

## **Análisis de características orientado a desarrollar un modelo automatizado para la identificación de montes en el norte de Entre Ríos**

Silvana Bulatovich<sup>1</sup>, Alejandro Hadad<sup>1</sup>, Alejandra Kemerer<sup>2,3</sup>

<sup>1</sup> Facultad de Ingeniería, UNER, Ruta 11 km 10.5,

<sup>2</sup> Facultad de Ciencias Agropecuarias, UNER, Ruta 11 km 10.5

<sup>3</sup> Estación Experimental Agropecuaria Paraná del INTA, Ruta 11 km 11.5

Oro Verde, Argentina

silvana.bulatovich@gmail.com

**Resumen.** En el presente trabajo se realiza un análisis exploratorio de características, tales como varianza; entropía y distintos índices espectrales reportados en la bibliografía como relacionados con la vegetación, de las distintas coberturas de monte presentes en predios ganaderos en los departamentos de Feliciano y La Paz, Entre Ríos mediante la utilización de imágenes satelitales provenientes del satélite Landsat 8 con el fin de en un futuro, construir un modelo que permita automatizar el proceso. La clasificación de las distintas coberturas del suelo, brinda información básica para la planificación, ordenación y gestión del uso del mismo.

### **1 Introducción**

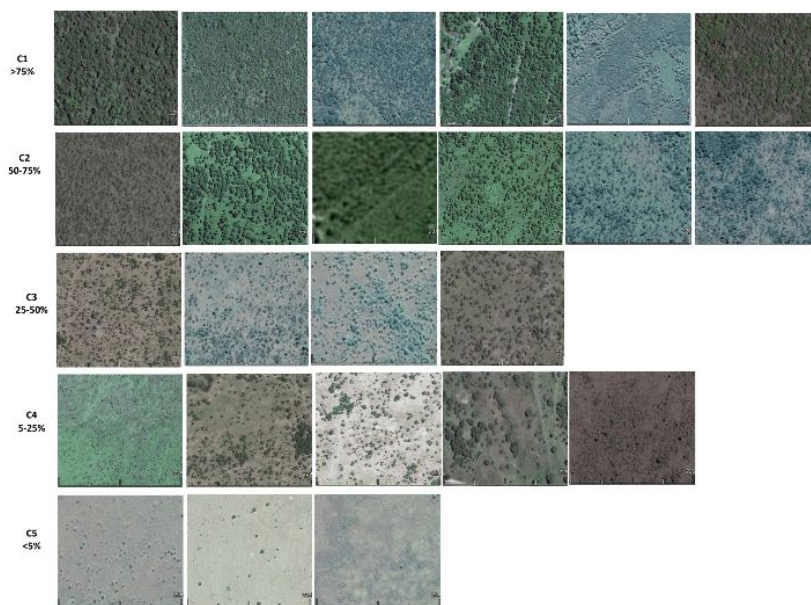
El monte nativo de la provincia de Entre Ríos reviste gran importancia desde el punto de vista socio-económico vinculado a la ganadería de cría y también por los numerosos servicios ecosistémicos que brinda. Por tal motivo, caracterizar y monitorear el estado y evolución de los montes resulta clave para garantizar la sustentabilidad de los mismos [1]. La provincia dispone de cartografía de las áreas de monte generadas solo por interpretación visual, lo que dificulta su uso por los usuarios para evaluar los cambios en el tiempo [2,3]. La posibilidad de contar con modelos numéricos-dinámicos a partir de información espectral permitiría caracterizar el tipo y nivel de degradación de áreas de monte.

La evaluación de la condición/estado de montes en el área de producción ganadera de cría normalmente se realiza mediante recorrido e inspección visual de las áreas medianamente accesibles, y presenta un carácter netamente subjetivo. En cuanto a la incorporación de tecnología, las aproximaciones a través de imágenes satelitales generalmente está limitada a la identificación de la presencia de montes y a la mensura del desmonte, con una fuerte interpretación visual, estática (en un instante de tiempo) y subjetiva (del observador). Esta situación plantea la necesidad de analizar/extraer indicadores que permitan evaluar la presencia y el estado del monte

de manera automatizada mediante el uso de imágenes satelitales. A tal fin se realizará un análisis de variables derivadas de imágenes satelitales Landsat 8[4] orientadas a desarrollar un modelo automatizado para la identificación de la presencia de montes y su condición.

## 2 Materiales y Métodos

Se evaluaron 13 establecimientos ganaderos de cría del Departamento Feliciano y 11 del Departamento La Paz, Entre Ríos. La superficie promedio de los establecimientos evaluados fue de 274 has, con una gran variabilidad, presentando el de menor superficie 8 has y el de mayor superficie 1139 has. En los establecimientos evaluados se delimitaron los ambientes presentes en base al porcentaje de cobertura arbórea generando archivos Shapefile. Se definieron cinco zonas o ambientes, desde Z1 con más de 75% de cobertura arbórea a Z5 con menos de 5% de cobertura arbórea. La identificación de los ambientes se realizó por interpretación visual a partir de Google Earth a una escala fija habiendo definido previamente los criterios de interpretación y generado un patrón comparativo (Figura 1), los que fueron corroborados a través un relevamiento presencial a campo determinando visualmente la proporción de cobertura del área evaluada. Asimismo, se relevaron seis puntos por establecimiento realizando un censo de los individuos presentes en un área de 90 m<sup>2</sup> y determinando la estructura de la vegetación, pero esta información será utilizada para validar el modelo definitivo en una etapa posterior.



**Figura 1.** Patrones para definir los ambientes en función del porcentaje de cobertura arbórea en base a Google Earth.

## 2.1 Selección y recorte de imágenes.

A fin de cubrir las variaciones que pueden generarse por estacionalidad y condiciones ambientales como precipitaciones, su distribución o cambios en las temperaturas, en las coberturas de monte se descargaron del portal de datos “EarthExplorer”[5] disponible en El Servicio Geológico de los Estados Unidos (USGS) un total de 180 imágenes del Satélite Landsat 8 tomadas con el sensor “Operational Land Imager” (OLI), con sus correspondientes bandas, de los años 2016 y 2017 . Los Path/ Row utilizados fueron 225/81 para cubrir los establecimientos del Departamento de Feliciano, 226/81 y 226/82 para cubrir los establecimientos del Departamento de La Paz.

Sobre las imágenes crudas se realizó la corrección atmosférica DOS1[6] y la transformación de escala digital a unidades físicas. Dicho preprocesamiento se realizó mediante el Sistema de Información Geográfica QGIS (QGIS 2018)[7].

Se procedió a recortar todas las imágenes utilizando los archivos Shapefile de los ambientes delimitados mediante el software Qgis, con un script de bash para automatizar el proceso.

## 2.2 Detección de nubes.

Para detectar las nubes se utilizó la banda de evaluación de calidad (BQA) provista por el USGS y se guardaron todos los resultados para poder utilizar diferentes umbrales de presencia de nubes por píxel de considerarse necesario.

La banda de 16 bits de la Evaluación de calidad Landsat Collection 1 Level-1 (QA) permite a los usuarios aplicar filtros por píxel a todos los productos de datos Landsat Collection 1 Level-1. Cada píxel en la banda de control de calidad contiene enteros sin signo que representan combinaciones de bits de condiciones de superficie, atmosféricas y de sensor que pueden afectar la utilidad general de un píxel determinado. Las bandas de QA Landsat Collection 1 Level-1 (.TIF) se incluyen en el producto Landsat Level-1 GeoTIFF Data descargado de EarthExplorer. Las descripciones de cada bit de QA y los valores de píxeles se muestran en la Figura 2.

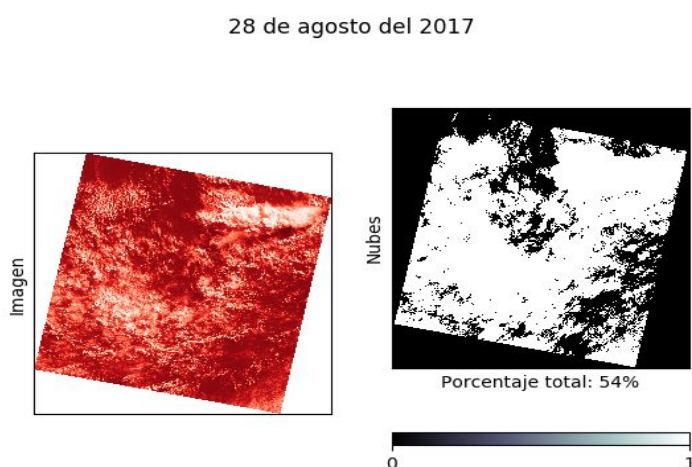
Landsat 8 OLI, OLI/TIRS Pre-Collection QA band bits: Read from RIGHT to LEFT, starting with Bit 0																
BIT	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Description	Cloud Confidence		Cirrus Confidence		Snow/Ice Confidence		Reserved (Unused)		Reserved (Unused)		Water Confidence		Reserved (Unused)	Terrain Occlusion	Dropped Frame	Designated Fill

**Figura 2.** Descripción de la información contenida en los píxeles de la banda de calidad de Landsat 8 16 bits.

Usados de manera efectiva, los bits de control de calidad ayudan a mejorar la integridad de las investigaciones científicas derivadas de Landsat al indicar qué píxeles podrían verse afectados por las condiciones de la superficie, la contaminación

de la nube o las condiciones de los sensores. Por ejemplo, NDVI calculado sobre píxeles que contienen nubes mostrará valores anómalos. Si dichos píxeles se incluyeran en un estudio de fenología, los resultados podrían no mostrar las verdaderas características de la superficie del crecimiento estacional de la vegetación. Los píxeles contaminados por las nubes reducirán los valores de NDVI, y las medidas como el tiempo de "verde" o la madurez máxima aparecerán más tarde de lo que realmente ocurrió.

Para el presente trabajo se tuvieron en cuenta los píxeles de nubes, sombras de nubes y cirrus, haciendo un promedio del porcentaje de los mismos por píxel para poder utilizarlo en las imágenes recortadas y en las enteras (Figura 3).



**Figura 3.** Representación del porcentaje de nubes para el 28/08/2017 para el Path 226 y Row 81.

### 2.3 Extracción de características

Previo al proceso de extracción de características se analizaron los recortes marcados para determinar qué tamaño de ventanas es conveniente utilizar para posteriores procesamientos en función de la cantidad de muestras que se dispone para cada tipo de cobertura. En la Tabla 1 se observa el significativo decrecimiento en el número de muestras disponibles a medida que aumenta el tamaño de ventana. En la Tabla 2 además se observa como las coberturas con mayor número de muestras aumentan su participación relativa a las coberturas con menor cantidad de muestras. Teniendo en cuenta estos aspectos los procesos de extracción de características que se describen a continuación se realizan sobre las muestras obtenidas a partir de ventanas de 3 por 3 píxeles.

**Tabla 1.** Número de muestras por tipo de cobertura en función del tamaño de ventana utilizado para recorrer las imágenes.

cobertura/ventana	<b>3x3</b>	<b>5x5</b>	<b>7x7</b>	<b>9x9</b>
<b>z1</b>	391	191	77	16
<b>z2</b>	4728	3492	2583	1882
<b>z3</b>	12748	10399	8384	6579
<b>z4</b>	15768	12905	10466	8595
<b>z5</b>	7396	5643	4255	3127
Total	<b>41031</b>	<b>32630</b>	<b>25765</b>	<b>20199</b>

**Tabla 2.** Porcentaje de participación de cada tipo de cobertura para los distintos tamaños de ventana.

cobertura/ventana	<b>3x3</b>	<b>5x5</b>	<b>7x7</b>	<b>9x9</b>
<b>z1</b>	0,95%	0,59%	0,30%	0,08%
<b>z2</b>	11,52%	10,70%	10,03%	9,32%
<b>z3</b>	31,07%	31,87%	32,54%	32,57%
<b>z4</b>	38,43%	39,55%	40,62%	42,55%
<b>z5</b>	18,03%	17,29%	16,51%	15,48%

Se calcularon distintos índices espectrales para los dos años considerados y comúnmente asociados a describir los distintos tipos de vegetación [9,10,11,12]. Luego a fin de disponer de una primera descripción desde el punto de vista estadístico se calcularon la varianza y la entropía para cada una de las bandas del espectro de la imagen Landsat. Dichos valores juntos con los índices espectrales antes mencionados conforman el conjunto inicial de descriptores/características.

Índice de vegetación de diferencia normalizada. [9]

$$NDVI = \frac{\text{infrarrojo Cercano} - \text{Rojo}}{\text{Infrarrojo Cercano} + \text{Rojo}}$$

Índice de vegetación de diferencia normalizada verde. [10]

$$GNDVI = \frac{\text{infrarrojo Cercano} - \text{Verde}}{\text{Infrarrojo Cercano} + \text{Verde}}$$

Índice de vegetación de amplio rango dinámico. [11]

$$NDVIm = \frac{0.62 * \text{infrarrojo Cercano} - \text{Rojo}}{0.62 * \text{Infrarrojo Cercano} + \text{Rojo}}$$

Generador de imágenes multiespectral. [12]

$$MSI = \frac{\text{infrarrojo Cercano} - \text{SWIR}}{\text{Infrarrojo Cercano} + \text{SWIR}}$$

Finalmente y a fin de evaluar la separabilidad de las zonas definidas en base a las características obtenidas hasta el momento, se realizó una primera clasificación con Random Forest[13] utilizando como variables de entrada el valor del píxel, la varianza, la entropía, el NDVI, el GNDVI, el NDVIm y el MSI. Para evaluar el clasificador se utilizaron los píxeles previamente marcados y la técnica de validación cruzada[14] promediando los resultados de 10 predicciones con distintas particiones de los datos (prueba y entrenamiento) hechas al azar para evaluar los resultados, garantizando que son independientes de la partición entre datos de entrenamiento y prueba.

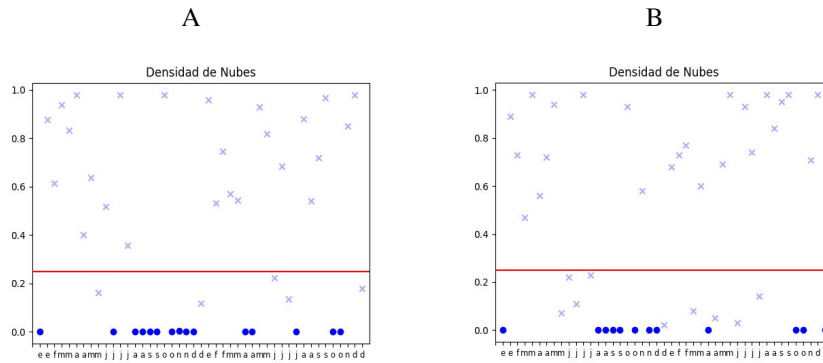
### 3 Resultados.

La proporción de monte en los establecimientos evaluados respecto a la superficie total en promedio representó el 89%, con un mínimo de 63% y el máximo la totalidad de la superficie bajo monte. La proporción de cada zona de cobertura arbórea definida en la superficie bajo monte fue variable, pero se observó una menor proporción de área de escasa cobertura arbórea (Z5) en los establecimientos de Feliciano respecto a La Paz (Tabla 3). Respecto a las zonas con mayor cobertura arbórea (Z1), los valores máximos más altos se observaron en La Paz, asociado a la presencia de selva en galería rodeando los cursos de agua, ambiente más escaso en Feliciano. Las zonas de coberturas intermedias estuvieron presentes en todos los casos y en proporciones variables. No se pudo establecer un patrón de gradiente, sino que en los diferentes establecimientos las zonas estuvieron presentes en proporciones variables. En la Tabla 3 no se indican los valores mínimos ya que para las diferentes zonas se presentaron casos de 0% en ambas regiones.

**Tabla 3.** Porcentaje de cada zona de cobertura arbórea presente en los establecimientos ganaderos de cría de Feliciano y La Paz, Entre Ríos.

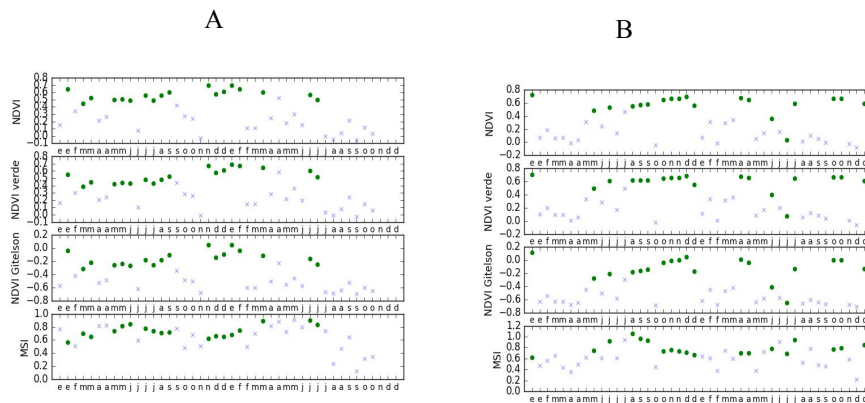
Zonas de cobertura arbórea	Feliciano		La Paz	
	% Promedio	% Máximo	% Promedio	% Máximo
Z1	20.7	55.3	14	100
Z2	35.8	72.1	12.7	100
Z3	17.3	77.9	8.6	53.8
Z4	21.1	100	35.8	100
Z5	5.2	42.5	24.7	67.1

De las 180 imágenes descargadas, solo 27 presentaron un umbral de 0% de nubes en la imagen completa y 41 un umbral inferior al 30% de nubes como se muestra en la Figura 4.



**Figura 4.** Densidad de nubes para el Path 226 y Row 81 (a), y para Path 226 y Row 82 (b), en los años 2016 y 2017 las cruces en celeste representan las fechas para las que los valores de nubes son mayores a 0% y la línea roja representa un umbral de corte de referencia al 30%.

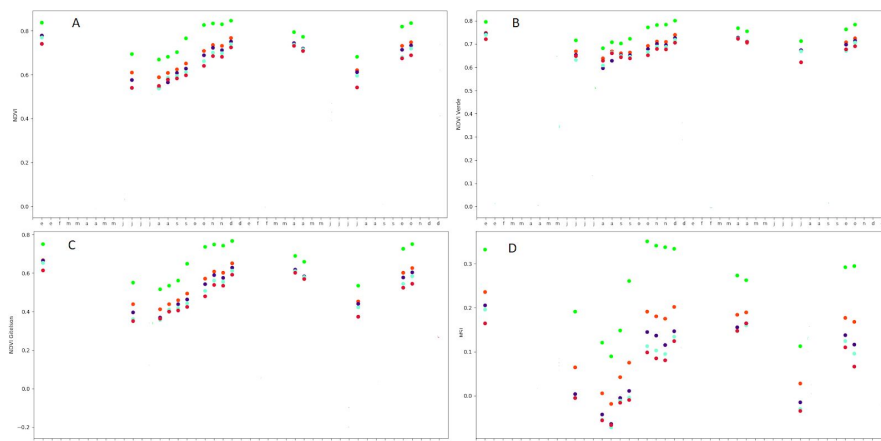
Los índices espectrales evaluados presentaron un comportamiento similar en las diferentes fechas. Debido a la numerosa cantidad de imágenes con nubes no fue posible identificar un patrón de estacionalidad en el período evaluado para ninguno de los índices considerados (Figura 5a y b).



**Figura 5.** Valor promedio para todas las zonas evaluadas de los índices NDVI, GNDVI, NDVI<sub>m</sub> y MSI en los años 2014-2015 (a) y 2016-2017 (b). Los puntos grises representan fechas para las que las imágenes tienen una alta cobertura de nubes y los datos no son confiables.

### 3.1 Índices por zona.

Al evaluar los índices por zona, el NDVI fue el que mostró la mayor capacidad para discriminar, al menos entre las zonas con porcentajes de coberturas arbórea contrastante (z1 vs z5). Las zonas con mayor cobertura arbórea (z1) presentaron valores más elevados de NDVI que las zonas con baja cobertura arbórea (z5), en los distintos momentos del año. Las zonas de coberturas intermedias resultaron más difíciles de discriminar en base al NDVI (Figura 6a). Los índices GNDVI y NDVI<sub>m</sub> presentaron un comportamiento variable y permitieron separar en menor medida los distintos tipos de zonas, con mezclas en los valores medios en la mayoría de las fechas evaluadas (Figura 6b y c). El MSI permitió separar las zonas contrastantes en la mayor parte de las fechas, pero los valores presentaron una marcada variabilidad entre fechas. Las zonas de menor cobertura presentaron mayores valores mezclados en comparación con los otros índices (Figura 6d).



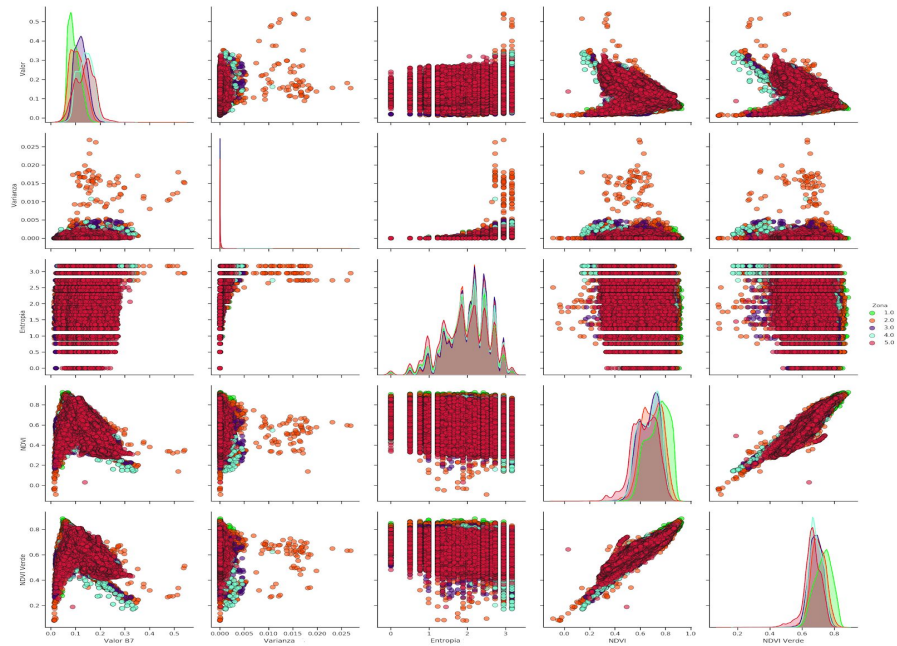
**Figura 6.** Evolución temporal en los años 2016 y 2017 para el Path 226 y Row 81 de los índices a) NDVI, b) NDVI verde, c) NDVI<sub>m</sub> y d) MSI para las Zonas 1 a 5 definidas en función del porcentaje de cobertura arbórea. (Zonas: 1) Verde, 2) Naranja, 3) Violeta, 4) Celeste, 5) Rosada).

### 3.2 Varianza y entropía.

Mediante representaciones gráficas de la varianza y la entropía fue posible caracterizar las zonas con diferente cobertura arbórea con mayor claridad respecto al uso de los índices espectrales. El comportamiento entre bandas difirió levemente, en todos los casos la zona 1 presentó valores de entropía y varianza mayores a los de las restantes zonas. Para el caso de las bandas 1 y 2 se observó mayor variabilidad en los valores de entropía de la zona 1 respecto a las demás bandas, mientras que la varianza presentó un rango amplio para todas las bandas en la zona 1.

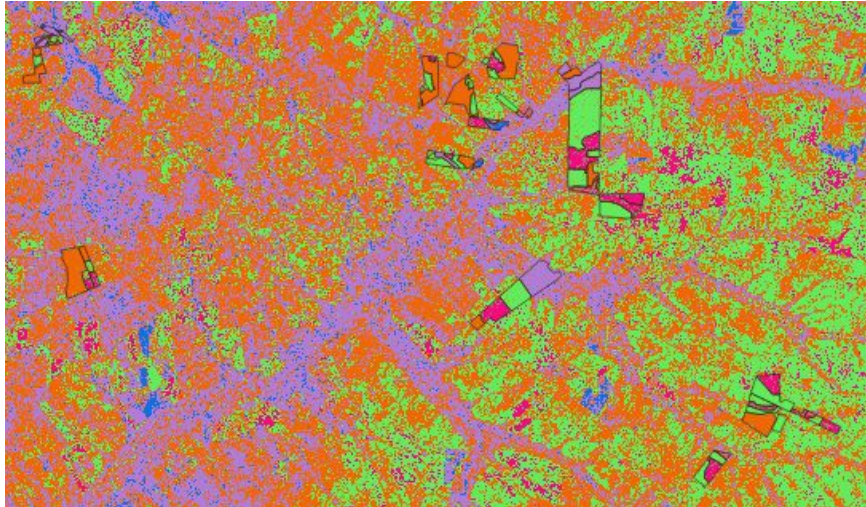
Se presenta un resumen de las mencionadas gráficas (Figura 7), donde se puede ver gráficamente que la separación entre zonas más disímiles es más fácil de realizar.



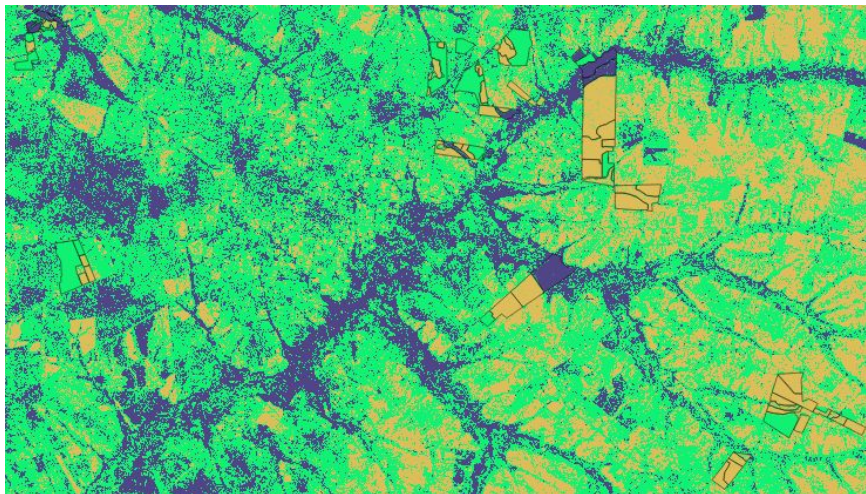


**Figura 7.** Valor de cada pixel, varianza, entropía NDVI y GNDVI de todas las imágenes de 2016 y 2017 para el Path 226 y Row 81 y 82 libres de nubes. (Zonas: 1) Verde, 2) Naranja, 3) Violeta, 4) Celeste, 5) Rosada)

Como se mencionó previamente y a modo exploratorio se realizaron algunos ensayos de segmentación de las imágenes solo a los efectos de visualizar y representar el tipo de resultado que se espera obtener. La clasificación considerando las cinco zonas de cobertura arbórea presentó una confiabilidad de 61%. Como era de esperar en base a lo observado previamente, las zonas con porcentajes de cobertura arbórea intermedia fueron más difícil de discriminar y en general se observó una importante proporción de mezcla entre ellas (Figura 8a). Para evaluar si se lograban mejores resultados se agruparon las zonas 1 y 2 y las zonas 4 y 5 intentando separar en tres zonas en vez de cinco, por considerar que el gradiente de corte es subjetivo lo que permitió mejorar el resultado global de la clasificación a una confiabilidad de 79% (Figura 8b).



**Figura 8a.** Representación de la clasificación de los cinco tipos de cobertura en distintos colores con los recortes con transparencia marcados por observación a terreno para un recorte del del Path 226 y Row 81.



**Figura 8b.** Representación de la clasificación de los tres tipos de cobertura en distintos colores con los recortes con transparencia marcados por observación a terreno para un recorte del del Path 226 y Row 81.

#### 4 Conclusiones.

A fin de caracterizar la condición/estado de montes en el área de producción ganadera de cría de la provincia de Entre Ríos se demarcaron, etiquetaron y evaluaron cinco niveles de cobertura arbórea a partir índices espectrales y estadísticas básicas de las bandas espectrales de Landsat. Las coberturas extremas z1 y z5 se pudieron diferenciar con claridad pero las coberturas intermedias z2, z3 y z4 mostraron valores comunes, en especial con la z5. En trabajos posteriores se profundizará el análisis sobre el conjunto de datos recientemente construido y se evaluarán diferentes alternativas para mejorar los resultados de la clasificación.

#### Referencias

- [1] Sabattini, R.; Sione, S.; Ledesma, S.; Sabattini, J.; Wilson, M. 2016. Estimación de la pérdida de superficie de bosques nativos y tasa de deforestación en la cuenca del Arroyo Estacas (Entre Ríos, Argentina). *Revista Científica Agropecuaria* 20(1-2): 45-56
- [2] Sabattini, R.; Ledesma, S.; Brizuela, A.; Sabattini, J.; Fontana, E.; Diez, J.; Muracciole, B. 2009. INFORME I. Zonificación de los bosques nativos en el Departamento La Paz (Entre Ríos) según las categorías de conservación. Presentado a Dirección General de Recursos Naturales de la Secretaría de la Producción del Gobierno de Entre Ríos. 21 pp y 1 Anexo con Imágenes.
- [3] Sabattini, R.; Ledesma, S.; Sione, S.; Sabattini, J. 2013. Ordenamiento territorial de bosques nativos en la cuenca del Arroyo Feliciano (Entre Ríos) (p. 313-331). En: Bases conceptuales y metodológicas para el Ordenamiento Territorial en el Medio Rural. Región Centro Argentina (Giayetto, O. Plevich J, Lallana, V y M Pilatti, comp.). 672 p.
- [4] Landsat 8. NASA. (2019). <https://landsat.gsfc.nasa.gov/landsat-8/landsat-8-bands/>
- [5] Earth Explorer. USGS. (2019). <https://earthexplorer.usgs.gov/>
- [6] QGIS Development Team. (2016). QGIS geographic information system. *Open Source Geospatial Foundation Project*.
- [7] H. Ding, J. Shi, Y. Wang, L. Wei. 2015. An improved dark-object subtraction method for atmospheric correction of Landsat 8
- [8] Banda de Calidad(BQA). USGS. (2019). <https://landsat.usgs.gov/collectionqualityband>
- [9] Rouse, J.; Haas, R.; Schell, J.; Deerin, D. 1973. Monitoring vegetation systems in the great plains with ERTS.N.SP-351. Ed Third ERTS Symposium. 1:309-317. Washington. NASA.
- [10] Gitelson, A.; Kaufman, Merzlyak, M. 1996. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment* 58:289-298.
- [11] Gitelson A. 2004. Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. *Journal of Plant Physiology* 161:165-173.
- [12] Hawkins, S. E., III, et al. 1997. Multi-spectral imager on the Near Earth Asteroid Rendezvous
- [13] Breiman, L. 2001. Random Forest.
- [14] Kohavi, Ron (1995). "A study of cross-validation and bootstrap for accuracy estimation and model selection". *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*. San Mateo, CA: Morgan Kaufmann. 2 (12): 1137-1143.