

2023



Instituto Geográfico
Nacional

Dirección Nacional de
Servicios Geográficos

Dirección de Información
Geoespacial

Departamento de
Aplicaciones
Geoespaciales

[Clasificación de cobertura del suelo en Argentina a partir de imágenes satelitales utilizando el algoritmo Random Forest]

Documento metodológico acerca de la clasificación del Uso del Suelo del territorio nacional
utilizando el algoritmo Random Forest.

Contenido

1 - Resumen	4
2 - Introducción.....	5
3 - Materiales y métodos.....	5
3.1 - Área de estudio	5
3.2 - Imágenes satelitales	7
3.3 - Categorías de cobertura	7
3.4 - Preprocesamiento de imágenes.....	9
3.5 - Puntos de entrenamiento y validación	13
3.6 - Algoritmo Random Forest.....	15
3.7 - Análisis de la clasificación	16
4 - Resultados.....	17
4.1 - Centro	17
4.2 - Cuyo	18
4.3 - Noroeste Argentino (NOA)	18
4.4 - Patagonia	18
4.5 - Noreste Argentino (NEA).....	19
4.6 - Atlántica.....	19
5 - Conclusiones.....	20
6 - Referencias	21

Índice de figuras

Figura 1: Área de estudio	6
Figura 2: División en regiones de la Argentina, de la región patagónica por procesamiento se separó a las islas del atlántico sur.....	7
Figura 3: Esquema de procesamiento, donde ee.Reducer () calcula la media, mínima y máxima. ..	10
Figura 4: Distribución de los puntos de entrenamiento y validación en el área de estudio	14
Figura 5: Representación de cómo funciona el algoritmo de clasificación supervisada Random Forest	16
Figura 6: Ejemplo de Matriz de Confusión, donde se observa los distintos tipos de precisión, de errores y las diferencias entre las clases analizadas.	17
Figura 7: Resultado final de cobertura del suelo de la República Argentina.....	19

Índice de tablas

Tabla 1: Productos, niveles de procesamiento y fechas utilizadas.....	13
Tabla 2: Cantidad de puntos de entrenamiento y validación por categoría de cobertura del suelo. .	15

1 - Resumen

El Instituto Geográfico Nacional de Argentina tiene la misión de entender la representación oficial del territorio nacional y elaborar marcos normativos para su realización, mediante la obtención de información geográfica precisa, oportuna y concisa. La gran disponibilidad de información satelital permite incorporar la dimensión espacial y temporal, dirigiendo los esfuerzos en el desarrollo de metodologías que permitan una actualización constante y el estudio de cambios cualitativos en las coberturas y usos de suelo en el tiempo, especialmente debidos a intervenciones humanas, que tienen efectos significativos en el ambiente. En este contexto, se propone una metodología para establecer, de manera gráfica y cuantitativa, los distintos tipos de cobertura del territorio argentino, para que puedan ser insumos de posteriores análisis en cuanto a su evolución espacial y temporal. Se dividió al territorio en 6 regiones y para ello se definieron 13 categorías de cobertura y se procedió a utilizar el algoritmo Random Forest en la plataforma basada en la nube Google Earth Engine. Los resultados otorgan índices de precisión general aproximados al 83% y un coeficiente Kappa aproximado de 66%.

Palabras Clave: cobertura, teledetección, random forest, ambiente, análisis.

2 - Introducción

Conocer y caracterizar la cobertura de suelo, y eventualmente monitorear su cambio en el tiempo constituyen insumos fundamentales para múltiples aplicaciones, que van desde la planificación y gestión del territorio y sus recursos naturales, hasta la prevención e intervención en materia de Riesgo de Desastres (Talukdar, S., 2020; Nguyen, K. A., Liou, Y. A., 2019).

El Comité de Expertos sobre la Gestión Mundial de la Información Geoespacial (UN-GGIM) es el órgano intergubernamental de Naciones Unidas encargado de debatir, mejorar y coordinar las actividades de gestión de la información geoespacial mundial con la participación de los Estados miembros y cuenta con su capítulo regional: UN-GGIM Américas. Su objetivo es maximizar el uso y los beneficios económicos, sociales y ambientales derivados de la información geoespacial, basado en el conocimiento y el intercambio de experiencias y tecnologías por parte de los Estados miembros (UN-GGIM, 2019). En este contexto, un grupo especializado del UN-GGIM realizó la identificación de una serie de 14 datos básicos y fundamentales para fortalecer la infraestructura de datos espaciales de un país, entre los cuales se encuentran las coberturas y usos de suelo.

Por su parte, el Instituto Geográfico Nacional tiene la misión de entender la representación oficial del territorio nacional y en la elaboración de los marcos normativos para su realización, mediante la obtención de información geográfica precisa, oportuna y concisa (Instituto Geográfico Nacional, 2022).

La gran disponibilidad de información satelital permite incorporar la dimensión espacial y la dimensión temporal, dirigiendo los esfuerzos en el desarrollo de metodologías que admiten una actualización constante y el estudio de cambios cuali-cuantitativos en las coberturas y usos de suelo en el tiempo, especialmente debidos a intervenciones antrópicas, que tienen efectos significativos en el ambiente (Liou, Y. A., Nguyen, K. A., 2017; Nguyen, K. A., Liou, Y. A., 2019)..

Se propone aquí explorar una metodología para establecer, de manera gráfica y cuantitativa, los distintos tipos de cobertura del territorio nacional, para lograr brindar un insumo que permita futuros análisis de su evolución espacial y temporal, compatibilizando las coberturas de suelo que pueden ser discernidas con técnicas de teledetección, y las clases establecidas en el Catálogo de Objetos Geográficos institucional (Instituto Geográfico Nacional, 2017).

3 - Materiales y métodos

3.1 - Área de estudio

El área de estudio seleccionada para implementar la metodología fue la República Argentina, exceptuando los territorios de la Antártida. En la figura 1 se observa el recorte del mapa ArgenMap (<https://mapa.ign.gob.ar>) con el área propiamente dicha. Argentina cuenta con un total de 18 ecorregiones, resultado de un conjunto amplio de características climáticas, topográficas, geológicas, biológicas, entre muchas otras que conforman un gran mosaico de coberturas (Waller, T., et. al., 2010).



Figura 1: Área de estudio

Para lograr optimizar el tiempo de procesamiento se dividió a la totalidad del área en las 6 regiones de la Argentina: Centro, Patagonia, NOA, NEA, Cuyo y Atlántica (Figura 2).

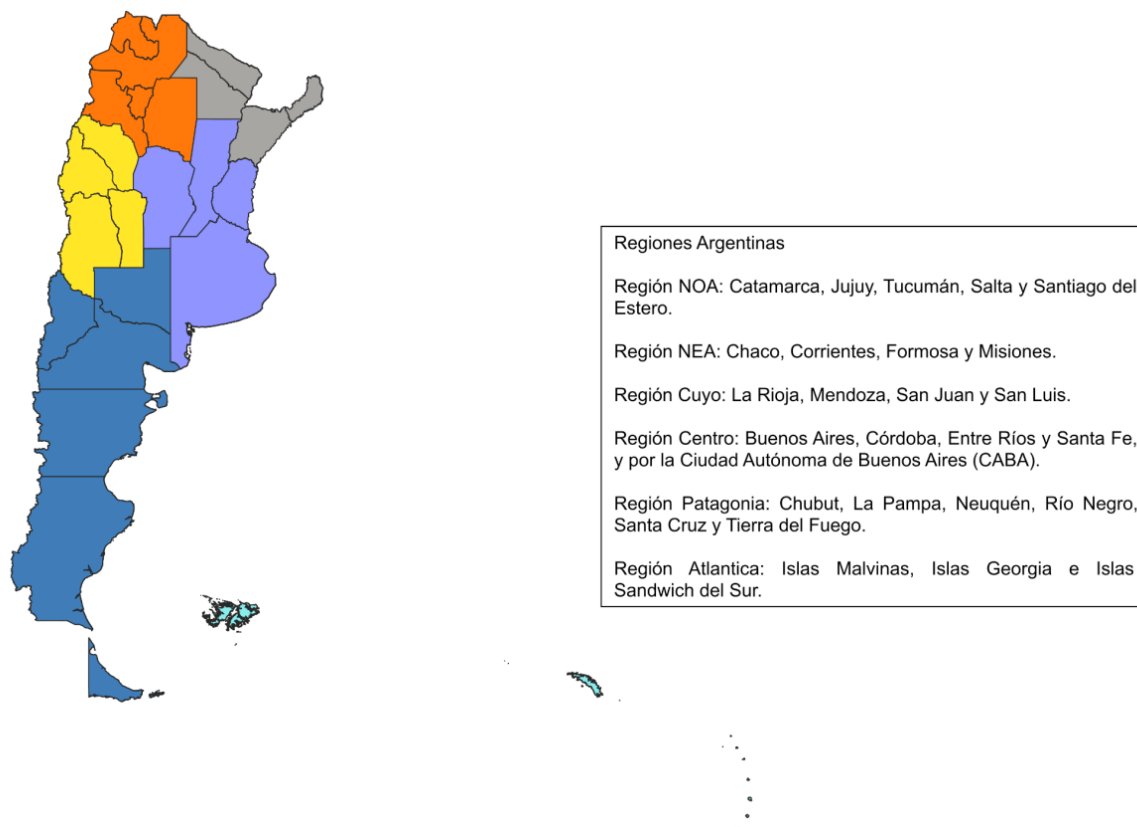


Figura 2: División en regiones de la Argentina, de la región patagónica por procesamiento se separó a las islas del atlántico sur

3.2 - Imágenes satelitales

Se utilizaron las imágenes satelitales de la misión SENTINEL-2 MSI, con nivel de preprocesamiento 1-C para todo el territorio a excepción de Tierra del Fuego que se utilizaron las 2-A, ambas con una resolución espacial de 10m y 20m dependiendo la banda utilizada. En la provincia de Tierra del Fuego se utilizó ese nivel de procesamiento debido a la alta presencia de nubes en esa zona. El nivel de procesamiento 1-C son imágenes ortorectificadas con nivel de reflectancia por debajo de la atmósfera (BOA por sus siglas en inglés), mientras que la 2-A son imágenes corregidas atmosféricamente. Del mismo proveedor, se utilizó el producto SENTINEL-1 SAR GRD: escala log, con una resolución de 10m, que brinda imágenes del tipo Radar de Apertura Sintética (SAR), banda C. La colección de imágenes se utiliza directamente desde la plataforma Google Earth Engine, y son provistas por Copernicus Sentinel Data (Ruiz-Ramos, J., et. al., 2020). El periodo analizado fue para los meses de enero a diciembre del año 2021, excepto en la región del NOA donde se utilizaron imágenes de junio 2021 a junio 2022 por disponibilidad de imágenes sin presencia de nubes.

Adicionalmente, se utilizó el Modelo Digital de Elevación (DEM por sus siglas en inglés) provisto por el instituto (Instituto Geográfico Nacional, s/f) con una resolución de 30m para proveer al clasificador de información auxiliar de altitud.

3.3 - Categorías de cobertura

Las categorías fueron definidas previamente a partir del conocimiento del terreno, información bibliográfica e información geoespacial disponible, así también con una perspectiva de trabajo en una

escala de 1:25.000 . Se buscó compatibilizar las clases presentes en el Catálogo de Objetos institucionales (Instituto Geográfico Nacional, 2017), que fueron relevadas en cartografía antecedente, con las clases consideradas por organismos nacionales e internacionales. Del análisis de antecedentes se concluyó que los trabajos realizados por el Instituto Nacional de Tecnología Agropecuaria (INTA), la Agencia Espacial Europea (ESA), la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO), MapBiomias y el mapa de las unidades de vegetación en la Argentina realizada por Oyarzabal et al. (2018) son las más pertinentes a considerar por su similitud con los antes mencionados Objetos del Catálogo (Oyarzabal, et al., 2018; Zanaga, D., et al., 2018; Land Cover Classification System, s/f; Volante, Jose, et al., 2009)

En base a todo lo antes mencionado, las categorías finales fueron:

- 1. Cuerpos de Agua
 - Abarca tanto los cuerpos lénticos como lóticos.
- 2. Nieve
 - Manchones de nieve estacionales como permanentes
- 3. Bosque, selva
 - Área cerrada o abierta con árboles naturales de importante densidad, de bajo, mediano o gran porte.
- 4. Estepa Arbustiva
 - Asociación de arbustos bajos, de tallos finos y con baja densidad de cobertura que puede dejar claros. Se encuentran ejemplos en Patagonia.
- 5. Tierra para cultivo
 - Terreno destinado a la siembra de cereales, hortalizas u otros vegetales que no son permanentes, y que pueden rotar de un año a otro.
- 6. Afloramiento rocoso
 - Manifestación en superficie de alguna litología de cualquier tipo y composición. Incluye macizo rocoso.
- 7. Humedal
 - Es la zona inundada con un régimen prácticamente permanente, regulada por factores climáticos. Incluye las extensiones de marismas, pantanos y turberas, o superficies cubiertas de aguas, sean éstas de régimen natural o artificial, permanentes o temporales, estancadas o corrientes, dulces, salobres o saladas, incluidas las extensiones de agua marina cuya profundidad en marea baja no exceda de seis metros.
- 8. Salina

- Zona baja arreica donde se produce acumulación de sales por evaporación (cloruros y sulfatos principalmente), constituyendo un depósito natural de sales, cuya denominación varía según el tipo de sal. Salina o salar se refiere a la acumulación de sales en solución acuosa. Salitral es el depósito natural de salitre. Boratera es una salina donde la sal predominante es borato.
- 9. Médano, duna
 - Acumulación de sedimentos sueltos, tamaño arena, que puede trasladarse por acción del viento. Pueden estar en ambiente continental o costero
- 10. Pastizal
 - Implica un área con dominio de gramíneas por presentar tallos de igual grosor y nudos salientes, hojas cintadas y envainadoras. Comúnmente se denomina estepa herbácea o pradera.
- 11. Monte
 - Superficie cubierta por hierbas, arbustos y árboles de no más de 5 metros de altura, de menor densidad de cobertura que en el caso del bosque
- 12. Planta Urbana*
 - Acumulación de sedimentos sueltos, tamaño arena, que puede trasladarse por acción del viento. Pueden estar en ambiente continental o costero
- 13. Glaciar*
 - Masa de hielo permanente y en movimiento, alimentada desde un ventisquero y que en forma de lengua ocupa un valle moviéndose cuesta abajo.

Es importante remarcar que las categorías planta urbana y glaciar no fueron categorizadas a través del algoritmo implementado, sino que se usó una máscara. En el caso de planta urbana a partir del shapefile proporcionado por la institución que se realizó a base de fotointerpretación por el departamento de Actualizaciones de Información Geoespacial perteneciente al Instituto Geográfico Nacional. En el caso de glaciar se utilizó la información vectorial provista por el Inventario Nacional de Glaciares (Inventario Nacional de Glaciares, 2018).

3.4 - Preprocesamiento de imágenes

Todo el procesamiento fue realizado en la nube, en la plataforma Google Earth Engine. Una vez seleccionados los productos y periodo de procesamiento, se realizó un filtrado de los metadatos para las imágenes Sentinel 2A/B - 1C que permite obtener píxeles con un porcentaje de nubosidad menor al 30%, ya que la presencia de nubes en imágenes que no son radar tienen impacto en el cálculo de índices y datos de reflectancia.

A partir de las imágenes seleccionadas para el periodo en cuestión (enero 2021 - diciembre 2021, y junio 2021 - junio 2022 únicamente para la región de NOA) y las distintas áreas de interés, se calculó

la media aritmética de los píxeles en el período considerado, tal como se observa en la figura 3 en el uso del algoritmo *ee.Reducer* de la plataforma de Google (figura 3). Este algoritmo calcula la media para cada píxel de todas las bandas, dentro del período seleccionado. A su vez, también se calculó de la misma manera la máxima y la mínima para cada píxel.

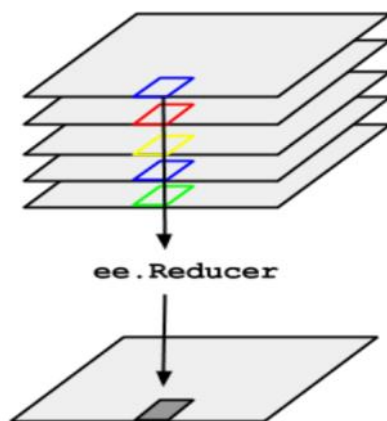


Figura 3: Esquema de procesamiento, donde *ee.Reducer* () calcula la media, mínima y máxima.

Posteriormente se seleccionaron las bandas espectrales a utilizar en la clasificación. Para el producto SENTINEL 2 se utilizaron las siguientes bandas:

- Banda 2: **Azul** (0,49 μm)
- Banda 3: **Verde** (0,56 μm)
- Banda 4: **Rojo** (0,65 μm)
- Banda 5: **Borde del rojo** (0,70 μm)
- Banda 8: Infrarrojo cercano (**NIR** por sus siglas en inglés) (0,84 μm)
- Banda 11: Infrarrojo onda corta (**SWIR** por sus siglas en inglés) (1,61 μm)

Para las imágenes SAR se utilizaron la banda de co-polarización simple, transmisión vertical/recepción vertical. Esta fue seleccionada debido a su fuerte interacción con volúmenes, ya que la vegetación, especialmente arbórea despolariza la onda, generando una respuesta de mayor intensidad (Shimada, M., et al., 2018; Pasolli, L., et al., 2015).

Los índices espectrales son herramientas fundamentales en la clasificación supervisada, ya que permiten aprovechar la información contenida en las diferentes bandas del espectro electromagnético para identificar y distinguir características específicas en una imagen. Estos índices proporcionan medidas cuantitativas de fenómenos biológicos, químicos y físicos, lo que facilita la discriminación entre las diversas categorías de cobertura. Al utilizar índices espectrales en la clasificación supervisada, podemos mejorar la precisión y la eficiencia del proceso, al tiempo que obtenemos una comprensión más profunda de la distribución de clases y características en la imagen. En este trabajo se utilizaron los siguientes:

- **NDVI** = $(\text{NIR} - \text{Rojo}) / (\text{NIR} + \text{Rojo})$

El NDVI (*Normalized Difference Vegetation Index*) está altamente asociado con la presencia de vegetación vigorosa, permitiendo su identificación gracias a su fuerte respuesta en el NIR, y absorción en el Rojo (Raj, K. J., Sivasathya, S., s/f). Por lo tanto, los valores altos de este índice estarán asociados a las categorías de bosque y cultivo, mientras que los bajos a zonas de afloramiento rocoso, arenal, agua y estepa.

- **NDWI** = $(\text{Verde} - \text{NIR}) / (\text{Verde} + \text{NIR})$

El Índice de Agua de Diferencia Normalizada (NDWI) se utiliza para resaltar el aspecto de las masas de agua en una imagen de satélite. Permitiendo que dichas masas de agua “destaquen” en la imagen (McFeeters, S. K., 1996)

- **BSI** = $[(\text{SWIR} + \text{Rojo}) - (\text{NIR} + \text{Azul})] / (\text{SWIR} + \text{Verde}) + (\text{NIR} + \text{Azul})]$

El BSI (*Bare Soil Index*) es un indicador numérico que combina bandas espectrales azules, rojas, infrarrojas cercanas e infrarrojas de onda corta para capturar las variaciones del suelo (Castelo-cabay, M., Piedra-Fernandez, J. A., Ayala, R., 2022). Este índice estará asociado a la detección de las categorías de estepa y afloramiento rocoso principalmente.

- **EVI2** = $2,4 * (\text{NIR} - \text{Rojo}) / (\text{NIR} + \text{Rojo} + 1)$

El Índice de Vegetación Mejorado (EVI2 por sus siglas en inglés) es un índice espectral utilizado para evaluar la salud y la vigorosidad de la vegetación en un área determinada. El EVI2 se basa en la relación entre las bandas del espectro electromagnético sensible a la vegetación, generalmente la banda roja (R) y la banda infrarroja cercana (NIR). El EVI2 proporciona una medida cuantitativa de la densidad y el estado de la vegetación, teniendo en cuenta tanto el contenido de clorofila como la estructura del dosel vegetal. Este índice es útil en aplicaciones de monitoreo ambiental y agrícola, ya que puede ayudar a detectar cambios en la cobertura vegetal, evaluar la salud de los cultivos y estimar la productividad de la vegetación. (Mondal, P., 2011)].

- **NDSI** = $(\text{Verde} - \text{SWIR}) / (\text{Verde} + \text{SWIR})$

El Índice Diferencial Normalizado de Nieve (NDSI) es un indicador numérico que muestra la cobertura de nieve en áreas terrestres. Las bandas espectrales de infrarrojos, verde y de onda corta (SWIR) se utilizan dentro de esta fórmula para trazar el mapa de la cubierta de nieve. Dado que la nieve absorbe la mayor parte de la radiación incidente en el SWIR mientras que las nubes no lo hacen, esto permite a NDSI distinguir la nieve de las nubes. Esta fórmula se utiliza comúnmente en la aplicación de mapas de cobertura de nieve y hielo, así como en el monitoreo de glaciares (Raster calculator, s/f).

- **SAVI** = $(\text{NIR} - \text{Rojo}) / (\text{NIR} + \text{Rojo} + 0,5) * (1 + 0,5)$

El Índice de Vegetación Ajustado al Suelo (SAVI, por sus siglas en inglés) es un índice espectral utilizado para evaluar la vegetación en áreas con cobertura vegetal densa, donde el brillo del suelo puede afectar la precisión de otros índices de vegetación. El SAVI ajusta la influencia del suelo en el

índice de vegetación, lo que lo hace especialmente útil en áreas con suelo expuesto o con vegetación muy densa. Se calcula a partir de la utilización de la banda NIR que representa el valor de reflectancia en la banda infrarroja cercana, R que representa el valor de reflectancia en la banda roja y L que es un factor de ajuste que puede variar de 0 a 1, dependiendo de las características del suelo y la vegetación del área. Generalmente se suele usar un valor de L igual a 0,5, pero en algunos casos puede ser necesario ajustar ese valor para obtener resultados más precisos dependiendo el objetivo de estudio (Landsat Soil Adjusted Vegetation Index, s/f, 2023).

- **NDMI** =
$$\frac{\text{NIR} - \text{SWIR}}{\text{NIR} + \text{SWIR}}$$

El Índice de Diferencia Normalizada de Humedad (NDMI) se utiliza para determinar el contenido de agua de la vegetación. Se calcula como una relación entre los valores NIR y SWIR de forma tradicional (Normalized Difference Moisture Index., s/f, 2023).

Para cada uno de estos índices, se calculó su media, máxima, mínima y el rango del período analizado a partir de la utilización de las bandas espectrales previamente reducidas por el algoritmo ee.Reducer() de Google Earth Engine. Además, se incorporó el modelo digital de elevación para proveer al clasificador de información sobre el relieve de la provincia. Toda esta información se agrupó en una sola imagen stack, donde cada banda espectral, índice espectral e información auxiliar se incorporó como una banda independiente dentro del stack. El resumen de todos los productos acumulados utilizados, con su nivel de procesamiento, más período analizado dentro de la imagen stack se encuentran en la Tabla 1:

Producto	Nivel de Procesamiento	Fechas
SENTINEL 2A/B	1-C	2021 excepto la Tierra del Fuego que fue junio 2021 a junio 2022
SENTINEL 2A/B	2-A	2021 excepto la Tierra del Fuego que fue junio 2021 a junio 2022
SENTINEL 1 SAR -GRD	Banda C	2021 excepto la Tierra del Fuego que fue junio 2021 a junio 2022

Bandas del espectro visible e infrarrojo		2021 excepto la Tierra del Fuego que fue junio 2021 a junio 2022
Índices Espectrales	<ul style="list-style-type: none"> • Media • Máxima • Mínima 	2021 excepto la Tierra del Fuego que fue junio 2021 a junio 2022
MDE-Ar v2		

Tabla 1: Productos, niveles de procesamiento y fechas utilizadas

3.5 - Puntos de entrenamiento y validación

A partir de la fotointerpretación de imágenes satelitales provistas por Google, así también como las imágenes SENTINEL-2 en falso color compuesto, se realizó la elección de un variado número de píxeles para cada una de las categorías, tratando de incorporar la heterogeneidad dentro de las clases como espacialmente dentro del área de estudio. Por otro lado, se utilizó la plataforma Collect Earth, donde se incorporó el shape de cartas topográficas de las distintas regiones y estableciendo un muestreo aleatorio dentro de cada una. Luego se exportó como .csv para ser editado y georeferenciado en QGIS. También se incorporó un muestreo no aleatorio y focalizado en las categorías con mayor representación en el territorio para obtener una mayor equitatividad de las mismas. Las zonas como Cuyo y la provincia de Misiones poseen un mayor número de muestras dado a la incorporación de puntos de muestreo de trabajos previos realizados en esas áreas.

A los efectos de garantizar un mínimo de puntos que resulten en una clasificación robusta y confiable, se identificaron como mínimo 10x píxeles, siendo x la cantidad de bandas utilizadas para la clasificación (Griffith, D. A., 2004). Se intentó distribuir los puntos de manera aleatoria siempre que sea posible, para reducir el riesgo de autocorrelación (High-resolution, L. C. L., Ramezan, C. A., Warner, T. A., Maxwell, A. E., Price, B. S., 2021).

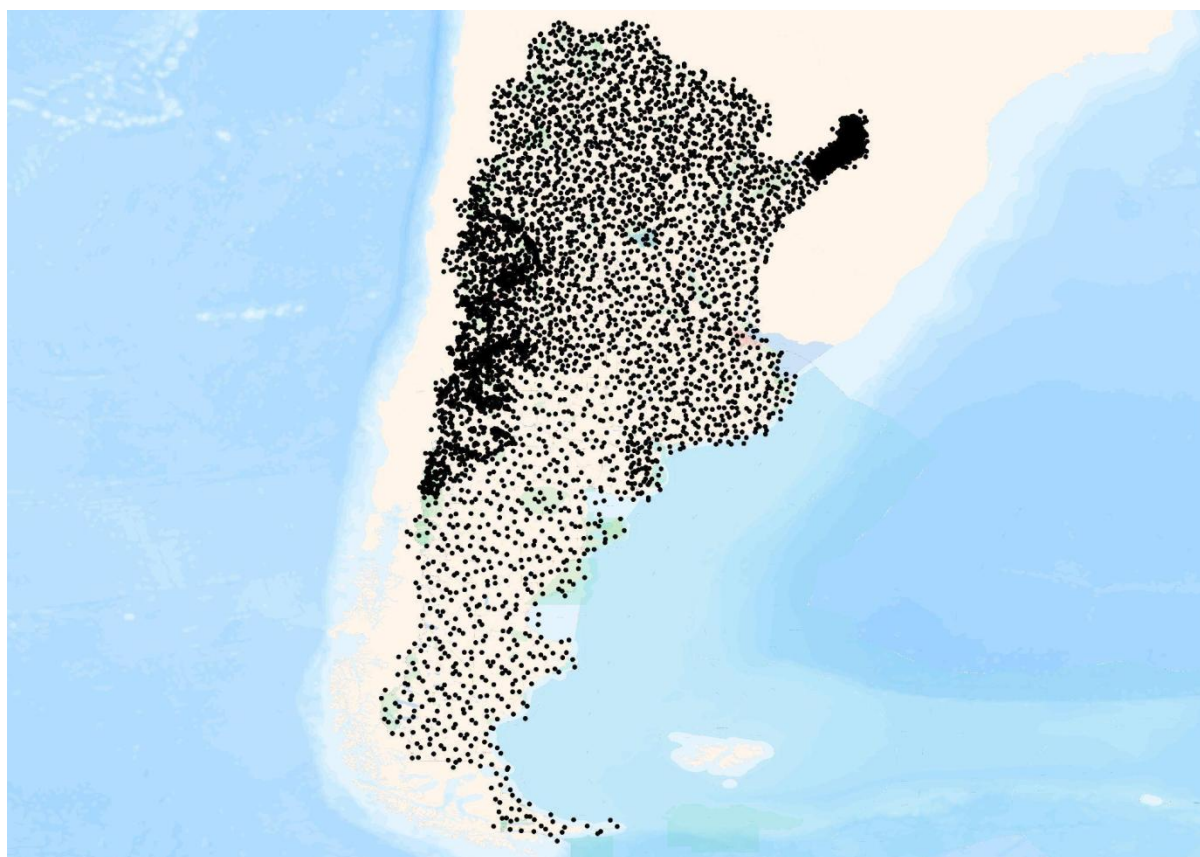


Figura 4: Distribución de los puntos de entrenamiento y validación en el área de estudio

Del total de puntos de entrenamiento y validación, se utilizó el 70% para entrenamiento del algoritmo, y el 30% para validación (Tabla 2). Todas las clases deben tener una buena proporción de datos en el grupo de entrenamiento y validación para subsanar los problemas de la división entrenamiento-validación, utilizada en la validación cruzada. Se pueden hacer sucesivas iteraciones, donde la muestra se divide aleatoriamente, y comparar cómo resultan los parámetros luego de esa serie de repeticiones (es decir, se dividen en 70/30 de manera aleatoria, pero se procesa varias veces el algoritmo, así cada división reparte los datos de manera diferente). La plataforma GEE reparte el 70% y 30% en cada iteración de manera aleatoria.

Categoría del punto	Cantidad
Cuerpos de Agua	2166
Nieve	503
Bosque, selva	2075
Estepa arbustiva	2047

Tierra para cultivo	1203
Afloramiento rocoso	935
Humedal	1212
Salina	987
Médano, duna	880
Monte	471
Pastizal	1378
Planta Urbana*	MÁSCARA
Glaciar*	MÁSCARA
TOTAL	13857

Tabla 2: Cantidad de puntos de entrenamiento y validación por categoría de cobertura del suelo.

3.6 - Algoritmo Random Forest

El Random Forest es un método de clasificación supervisada que ejecuta varios algoritmos de árboles de decisiones para clasificar un nuevo objeto basado en atributos, bandas del stack en este caso que incluye las bandas espectrales de las imágenes, las imágenes radar, los índices espectrales y el DEM. Cada árbol de decisión da una clasificación y finalmente la decisión con mayor “votos” es la predicción del algoritmo. Uno de los beneficios que más llama la atención es el de poder manejar grandes cantidades de datos con mayor dimensionalidad. Puede manejar miles de variables de entrada e identificar las variables más significativas, por lo que se considera uno de los métodos de reducción de dimensionalidad. Además, el modelo muestra la importancia de la variable, que puede ser una característica muy útil a la hora de analizar las bandas del stack que se usó en la clasificación.

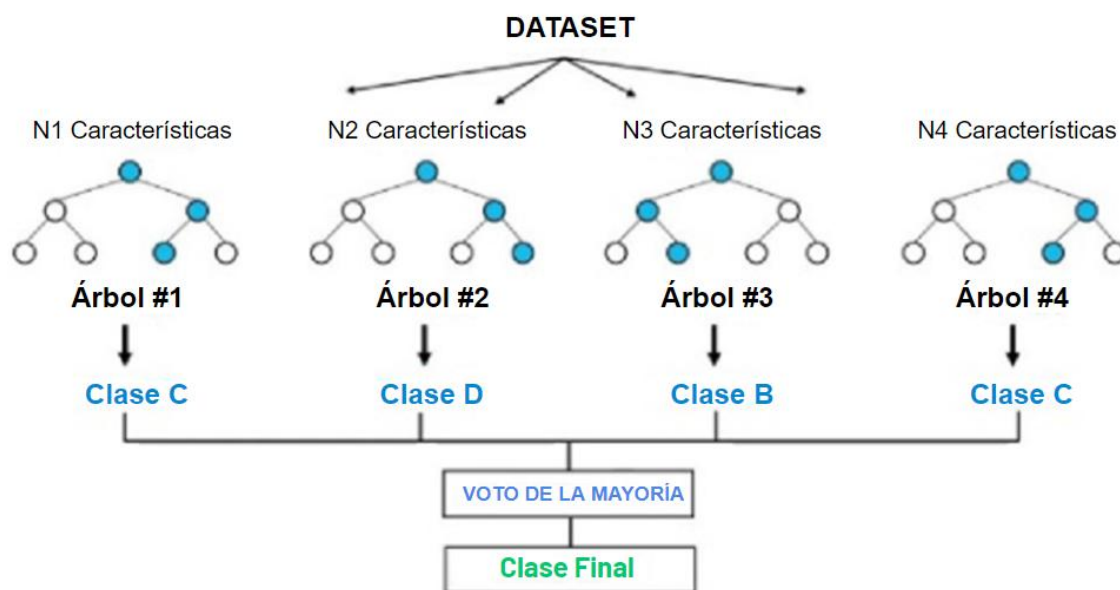


Figura 5: Representación de cómo funciona el algoritmo de clasificación supervisada Random Forest

Se utilizó esta metodología para realizar la clasificación supervisada ya que ofrece una buena respuesta con muestras reducidas, aunque si los puntos de muestreo son escasos puede tener problemas con la identificación de clases de menor peso relativo (Bialas, J., Oommen, T., Havens, T. C., 2019). Entonces, entre las características deseables del clasificador RF podemos nombrar (Breiman, L., 2001):

- Es relativamente robusto a valores atípicos y al ruido.
- Brinda información útil sobre estimaciones internas del error, la fuerza, correlación e importancia de variables.
- Es simple y fácilmente paralelizado.
- Es capaz de tomar numerosas variables de entrada y ejecutar clasificaciones rápidas, sin ser severamente afectado por el "curse of dimensionality" (Rodríguez-Galiano, V. F., Chica-Olmo, M., Abarca-Hernandez, F., Atkinson, P. M., Jeganathan, C., 2012).

3.7 - Análisis de la clasificación

Finalmente, con la zona de estudio ya clasificada, se analizó la precisión del modelo sobre los puntos de validación, a partir de los índices de precisión general, precisión del usuario, precisión del productor y el índice Kappa. A su vez, se realizó la matriz de confusión para evidenciar las precisiones del clasificador en cada categoría. Cabe destacar que estos índices de precisión y matriz son propios de cada región.

Clasificación	Observaciones a campo			Errores por inclusión	
	Agua	Vegetación	Suelo desnudo	Suma	Precisión del usuario
Clase 1	245	0	8	253	0,97
Clase 2	23	356	5	384	0,93
Clase 4	0	6	156	162	0,96
Suma	268	362	169	799	
Precisión del productor	91,42	98,34	92,31		0,95
Errores por exclusión			Precisión general		

Figura 6: Ejemplo de Matriz de Confusión, donde se observa los distintos tipos de precisión, de errores y las diferencias entre las clases analizadas.

Se obtuvieron las matrices y resultados de precisión desde el plug-in de QGIS “Semi-Automatic Classification”. Para utilizarlo, se procesaron las muestras de puntos con un buffer de 30 metros, aproximadamente el tamaño de un píxel de clasificación, para transformarlas a polígonos. Luego el plugin toma como inputs la clasificación y las muestras para corroborar la verdad terrestre con la brindada por el clasificador. Se tuvieron en cuenta dos índices de precisión, el general y el Kappa. El Índice Kappa evalúa la concordancia entre las clasificaciones realizadas por un modelo y las clases de referencia, teniendo en cuenta la concordancia esperada al azar. Es una medida útil para evaluar la confiabilidad y el acuerdo más allá de lo que se podría esperar por azar. El Índice Kappa tiene en cuenta tanto los acuerdos correctos como los desacuerdos, lo que puede ser especialmente importante cuando hay clases desbalanceadas o cuando los errores en algunas clases son más críticos que en otras. Por otro lado, la precisión general mide la proporción de clasificaciones correctas en relación con el total de observaciones clasificadas. Es una medida más simple y directa de la precisión global del modelo, sin tener en cuenta los desacuerdos entre clases individuales.

4 - Resultados

Como resultado se obtuvo un producto que representa la cobertura del suelo del territorio nacional para las 13 categorías propuestas anteriormente, para una utilización en escala 1:25.000. El tamaño del píxel es de 30m y se realizaron en las proyecciones EPSG:4326 y EPSG:3857. Se adecuaron los estilos a las categorías correspondientes y se realizaron los procesos de post-producción (compresión del archivo y creación de pirámides).

A continuación se muestran los índices de precisión, general y Kappa, para cada una de las regiones.

4.1 - Centro

→ Precisión general (%) = 81

→ Índice Kappa = 0.6440

Matriz de Confusión		Verdad Terrestre											Precisión	
		Cuerpo de agua	Bosque, selva	Estepa Arbustiva	Cultivo	Afloramiento rocoso	Humedal	Salina	Arenal	Pastizal	Monte	Total	P. Usuario	
C l a s i f i c a c i ó n	Cuerpo de agua	70	0	0	0	0	0	0	1	0	0	73	0.96	
	Bosque, selva	0	11	0	0	0	0	0	0	0	0	14	0.79	
	Estepa Arbustiva	0	0	23	1	0	3	1	0	1	0	30	0.77	
	Cultivo	5	0	1	87	0	3	0	0	12	2	110	0.79	
	Afloramiento rocoso	0	0	0	0	8	0	0	0	0	0	8	1.00	
	Humedal	0	2	0	0	0	59	1	0	9	1	72	0.82	
	Salina	0	0	0	0	0	1	23	0	0	0	24	0.96	
	Arenal	0	0	0	0	0	0	1	23	0	0	24	0.96	
	Pastizal	0	0	2	17	1	6	0	1	49	1	77	0.64	
	Monte	0	4	1	0	0	0	0	0	1	18	24	0.75	
Precisión	Total	75	17	27	106	10	72	27	24	74	24	456	-	
	P. Productor	0.93	0.65	0.85	0.82	0.80	0.82	0.85	0.96	0.66	0.75	-	0.81	

4.2 - Cuyo

→ Precisión general (%) = 77

→ Índice Kappa = 0.5661

Matriz de Confusión		Verdad terrestre											Precisión	
		Cuerpo de agua	Nieve	Estepa arbustiva	Tierra para cultivo	Afloramiento rocoso	Humedal	Salina	Médano, duna	Pastizal	Monte	Total	P. Usuario	
C l a s i f i c a c i ó n	Cuerpo de agua	108	0	0	0	0	0	0	0	16	0	124	0.87	
	Nieve	0	142	0	0	0	9	0	0	0	0	151	0.94	
	Bosque, selva	0	0	102	15	0	8	7	0	0	0	124	0.76	
	Estepa arbustiva	16	0	25	727	51	62	76	10	148	8	1142	0.64	
	Tierra para cultivo	0	0	0	0	123	0	0	0	0	4	127	0.97	
	Afloramiento rocoso	3	9	5	17	0	191	7	0	0	0	232	0.82	
	Humedal	5	0	2	0	4	0	84	7	0	0	102	0.82	
	Salina	0	0	0	0	0	0	4	155	0	0	159	0.97	
	Médano, duna	0	0	0	10	0	0	0	1	187	0	198	0.94	
	Pastizal	0	0	0	0	0	0	0	0	18	0	18	1.00	
Precisión	Total	132	151	143	783	178	270	179	173	335	42	88	2474	-
	P. Productor	0.82	0.94	0.71	0.93	0.69	0.71	0.47	0.90	0.56	0.43	0.72	-	0.77

4.3 - Noroeste Argentino (NOA)

→ Precisión general (%) = 79

→ Índice Kappa = 0.6707

Matriz de Confusión		Verdad terrestre											Precisión	
		Cuerpo de agua	Nieve	Bosque, selva	Estepa arbustiva	Tierra para cultivo	Afloramiento rocoso	Humedal	Salina	Médano, duna	Pastizal	Monte	Total	P. Usuario
C l a s i f i c a c i ó n	Cuerpo de agua	64	0	0	0	0	0	0	0	0	1	0	65	0.98
	Nieve	0	6	0	0	0	0	1	0	0	0	0	7	0.86
	Bosque, selva	20	0	60	5	2	1	1	0	0	1	6	96	0.63
	Estepa arbustiva	0	0	0	27	0	0	3	5	0	1	0	36	0.75
	Tierra para cultivo	0	0	1	1	26	0	1	1	0	0	1	30	0.87
	Afloramiento rocoso	0	0	1	9	0	36	2	1	1	0	0	50	0.72
	Humedal	0	0	0	2	0	0	13	0	0	0	0	15	0.87
	Salina	0	0	0	0	0	0	0	17	0	0	0	17	1.00
	Médano, duna	0	0	0	0	0	0	0	0	1	0	0	1	1.00
	Pastizal	0	0	0	0	0	0	0	0	0	1	0	1	1.00
Precisión	Total	84	6	64	44	28	41	22	18	3	3	23	336	-
	P. Productor	0.76	1.00	0.94	0.61	0.93	0.88	0.59	0.94	0.33	0.33	0.70	-	0.79

4.4 - Patagonia

→ Precisión general (%) = 87

→ Índice Kappa = 0.6027

Matriz de confusión		Verdad terrestre											Precisión	
		Cuerpo de agua	Nieve	Bosque, selva	Estepa arbustiva	Tierra para cultivo	Afloramiento rocoso	Humedal	Salina	Médano, duna	Pastizal	Monte	Total	P. Usuario
C l a s i f i c a c i ó n	Cuerpo de agua	169	0	0	0	0	1	0	0	0	14	0	184	0.92
	Nieve	5	16	0	0	0	1	0	0	0	1	0	23	0.70
	Bosque, selva	1	0	97	0	2	0	1	0	0	6	0	107	0.91
	Estepa arbustiva	2	0	2	141	1	9	7	11	0	3	1	177	0.80
	Tierra para cultivo	0	0	0	2	33	1	2	0	0	1	0	39	0.85
	Afloramiento rocoso	0	0	1	1	0	31	0	3	0	1	0	37	0.84
	Humedal	0	0	0	2	0	0	37	2	0	1	0	42	0.88
	Salina	3	0	0	0	0	0	0	78	2	1	0	84	0.93
	Médano, duna	0	0	0	0	0	0	0	1	41	1	0	43	0.95
	Pastizal	1	0	0	0	0	0	3	0	0	28	0	32	0.88
Monte	0	0	1	0	0	0	0	0	0	0	6	7	0.80	
Precisión	Total	181	16	101	146	37	43	50	95	43	56	7	775	-
	P. Productor	0.93	1.00	0.96	0.97	0.89	0.72	0.74	0.82	0.95	0.50	0.86	-	0.87

4.5 - Noreste Argentino (NEA)

→ Precisión general (%) = 77

→ Índice Kappa = 0.6613

Matriz de Confusión		Verdad terrestre							Precisión	
		Cuerpo de Agua	Bosque, selva	Estepa arbustiva	Tierra para cultivo	Humedal	Pastizal	Monte	Total	P. Usuario
C l a s i f i c a c i ó n	Cuerpo de Agua	37	0	0	0	0	0	0	37	1.00
	Bosque, selva	1	164	0	8	4	21	4	202	0.81
	Estepa arbustiva	0	0	0	0	0	0	0	0	0.00
	Tierra para cultivo	0	0	0	13	0	0	1	14	0.93
	Humedal	4	4	2	0	30	2	1	43	0.70
	Pastizal	0	17	0	15	2	45	0	79	0.57
	Monte	0	0	1	0	1	0	11	13	0.85
	Total	42	185	3	36	37	68	17	388	-
Precisión	P. Productor	0.88	0.89	0.00	0.36	0.81	0.66	0.65	-	0.77

4.6 - Atlántica

→ Precisión general (%) = 97

→ Índice Kappa = 0.8290

Matriz de Confusión		Verdad Terrestre				Precisión	
		Nieve	Afloramiento rocoso	Humedal	Pastizal	Total	P. Usuario
C l a s i f i c a c i ó n	Nieve	9	0	0	0	9	1.00
	Afloramiento rocoso	0	2	0	0	2	1.00
	Humedal	0	0	9	0	9	1.00
	Pastizal	0	0	1	16	17	0.94
Precisión	Total	9	2	10	16	37	-
	P. Productor	1.00	1.00	0.90	1.00	-	0.97

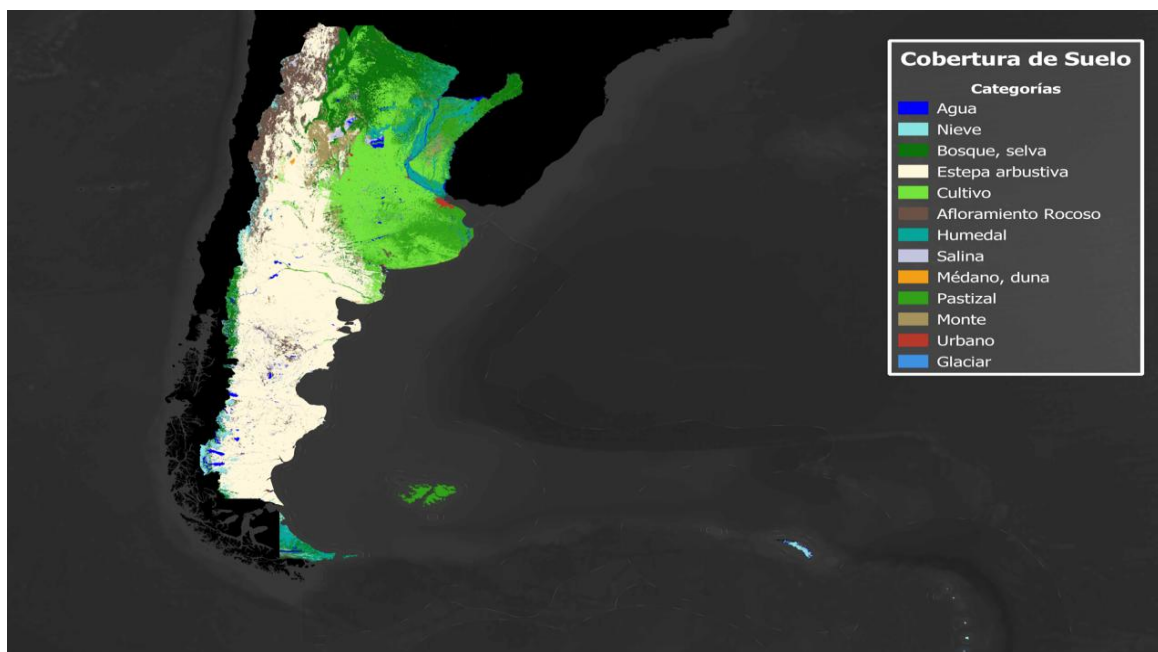


Figura 7: Resultado final de cobertura del suelo de la República Argentina

5 - Conclusiones

La clasificación de cobertura de suelo para el territorio Argentino constituye un insumo nacional de calidad para su futura implementación en análisis temporales y espaciales de las mismas. El contar con un producto nacional y de libre acceso, brinda la posibilidad de generar una inmensa cantidad de análisis o subproductos posteriores tales como ordenamientos territoriales, gestión de riesgos, gestión ambiental, entre muchos otros.

Durante el trabajo se detectaron ciertas oportunidades de mejora, tales como, mejorar el tamaño de píxel para trabajar en menores escalas, la toma de muestras a campo para la clasificación, sub-categorías en las clases y la unión entre regiones para que la continuidad sea más fluida.

6 - Referencias

- S. Talukdar et al., “Land-use land-cover classification by machine learning classifiers for satellite observations-A review,” *Remote Sens.*, vol. 12, no. 7, 2020, doi: 10.3390/rs12071135.
- Y. A. Liou, A. K. Nguyen, and M. H. Li, “Assessing spatiotemporal eco-environmental vulnerability by Landsat data,” *Ecol. Indic.*, vol. 80, no. May, pp. 52–65, 2017, doi: 10.1016/j.ecolind.2017.04.055.
- K. A. Nguyen and Y. A. Liou, “Global mapping of eco-environmental vulnerability from human and nature disturbances,” *Sci. Total Environ.*, vol. 664, pp. 995–1004, 2019, doi: 10.1016/j.scitotenv.2019.01.407.
- UN-GGIM, UN-GGIM: Américas. Estatutos, Comité Regional de Naciones Unidas Sobre Gestión Global de la Información Geoespacial para las Américas (UN-GGIM: Américas). 2019, pp. 1–12.
- Instituto Geográfico Nacional, “DETERMINACIÓN DE LA SUPERFICIE CORRESPONDIENTE AL TERRITORIO CONTINENTAL, ANTÁRTICO E INSULAR DE LA REPÚBLICA ARGENTINA,” 2022. [Online]. Available: <https://www.ign.gob.ar/NuestrasActividades/InformacionGeoespacial/UnidadesTerritoriales/DocumentacionTecnica/>.
- Instituto Geográfico Nacional, “Descripción del Diccionario de Conceptos y Catálogo de Objetos Geográficos del Instituto Geográfico Nacional República Argentina,” 2017.
- Waller, Tomas & Micucci, Patricio & Draque, Juan & Barros, Mariano & Estavillo, Candelaria. (2010). CONSERVACIÓN DE LA BOA AMPALAGUA (*Boa constrictor occidentalis*) EN LA REPÚBLICA ARGENTINA.
- Ruiz-Ramos, J., Marino, A., Boardman, C., & Suarez, J. (2020). Continuous Forest Monitoring Using Cumulative Sums of Sentinel-1 Timeseries. *Remote Sensing*, 12(18), 3061. MDPI AG. Retrieved from <http://dx.doi.org/10.3390/rs12183061>
- MDE-Ar | Instituto Geográfico Nacional. (s. f.). <https://www.ign.gob.ar/category/tem%C3%A1tica/geodesia/mde-ar>
- Oyarzabal, M., Clavijo, J., Oakley, L., Biganzoli, F., Tognetti, P., Barberis, I., Maturo, H. M., Aragón, R., Campanello, P. I., Prado, D., Oesterheld, M., & León, R. J. (2018). Unidades de vegetación de la Argentina. *Ecología Austral*, 28(1), 040–063. <https://doi.org/10.25260/EA.18.28.1.0.399>
- Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., Lesiv, M., Herold, M., Tsensbazar, N.E., Xu, P., Ramoino, F., Arino, O., 2022. ESA WorldCover 10 m 2021 v200. <https://doi.org/10.5281/zenodo.7254221>
- LAND COVER CLASSIFICATION SYSTEM. (s/f). Fao.org. Recuperado el 23 de febrero de 2023, de <https://www.fao.org/3/x0596e/x0596e00.htm>

Volante, José & Collado, Alfredo & Ferreyra, Eduardo & López, Carlos & Navarro Rau, Maria & Pezzola, Alejandro & Puentes, María & Olmedo, Guillermo & al., (2009). Monitoreo de la cobertura y el uso del suelo a partir de sensores remotos.

IANIGLA - Inventario Nacional de Glaciares. 2018. Resumen ejecutivo de los resultados del Inventario Nacional de Glaciares. IANIGLA-CONICET, Ministerio de Ambiente y Desarrollo Sustentable de la Nación. Pp. 27.

M. Shimada, T. Itoh, T. Motooka, M. Watanabe, and T. Shiraishi, "New global forest / non-forest maps from ALOS PALSAR data (2007 – 2010) Remote Sensing of Environment New global forest / non-forest maps from ALOS PALSAR data (2007 – 2010)," Remote Sens. Environ., no. February 2018, 2014, doi: 10.1016/j.rse.2014.04.014.

L. Pasolli et al., "Estimation of soil moisture in mountain areas using SVR technique applied to multiscale active radar images at C-band," IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., vol. 8, no. 1, pp. 262–283, 2015, doi: 10.1109/JSTARS.2014.2378795.

K. J. Raj and S. Sivasathya, "SVM and Random Forest Classification of Satellite Image with NDVI as an Additional Attribute to the Dataset," doi: 10.1007/978-81-322-1771-8.

S. K. McFEETERS (1996) The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features, International Journal of Remote Sensing, 17:7, 1425-1432, DOI: 10.1080/01431169608948714

M. Castelo-cabay, J. A. Piedra-fernandez, and R. Ayala, "Deep learning for land use and land cover classification from the Ecuadorian Paramo .," 2022, doi: 10.1080/17538947.2022.2088872.

Mondal, P. (2011). Quantifying surface gradients with a 2-band Enhanced Vegetation Index (EVI2). Ecological Indicators, 11(3), 918–924. <https://doi.org/10.1016/j.ecolind.2010.10.006>

Raster calculator. (s/f). Bluemarblegeo.com. Recuperado el 23 de febrero de 2023, de https://www.bluemarblegeo.com/knowledgebase/global-mapper-19/Raster_Calculator.htm

Landsat Soil Adjusted Vegetation Index. (s/f). Usgs.gov. Recuperado el 23 de febrero de 2023, de <https://www.usgs.gov/landsat-missions/landsat-soil-adjusted-vegetation-index>

Normalized Difference Moisture Index. (s/f). Usgs.gov recuperado el 28 de febrero de 2023, de [Normalized Difference Moisture Index | U.S. Geological Survey \(usgs.gov\)](https://www.usgs.gov/landsat-missions/landsat-soil-adjusted-vegetation-index)

D. A. Griffith, "Effective Geographic Sample Size in the Presence of Spatial Autocorrelation," vol. 95, no. December 2004, pp. 740–760, 2005.

L. C. L. High-resolution, C. A. Ramezan, T. A. Warner, A. E. Maxwell, and B. S. Price, "Effects of Training Set Size on Supervised Machine-Learning Remotely Sensed Data," Remote Sens., pp. 1–27, 2021.

J. Bialas, T. Oommen, and T. C. Havens, "Int J Appl Earth Obs Geoinformation Optimal segmentation of high spatial resolution images for the classification of buildings using random

forests,” *Int J Appl Earth Obs Geoinf.*, vol. 82, no. June, p. 101895, 2019, doi: 10.1016/j.jag.2019.06.005.

Breiman, L. (2001). Random Forest. *Machine learning*, 45(1), 5–32.
<https://doi.org/10.1023/a:1010933404324>

Rodriguez-Galiano, V. F., Chica-Olmo, M., Abarca-Hernandez, F., Atkinson, P. M., & Jeganathan, C. (2012). Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. *Remote Sensing of Environment*, 121, 93–107.
<https://doi.org/10.1016/j.rse.2011.12.003>