

Estimation of above-ground carbon in the Pereyra Iraola Biosphere Reserve (Buenos Aires, Argentina) through integration of GEDI, Sentinel-2 and artificial intelligence

Darío Nicolás Sánchez Leguizamón

*Universidad Nacional Arturo Jauretche (UNAJ), Florencio Varela, Buenos Aires, Argentina
sdarionicolos@gmail.com*

Abstract. This paper presents a modular and reproducible computational pipeline for the estimation and cartography of above-ground carbon stocks in the Pereyra Iraola Biosphere Reserve, a peri-urban ecological corridor between Greater Buenos Aires and Greater La Plata. The system integrates three digital technology layers: orbital LiDAR data from NASA's GEDI sensor (Global Ecosystem Dynamics Investigation), multispectral imagery from the Sentinel-2 constellation processed in Google Earth Engine (GEE), and a Random Forest regression model implemented in Python (scikit-learn) on Google Colab. The workflow, structured in independent and interchangeable modules, processed 873,668 valid samples to train a machine learning model that maps above-ground carbon at 10 m spatial resolution. The model achieved $R^2 = 0.71$, RMSE = 7.62 tC/ha, and MAE = 5.14 tC/ha. The SWIR1 band (B11) emerged as the dominant predictor (59% of feature importance), outperforming traditional vegetation indices such as NDVI, which saturates in high-biomass ecosystems. Results show a total estimated stock of 4.6 kilotons of above-ground carbon across the reserve. The modular architecture is fully transferable to other peri-urban ecosystems with GEDI and Sentinel-2 coverage, making it a scalable tool for environmental monitoring in the context of Agriculture 4.0 and Digital Agriculture.

Keywords: GEDI; Sentinel-2; Random Forest; Google Earth Engine; above-ground carbon; remote sensing; reproducible pipeline

Estimación del carbono aéreo en la Reserva de Biósfera Pereyra Iraola (Buenos Aires, Argentina) mediante integración de datos GEDI, Sentinel-2 e inteligencia artificial

Resumen. Este trabajo presenta un pipeline computacional modular y reproducible para la estimación y cartografía de reservas de carbono aéreo en la Reserva de Biósfera Pereyra Iraola, un corredor ecológico periurbano entre el Gran Buenos Aires y el Gran La Plata. El sistema integra tres capas tecnológicas digitales: datos LiDAR orbital del sensor GEDI de la NASA, imágenes multiespectrales de la constelación Sentinel-2 procesadas en Google Earth Engine (GEE) y un modelo de regresión Random Forest implementado en Python (scikit-learn) sobre Google Colab. El flujo, estructurado en módulos independientes e intercambiables, procesó 873.668 muestras válidas para entrenar un modelo de aprendizaje automático que mapea el carbono aéreo con resolución espacial de 10 m. El modelo alcanzó $R^2 = 0,71$, RMSE = 7,62 tC/ha y MAE = 5,14 tC/ha. La banda SWIR1 (B11) resultó el predictor dominante (59 % de

importancia de variable), superando a índices tradicionales como el NDVI, que se satura en ecosistemas de alta biomasa. Los resultados estiman un stock total de 4,6 ktoneladas de carbono distribuidas en la reserva. La arquitectura modular es transferible a otros ecosistemas periurbanos con cobertura GEDI y Sentinel-2, posicionándola como herramienta escalable para monitoreo ambiental en el marco de la Agricultura 4.0.

Palabras clave: GEDI; Sentinel-2; Random Forest; carbono aéreo; teledetección

1. Introducción

La cuantificación del carbono almacenado en la vegetación constituye un problema de procesamiento de datos geospaciales de alta dimensionalidad. Implica fusionar fuentes heterogéneas —sensores activos, pasivos y modelos digitales de elevación—, manejar volúmenes de información a escala de millones de píxeles y aplicar modelos estadísticos capaces de capturar relaciones no lineales entre la firma espectral y la estructura tridimensional del dosel.

Los inventarios forestales tradicionales, basados en parcelas de campo, ofrecen mediciones de alta precisión local pero resultan insuficientes ante las escalas espaciales y las frecuencias de actualización que requieren los sistemas de monitoreo ambiental moderno. En Argentina, los inventarios nacionales de bosques nativos operan con ciclos de relevamiento extensos y resolución espacial gruesa (Secretaría de Ambiente y Desarrollo Sustentable de la Nación, 2007), lo que limita su capacidad para detectar cambios estructurales en escalas temporales cortas. Esta brecha es particularmente crítica en corredores biológicos periurbanos como la Reserva Pereyra Iraola, donde la presión antrópica es elevada y la información sistematizada de biomasa y carbono aéreo es escasa o inexistente (Paszkievicz, 2010). La teledetección satelital, combinada con algoritmos de aprendizaje automático, surge como alternativa para cerrar esta brecha mediante estimaciones espacialmente explícitas, actualizables y de bajo costo marginal.

En respuesta a estos desafíos, el presente trabajo propone un pipeline computacional modular diseñado explícitamente para la reproducibilidad y la escalabilidad. Todo el flujo corre íntegramente en infraestructura cloud —Google Colab y Google Earth Engine (GEE)—, sin necesidad de instalaciones locales, lo que elimina barreras de acceso y garantiza que cualquier investigador con una cuenta Google pueda replicar los resultados. Esta característica es un aporte central al paradigma de Ciencia Abierta y a las prácticas de Agricultura Digital y Agricultura 4.0, donde la democratización de herramientas de análisis territorial es un objetivo estratégico, y posiciona al pipeline como una herramienta de monitoreo ambiental transferible aplicable a cualquier área natural periurbana con cobertura GEDI y Sentinel-2 disponible.

Dos plataformas de teledetección son tecnológicamente determinantes para el sistema. Por un lado, la misión GEDI (Global Ecosystem Dynamics Investigation) de la NASA, operada desde la Estación Espacial Internacional, provee mediciones LiDAR de onda completa con huellas de ~25 m que permiten derivar métricas de altura del dosel a escala global (Dubayah et al., 2020). Por otro, la constelación Sentinel-2 ofrece imágenes multiespectrales de 10–20 m de resolución con frecuencia de revisita de cinco días, accesibles mediante GEE. La integración digital de ambas fuentes plantea un problema de regresión multivariada de alta dimensión que los algoritmos de aprendizaje

automático como Random Forest resuelven eficientemente, sin asumir linealidad ni distribuciones paramétricas (Belgiu y Drăguț, 2016).

La Reserva de Biósfera Pereyra Iraola (~10.248 ha, noreste de Buenos Aires) fue seleccionada como caso de uso por su heterogeneidad estructural y su relevancia como enclave ecológico periurbano reconocido por la UNESCO (UNESCO, 2008). Su mosaico de coberturas —bosques densos, talaes relictuales, pastizales y humedales— ofrece un espectro amplio de valores de carbono que maximiza la capacidad discriminante del modelo y la robustez de la evaluación. La zonificación histórica del parque y la coexistencia de masas forestales nativas y exóticas (Paszkievicz, 2010) refuerzan su valor como sitio de prueba para metodologías de estimación de biomasa en paisajes mixtos.

El objetivo general es desarrollar, implementar y validar un pipeline computacional modular para mapear el carbono aéreo con resolución de 10 m, estableciendo una línea de base cuantitativa replicable para el monitoreo estructural de ecosistemas periurbanos. Los objetivos específicos son:

1. Construir un flujo de ingesta, filtrado y preprocesamiento de datos GEDI L2A y Sentinel-2 L2A desde GEE.
2. Implementar la estimación de biomasa y carbono aéreo mediante ecuaciones alométricas parametrizadas por tipo de cobertura.
3. Entrenar, validar y diagnosticar un modelo Random Forest multibanda para la predicción espacial del carbono.
4. Evaluar la modularidad, reproducibilidad y transferibilidad del pipeline a otros ecosistemas.

2. Materiales y Métodos

El pipeline se organiza en cuatro etapas secuenciales e independientes: (i) adquisición y filtrado de datos satelitales, (ii) preprocesamiento y construcción del dataset multibanda, (iii) modelado estadístico y entrenamiento del algoritmo de aprendizaje automático, y (iv) generación de productos cartográficos finales. Esta arquitectura modular permite sustituir cualquier bloque —por ejemplo, reemplazar Random Forest por XGBoost o integrar datos LiDAR aéreo— sin afectar el resto del flujo.

2.1 Entorno computacional y plataformas digitales

El flujo fue desarrollado íntegramente sobre infraestructura cloud. La Tabla 1 resume los componentes tecnológicos del sistema.

Tabla 1. Componentes tecnológicos del pipeline.

Componente	Tecnología / Plataforma	Rol en el pipeline
Procesamiento satelital	Google Earth Engine (GEE)	Ingesta, filtrado y exportación de colecciones GEDI y Sentinel-2
Entorno de ejecución	Google Colab (Python 3.10)	Desarrollo, entrenamiento del modelo y visualización

Componente	Tecnología / Plataforma	Rol en el pipeline
Álgebra de rasters	rasterio, numpy	Lectura, reproyección y operaciones sobre capas GeoTIFF
Análisis geoespacial	geopandas, geemap	Manejo de geometrías vectoriales e integración con GEE
Aprendizaje automático	scikit-learn (RandomForestRegressor)	Entrenamiento, validación e importancia de variables
Manipulación tabular	pandas	Gestión del dataset de muestras y métricas de error
Visualización	matplotlib	Mapas de carbono, residuos y diagnósticos espaciales

El código completo del pipeline está disponible en <https://github.com/sdarionicolasioop/carbono-pereyra-gedi>

El entorno fue autenticado en GEE mediante `ee.Authenticate()` y `ee.Initialize()`. Todo el código es ejecutable sin instalaciones locales, garantizando reproducibilidad completa.

2.2 Adquisición y preprocesamiento de datos

2.2.1. Sensor GEDI L2A (LiDAR orbital)

GEDl es un sensor LiDAR de onda completa montado en la Estación Espacial Internacional que opera a 1064 nm, con ocho haces láser simultáneos (cuatro de alta potencia) que generan huellas de ~25 m de diámetro. El producto L2A (V002) provee la métrica RH95 —percentil 95 de energía de retorno—, indicador robusto de la altura dominante del dosel (Dubayah et al., 2021). El filtrado siguió el protocolo de Saatchi y Favrichon (2023) y Lei et al. (2024): `quality_flag = 1`, `degrade_flag = 0`, sensibilidad > 0,95, adquisición nocturna, diferencia altimétrica GEDl–SRTM < 50 m y uso exclusivo de power beams. El protocolo fue implementado como consulta sobre la colección GEDl02_A en GEE.

2.2.2. Sensor Sentinel-2 L2A (multiespectral óptico)

Las imágenes Sentinel-2 (nivel L2A) se obtuvieron desde la colección COPERNICUS/S2_SR_HARMONIZED en GEE para el período 2023, filtrando cobertura nubosa < 10 % mediante la banda QA60. Se generó un mosaico anual por mediana espectral píxel a píxel. El stack resultante (Sentinel2_Pereyra_Stack.tif) integra las capas de la Tabla 2, todas proyectadas a 10 m y EPSG:4326.

Tabla 2. Bandas e índices del stack Sentinel-2.

Banda	Descripción espectral	Resolución nativa	En stack
B2	Azul (490 nm)	10 m	10 m
B3	Verde (560 nm)	10 m	10 m
B4	Rojo (665 nm)	10 m	10 m

Banda	Descripción espectral	Resolución nativa	En stack
B8	NIR (842 nm)	10 m	10 m
B11	SWIR1 (1610 nm)	20 m	10 m (bilineal)
B12	SWIR2 (2190 nm)	20 m	10 m (bilineal)
NDVI	$(B8-B4)/(B8+B4)$	—	Calculado en GEE
NDWI	$(B3-B8)/(B3+B8)$	—	Calculado en GEE

2.3 Estimación de biomasa y carbono aéreo

La variable objetivo se derivó en tres pasos. Primero, a partir de RH95 se construyó un modelo regional altura–diámetro (H:D) mediante regresión potencial. Segundo, la biomasa aérea se estimó con la ecuación alométrica de Chave et al. (2014):

$$AGB_{est} = 0,0673 \times (\rho \times D^2 \times H)^{0,976}$$

donde ρ es la densidad de madera ($g\ cm^{-3}$, asignada por tipo de hábitat según el mapa UNESCO), D el diámetro estimado (cm) y H la altura derivada de GEDI (m). Tercero, la biomasa fue convertida a carbono mediante el factor estándar de 0,47, siguiendo los protocolos internacionales de cálculo de reservas de carbono forestal (Winrock International, 2018), generando el raster Carbono_Pereyra.tif a 10 m. Este raster constituye la variable objetivo (y) del modelo.

2.4 Implementación del modelo Random Forest

2.4.1. Diseño del pipeline de aprendizaje automático

El modelo implementado fue RandomForestRegressor de scikit-learn. La elección responde a tres características técnicas: (i) manejo nativo de la multicolinealidad entre bandas espectrales sin preprocesamiento PCA; (ii) capacidad para modelar relaciones no lineales; y (iii) provisión de feature importance por impureza de Gini, que permite interpretar la contribución de cada banda. El dataset se construyó extrayendo los valores espectrales del stack Sentinel-2 en cada píxel con muestra GEDI válida, produciendo 873.668 muestras. La Tabla 3 detalla los hiperparámetros del modelo y su justificación técnica.

Tabla 3. Parámetros del modelo Random Forest.

Parámetro	Valor	Justificación técnica
División entren./validación	70 % / 30 %	Partición aleatoria por clase de altura
n_estimators	200 árboles	Convergencia del error OOB verificada desde 150 árboles
max_depth	20 niveles	Balance entre capacidad y regularización implícita
random_state	42	Reproducibilidad exacta del modelo

Parámetro	Valor	Justificación técnica
n_jobs	-1	Paralelización completa sobre núcleos Colab
Muestras totales	873.668 píxeles	Cobertura completa de la reserva tras filtrado GEDI

2.4.2. Variables predictoras y métricas de evaluación

Las variables predictoras (X) son las ocho capas del stack Sentinel-2: B2, B3, B4, B8, B11, B12, NDVI y NDWI. La variable objetivo (y) es el carbono en tC/ha. No se aplicó normalización previa dado que Random Forest es invariante a escala. El desempeño fue evaluado mediante R^2 , RMSE y MAE sobre el conjunto de validación, complementado con análisis de residuos espaciales y evaluación estratificada del error por clase de altura del dosel.

3. Resultados

3.1 Desempeño del modelo Random Forest

El modelo evidenció desempeño robusto sobre el conjunto de validación (Tabla 4).

Tabla 4. Métricas de desempeño del modelo Random Forest.

Métrica	Valor	Interpretación
R^2	0,712	El modelo explica el 71,2 % de la variabilidad del carbono aéreo
RMSE	7,62 tC/ha	Equivalente a ± 17 % del rango operativo de valores
MAE	5,14 tC/ha	Desviación absoluta media entre predicciones y observaciones

Estos valores son coherentes con los reportados por Li et al. (2022) y Belgiu y Drăguț (2016) para modelos Random Forest aplicados a estimación de biomasa con Sentinel-2, confirmando que el pipeline opera dentro del rango de desempeño esperado para ecosistemas templados con heterogeneidad estructural moderada.

El gráfico de dispersión entre carbono observado y predicho (Figura 1) permite inspeccionar el comportamiento del modelo a lo largo de todo el rango de biomasa. La nube de puntos se ajusta a la línea 1:1 en los valores intermedios, mientras que en los rangos altos (> 80 tC/ha) se observa una leve subestimación —comportamiento típico de Random Forest por la naturaleza promediadora de sus árboles— y mayor dispersión en los valores bajos asociada a la mezcla espectral suelo–vegetación.

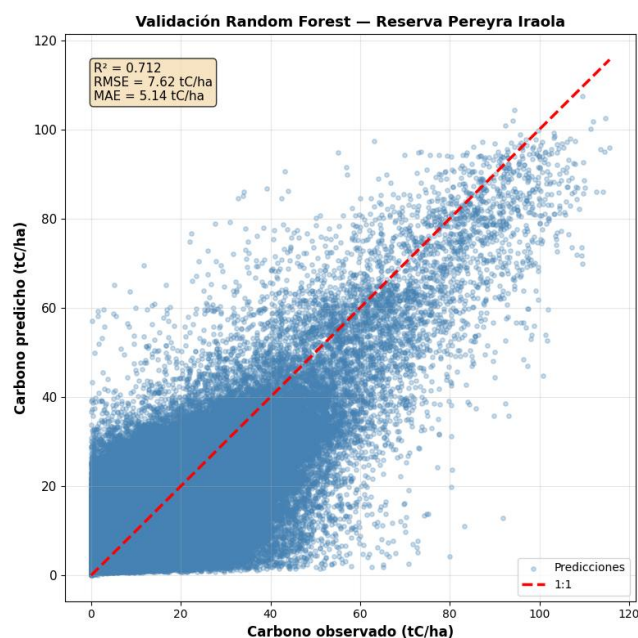


Figura 1. Validación del modelo Random Forest: carbono observado vs. predicho (tC/ha). La línea discontinua representa la relación 1:1. $R^2 = 0,712$; RMSE = 7,62 tC/ha; MAE = 5,14 tC/ha.

3.2 Importancia de variables espectrales: el rol determinante de SWIR1

El análisis de feature importance revela el hallazgo técnico central del trabajo: la banda SWIR1 (B11) concentra el 59,1 % de la importancia total del modelo, superando ampliamente a todas las demás variables espectrales (Tabla 5).

Tabla 5. Importancia de variables del modelo Random Forest.

Variable	Importancia (%)	Interpretación física
B11 (SWIR1)	59,1 %	Sensibilidad directa a biomasa leñosa y contenido de agua estructural
B3 (Verde)	13,2 %	Diferenciación entre especies caducifolias y siempreverdes
NDWI	6,2 %	Humedad superficial y densidad del dosel
B4 (Rojo)	5,4 %	Absorción clorofílica; útil en zonas degradadas
B2 (Azul)	4,9 %	Reflectancia superficial y albedo
B12 (SWIR2)	4,3 %	Complemento espectral del SWIR1
B8 (NIR)	3,8 %	Vigor foliar y cobertura del dosel
NDVI	3,2 %	Saturado en áreas de alta biomasa; escasa utilidad discriminante

La baja ponderación del NDVI (3,2 %) es técnicamente significativa: confirma su saturación espectral en ecosistemas de alta cobertura y su limitada utilidad para discriminar diferencias estructurales. En contraste, el SWIR1 captura información sobre la estructura interna de la vegetación y el contenido de agua del tejido leñoso, variables directamente relacionadas con la acumulación de biomasa. Este hallazgo es consistente con los resultados de Tolan et al. (2024) y Lei et al. (2024) para ecosistemas templados.

3.3 Análisis estratificado del error por clase de altura

Los residuos del modelo fueron analizados por clase de altura del dosel (Tabla 6).

Tabla 6. Error del modelo por clase de altura del dosel.

Clase de altura (m)	N muestras	Error medio (tC/ha)	Desvío estándar	RMSE aprox.
< 1	342.558	-2,43	2,64	3,59
1-5	222.549	-1,28	4,57	4,75
5-10	175.285	+0,85	6,48	6,54
10-15	105.578	+6,05	6,83	9,12
15-20	20.266	+11,06	9,64	14,67
> 20	7.432	+10,02	11,28	15,08

El gradiente de error revela subestimación en estratos bajos (< 5 m) y sobreestimación creciente en doseles altos (> 15 m), comportamiento consistente con la saturación del SWIR en coberturas continuas densas (Lei et al., 2024; Potapov et al., 2021) y con la mezcla espectral suelo-vegetación en pastizales.

3.4 Validación visual y balance general

El mapa final de carbono aéreo predicho (Figura 2) diferencia con nitidez los núcleos forestales de alta biomasa (> 40 tC/ha, en verde) de las coberturas herbáceas y humedales (< 5 tC/ha, en tonos rojizos). La distribución espacial muestra una densidad media de 10,89 tC/ha (DE = 12,11; mín. = 0,04; máx. = 115,10 tC/ha) sobre el área analizada, valor relativamente bajo a escala de reserva debido al predominio de pastizales y coberturas abiertas en el mosaico paisajístico. El sesgo medio entre el mapa observado y el predicho fue +1,5 tC/ha, sin agrupamientos sistemáticos, respaldando la estabilidad espacial del modelo. El stock total estimado es de 4,6 ktoneladas de carbono aéreo en la reserva.

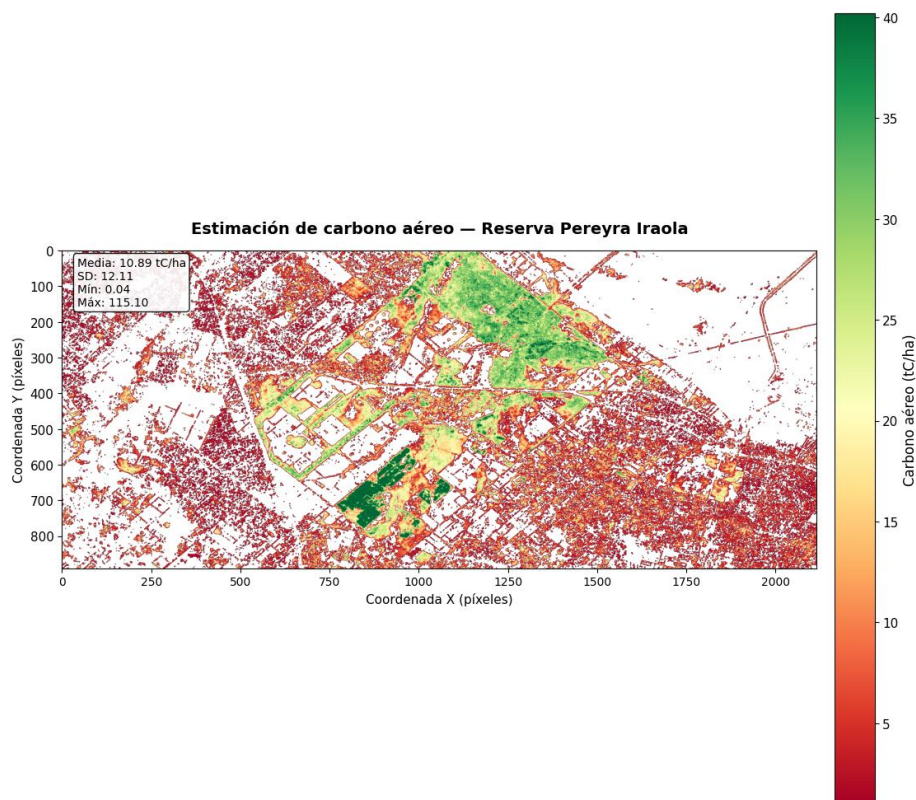


Figura 2. Mapa de estimación de carbono aéreo (tC/ha) en la Reserva de Biósfera Pereyra Iraola. Los núcleos forestales (verde) alcanzan valores > 40 tC/ha; las coberturas herbáceas y humedales (rojo) presentan valores < 5 tC/ha.

4. Discusión

4.1 Pipeline modular, reproducible y escalable

Una contribución técnica central de este trabajo es la arquitectura modular del pipeline. Al separar claramente la ingesta GEDI (Módulo 1), el preprocesamiento Sentinel-2 en GEE (Módulo 2), la estimación alométrica (Módulo 3) y el entrenamiento RF (Módulo 4), el flujo resulta adaptable a nuevos contextos sin reescritura sustancial. La sustitución del área de estudio requiere únicamente redefinir el polígono georreferenciado en GEE; el resto del pipeline opera sin modificaciones. Esta característica lo convierte en una herramienta práctica directamente aplicable a la gestión ambiental de otras reservas periurbanas con cobertura GEDI y Sentinel-2.

La reproducibilidad está garantizada por tres decisiones de diseño: (i) `random_state = 42` fija el estado aleatorio del modelo, produciendo resultados idénticos en cada ejecución; (ii) todos los datos de entrada provienen de colecciones públicas y versionadas en GEE; y (iii) el código corre en Google Colab sin instalaciones locales, accesible a cualquier investigador con cuenta Google. Para escalar a nivel continental,

el pipeline puede migrar íntegramente a GEE mediante Earth Engine Python API, eliminando la dependencia de Google Drive como almacenamiento intermedio.

4.2 SWIR1 como predictor dominante: bases físicas e implicaciones

La dominancia de B11 (SWIR1, 59 % de importancia) sobre el NDVI (3,2 %) tiene una base física precisa. El NDVI exhibe saturación para valores de índice de área foliar (LAI) superiores a $\sim 3 \text{ m}^2/\text{m}^2$, condición frecuente en coberturas boscosas densas: a partir de ese umbral, la reflectancia en el rojo y el infrarrojo cercano se estabiliza y el índice deja de responder a incrementos de biomasa. En contraste, la región del infrarrojo de onda corta (SWIR1, $\sim 1610 \text{ nm}$) es sensible al contenido de agua del tejido leñoso y a la estructura interna del dosel, propiedades que continúan variando con la acumulación de biomasa incluso bajo coberturas continuas. Por ello, el SWIR1 conserva poder discriminante donde los índices de verdor se saturan (Lei et al., 2024; Tolan et al., 2024).

Este resultado implica que los sistemas de monitoreo de carbono basados exclusivamente en índices de verdor (NDVI, EVI) subestiman sistemáticamente la variabilidad del carbono en ecosistemas de alta cobertura. La incorporación explícita de bandas SWIR en los modelos de regresión debería considerarse un estándar mínimo en plataformas de monitoreo continuo integradas en infraestructuras de Agricultura de Precisión y gestión ambiental digital. El pipeline aquí presentado demuestra que esta integración es técnicamente viable y computacionalmente accesible mediante herramientas cloud de uso libre.

4.3 Validación biológica: coherencia con estudios regionales

Para evaluar la coherencia biológica de las estimaciones, los valores obtenidos se contrastaron con inventarios de campo de ecorregiones comparables del centro-este argentino. Los núcleos forestales de la reserva, que en el mapa predicho superan los 40 tC/ha, se ubican en el mismo orden de magnitud que los talares del noreste bonaerense —la principal comunidad boscosa nativa de la región, dominada por *Celtis tala* (Arturi y Goya, 2004)— y resultan algo inferiores a los stocks medidos en bosques nativos semixerofíticos del Espinal entrerriano, donde Sione et al. (2021) reportan un promedio de $43,99 \pm 10,43 \text{ tC/ha}$ (rango 26,02–52,78 tC/ha) mediante modelos alométricos locales. La fisonomía semixerofítica del Espinal es comparable a la de los parches de talar de Pereyra Iraola, por lo que esta coincidencia respalda la validez del rango superior estimado por el modelo.

La densidad media a escala de reserva (10,89 tC/ha) es marcadamente inferior a estos valores forestales porque integra el mosaico completo de coberturas —pastizales, humedales y suelos abiertos—, que dominan la superficie del parque y presentan stocks muy bajos. Esta diferencia entre la media territorial y los valores de los núcleos forestales es esperable y coherente con la heterogeneidad estructural del corredor periurbano. La ausencia de inventarios sistematizados específicos para Pereyra Iraola (Paszkievicz, 2010) impide, no obstante, una validación directa con parcelas permanentes en el sitio, lo que constituye la principal vía de mejora de la calibración alométrica.

4.4 Limitaciones técnicas y líneas de mejora

El modelo presenta cuatro limitaciones identificadas. Primero, la autocorrelación espacial entre píxeles adyacentes (separación de 10 m) infla las métricas de validación reportadas, ya que píxeles vecinos de entrenamiento y prueba comparten información estructural; en consecuencia, el $R^2 = 0,71$ debe interpretarse como un límite superior del desempeño real sobre píxeles espacialmente independientes. La implementación de validación cruzada espacial —separando bloques geográficos disjuntos de entrenamiento y prueba— es la mejora prioritaria y permitiría estimar el desempeño de generalización de forma no sesgada. Segundo, la fusión de datos a distintas resoluciones (GEDI ~25 m vs. Sentinel-2 10 m) introduce simplificación espacial intraparche y un muestreo no uniforme determinado por la huella GEDI. Tercero, el modelo alométrico opera con densidades de madera asignadas por clase de cobertura sin mediciones de campo; la calibración con parcelas permanentes reduciría la incertidumbre. Cuarto, la arquitectura requiere reentrenamiento para cada nuevo ecosistema; el desarrollo de modelos de transfer learning representa la principal línea de evolución técnica futura.

5. Conclusiones

Este trabajo desarrolló y validó un pipeline computacional modular y reproducible para la estimación espacialmente explícita del carbono aéreo, integrando datos LiDAR orbital GEDI, imágenes multiespectrales Sentinel-2 y un modelo Random Forest implementado en Python sobre Google Colab. Los resultados demuestran la viabilidad técnica del enfoque ($R^2 = 0,71$, RMSE = 7,62 tC/ha, stock estimado: 4,6 ktoneladas de C).

Las contribuciones principales son: (i) un pipeline modular completamente ejecutable en infraestructura cloud, sin instalaciones locales, reproducible y transferible a cualquier ecosistema con cobertura GEDI y Sentinel-2; (ii) la identificación de SWIR1 como predictor dominante de la biomasa leñosa (59 %) frente al NDVI saturado (3,2 %), con bases físicas en la sensibilidad del infrarrojo de onda corta al contenido de agua del tejido leñoso, e implicaciones directas para el diseño de herramientas de monitoreo ambiental en el marco de la Agricultura Digital y la Agricultura 4.0; y (iii) un análisis estratificado del error y una validación biológica frente a inventarios regionales (talares bonaerenses y bosques del Espinal) que sitúan las estimaciones en rangos ecológicamente coherentes. Las líneas de mejora prioritarias incluyen la validación cruzada espacial, la calibración alométrica con datos de campo y la exploración de modelos de transfer learning para generalizar el pipeline entre regiones. En conjunto, este trabajo ilustra cómo la convergencia entre teledetección estructural, aprendizaje automático e infraestructura cloud configura una nueva generación de herramientas digitales para la gestión ambiental periurbana, alineada con los objetivos de transformación digital del agro en el marco de la Agricultura 4.0.

Referencias bibliográficas

Arturi, M. F. y Goya, J. F. (2004). Estructura, dinámica y manejo de los talares del NE de Buenos Aires. En M. F. Arturi, J. L. Frangi y J. F. Goya (Eds.), *Ecología y manejo de los bosques de Argentina* (pp. 1–24). Editorial de la Universidad Nacional de La Plata. <http://sedici.unlp.edu.ar/handle/10915/15915>

- Belgiu, M. y Drăguț, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>
- Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M. S., Delitti, W. B., ... y Vieilledent, G. (2014). Improved allometric models to estimate the aboveground biomass of tropical trees. *Global Change Biology*, 20(10), 3177–3190. <https://doi.org/10.1111/gcb.12629>
- Dubayah, R., Blair, J. B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., ... y Silva, C. (2020). The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. *Science of Remote Sensing*, 1, 100002. <https://doi.org/10.1016/j.srs.2020.100002>
- Dubayah, R., Hofton, M., Blair, J., Armston, J., Tang, H. y Luthcke, S. (2021). *GEDl L2A Elevation and Height Metrics Data Global Footprint Level V002*. NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/GEDI/GEDI02_A.002
- Lei, Y., Wang, Y., Wang, G., Song, C., Cao, H. y Xiao, W. (2024). Estimating forest canopy height based on GEDI LiDAR data and multi-source remote sensing images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-1, 297–303. <https://doi.org/10.5194/isprs-archives-XLVIII-1-2024-297-2024>
- Li, X., Zhao, K. y Zhang, H. (2022). Integrating Sentinel-2 imagery and machine learning algorithms to estimate aboveground biomass across heterogeneous landscapes. *Remote Sensing*, 14(8), 1913. <https://doi.org/10.3390/rs14081913>
- Paszkievicz, M. S. (2010). *El Parque Pereyra Iraola como reserva de biósfera: conflictos y potencialidades*. Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC). <https://digital.cic.gba.gob.ar/handle/11746/832>
- Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M. C., Kommareddy, A., ... y Hofton, M. (2021). Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sensing of Environment*, 253, 112165. <https://doi.org/10.1016/j.rse.2020.112165>
- Saatchi, S. S. y Favrichon, S. (2023). *Global Vegetation Height Metrics from GEDI and ICESat-2*. ORNL DAAC. <https://doi.org/10.3334/ORNLDAAC/2294>
- Secretaría de Ambiente y Desarrollo Sustentable de la Nación. (2007). *Primer Inventario Nacional de Bosques Nativos: Informe Regional Espinal*. Ministerio de Salud y Ambiente de la Nación, Argentina.
- Sione, S. M., Andrade-Castañeda, H. J., Ledesma, S. G., Rosenberger, L. J., Oszust, J. D. y Wilson, M. G. (2021). Estimación del contenido y captura potencial de carbono en la biomasa arbórea de bosques nativos del Espinal (Entre Ríos, Argentina). *FAVE Sección Ciencias Agrarias*, 20(1), 331–349. https://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S1666-77192021000100331
- Tolan, J., Yang, H. I., Nosarzewski, B., Couairon, G., Vo, H. V., Brandt, J., ... y Couprie, C. (2024). Very high-resolution canopy height maps from RGB imagery

using self-supervised vision transformer and convolutional decoder trained on aerial LiDAR. *Remote Sensing of Environment*, 300, 113888. <https://doi.org/10.1016/j.rse.2024.113888>

UNESCO. (2008). *Proyecto de Reserva de Biosfera Pereyra Iraola: Documento de postulación para la Red Mundial de Reservas de Biosfera (MAB-UNESCO)*. Ministerio de Asuntos Agrarios de la Provincia de Buenos Aires.

Winrock International. (2018). *Calculating carbon stocks: A guidance module*. Winrock International, Ecosystem Services Unit.

Apéndice – Uso de Inteligencia Artificial en la redacción

Durante la preparación de este trabajo, el autor utilizó herramientas de inteligencia artificial (modelos de lenguaje de gran escala) con el fin de mejorar el lenguaje y la legibilidad del manuscrito, así como para la documentación y revisión de los scripts de procesamiento. Tras utilizar estas herramientas, el autor revisó y editó el contenido según fue necesario y asume plena responsabilidad por el contenido de la publicación. La totalidad del diseño experimental, el flujo de análisis, la interpretación de datos y la estructura conceptual fueron desarrollados por el autor. Los algoritmos ejecutados corresponden a bibliotecas científicas abiertas (scikit-learn, rasterio, numpy, pandas); no se generaron modelos predictivos mediante IA generativa.