

# Obtención del mapa de cobertura de la zona del Páramo de Sumapaz

## Obtaining a land coverage map of Páramo de Sumapaz area

DIANA ROCÍO GALINDO<sup>1,a</sup>, CAMPO ELÍAS PARDO<sup>1,b</sup>, ORLANDO RIAÑO-MELO<sup>2,c</sup>

<sup>1</sup>DEPARTAMENTO DE ESTADÍSTICA, FACULTAD DE CIENCIAS, UNIVERSIDAD NACIONAL DE COLOMBIA, BOGOTÁ, COLOMBIA

<sup>2</sup>FACULTAD DEL MEDIO AMBIENTE Y RECURSOS NATURALES, UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS, BOGOTÁ, COLOMBIA

---

### Resumen

Se muestra una aplicación del uso de métodos multivariados para obtener información espectral de la zona del Páramo de Sumapaz. El procedimiento comprende dos etapas: transformación de la información espectral redundante por medio de un análisis de componentes principales y una clasificación sobre las coordenadas principales obtenidas. Una vez interpretados visualmente cada uno de los grupos resultantes, se definieron tanto geográfica como espectralmente las clases de coberturas terrestres presentes en la zona de estudio. La base de datos corresponde a una subescena de una imagen del satélite Landsat 5 TM tomada el 14 de febrero de 1997 con path 8 - row 57, que cubre parte del Páramo de Sumapaz en formato GeoTIFF con sistema de referencia WGS84 UTM Zona 18N. Para el procesamiento de la imagen se utilizaron programas de uso libre: las librerías de python *numpy*, para la manipulación de los valores matriciales en niveles digitales de imagen y *gdal*, para definir el sistema de referencia de la imagen resultado basado en la imagen inicial; y los paquetes *FactoClass* y *ade4* de R para los métodos multivariados.

**Palabras clave:** Análisis en componentes principales, clasificación, análisis multivariado, procesamiento digital de imágenes.

### Abstract

This paper shows an application about how to obtain information from the Paramo de Sumapaz area using multivariate statistical methods. The procedure comprises two stages: firstly, the application of a redundant spectral information transformation through a principal component analysis and secondly, a cluster analysis on the principal coordinates obtained in the first stage. Once the obtained clusters were characterized, terrestrial coverage classes were defined both geographically and spectrally in the studied zone. The database corresponds to a subscene from a Landsat 5 TM satellite image took on February 14, 1997 with path - row 008-057 covering part of The Paramo de Sumapaz in GeoTIFF format with reference system WGS84 UTM Zone 18N. Open source software was used in order to process the available information: the *numpy* python library for the conversion of matrix values into pixel image data, the *gdal* python library to define the reference system of the new image based on the input image and the R packages *FactoClass* and *ade4* to perform the required multivariate methods in the application.

**Key words:** Principal Component Analysis, Cluster Analysis, Multivariate Analysis, Digital Image Processing.

---

<sup>a</sup>Estudiante Especialización en Estadística. E-mail: drgalindog@unal.edu.co

<sup>b</sup>Profesor asociado. E-mail: cepardot@unal.edu.co

<sup>c</sup>Profesor asociado. E-mail: oriano@udistrital.edu.co

## 1. Introducción

El procesamiento digital de imágenes de satélite se ha convertido en una de las herramientas más importantes en el análisis y seguimiento de los recursos naturales. Esta disciplina tiene entre otras bases el análisis multivariado de datos, puesto que los millones de píxeles capturados en una sola imagen son considerados individuos y su comportamiento en las diferentes regiones del espectro electromagnético sus variables asociadas. Aunque existe software especializado en el procesamiento digital de imágenes de satélite, se hace necesario explorar herramientas estadísticas de uso libre como alternativa de acceso a la información de la imagen.

Este trabajo muestra una aplicación de métodos multivariados para obtener información de la zona del Páramo de Sumapaz usando el procesamiento de la información espectral proveniente de una imagen del satélite Landsat en el software R (R Development Core Team 2009) con herramientas de los paquetes FactoClass (Pardo & DelCampo 2007) y ade4 (Dray & Dufour 2007). Inicialmente se presentan conceptos básicos de los temas relacionados con el fin de contextualizar el desarrollo del trabajo, posteriormente se muestran los objetivos, se describe la base de datos y se desarrolla el análisis multivariado. Finalmente se presentan los resultados obtenidos.

El análisis multivariado comprende dos etapas: transformación de la información espectral redundante por medio de un análisis de componentes principales a partir de la matriz de correlaciones de las bandas espectrales, que permite compactar la imagen en las componentes por su alta correlación, lo que optimiza el procesamiento visual y matemático con respecto a la imagen original. La figura 1 muestra un diagrama de flujo resumiendo el procedimiento realizado.

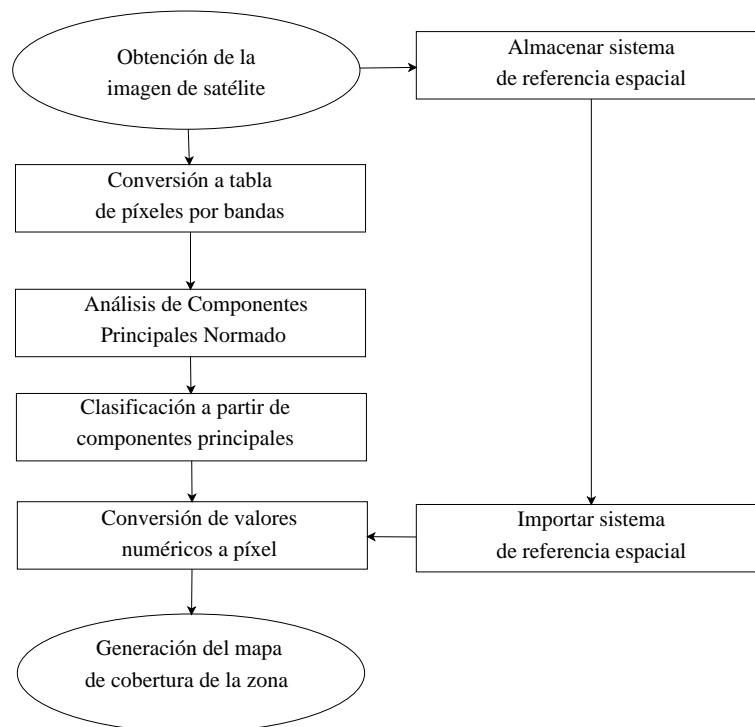


FIGURA 1: Diagrama de flujo del procedimiento realizado.

En la segunda etapa se realiza una clasificación no supervisada (por agrupamiento), sobre las coordenadas factoriales obtenidas. Una vez interpretados visualmente cada uno de los grupos resultantes, se

definen tanto geográfica como espectralmente las clases de coberturas terrestres presentes en la zona de estudio. A continuación se resumen algunos aspectos del contexto en que se enmarca esta aplicación.

### 1.1. Percepción remota

Es una técnica que permite obtener información a distancia de los objetos situados sobre la superficie terrestre. Un sistema de percepción remota requiere al menos de tres componentes: la fuente de energía, la superficie terrestre y un sensor. La superficie terrestre es iluminada por la radiación solar, esta es reflejada caracterizando los objetos presentes y es capturada por los sensores ubicados en el satélite, siendo parcialmente procesada a bordo de este y retransmitida a estaciones receptoras para su posterior procesamiento y análisis (Chuvieco 1993), y se convierte en una herramienta importante en la toma de decisiones sobre los recursos monitoreados.

Existen tres formas de adquirir la información en un sensor remoto a partir de la radiación: por reflexión, emisión y emisión-reflexión. La reflexión es la más importante para los sensores remotos debido a que la superficie terrestre al ser iluminada por el sol, principal fuente de energía, refleja diferente cantidad de energía en función del tipo de cubierta. La información se captura en bandas o rangos de longitudes espectrales, dado que la propagación de la energía se da en forma de ondas electromagnéticas caracterizadas por la longitud de onda y la frecuencia.

La figura 2 <sup>1</sup> muestra las curvas relacionadas con los comportamientos espectrales de algunas superficies terrestres de acuerdo a la longitud de onda. Se puede notar que la nieve tiene una alta reflectividad en las primeras regiones y va disminuyendo a lo largo del espectro electromagnético. El suelo presenta un comportamiento poco variable, alcanza su mayor punto en el infrarrojo medio. La vegetación tiene su mayor reflectividad en la región del infrarrojo cercano. El agua tiene una alta reflectividad en la región del visible disminuyendo en la región del infrarrojo. Éstas firmas espectrales varían de acuerdo a las diferentes condiciones de las coberturas mencionadas, como por ejemplo, la sedimentación del agua o el estado de la vegetación.

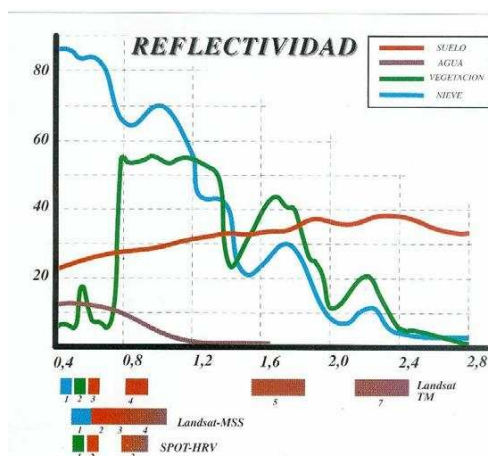


FIGURA 2: Curvas de reflectancia por cobertura.

El procesamiento digital de imágenes es una disciplina que desarrolla las bases teóricas y algorítmicas mediante las cuales pueda extraerse información del mundo real, de manera automática a partir de una imagen observada. Tal información puede relacionarse con el reconocimiento de objetos, posición y orientación de los mismos (Haralick 1992).

La interpretación de las imágenes de sensores remotos consiste en la utilización de técnicas, sistemas y procesos de análisis de imágenes por personal calificado para dar información segura y detallada acerca de los objetos naturales o artificiales contenidos en la superficie cuya imagen se analiza, y determinar los

<sup>1</sup>Tomada de: <http://www.fao.org/docrep/003/T0446S/T0446S04.htm>, octubre de 2009

factores que implican la presencia, condición y uso de ellos (Aguilar 2002). La interpretación puede ser directa cuando los objetos son visibles y fáciles de identificar aunque existen fases para la interpretación de las imágenes: lectura, análisis y clasificación (Melo 2004).

## 1.2. Cobertura y uso de la tierra

El levantamiento de la cobertura y uso de la tierra se puede definir como la obtención, análisis y clasificación de los diferentes tipos de cobertura y usos asociados que el hombre practica en una zona o región determinada (Melo 2004). Su importancia se fundamenta en que la información obtenida durante el estudio puede contribuir a la solución de diversos problemas de interés para el hombre y su bienestar, siempre y cuando se utilicen métodos adecuados para levantar la información. De esta manera, un mapa de este tipo debe proporcionar información fundamental para el desarrollo de estudios posteriores orientados a la planificación del uso de la tierra (IGAC 2005).

El término cobertura comprende los atributos de la tierra, los cuales por estar localizados sobre ésta ocupan una porción de su superficie (Melo 2004). Puede originarse de ambientes naturales producto de la evolución geológica (bosques, sabanas, lagunas, etc.) o a partir de ambientes artificiales creados y mantenidos por el hombre (cultivos, ciudades, represas, etc.).

## 1.3. Páramo de Sumapaz

El macizo del Páramo de Sumapaz<sup>2</sup> corresponde a un nudo orográfico de la cordillera oriental, con una altura media que oscila entre 3.500 m.s.n.m y 4.000 m.s.n.m. Se encuentra ubicado entre los departamentos de Cundinamarca, Huila y Meta y presenta, en su mayor parte, características montañosas con diversas lagunas. Es el páramo más grande del mundo y dentro de él nacen diversos ríos, constituyéndose en una gran fuente hídrica para Colombia (Ospina 2003).

Las plantas que más caracterizan la vegetación del páramo son el frailejón, la paja, lo mismo que los chusques y las cañuelas en las partes más bajas. En el páramo son frecuentes las formas vegetales propias de la alta selva andina, que suben hasta los 3500 m.s.n.m. Dichos bosques están formados por encenillos y palmas, entre otros. Para el objetivo del presente trabajo, al conjunto de la vegetación se denominará *vegetación de páramo*.

La principal función que desempeña un páramo es ser una fábrica de agua freática que da lugar al nacimiento de innumerables ríos, quebradas o pequeñas fuentes que alimentan todo el sistema hídrico de la cuenca de Orinoco y la parte media de la cuenca del Magdalena. Al páramo le corresponde sostener el régimen hidrológico, es decir, regular el ciclo del agua en la cabecera de los ríos. Pero esta función esta alterada y seriamente amenazada por la acción del hombre que quema y destruye el páramo hasta tal grado que hoy los páramos hacen parte de las regiones geográficas más devastadas del país, de allí la importancia de su estudio y monitoreo.

## 2. Objetivos

1. Extraer la información espectral relevante y no correlacionada de la zona estudiada a partir de análisis de componentes principales de la imagen Landsat 5TM.
2. A partir de la información espectral extraída de la imagen obtener un mapa de cobertura de la zona del Páramo de Sumapaz por medio de una clasificación no supervisada.
3. Realizar el procesamiento de la información en R y su visualización con software de uso libre.

---

<sup>2</sup>[http://www.paramo.org/portal/files/recursos/14\\_cruz\\_sumapaz.pdf](http://www.paramo.org/portal/files/recursos/14_cruz_sumapaz.pdf)

### 3. Descripción de los datos

La base de datos corresponde a una porción de la imagen de satélite Landsat 5 TM tomada el 14 de febrero de 1997, con *path 8 - row 57* que cubre parte del Páramo de Sumapaz y contiene 1.001.064 píxeles por banda espectral, es una imagen en formato GeoTIFF con sistema de referencia UTM WGS84 Zona 18N, descargada del sitio gratuito del *Global Land Cover Facility*.

La imagen se compone de seis bandas (ver tabla 1): tres del espectro visible y tres del infrarrojo. En la figura 3 puede observarse la composición en verdadero color de la zona.

En este caso, los individuos son cada uno de los niveles digitales o píxeles de la imagen y las variables explicativas son las bandas espectrales.

TABLA 1: Bandas espectrales de la imagen. Tomada de: <http://www.imagenesgeograficas.com/Landsat.html>. - julio de 2010

Banda espectral (longitud de onda)	Características	Región del espectro
Banda 1: azul ( $0.45\mu m - 0.52\mu m$ )	Diseñada para penetración en cuerpos de agua, es útil para el mapeo de costas, para diferenciar entre suelo y vegetación y para clasificar distintos cubrimientos boscosos, por ejemplo coníferas y latifoliadas. También es útil para diferenciar los diferentes tipos de rocas presentes en la superficie terrestre.	visible
Banda 2: verde ( $0.52\mu m - 0.60\mu m$ )	Especialmente diseñada para evaluar el vigor de la vegetación sana, midiendo su pico de reflectancia (o radiancia) verde. También es útil para diferenciar tipos de rocas.	visible
Banda 3: rojo ( $0.61\mu m - 0.69\mu m$ )	Es una banda de absorción de clorofila, muy útil para la clasificación de la cubierta vegetal. También sirve en la diferenciación de las distintas rocas	visible
Banda 4: infrarrojo cercano ( $0.76\mu m - 0.90\mu m$ )	Es útil para determinar el contenido de biomasa, para la delimitación de cuerpos de agua y para la clasificación de las rocas.	Infrarrojo cercano
Banda 5: infrarrojo medio ( $1.55\mu m - 1.75\mu m$ )	Indicativa del contenido de humedad de la vegetación y del suelo. También sirve para discriminar entre nieve y nubes.	Infrarrojo medio
Banda 7: infrarrojo medio ( $2.08\mu m - 2.35\mu m$ )	Especialmente seleccionada por su potencial para la discriminación de rocas y para el mapeo hidrotermal.	Infrarrojo medio

### 4. Análisis en componentes principales

Como una herramienta para obtener la información relevante de la imagen se lleva a cabo el análisis en componentes principales (ACP). No se introducen valores de ponderación, escala o centro, porque es información desconocida y porque no se requieren para el objetivo planeado en este análisis.

El análisis se plantea usando como individuos cada uno de los píxeles o niveles digitales de la imagen y como variables las bandas espectrales de la misma. Se retuvieron los dos primeros ejes para el análisis, de acuerdo los siguientes criterios (ver figura 4):

- Existe una marcada diferencia en el segundo de los ejes del análisis con relación a los cambios experimentados en los siguientes ejes.
- El porcentaje de inercia acumulado retenido por los dos primeros ejes es del 90 %.

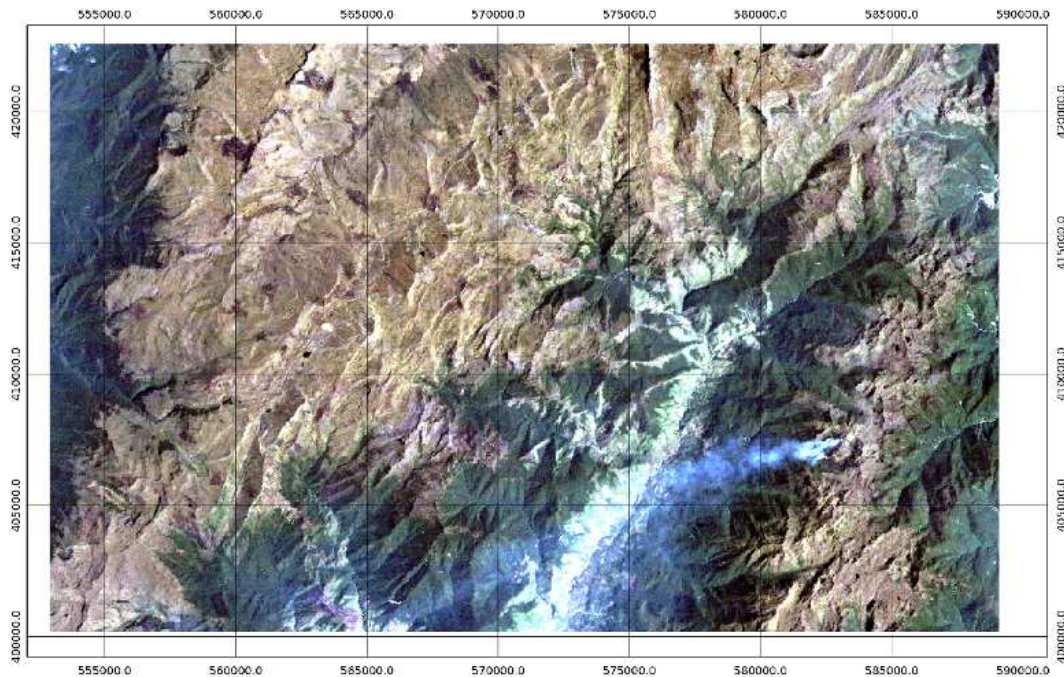


FIGURA 3: Composición de color: TM321 (RGB) Subescena imagen Landsat 5 TM - Páramo de Sumapaz.  
Fuente: <http://glcf.umiacs.umd.edu/data/> - Octubre de 2009

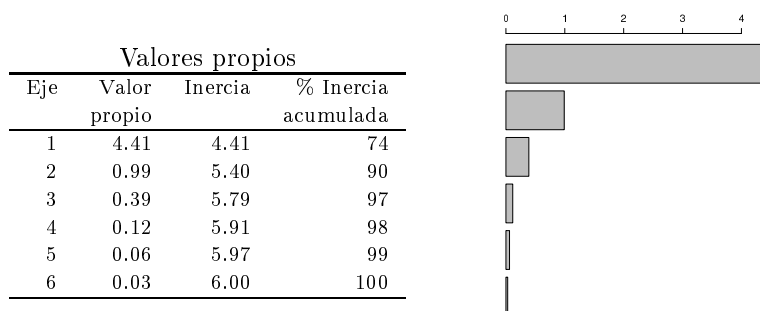


FIGURA 4: Histograma de valores propios del ACP

- Solamente el primer valor propio se encuentra muy por encima del promedio y el segundo coincide con el promedio.

De acuerdo con la teoría, es usual encontrar una alta correlación entre las bandas de la región visible del espectro electromagnético: azul, verde y roja (1,2 y 3 respectivamente), y como en este caso, con las bandas de la región del infrarrojo (bandas 5 y 7), se debe a por la alta presencia de cobertura vegetal en la zona. Las bandas menos relacionadas son la banda del infrarrojo cercano (banda 4) y el infrarrojo medio (banda 7), posiblemente porque la banda 7 es útil para la clasificación de rocas y la banda 4 es usada principalmente para determinar contenidos de biomasa (ver matriz de correlaciones en la figura 5).

El primer eje se encuentra caracterizado por las bandas azul, verde, rojo e infrarrojo medio (1, 2, 3 y 5 y 7 respectivamente), constituyéndose en un índice resumen de las coberturas de la zona de estudio. El segundo recibe su mayor aporte de la banda 4 que se utiliza principalmente para determinar contenido de biomasa. Entonces, el segundo eje permite a definir un perfil de las coberturas vegetales presentes en la

zona (ver figura 5). En el procesamiento digital, a este resultado suele denominarse brillo. En consecuencia este componente recoge características espectrales comunes a estas bandas.

En la zona de estudio se presentan solamente coberturas asociadas a vegetación o a cuerpos de agua, su comportamiento es similar en las diferentes regiones el espectro y de ahí los resultados que se observa en las correlaciones de las bandas espectrales.

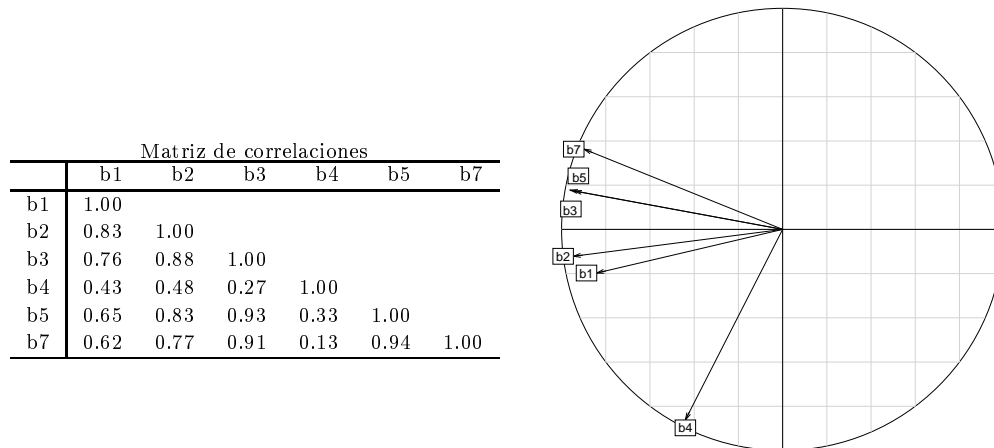


FIGURA 5: Matriz y círculo de correlaciones de las bandas. Nótese que las bandas 3 y 5 presentan la mayor correlación y en el primer plano factorial están representados por la misma flecha

El ACP realizado arroja una matriz de 787 por 1272 datos de las coordenadas factoriales. Para visualizarla como imagen se utilizó el lenguaje de programación *python* y las librerías *numpy* (Oliphant 2006) y *gdal* (Warmerdam 2010), la primera para el óptimo manejo de matrices de gran tamaño y la segunda para la creación y georreferenciación del archivo con formato raster convencional (para este caso GeoTIFF).

El proceso se realizó en dos fases: en la primera, se extrajeron los parámetros de georreferenciación de la imagen inicial que son los mismos para las imágenes resultado, para esto se utilizó el siguiente código:

```
# Obtener parametros de georreferenciación
dataset = gdal.Open("/ruta/archivo_original.tif", gdal.GA_ReadOnly)
parameters = ds.GetGeoTransform()
```

En la segunda fase se transformó el arreglo numérico obtenido a valores de imagen asignando el sistema de referencia espacial almacenado en la fase anterior. Lo cual se consiguió con el siguiente código:

```
import numpy
from osgeo import gdal
from osgeo import osr

# Cargue del archivo que contiene la matriz
array = numpy.loadtxt("/ruta/archivo_cp1", delimiter=';')

# Formato de salida
out_driver = gdal.GetDriverByName("GTiff")

# Creación del DataSet de salida
outds = out_driver.Create("/ruta/archivo_salida.tif", 1272, 787, 1,
    gdal.GDT_Float32, options=["COMPRESS=PACKBITS", "TFW=YES"] )

# Fijar parametros de georreferenciación obtenidos de la imagen original
```

```

outds.SetGeoTransform([552910.963270999996,28.5,0.0,422591.156340999999,0.0,-28.5])

# Definir el sistema de referencia de la imagen de salida
srs = osr.SpatialReference()
srs.ImportFromEPSG( 32618 )
outds.SetProjection( srs.ExportToWkt() )

# Escribir la matriz 'array' en la primera banda de la imagen de salida
outband = outds.GetRasterBand( 1 )
outband.WriteArray( array )

```

En la primera componente principal (figura 6) las cuatro coberturas: bosques, vegetación de páramo, pastos y cuerpos de agua se encuentran realizadas con respecto a la imagen original y en la segunda componente principal (figura 7) se explica la información no determinada directamente en la primera. Esto quiere decir, que si alguna cobertura tiene reflectancia similar a otra en la primera componente principal, la segunda será quien determine a cual de ellas pertenece verdaderamente. Un caso por ejemplo, es de las sombras que se pueden confundir con los cuerpos de agua; su nivel digital es similar pero su comportamiento espectral varía de acuerdo a la banda en la que están siendo observados tal como se describió en la tabla 1.

Al comparar las componentes con las bandas originales (figura 8) se puede ver mejor el resultado: en la primera fila se puede observar la primera componente (a) y las bandas que se encuentran mejor representadas en ella: azul (b), rojo (c) y verde (d); en la segunda fila se observa la segunda componente a la izquierda (a), la banda 4 (b) que es la mejor representada y las bandas 1 (c) y 7 (d) que también le aportan información.

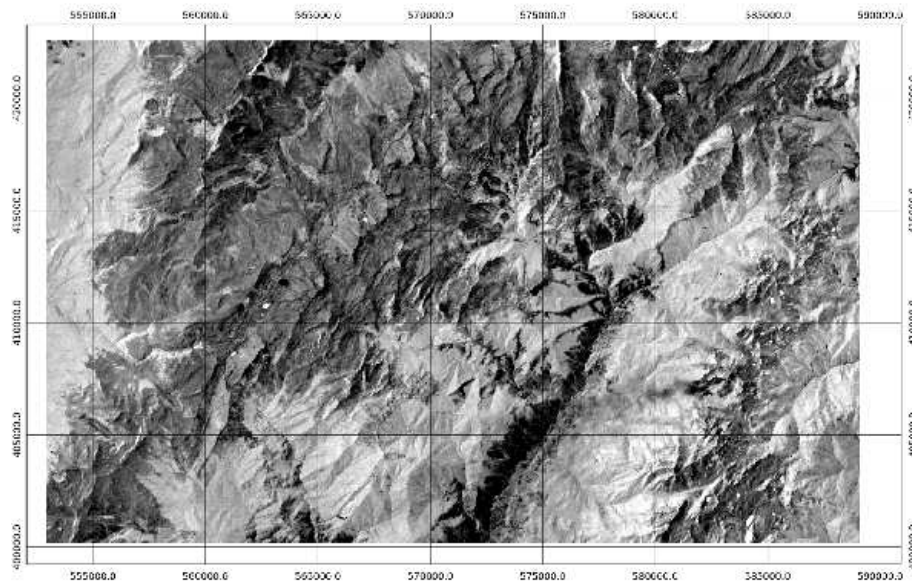


FIGURA 6: Primera componente principal en niveles digitales.

## 5. Clasificación no supervisada (agrupamiento)

Se hace una clasificación de píxeles, siguiendo el procedimiento descrito en Lebart et al. (1995), a partir de las dos primeras componentes principales para obtener el mapa de cobertura de la zona del



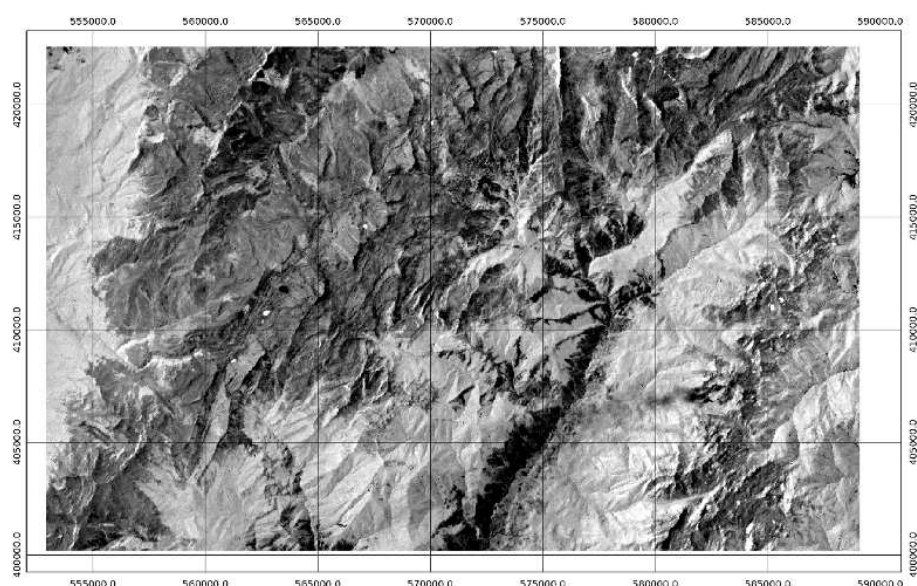


FIGURA 7: Segunda componente principal en niveles digitales.

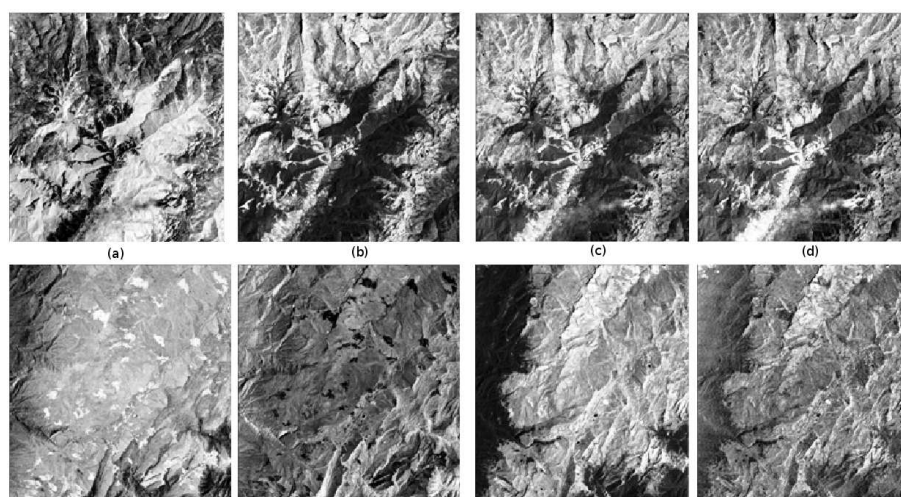


FIGURA 8: Comparación entre la componente principal (izquierda) y las bandas originales que cada una resalta (derecha).

Páramo de Sumapaz. Se definieron 4 clases de cobertura. El proceso de consolidación con *K-means* indica que la partición en cuatro clases es relativamente estable debido a que los cambios son menores, en efecto la clase 2 es la de más cambios con una pérdida del 7 % de sus píxeles (tabla 2). Las clases se pueden describir como: vegetación de páramo, cuerpos de agua, bosques y pastos y las vías y/o construcciones. Esta última no es relevante ya que dichos píxeles no son representativos dentro del área, como se puede apreciar en la imagen mostrada en la figura 3. La clasificación se llevó a cabo con el paquete *FactoClass* (Pardo & DelCampo 2007) de R (R Development Core Team 2009) y usando los dos ejes retenidos del ACP normado.

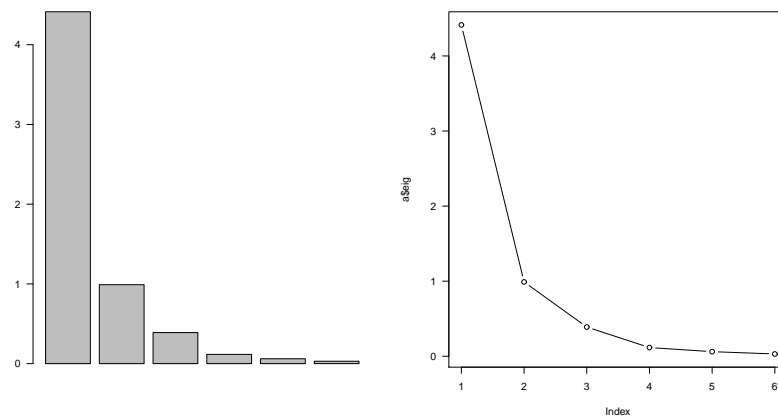


FIGURA 9: Histograma de índices de nivel de la clasificación jerárquica con el método de Ward utilizando las coordenadas sobre los dos primeros ejes del ACP. Se observan buenos cortes para 2, 3 y 4 clases

TABLA 2: Cambios en el proceso de consolidación de las clases obtenidas

Clase	Frecuencia antes	Frecuencia después	Inercia antes	Inercia después	Peso antes	Peso después
1	241751	254552	1.92	1.97	0.24	0.25
2	350722	345533	0.77	0.70	0.35	0.35
3	224455	211005	0.91	0.87	0.22	0.21
4	184136	189974	2.40	2.45	0.18	0.19
TOTAL	1001064	1001064	6.00	6.00	1.00	1.00

Convirtiendo el resultado a niveles digitales, tal como se muestra en la figura 10, se obtiene una nueva imagen clasificada por características similares en la cobertura de la zona.

De acuerdo con la definición de las clases, en esta imagen la vegetación de páramo se puede ver en color cian, los bosques en color rojo, los pastizales en color verde y los cuerpos de agua y las sombras por relieve en color magenta, lo que permite inventariar la zona con respecto a sus coberturas terrestres.

## 6. Conclusiones

El análisis en componentes principales (ACP) para una imagen de sensor remoto Landsat 5TM extrae la información espectral relevante y no correlacionada de las coberturas del suelo, tan solo los dos primeros ejes explican el 90 % de la información, reduciendo considerablemente la dimensionalidad de la imagen. Esto es de gran utilidad en el procesamiento digital de las imágenes, además los componentes principales se convierte en un apoyo visual para el intérprete de la misma, porque el resultado del análisis permite discriminar mejor algunas de las coberturas que con las bandas originales.

Fue posible obtener un mapa preliminar de cobertura de la zona del Páramo de Sumapaz siguiendo un procedimiento que incluye un ACP normado y una clasificación no supervisada. Con éste se tienen resultados que se aproximan a la naturaleza de la superficie, en general, este proceso se puede complementar, con correcciones de tipo atmosférico, por topografía, sombras y relieve previos al ACP con el fin de lograr el mayor aprovechamiento de la imagen.

Al llevar a cabo el ACP y la clasificación no supervisada de la imagen mediante el software estadístico R con los paquetes *FactoClass* (Pardo & DelCampo 2007) y *ade4* (Dray & Dufour 2007), se logran

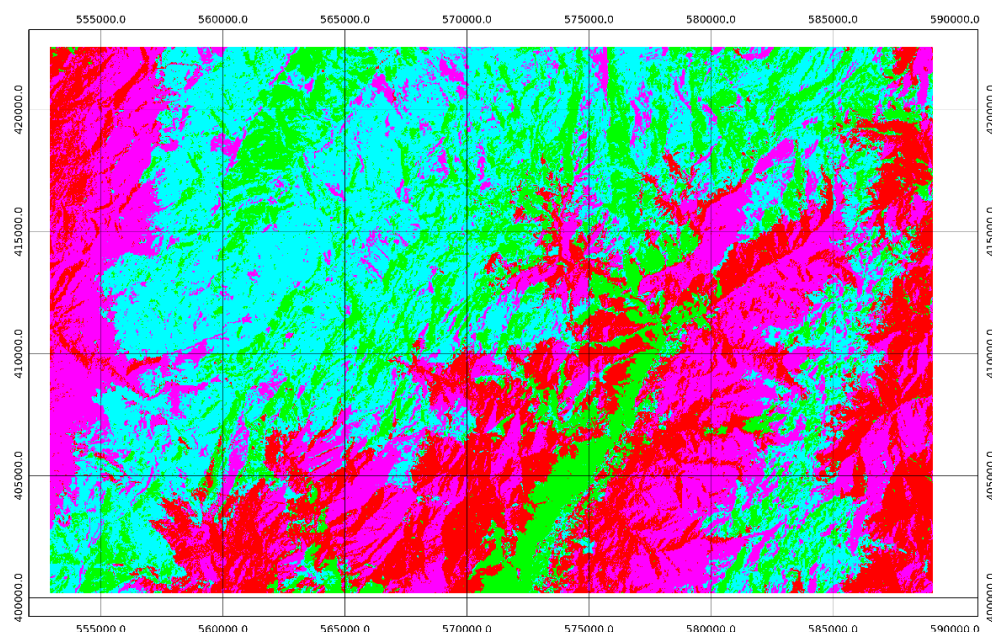


FIGURA 10: Imagen clasificada. En color cian la vegetación de páramo, en magenta los cuerpos de agua y las sombras, en verde los pastizales y en rojo los bosques.

resultados similares a los obtenidos con los softwares especializados en procesamiento digital de imágenes. Sin embargo, el tratamiento espacial de la información se debe complementar con herramientas o librerías, en este caso de uso libre, tal es el caso de la conversión de datos numéricos a formatos ráster para visualizar la imagen y la definición del sistema de referencia de las imágenes resultado del proceso.

Esta clasificación no supervisada es una aproximación de la cobertura de la zona. Se considera que es una base para establecer las zonas de entrenamiento necesarias en un análisis discriminante (clasificación supervisada) que permita obtener una clasificación más cercana a la realidad.

## Referencias

- Aguilar, M. (2002), *Fotointerpretación*, Universidad de la República de Uruguay.
- Chuvieco, E. (1993), *Fundamentos de Teledetección Espacial*, Editorial Rialp, Madrid.
- Dray, S. & Dufour, A. (2007), 'The ade4 package: implementing the duality diagram for ecologists', *Journal of Statistical Software* **22**(4), 1–20.
- Haralick, R. (1992), *Digital Image Processing Methods*, Marcel Dekker Inc., Estados Unidos.
- IGAC (2005), *Interpretación visual de imágenes de sensores remotos y su aplicación en levantamientos de cobertura y uso de la tierra*, Instituto Geográfico Agustín Codazzi, Bogotá.
- Lebart, L., Morineau, A. & Piron, M. (1995), *Statistique exploratoire multidimensionnelle*, Dunod, Paris.
- Melo, H. (2004), *Cobertura y usos de la tierra. Notas de clase*, Centro de Investigación y Desarrollo de Información Geográfica, Bogotá.
- Oliphant, T. (2006), 'Guide to NumPy', *Spanish Fork, UT*, Trelgol Publishing .

- Ospina, M. (2003), *El Páramo de Sumapaz en un ecosistema estratégico para Bogotá*, Sociedad Geográfica de Colombia, Academia de Ciencias Geográficas, Bogotá.
- Pardo, C. & DelCampo, P. (2007), 'Combinacion de metodos factoriales y de analisis de conglomerados en R: el paquete FactoClass', *Revista Colombiana de Estadística* **30**(2), 231–245.  
\*[www.estadistica.unal.edu.co/revista](http://www.estadistica.unal.edu.co/revista)
- R Development Core Team (2009), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.  
\*<http://www.R-project.org>
- Warmerdam, F. (2010), *GDAL API Tutorial*. Citado: abril de 2010.  
\*<http://www.gdal.org>