales, y cambios en el tiempo a una escala de 1:100.000, con base en la interpretación de mosaicos de imágenes de Landsat y con la ayuda de imágenes de alta resolución de otros sensores como apoyo a la clasificación. Luego de un trabajo de diagnóstico, se pudo evidenciar la rigurosidad de los protocolos usados para la construcción del mapa de coberturas en la jurisdicción de PNN y la calidad de los mapas. La calidad de estos mapas se ajusta a las necesidades de información para la construcción de indicadores y la toma de decisiones. Sin embargo, este trabajo es dispendioso y por tanto la implementación de algoritmos automatizados pueden mejorar la eficiencia en la elaboración de dichos productos con un potencial de mejorar la caracterización de áreas de diferentes coberturas mediante el cambio de paradigma de clasificación de polígono a pixel.

El objetivo del estudio es evaluar la viabilidad de métodos de clasificación automatizados para mejorar la eficiencia y precisión en el mapeo de cambios de ecosistemas en áreas protegidas de Colombia. Luego de iniciado el trabajo, el IDEAM anunció la consecución de recursos para actualizar el mapa nacional de coberturas para los años 2020, 2022, and 2023. En reuniones de socialización, el IDEAM informó sobre su interés en explorar la viabilidad de incorporar métodos automatizados de clasificación para informar la producción futura de mapas nacionales. Por tanto, la elaboración de este producto tiene un potencial importante para informar metodologías y procedimientos para este fin.

### **Áreas de estudio**

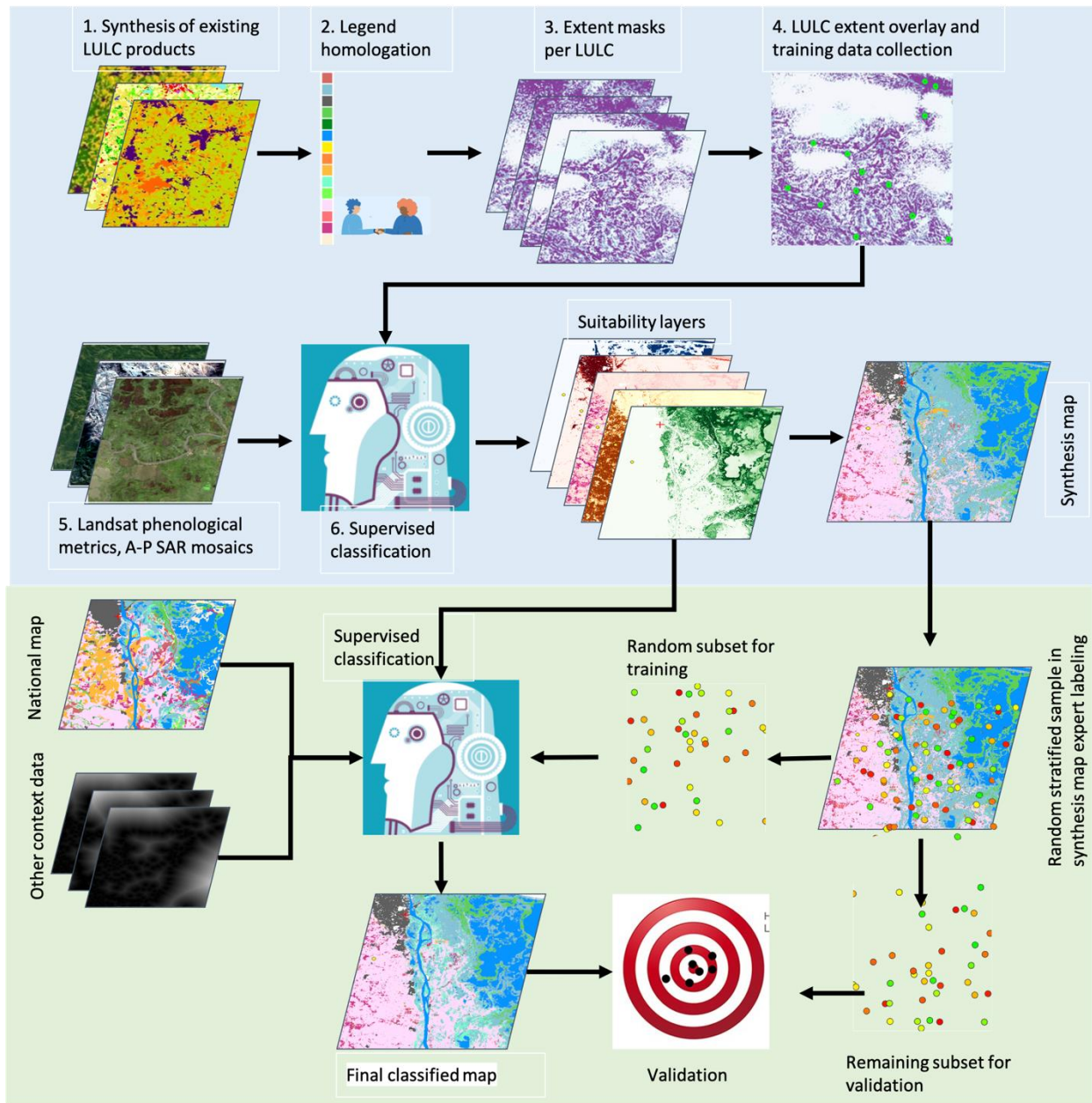
El estudio tiene como alcance las áreas en jurisdicción del sistema de Parques Nacionales Naturales. Con el fin de probar la viabilidad de los métodos propuestos, se definieron cuatro áreas críticas, que incluyen cinco parques nacionales naturales (Figura 1). Las áreas se seleccionaron en común acuerdo con personal de PNN ya que representan de manera óptima la variabilidad de condiciones ambientales en las áreas protegidas de Colombia en términos de tipos de ecosistemas y de procesos de transformación.



**Figura 1.** Áreas de estudio.

### **Métodos**

Los métodos de estudio se definieron con el fin de optimizar el poder predictivo de puntos de entrenamiento a usar para entrenar en algoritmo de clasificación. Para este fin, el trabajo se realizó en dos fases (Figura 2).



**Figura 2.** Métodos usados para la clasificación de imágenes. Azul corresponde a los pasos de la fase 1 y verde a los pasos de la fase 2

La primera fase consiste en la colección de una variedad de datos espaciales de coberturas vegetales disponibles a nivel nacional y global (región azul en Figura 2). Luego de una discusión entre el equipo de trabajo, se definió una leyenda común que balancea las coberturas discernibles a partir de datos disponibles y las necesidades de información para la toma de decisiones. Esta leyenda común se utilizó para homologar las bases de datos de tal manera que representen coberturas afines. Luego de la homologación, se produjeron mascarar que representan la extensión de cada cobertura por cada una de las capas de información de entrada colectada. Luego, se superpusieron las diferentes mascarar por cobertura para definir una máscara común que representa las áreas clasificadas por todos los datos de entrada como

la misma cobertura. Para cada cobertura, se sortearon puntos aleatorios y luego se usaron para producir una clasificación supervisada representando el porcentaje de afinidad de cada pixel para la cobertura de interés. Como datos predictores, se usaron mosaicos representando la mediana y el intervalo intercuartil de información espectral del satélite Landsat para el año 2018. Estos mosaicos se construyeron mediante el procesamiento de imágenes de Landsat listas para análisis (*Analysis Ready*) por el laboratorio GLAD de la Universidad de Maryland. Dichas imágenes son sometidas a un proceso de control de calidad que elimina pixeles de baja confiabilidad debido a distorsiones asociadas a características geométricas de iluminación u observación o a condiciones atmosféricas o de cubrimiento de nubes indeseadas (Potapov et al 2020<sup>1</sup>). También se incluyeron datos de mosaicos anuales de imágenes de radar producidas por el sensor ALOS-PALSAR. Finalmente las capas de afinidad para cada cobertura se sintetizaron en una capa única representando la cobertura con el valor de afinidad más alto para cada pixel.

En la segunda fase, se diseñó e implementó un protocolo para la colección y rotulación de puntos de muestreo con el fin de validar el mapa de coberturas (región verde en Figura 2). Dicho protocolo consistió en la adaptación de los protocolos usados en PNN para la producción de mapas de coberturas. El proceso consiste en el uso de la capa sintetizada producida en la primera fase para sortear puntos de muestreo en las diferentes coberturas bajo un diseño aleatorio estratificado, usando los tipos de coberturas como estratos. Los puntos muestreados fueron interpretados en un taller de trabajo realizado en el mes de Julio en las instalaciones de PNN con un grupo de expertos en interpretación de imágenes que trabajan en la construcción del mapa de coberturas a escala 1:100,000 (Figura 3). La interpretación se implementó usando el Plug-In ACATAMA que es diseñado para facilitar la interpretación de datos. La aplicación permite el despliegue de diferentes capas de información de manera simultánea para cada punto a validar. En la implementación se utilizó como imagen de referencia, el mosaico de Landsat usado para la clasificación en la primera fase, desplegado como semi-transparencias en colores RGB. Como imágenes de apoyo se usaron mosaicos semestrales de imágenes de alta resolución obtenidas mediante sensores *Planet Scope* a 3 m de resolución espacial e imágenes de alta resolución de *Google satellite* y *Microsoft Bing* (Figura 4). Cada punto fue interpretado por tres expertos y se consideraron como validos si fueron clasificados como una misma cobertura por al menos dos de los intérpretes. Aquellos puntos que no cumplieron esta condición, fueron revisitados por grupos de tres expertos para resolver inconsistencias en dos actividades posteriores al taller en Bogotá.

Inicialmente los puntos colectados estaban destinados a la validación de los mapas pero luego de recibir retroalimentación por parte de expertos (Figura 3), se vio la necesidad de destinar observaciones para refinar la clasificación. Para esto, se dividieron los puntos muestreados de manera aleatoria en dos conjuntos. El primero conjunto se usó para calibrar un modelo de clasificación y el otro para validación. En cada sitio, los datos de las coberturas con una

---

<sup>1</sup> Potapov, P., Hansen, M.C., Kommareddy, I., Kommareddy, A., Turubanova, S., Pickens, A., Adusei, B., Tyukavina, A., Ying, Q., 2020. Landsat analysis ready data for global land cover and land cover change mapping. *Remote Sensing* 12, 426.

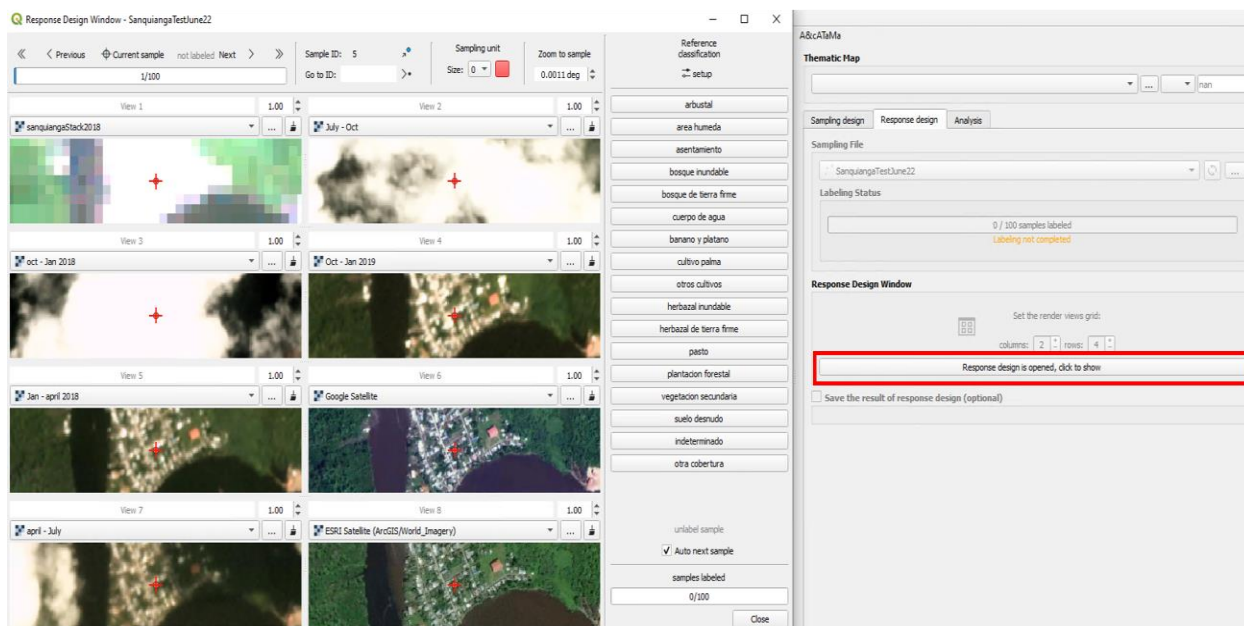


intensidad de muestreo mayor a 70 observaciones, se dividió en dos grupos con un número de observaciones igual al 50% de la muestra. El primer grupo se usó para calibración y el resto para validación. Para coberturas con un tamaño de muestreo menor a 70, se usó la totalidad de los datos para clasificar ya que se considera este tamaño de muestra bastante bajo para producir una clasificación robusta y al mismo tiempo para validar.



**Figura 3.** Taller de trabajo para la rotulación de puntos de muestreo en las áreas de estudio.

El subconjunto de datos de calibración se usó para clasificar un mapa de coberturas usando como datos de entrada, las capas de afinidad producidas en la fase anterior, el mapa homologado de coberturas adaptado de CORINE y capas representando variables de contexto tales como distancia a cuerpos de agua, vías y centros poblados, altitud, y aptitud de uso. El producto final se validó para coberturas con un número de observaciones mayor a 70. La producción de capas de afinidad y del mapa final de coberturas se realizó mediante el entrenamiento de un algoritmo de bosques aleatorios ya que múltiples estudios han demostrado la robustez de dicho algoritmo para la producción de mapas de coberturas.

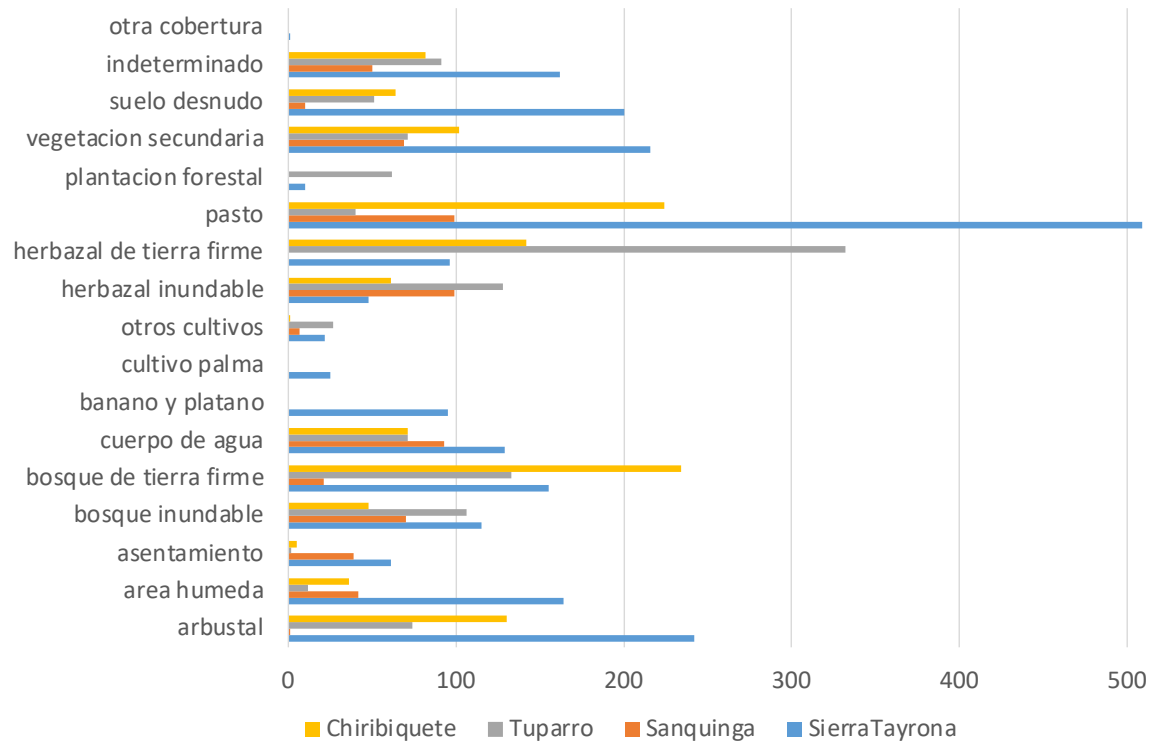


**Figura 4.** Despliegue de datos para la interpretación de puntos de muestreo usando el plug-in ACATAMA en QGIS

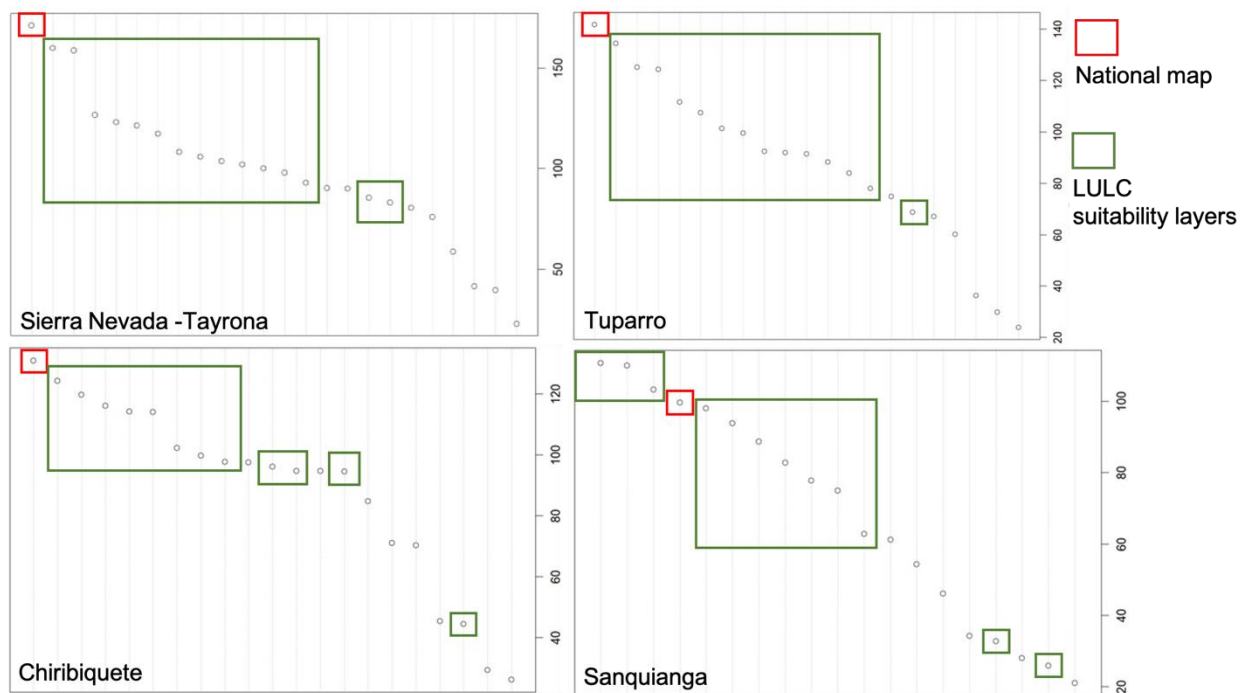
## Resultados

En total se rotularon 5250 puntos de muestreo. La cantidad de puntos muestreados por cobertura fue bastante heterogeneo (Figura 5).

El algoritmo de clasificación usado proporciona estimaciones de la importancia de las variables de entrada para la clasificación. Los resultados muestran que el mapa nacional de CORINE constituye el predictor más importante para la mayoría de las áreas de estudio, seguido de las capas de afinidad para las diferentes coberturas (Figura 6). Los resultados de la validación muestran que el método de clasificación mejora la capacidad de representación para la mayoría de las categorías validadas en las cuatro áreas de estudio (Figura 7). Ejemplos del desempeño del algoritmo se pueden apreciar en la figura 8.

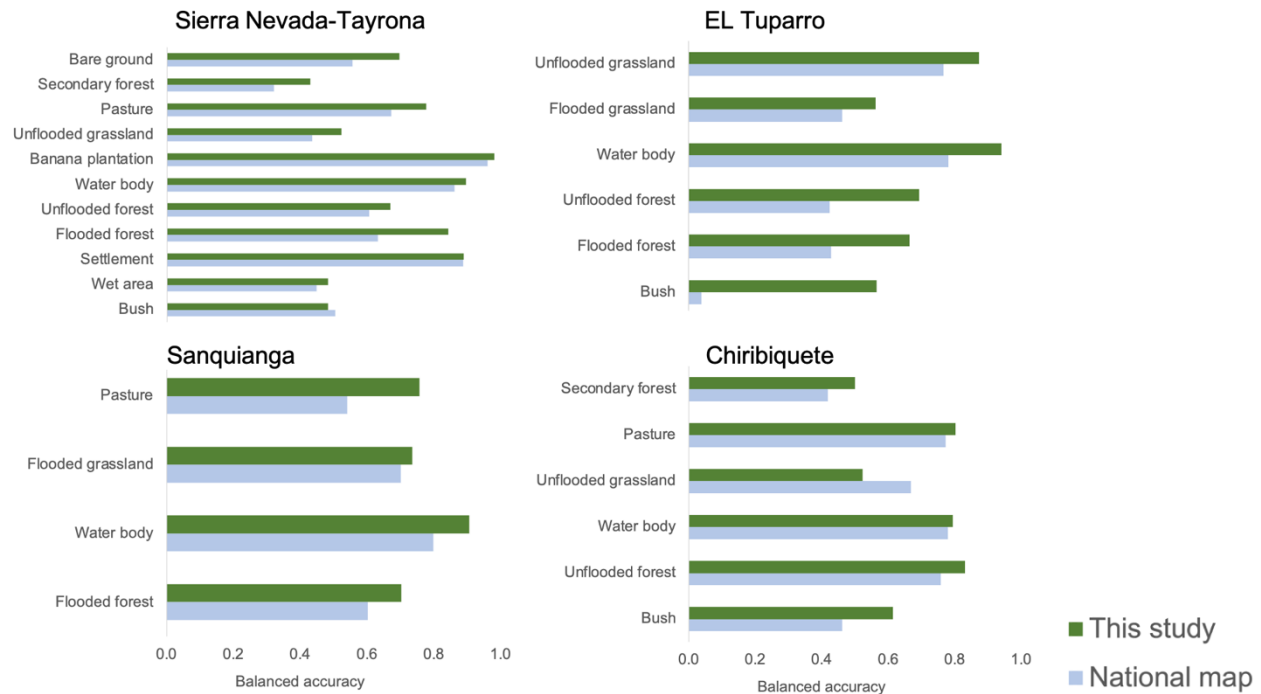


**Figura 5.** Número de puntos colectados por categoría en cada una de las áreas de estudio.

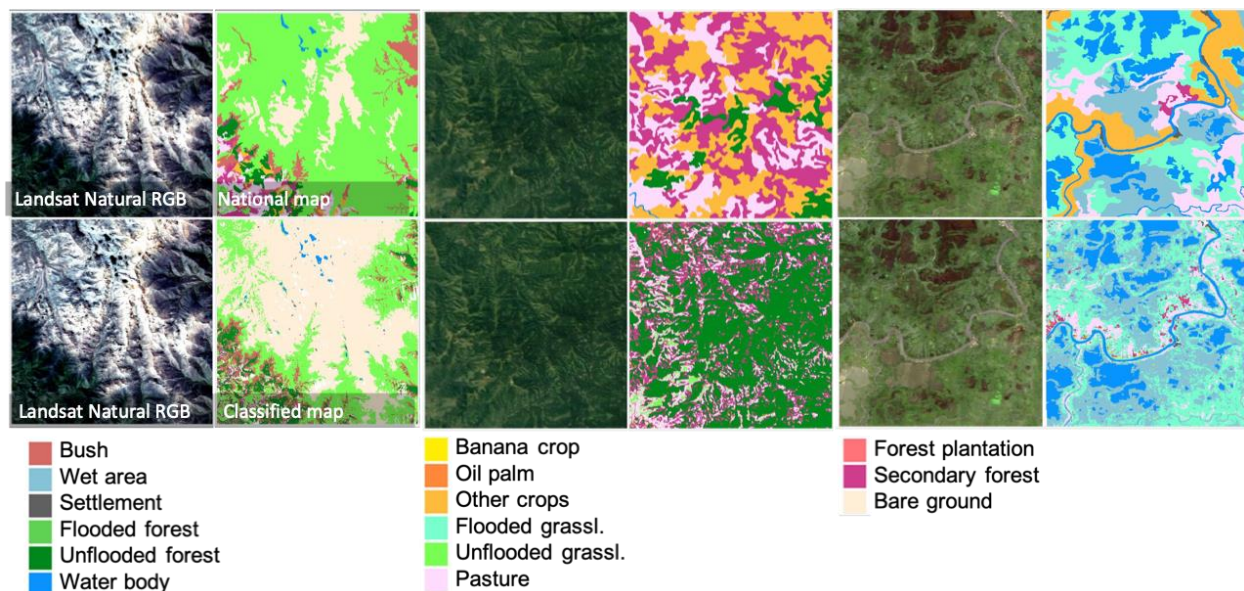


**Figura 6.** Importancia de variables para las cuatro áreas de estudio.





**Figura 7.** Resultado de validación para coberturas seleccionadas por sitio representados como precisión balanceada.



**Figura 8.** Ejemplos del desempeño del algoritmo.

## Conclusiones y pasos a seguir

- Los resultados demuestran que un método híbrido que combina conocimiento de expertos con datos de coberturas existentes y métodos automatizados puede mejorar la

eficiencia y precisión en el mapeo de la cobertura terrestre, de manera consistente en lugares condiciones ecológicas y de uso de la tierra contrastantes.

- La síntesis de mapas existentes proporciona conocimiento previo que es útil para optimizar la capacidad predictiva con datos de entrenamiento limitados para el mapeo de cobertura terrestre.
- La recopilación de puntos de referencia adicionales es necesaria para una evaluación de precisión completa. Por tanto el paso más prioritario a seguir es la recopilación de puntos de validación para categorías submuestreadas (Figura 4) con el objetivo de lograr un número de no menos de 100 puntos muestreados por sitio para producir estimativos de precisión general. La colección de dichos puntos está sujeta a la disponibilidad de expertos para etiquetar datos de referencia adicionales.
- Se requiere realizar discusiones entre los colaboradores para definir prioridades para los próximos 6 meses. Como opciones están 1) el desarrollo de un mapa de referencia para todas las áreas protegidas a nivel nacional (preferido) y 2) el análisis de cambios en las áreas piloto consideradas para este estudio.
- También es prioritario continuar la discusión con el IDEAM para evaluar la factibilidad de considerar nuestros métodos cartográficos para la producción de próximos mapas oficiales de cobertura terrestre en Colombia.