

Introducción al procesamiento de imágenes usando GRASS GIS 7.0 - Versión 0.3

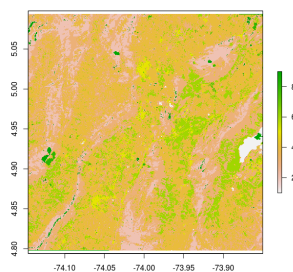
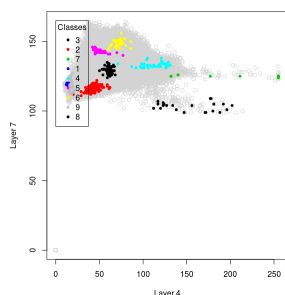
Ivan Lizarazo

Febrero de 2013



Attribution-ShareAlike
CC BY-SA

<http://creativecommons.org/licenses/by/3.0/>



Resumen

Esta es la tercera versión de un tutorial de procesamiento de imágenes de sensores remotos usando GRASS GIS versión 7.0. Su objetivo es mostrar que este programa de software libre permite realizar de manera eficiente un amplio rango de procesos digitales útiles para obtener información relacionada con las propiedades de la superficie terrestre. El autor usa datos e imágenes que cubren una zona rural cercana a Bogotá (Colombia), los cuales pueden ser descargados por el lector, para explicar paso a paso cómo realizar diversos procesos digitales tales como correcciones radiométricas, índices de vegetación, análisis de componentes principales y clasificación no supervisada y supervisada de la cobertura del suelo. En esta nueva versión, además de ampliar las secciones incluidas en la versión anterior, se profundiza en la interacción entre GRASS GIS y el software estadístico R.

1. Introducción

El lector puede utilizar cualquier imagen Landsat para realizar las operaciones que se indican en este tutorial. Sin embargo, si el usuario no está familiarizado con GRASS puede ser más conveniente que descargue la imagen utilizada por el autor, la cual corresponde a una pequeña ventana de Landsat 7 ETM+ que cubre una zona en el norte de Bogotá (Colombia). La imagen original es una escena completa, identificada con path 08 y row 57, tomada en enero de 2003. Dicha imagen fue descargada del sitio Global Land Cover Facility (GLCF) de la Universidad de Maryland. La ventana está en coordenadas geográficas WGS84. El identificador de dicha referencia espacial es el código EPSG 4326. La imagen de trabajo tiene 1091 columnas y 1093 filas y comprende siete bandas (tres en el espectro visible, dos en el infrarrojo cercano y una en el infrarrojo térmico). La resolución espacial de todas las bandas, excepto la correspondiente al infrarrojo térmico, es de 30 metros. La resolución radiométrica de la imagen es de 8 bits, es decir sus niveles digitales pueden ocupar el rango entre 0 y 255. El formato original es TIFF. El archivo de la imagen de trabajo se puede descargar del siguiente enlace:

<https://docs.google.com/open?id=0BzEwvK1H17qecEViN0t1bmpfWk0>

Para ejecutar las instrucciones de este tutorial hay que tener instalado el software GRASS GIS el cual se puede descargar desde el sitio <http://grass.osgeo.org>. Existen versiones de GRASS para los sistemas operativos MS Windows, GNU Linux y MAC OSX. La versión de GRASS GIS utilizada para realizar este tutorial es la *7.0-svn* la cual se instaló en Ubuntu 12.04.1 LTS. La versión estable más reciente es *6.4.2* la cual es probable que pueda ser utilizada para desarrollar estos ejercicios.

Este tutorial asume que el lector tiene un conocimiento básico de percepción remota y de procesamiento digital de imágenes. Por ello, no se explicarán de manera detallada conceptos relacionados con esos temas. El objetivo de este tutorial es mostrar que el software GRASS GIS ofrece funcionalidades apropiadas para diferentes proyectos y que representa una alternativa libre que es muy competitiva en comparación con programas comerciales.

En esta versión del tutorial, los temas desarrollados son:

- creación de un espacio de trabajo
- importación de imágenes
- creación de composición a color
- histogramas y diagramas de dispersión
- cálculo de factores de índice óptimo (OIF)
- calibración radiométrica
- cálculo de reflectancia absoluta
- importación de datos de elevación

- análisis de componentes principales
- índices de vegetación
- clasificación no supervisada
- clasificación supervisada
- integración entre R y GRASS (*nuevo*)

2. Organización de datos en GRASS

Los datos de GRASS se guardan en un directorio que internamente se denomina *GISDBASE*. El nombre usual de este directorio es *grassdata*. Dicho directorio se debe crear antes de empezar a trabajar con GRASS. Este directorio puede ser creado usando el comando *mkdir* o bien un gestor de archivos en el directorio de inicio o usando una carpeta compartida de una red (por ejemplo, en un sistema de archivos NSF). Dentro de dicho directorio, los datos de GRASS están organizados por proyectos que se almacenan en subdirectorios llamados *LOCATION* como se puede ver en la figura 1.

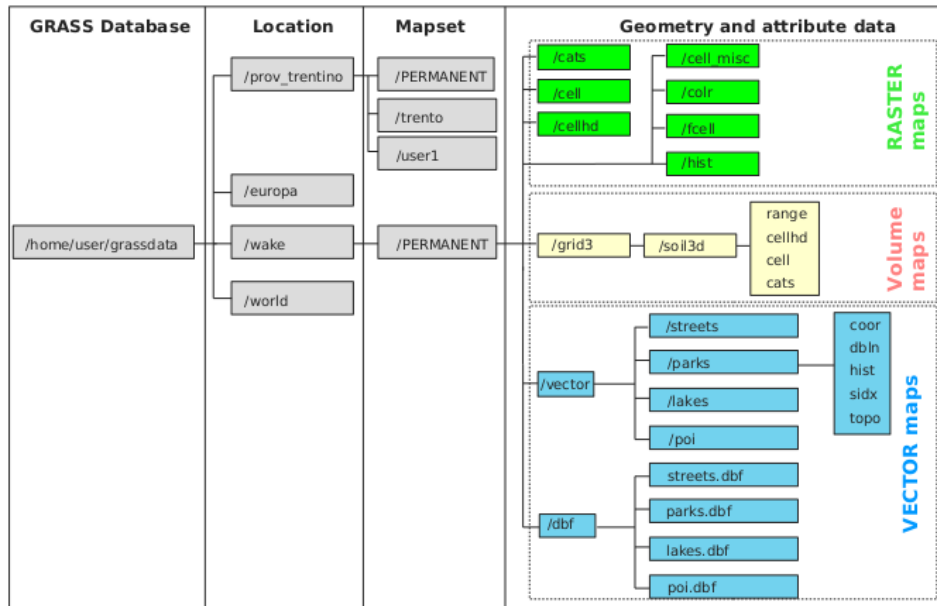


Figura 1. Organización de datos en GRASS

Cada *LOCATION* está definida por un sistema de coordenadas, una proyección cartográfica y unos límites geográficos. Los subdirectorios y archivos que definen una *LOCATION* se crean automáticamente cuando GRASS se inicia por

primera vez con una nueva LOCATION. Cada LOCATION puede tener varios MAPSETs (ver figura 1) que se usan para subdividir el proyecto en diferentes temas, sub-regiones o espacios de trabajo para integrantes individuales del grupo. Además de poder acceder a su propio MAPSET, cada usuario puede leer datos de otros MAPSETs pero únicamente puede modificar o remover los datos que están en su propio MAPSET. Todos los MAPSETs incluyen un archivo WIND que almacena los valores de coordenadas límites y la resolución raster que está actualmente seleccionada.

Cuando se crea una nueva LOCATION, GRASS crea automáticamente un archivo especial que se llama PERMANENT MAPSET. Ese MAPSET permite almacenar los datos centrales del proyecto, su extensión espacial default (dentro del archivo *DEFAULT_WIND*) y las definiciones del sistema de coordenadas. Únicamente el propietario de ese MAPSET puede modificar o remover ese archivo.

El archivo PERMANENT MAPSET es útil para proveer a un grupo de trabajo datos espaciales de referencia tales como elevación, vías y ríos al mismo tiempo que asegura su protección. Para importar datos en PERMANENT, simplemente hay que iniciar GRASS con la LOCATION relevante y el PERMANENT MAPSET. La organización interna y el manejo de LOCATION, MAPSETs y mapas se le deben dejar a GRASS. Es conveniente que cualquier operación que el usuario requiera tales como renombrar o copiar datos (que GRASS entiende como mapas) se debe realizar siempre usando los comandos de GRASS y no las funcionalidades del sistema operativo.

3. El concepto de región en GRASS

Una región es un concepto clave en GRASS. Una región define el área geográfica en la cual el programa debe trabajar. Una región se caracteriza por los siguientes parámetros:

- proyección geográfica (por ejemplo UTM, latitud-longitud, Gauss-Krueger, etc)
- extensión geográfica, es decir los límites North/South/East/West de la zona de estudio
- número de columnas y número de filas de los datos raster
- resolución espacial, es decir la extensión dividida entre el número de filas (resolución N-S) o de columnas (resolución E-W).

Los valores default de dichos parámetros para una *location* determinada se almacenan en el archivo *DEFAULT_WIND* dentro del mapset PERMANENT. Las definiciones de la región actual se almacenan en el archivo WIND dentro del *mapset* actual. Los valores almacenados permanecen válidos incluso si el usuario sale del programa y lo reinicia.

Cuál es la importancia de definir apropiadamente una región? Como se indicó anteriormente, la región define la extensión y la resolución de los datos dentro de la cual deben trabajar los comandos de GRASS. Eso significa, por ejemplo, que si la región está definida para ocupar un área menor que el área ocupada por los datos con los cuales está trabajando el usuario, el efecto de un comando que permite desplegar esos datos (por ejemplo el comando `d.rast`) únicamente mostrará aquella porción del mapa que está contenida dentro de esa región.

Muchos comandos de GRASS trabajan únicamente dentro del área que está definida en la región, por ejemplo, los comandos de exportación o los módulos de procesamiento raster. Esta característica de GRASS permite que el usuario trabaje solamente en la región de interés y que no utilice recursos de computador en zonas geográficas que no son relevantes para sus propósitos.

De una manera similar, el usuario puede reducir la resolución de la región con el objeto de utilizar menos recursos de máquina. Por ejemplo, si el usuario requiere convertir datos vectoriales en datos raster utilizando una resolución menor a la definida en la región, simplemente puede cambiar la definición de la región, de manera que no utilice recursos innecesarios.

En general, es recomendable que el usuario revise las propiedades de la región de trabajo antes de realizar cualquier procesamiento. Un problema común en los usuarios que se inician en GRASS ocurre cuando intentan importar unos datos y luego, cuando quieren desplegarlos, solamente ven una ventana vacía. Ello se debe, casi siempre, a que se han definido los parámetros de la región de manera incorrecta, de manera tal que los datos importados cubren un área que está por fuera de la región.

Idealmente, la región default de una 'location' debería incluir el área que eventualmente ocupan los datos que se van a trabajar en dicha 'location'. De esta manera, el usuario puede trabajar rutinariamente en el área que es su interés y, eventualmente, modificar los parámetros de la región para poder importar datos que cubren un área mayor.

4. Creación de una nueva LOCATION

Para ejecutar GRASS GIS en Linux se debe abrir una terminal. Teclee la expresión `grass70` como se indica en la figura 2:

En la ventana que se despliega, seleccione el directorio en el cual se almacenarán los datos, por ejemplo `grassdata`. Luego, haga clic en el botón *Location wizard* como se indica en la figura 3.

GRASS responde desplegando una serie de ventanas que permiten que el usuario defina una nueva LOCATION. En este caso, se puede utilizar la opción de indicar la referencia espacial utilizando el código EPSG. Como se indicó al principio de este documento, el código EPSG de la referencia espacial de nuestra imagen es 4326. Si el usuario desconoce cual es el código que identifica la referencia espacial de su imagen, puede buscarlo en el sitio <http://spatialreference.org>. Alternativamente, puede usar las otras opciones de

```
ivan@ubuntu:~$ grass70
WELCOME TO GRASS 7.0.svn

1) Have at your side all available GRASS tutorials

2) When working on your location, the following materials
are extremely useful:
- A topo map of your area
- Current catalog of available computer maps

3) Check the GRASS webpages for feedback mailinglists and more:
http://grass.osgeo.org
http://www.grass-gis.org

Hit RETURN to continue
```

Figura 2. Ejecucion de GRASS desde una terminal de Linux



Figura 3. Ventana de despliegue inicial de GRASS

creación de una LOCATION que ofrece GRASS que incluyen:

- seleccionar de una lista los parámetros del sistema de coordenadas
- leer los parámetros de proyección y datum de un archivo georeferenciado
- leer los parámetros de proyección y datum de un archivo WKT o PRJ
- especificar los parámetros de proyección y datum usando variables PROJ.4 personalizadas

Al finalizar la creación de una nueva LOCATION, se puede proceder a ejecutar GRASS GIS haciendo clic en el botón *Start GRASS*.

5. Importación de la imagen en el PERMANENT MAPSET

El programa inicia su ejecución mediante el despliegue de dos ventanas, una denominada *Layer Manager* y la otra *Map Display*. En la segunda fila de la ventana *Layer Manager*, en el tercer botón de izquierda a derecha, se encuentra la opción *Import raster or vector data*. Ese botón se puede utilizar para importar la imagen TIF descargada para trabajar con este tutorial. Al hacer clic en ese botón, el programa despliega una ventana denominada *Import raster data* en donde el usuario debe ingresar el formato y el nombre del archivo que se desea importar tal como se muestra en la figura 4.

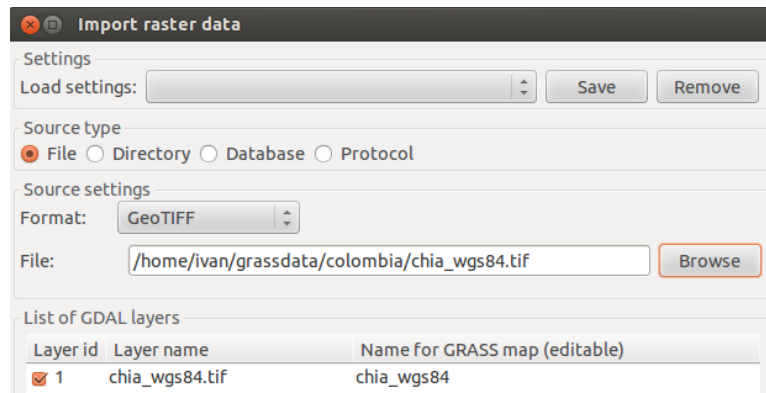


Figura 4. Importación de la imagen

```
(Sat Sep 22 11:44:17 2012)
r.in.gdal input=/home/ivan/grassdata/colombia/chia_wgs84.tif output=chia_wgs84
Datum <unknown> not recognised by GRASS and no parameters found
Projection of input dataset and current location appear to match
Raster map <chia_wgs84.1> created.
Raster map <chia_wgs84.2> created.
Raster map <chia_wgs84.3> created.
Raster map <chia_wgs84.4> created.
Raster map <chia_wgs84.5> created.
Raster map <chia_wgs84.6> created.
Raster map <chia_wgs84.7> created.
r.in.gdal complete.
(Sat Sep 22 11:44:19 2012) Command finished (1 sec)
```

Figura 5. Resultados de importar la imagen

En la figura 5 se pueden ver los resultados del proceso de importación. Se pueden observar dos cosas: (i) se han importado exitosamente cada una de las bandas de la imagen y se han creado siete mapas raster; y (ii) el programa GRASS no reconoce correctamente algunos parámetros asociados al datum de la imagen. Observe también que todos los mapas raster tienen un prefijo común que es *wgs_84*. Esto significa que, para GRASS, las imágenes forman parte de un mismo 'grupo'.

Para resolver el problema de indefinición de los parámetros de la referencia espacial, es conveniente revisar las características de la región creada por GRASS. Para ello, se debe usar la pestaña *Search module* del *Layer Manager*. En el árbol de menús, haga clic en *Settings/Region/Display Region*. El sistema muestra las definiciones que existen actualmente. Observe que el número de filas (rows) y de columnas (cols), entre otros parámetros, es incorrecto. Para cambiar esa definición, vuelva a la pestaña *Search module* y seleccione la opción *Set region*. El sistema responde desplegando la ventana *g.region* que tiene varias pestañas, entre ellas *Existing*, *Bounds* y *Resolution*. En la ventana *Existing*, en la opción *[multiple]Set region to match raster map*, seleccione cualquiera de siete bandas almacenadas en el PERMANENT MAPSET. Luego, en la ventana *Bounds*, active la opción *Align region to resolution*. Enseguida, en la pestaña *Resolution*, ingrese el número de filas y de columnas que tiene la imagen importada.

```
(Sat Sep 22 11:59:33 2012) Command finished (0 sec)
(Sat Sep 22 12:07:56 2012)
g.region -p
projection: 3 (Latitude-Longitude)
zone: 0
datum: wgs84
ellipsoid: wgs84
north: 5:05:39.292578N
south: 4:47:51.515295N
west: 74:08:46.459894W
east: 73:51:00.636458W
nsres: 0:00:00.976923
ewres: 0:00:00.976923
rows: 1093
cols: 1091
cells: 1192463
(Sat Sep 22 12:07:57 2012) Command finished (0 sec)
```

Figura 6. Características de la región de trabajo corregida

Para verificar que la región tiene los parámetros apropiados, ejecute nuevamente el comando *Display Region*. Usted debería obtener resultados similares a los que se indican en la figura 6.

6. Creación de una composición a color

Para visualizar la imagen en una composición a color, haga clic en el botón situado en el séptimo lugar de la primera fila y que se llama *Add various raster map layers*. En el menú que se despliega seleccione *Add RGB map layer*. El sistema despliega la ventana de ejecución del comando *d.rgb*. En esa ventana, se debe seleccionar cuál es la banda (raster map, en el lenguaje de GRASS) asignada a cada canal de color, *red*, *green* y *blue*. Como resultado, GRASS adiciona esa composición a color en la ventana *Layer Manager*. Allí, haga clic derecho en esa composición y seleccione la opción *Zoom to selected map(s)*. Luego, vaya a la ventana *Map Display* y haga clic en el primer botón denominado *Display map* para desplegar la composición a color. En la figura 7 se muestra parcialmente la imagen usando una composición a color RGB541.

En caso que la composición a color no produzca los resultados de visualización esperados (es decir que no se obtenga una imagen atractiva y fácil de



Figura 7. Composición a color RGB541

interpretar), el usuario puede modificar la banda espectral asignada a cada canal de color haciendo clic derecho sobre el layer que representa dicha composición en la ventana *Layer Manager* y usando la opción *Properties*.

7. Obtención de histogramas y diagramas de dispersión

Un histograma es un gráfico que muestra la frecuencia asociada a cada uno de los niveles digitales presentes en una banda individual. El histograma permite inferir, entre otras cosas, el tono predominante al visualizar una banda específica, lo mismo que el nivel de contraste correspondiente. Un diagrama de dispersión, también conocido como ploteo bidimensional o ploteo en el espacio espectral, es un gráfico que muestra los niveles digitales de todos los píxeles de la imagen en dos bandas específicas. El diagrama de dispersión permite observar la correlación existente entre las bandas en cuestión.

En la ventana *Map Display*, el cuarto botón de derecha a izquierda, denominado *Analyze map*, permite obtener histogramas y diagramas de dispersión, entre otras cosas. Haga clic en ese botón y luego en *Create histogram of raster map*. En la ventana que se despliega, seleccione la opción *Histogram imagery group* y luego, en la opción *Select image group* indique que el grupo de bandas

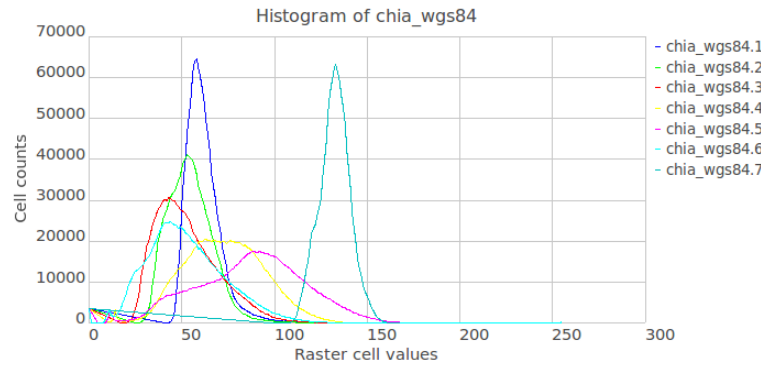


Figura 8. Histograma de todas las bandas de la imagen

de entrada se llama *chia_wgs84*. Haga clic en *OK* para obtener los histogramas de todas las bandas de la imagen tal como se ve en la figura 8.

Para obtener plots de dispersión, haga clic en el botón *Analyze map* y use la opción *Create bivariate scatterplot of raster maps*. En la ventana que se abre, seleccione las dos bandas de interés, por ejemplo las bandas 1 y 4 del grupo *chia_wgs84*, y haga clic en *OK*. El resultado debe ser similar al que se muestra en la figura 9.

8. Obtención de valores OIF

El Factor de Índice Optimo (OIF, por sus siglas en inglés) es una variable que expresa la cantidad de información no redundante asociada a cualquier combinación de tres bandas. En general, entre más alto es el valor de OIF para una combinación específica de tres bandas, mejor es la composición a color que se obtiene mediante la combinación de dichas bandas.

En la ventana *Layer Manager* seleccione el tabulador *Search Module*. En el árbol de menús, haga clic en *Imagery*. Luego, haga clic en *Reports and Statistics* y posteriormente en la opción *OIF for Landsat TM*. En la ventana que se despliega, denominada *i.oif* indique el nombre del mapa raster que representa cada una de las bandas Landsat-TM, de una manera similar a la indicada en la figura 10.

Observe que, en este caso particular, el mapa raster *chia_wgs84.6* corresponde a la banda 7 de Landsat-TM. Esto se debe al orden específico que el autor utilizó para unir en un solo archivo las bandas individuales que se descargaron del sitio GLCF (este orden puede ser útil para visualizar e interpretar correctamente perfiles espectrales pero puede traer complicaciones para realizar otros procesos). Luego de realizar la asignación requerida, haga clic en *Run*. Los resultados se muestran parcialmente en la figura 11.

El siguiente enlace proporciona algunas ‘sugerencias’ que pueden ser útiles para definir las combinaciones de bandas más apropiadas para discriminar vi-

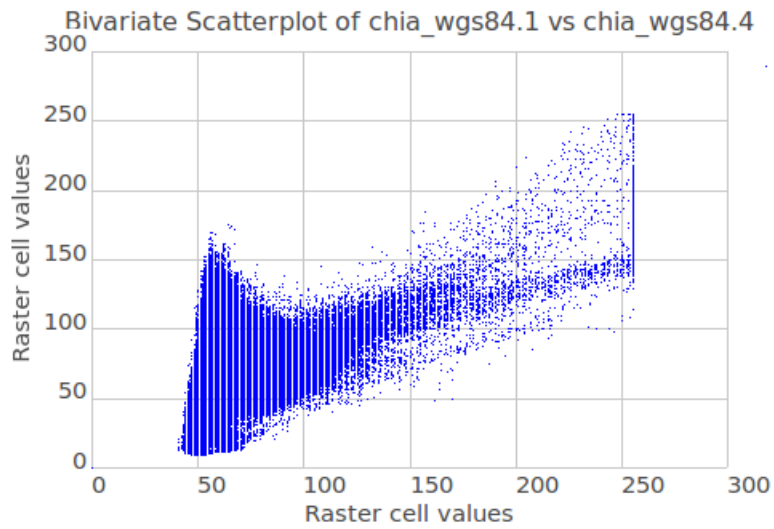


Figura 9. Gráfico de dispersión entre las bandas 1 y 4

sualmente diversos tipos de cobertura del suelo: <http://web.pdx.edu/~emch/ip1/bandcombinations.html>

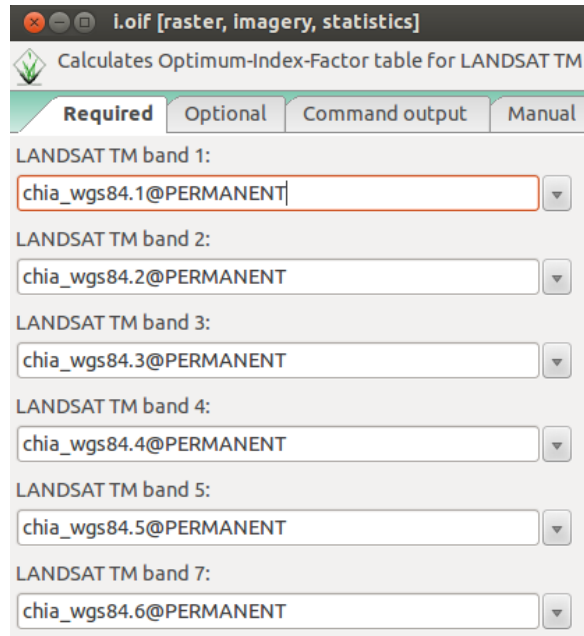


Figura 10. Asignación de mapas raster a cada una de las bandas Landsat

```
(Sat Sep 22 14:49:47 2012)
i.oif image1=chia_wgs84.1@PERMANENT image2=chia_wgs84.2@PERMANENT
Calculating Standard deviations for all bands...
Calculating Correlation Matrix...
r.covar: complete ...
Calculating OIF for the 20 band combinations...
347: 42.5344
345: 42.0629
145: 41.5852
457: 41.3590
147: 41.2375
247: 37.3572
245: 36.7323
134: 34.0029
234: 31.6122
157: 28.7791
357: 28.2516
124: 27.6301
257: 27.5799
135: 26.9503
```

Figura 11. Valores OIF correspondientes a la imagen *chia_wgs84*

9. Calibración radiométrica

La calibración radiométrica consiste en la conversión de los niveles digitales de cada banda de una imagen Landsat en radiancia espectral en el sensor. Este proceso se puede realizar en GRASS utilizando el comando *i.landsat.toar*. Para acceder a las funcionalidades de este comando se puede utilizar la pestaña *Search module* de *Layer Manager*. Otra manera de acceder a la ventana de ejecución del comando *i.landsat.toar* es ir al árbol de menús, hacer clic en *Imagery* y luego en *Satellite images tools*. Luego, se debe hacer clic en *Landsat DN to radiance/reflectance*.

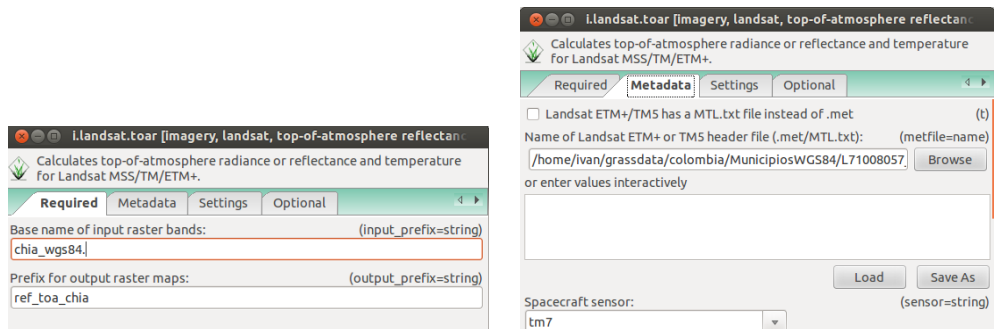


Figura 12. Parámetros suministrados por el autor para realizar calibración radiométrica

Para realizar la calibración radiométrica de una imagen se requiere conocer los datos del sensor y algunos metadatos, tales como fecha y hora de adquisición, lo mismo que las funciones de calibración que relacionan los niveles digitales (ND) y los valores de radiancia espectral de cada banda. Dichos datos son usualmente proporcionados por los proveedores de imágenes durante el suministro de las mismas. En este tutorial, se utilizó el archivo de metadatos obtenido del sitio GLCF. Dicho archivo se puede descargar en este enlace <http://db.tt/IJXSfNqz>. En la figura 12 se muestran los parámetros suministrados por el autor para realizar la calibración radiométrica de la imagen que estamos trabajando. En la figura 13 se observa la banda 1 original y la banda 1 calibrada.

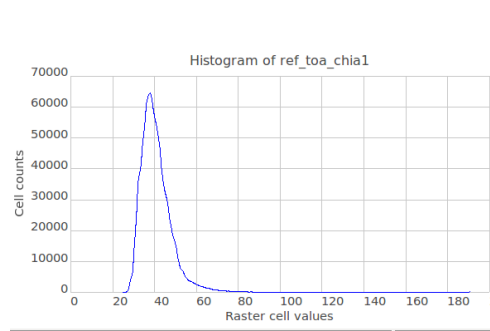
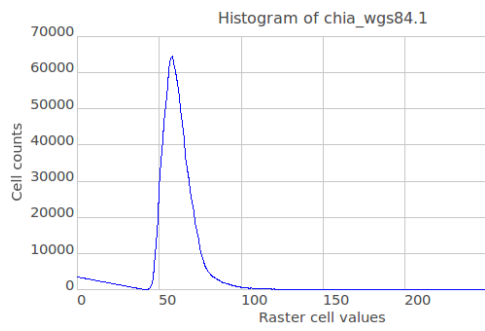


Figura 13. Comparación entre la banda 1 original y banda 1 calibrada

10. Corrección atmosférica

La corrección atmosférica de una imagen permite obtener valores de reflectancia de la superficie del terreno. La reflectancia es útil en estudios dirigidos a la obtención de variables biofísicas a partir de imágenes de sensores remotos. Aunque la corrección atmosférica no es estrictamente necesaria cuando se desean obtener variables cualitativas a partir de una imagen de una sola fecha, como por ejemplo cuando se requiere inferir la cobertura del suelo, aquella es indispensable cuando se desea realizar estudios multitemporales.

La corrección atmosférica se puede realizar utilizando el comando *i.atcor* que utiliza el algoritmo 6S (Second Simulation of a Satellite Signal in the Solar Spectrum). Para ejecutar dicho comando, vaya al árbol de menús, haga clic en *Imagery* y luego en *Satellite images tools*. Luego, haga clic en *Atmospheric correction*. Enseguida, se despliega la ventana asociada al comando *i.atcor*.

Para ejecutar dicho comando se deben ingresar, a través de un archivo de texto o de manera interactiva en una caja de texto, algunos parámetros relacionados con el sensor, las condiciones de adquisición de la imagen (día, hora, coordenadas del centro de la imagen), el modelo atmosférico, altitud del sensor, altitud del terreno, lo mismo que la banda de interés. En este caso, los parámetros utilizados para obtener el valor de reflectancia absoluta en la banda 1, a partir de la radiancia en el sensor en la misma banda, fueron los siguientes:

- 8 # indica que es una imagen adquirida por el sensor ETM+
- 01 11 17.67 -74.000 5.000 # imagen tomada el 11 de enero, 17 : 42 GMT, long. 74W y lat. 5N
- 1 # modelo atmosférico tropical
- 1 # modelo de aerosoles continental
- 50 # visibilidad para el modelo de aerosoles [km]
- -2.60 # altura del terreno [km] * -1
- -1000 # imagen tomada desde un sensor satelital
- 61 # banda espectral 1

En la figura 14 se muestra la imagen de reflectancia absoluta de la banda 1 con su correspondiente histograma. Observe que los valores de reflectancia han sido rescalados de sus valores originales en el rango [0,1] a un rango de enteros sin signo de 8 bits [0,255].

En el siguiente enlace se pueden consultar cuáles son los parámetros requeridos para ejecutar el commando *i.atcor* con diferentes tipos de imágenes y sensores satelitales: http://grass.osgeo.org/gdp/html_grass70/i.atcorr.html

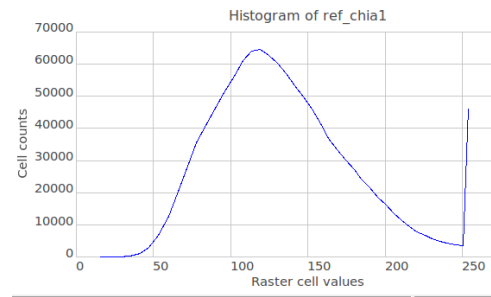


Figura 14. Imagen de reflectancia absoluta en la banda 1 con su correspondiente histograma

11. Importación de datos de elevación del terreno

En la siguiente dirección se pueden descargar datos raster de elevación de Suramérica: http://dds.cr.usgs.gov/srtm/version2_1/SRTM3/South_America/

Dichos datos fueron obtenidos por NASA dentro del proyecto Shuttle Radar Topographic Mission (SRTM) realizado en Febrero de 2000. Estos datos tienen una resolución espacial de 90 metros y una exactitud de posición vertical de 15 metros para el 90% de los datos.

Los datos SRTM están organizados en hojas (*tiles* en inglés) de dimensión un grado por un grado. Los archivos con datos SRTM contienen las coordenadas del centro del pixel de la esquina inferior izquierda (por ejemplo, el archivo N51E010 tiene su celda inferior izquierda en longitud 10E y latitud 51N). Para saber cuál es el nombre de las hojas SRTM que cubren la región de interés, se puede usar la siguiente instrucción que permite visualizar una grilla de un grado por un grado: *d.grid size=1*

Al examinar la grilla desplegada se puede saber cuáles son los archivos SRTM que se deben descargar del sitio mencionado anteriormente. En el caso de nuestra zona de estudio, los archivos a descargar son los siguientes: N04W074.hgt.zip, N04W075.hgt, N05W074.hgt y N05W075.hgt. Luego de realizar su descarga, los datos deben descomprimirse en el directorio de trabajo.

Para importar los datos SRTM HGT en GRASS se utiliza el comando *Import/link rasterorvectordata* (tercer botón desde la izquierda en la segunda fila de GRASS GIS Layer Manager). Para efecto, se debe hacer clic en ese botón, y luego en *Import raster data*. En la ventana que se despliega se debe indicar, entre otros parámetros, que el formato de los datos es *SRTM HGT*.

Luego de ejecutar el comando de importación, se pueden visualizar dichos datos. Tenga en cuenta que el área cubierta por cada *tile* importado es mucho mayor que el área cubierta por la imagen de trabajo. Para cubrir solamente el área de interés se debe elaborar un mosaico que, primero agrupe los cuatro *tiles* y, luego, extraiga únicamente los datos correspondientes a la región de trabajo. Para ello, se puede utilizar el comando *i.image.mosaic* que se ejecuta yendo al árbol de menús de la pestaña *Search module*, haciendo clic en *Imagery* y luego en *Mosaic images*.

En la figura 15 se observa el mosaico obtenido, el cual cubre únicamente la región de interés. El mosaico, que tiene un nivel de opacidad del 60%, está superpuesto sobre la imagen Landsat. Este ejemplo muestra la ventaja del concepto de *region* utilizado por el software GRASS GIS.

Los metadatos del mosaico de elevación se pueden obtener usando el comando *r.info* al cual se puede acceder en la pestaña *Map layers*, haciendo clic derecho en la capa *strm_chia*. Como se observa en la figura 16, el mosaico tiene el mismo número de columnas y filas (y, por ende, el mismo tamaño de pixel) que las bandas de la imagen de trabajo *chia_wgs84*.

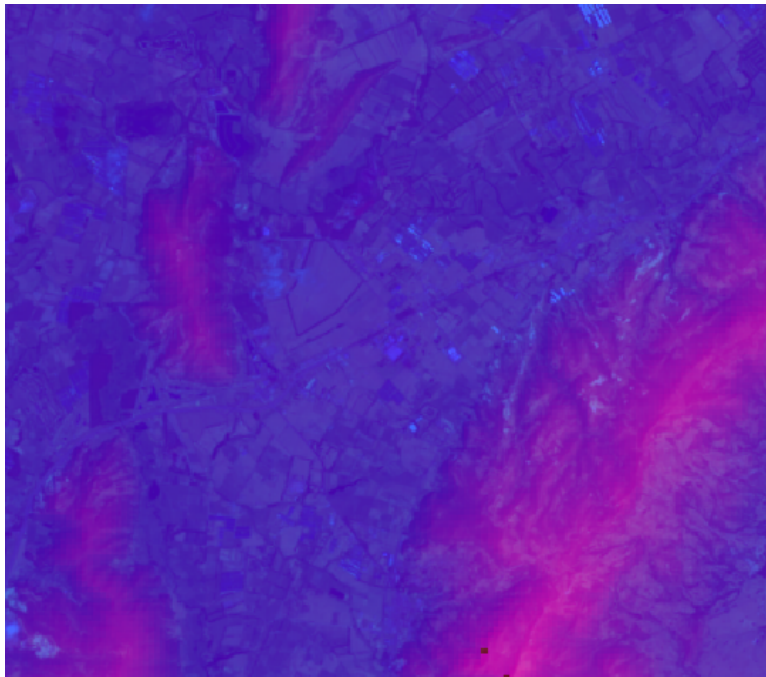


Figura 15. Mosaico de datos de elevación en la zona de estudio superpuesto sobre la imagen Landsat.

```
r.info map=strm_chia@PERMANENT
-----
| Layer:      strm_chia@PERMANENT          Date: Wed Oct  3 19:50:00 2012
| Mapset:     PERMANENT                   Login of Creator: ivan
| Location:   bogota
| DataBase:  /home/ivan/grassdata
| Title:     ( strm_chia )
| Timestamp: none
|-----
| Type of Map: raster                      Number of Categories: 0
| Data Type:  DCELL
| Rows:      1093
| Columns:   1091
| Total Cells: 1192463
| Projection: Latitude-Longitude
|             N: 5:05:39.292578N   S: 4:47:51.515295N   Res: 0:00:00.976923
|             E: 73:51:00.636458W   W: 74:08:46.459894W   Res: 0:00:00.97692
| Range of data:  min = 2533  max = 15879
|-----
```

Figura 16. Metadatos del mosaico de datos de elevación.

12. Análisis de componentes principales

El análisis de componentes principales (PCA, por sus siglas en inglés) tiene como propósito obtener nuevas bandas que contengan información no redundante (es decir, que no tengan ninguna correlación entre sí). En general, es posible afirmar que los tres primeros componentes principales de una imagen representan aproximadamente el 95% de la información contenida en las bandas originales. PCA es una técnica muy útil para reducir el volumen de información original de una imagen. Igualmente, es una técnica muy popular en la realización de estudios multi-temporales.

La figura 17 muestra los parámetros utilizados para obtener los componentes principales de la imagen usada en este tutorial. La figura 18 muestra los autovalores (*eigenvalues*), los autovectores (*eigenvalues*) y el porcentaje de varianza asociado a cada uno de los siete componentes principales. Observe que los tres primeros componentes principales representan aproximadamente el 96.5% de la información original. La figura 19 muestra una composición a color con los tres primeros componentes principales. Se puede observar que los colores de la escena son diferentes a los obtenidos mediante composiciones basadas en las bandas originales.

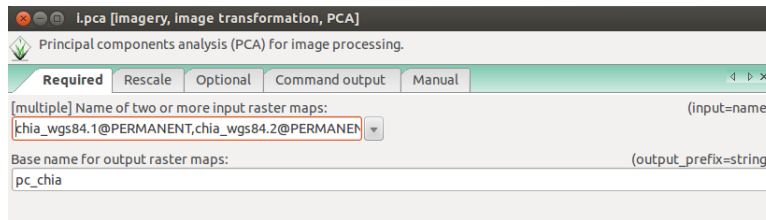


Figura 17. Parámetros de entrada para obtener componentes principales.

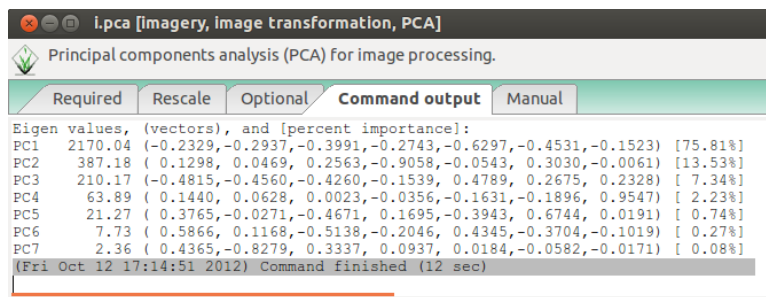


Figura 18. Resultados del análisis de componentes principales.

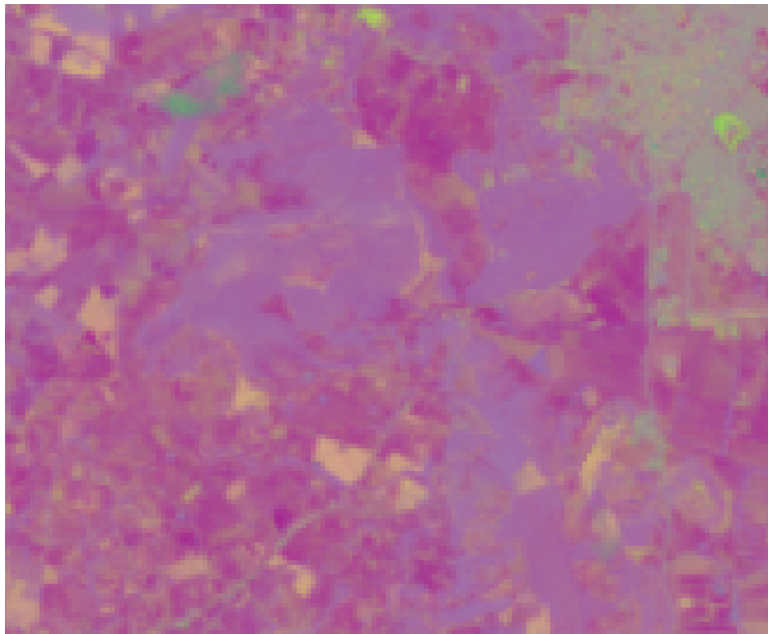


Figura 19. Composición a color RGB321 usando componentes principales.

13. Creación de una banda auxiliar de elevación

La disponibilidad de una capa raster de elevación que cubre la región de trabajo (el area de interés) con el mismo tamaño de pixel de las bandas originales puede ser aprovechada para crear un ‘grupo’ de bandas que agregue, junto a las bandas originales o a las bandas obtenidas mediante componentes principales, una banda auxiliar de elevación. Esta banda auxiliar, aunque no contiene información espectral, puede ser útil para discriminar coberturas del suelo que son espectralmente similares pero que ocurren en diferentes pisos altitudinales. En este tutorial, vamos a crear un grupo que incluya los componentes principales y la banda de elevación. Para el efecto, se usará el comando *i.group* el cual se encuentra en el árbol de menús, haciendo clic en *Imagery* y luego en *Create/edit group*. En la figura 20 se pueden ver los parámetros utilizados para crear un grupo denominado *chia8*. Este grupo puede emplearse posteriormente en la realización del proceso de clasificación de la cobertura del suelo.

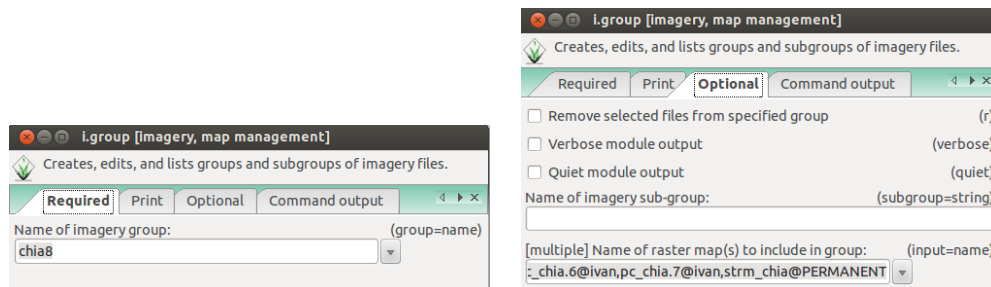


Figura 20. Parámetros utilizados para crear un grupo con siete componentes principales y una banda auxiliar de elevación

14. Índices de vegetación

Un índice de vegetación es una transformación espectral que busca representar en una sola banda la presencia y/o el vigor de la vegetación, aprovechando las diferencias en reflectancia de la vegetación en las bandas roja e infrarroja cercana. Uno de los índices de vegetación más utilizado es el denominado índice de vegetación de diferencias normalizado (NDVI, por sus siglas en inglés). Para obtener un NDVI apropiado es conveniente utilizar una imagen que tenga corrección atmosférica.

El módulo ‘map calculator’ de GRASS permite acceder a la funcionalidad *r.mapcalc* que es útil para construir modelos computacionales basados en capas raster (incluyendo las bandas de una imagen). Este módulo será usado en este tutorial para ingresar la fórmula correspondiente a NDVI: $NDVI = (NIR - Red)/(NIR + Red)$, en donde NIR es la reflectancia en el infrarrojo cercano (banda 4 de Landsat TM), y Red es la reflectancia en el rojo (banda 3 de Landsat TM).

En el menu Raster de Layer Manager, seleccione Raster map calculator. En la ventana resultante, ingrese *chia.ndvi* como el nombre de la capa raster que se va a crear, luego use los otros botones para crear la siguiente fórmula:

$$\text{float}(\text{ref_chia_4} - \text{ref_chia_3})/\text{float}(\text{ref_chia_4} + \text{ref_chia_3}) \quad (1)$$

Enseguida presione el botón Run para ejecutar el proceso. La figura 21 muestra el índice NDVI obtenido al igual que sus estadísticas (obtenidas haciendo clic derecho en dicha capa y seleccionando la opción Univariate raster statistics). Observe que entre más densidad o vigor de la vegetación existe, más alto es el valor de NDVI. Observe también que, en este caso, los valores NDVI se encuentran en el rango -0.32 a 0.925.

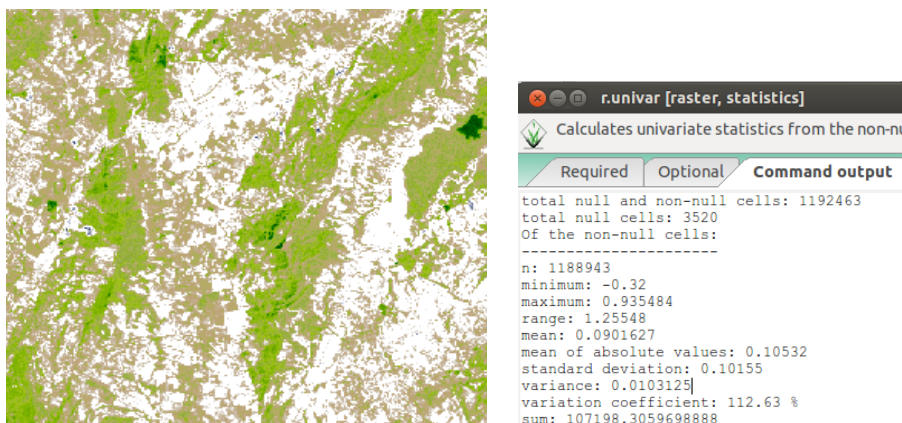


Figura 21. Índice de vegetación NDVI con sus correspondientes estadísticas.

15. Clasificación no supervisada

Una clasificación no supervisada es un proceso de agrupación de píxeles en ‘clusters’ o clases espectrales. Estos clusters no tienen significado real en sí mismos, pero el usuario puede establecer cuál es su relación con las categorías temáticas de la zona de estudio mediante un proceso de interpretación y análisis visual. En el siguiente ejercicio se realizará un ejemplo de obtención de clusters a partir de un subconjunto de las bandas de imagen, y luego, se utilizará el algoritmo de probabilidad máxima para asignar cada píxel de la imagen al cluster con el cual tenga más similitud espectral. Para el efecto, se deben realizar los siguientes pasos:

1. en el menú *Imagery*, seleccione *Develop images y groups—Create/edit group*. En la ventana *Required* ingrese el nombre del grupo de imágenes que se va a utilizar, por ejemplo *chia8*; luego, en el tabulador opcional, ingrese el nombre del subgrupo que se va a crear, por ejemplo *4bandas*; enseguida, ingrese los nombres de los tres primeros componentes principales y de la banda auxiliar de elevación. Finalmente, ejecute el proceso.
2. en el menú *Imagery*, seleccione el proceso *unsupervised classification* y luego, en la pestaña *Classify image—Clustering input*, indique que *chia8* es el nombre del *imagery group*, *4bandas* el nombre del subgroup, *4bandas_no_sup* el nombre del archivo de firmas espectrales que se obtendrá, y 10 el número de clusters deseado. Observe los parámetros default que se encuentran en la pestaña *Settings* y, en caso necesario, realice modificaciones. Finalmente, ejecute el proceso correspondiente. Observe que el módulo de GRASS que ejecuta ese proceso se llama *i.cluster*
3. en el menú *Imagery*, seleccione *Classify image—Maximum likelihood classification (MLC)*. En la pestaña *Required*, indique el nombre del imagery group, por ejemplo *chia8*, el nombre del subgroup, por ejemplo *4bandas*, el nombre del archivo de firmas, por ejemplo *4bandas_no_sup*, y el nombre del mapa raster en el cual se van a almacenar los resultados de la clasificación, por ejemplo *10clusters*. Finalmente, ejecute el proceso correspondiente. Observe que el módulo de GRASS que ejecuta ese proceso se llama *i.maxlik*.

Ahora es tiempo de examinar los resultados de la clasificación, los cuales se muestran en la figura . Para ser justos, no se deben esperar resultados muy buenos, toda vez que en la ejecución de una clasificación no supervisada no se realiza ninguna selección de muestras de entrenamiento que representen categorías reales. Es bastante probable, entonces, que estos resultados escasamente puedan ser calificados como aceptables.

En la práctica, una clasificación no supervisada es muy útil cuando no se tiene una buena referencia de las clases existentes en el terreno. Por ello, después de obtener una clasificación no supervisada, el usuario debe establecer una correspondencia entre cada clase espectral y las clases ‘reales’. Si dicha relación es

‘uno a uno’ o ‘muchos a uno’, el asunto es trivial. Sin embargo, si dicha relación es uno a muchos (es decir, una clase espectral representa muchas clases reales), es necesario enmascarar los píxeles que pertenecen a las clases ‘complicadas’ y realizar con ellas una clasificación no supervisada adicional. El resultado final de la clasificación será el resultado de unir las dos clasificaciones.

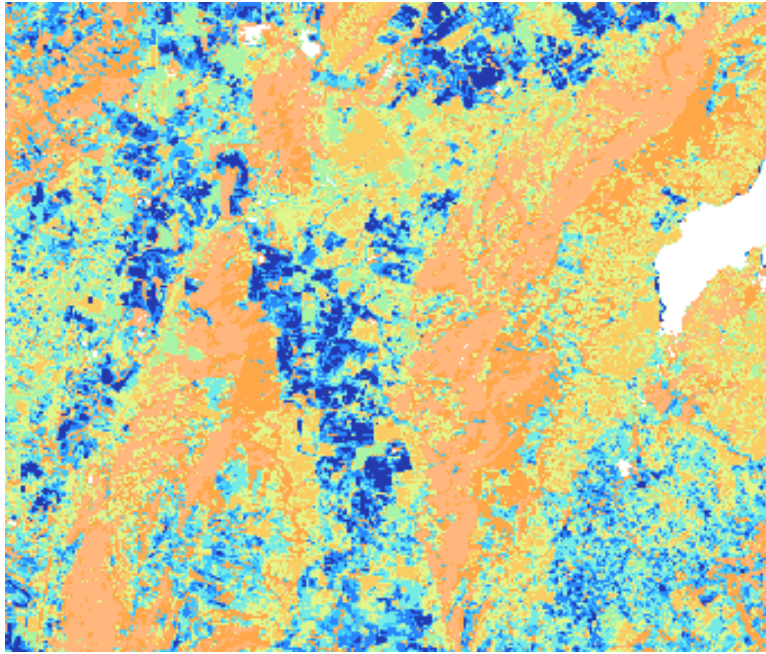


Figura 22. Resultados de ejecutar una clasificación no supervisada usando tres componentes principales y la banda auxiliar de elevación.

16. Clasificación supervisada

Para realizar una clasificación supervisada, es necesario crear primero muestras de entrenamiento por cada categoría de interés, es decir, dibujar polígonos que representen los píxeles típicos asociados a cada una de las diferentes clases de cobertura del suelo, por ejemplo, bosques, pastos, cultivos, agua, construcciones, etc. Estos polígonos de entrenamiento permiten crear firmas espectrales de cada clase que posteriormente se utilizan por algún algoritmo para asignar a cada píxel de la imagen alguna de las categorías temáticas de interés.

Para crear estos polígonos de entrenamiento, un usuario solamente tiene dos opciones: (i) el usuario realiza una visita de campo para reconocer cuáles son las diferentes coberturas existentes y, eventualmente, levantar los polígonos de entrenamiento, (ii) el usuario revisa información temática que exista en la zona de estudio y la utiliza como referencia principal para definir sus zonas de entrenamiento.

En GRASS, la clasificación supervisada se puede realizar utilizando alguna de las siguientes opciones:

- utilización de los comandos *i.gensig* e *i.maxlik* (utilizando un archivo raster con muestras de entrenamiento)
- utilización de los comandos *i.class* e *i.maxlik* (modo interactivo: creación de firmas espectrales por el usuario)

16.1. Clasificación utilizando un raster de entrenamiento

Enseguida se muestra como realizar el proceso de clasificación supervisada utilizando un archivo raster con muestras de entrenamiento. Se supone que este archivo raster ha sido obtenido de fuentes externas y que es una representación confiable de la cobertura del suelo existente en la zona de estudio. Para realizar la clasificación se deben ejecutar los siguientes pasos:

1. descargue un archivo con categorías de cobertura del suelo de la zona central andina de Colombia de la siguiente dirección: <https://docs.google.com/open?id=0BzEwvK1H17qeejNOZzhiVWxEcm8>
2. importe el archivo descargado en GRASS (este archivo se muestra en la figura 23).
3. utilice el comando *i.gensig* para obtener las muestras de entrenamiento a partir del archivo raster (este comando se puede ejecutar desde la consola de comandos del Layer Manager)
4. utilice el comando *i.maxlik* para realizar la clasificación de cobertura. Los resultados de la clasificación realizada se muestran en la figura 24)

Para información del lector, el archivo raster utilizado para obtener las muestras de entrenamiento corresponde a un subconjunto de los datos globales de cobertura

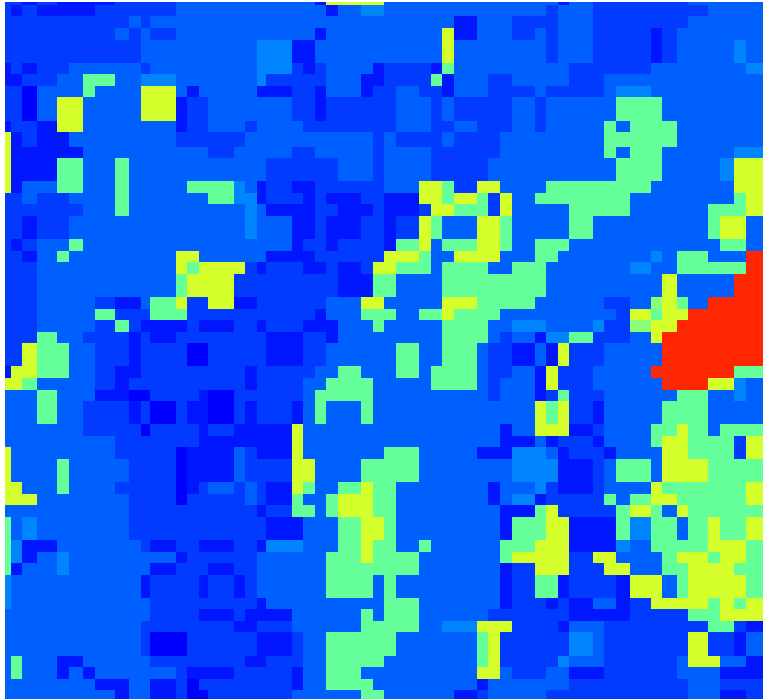


Figura 23. Vista parcial del archivo raster con categorías de cobertura.

del suelo denominados *GlobCover 2009*, los cuales fueron producidos por la agencia espacial europea (ESA). Dichos archivos se pueden descargar del siguiente enlace: <http://ionia1.esrin.esa.int/>

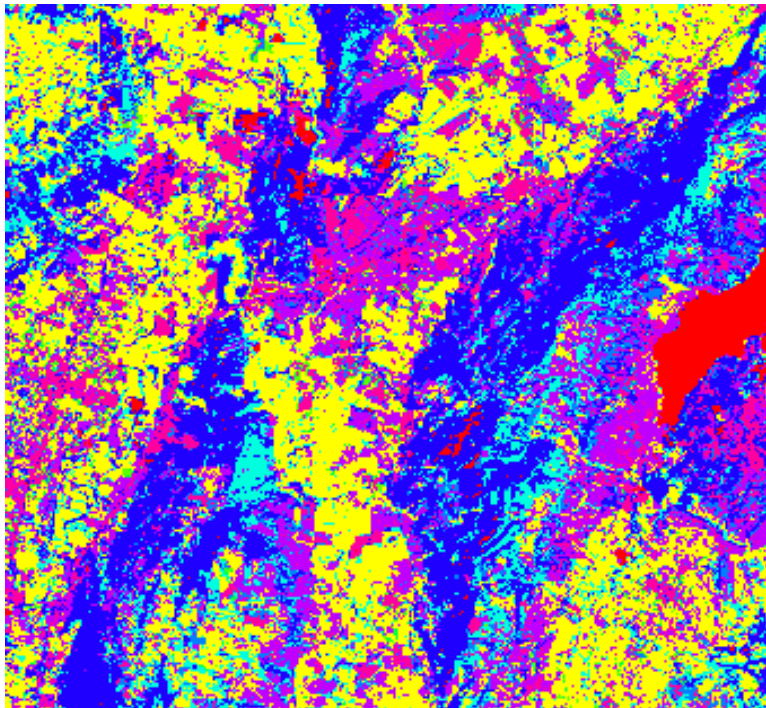


Figura 24. Clasificación de la cobertura del suelo usando un archivo raster de entrenamiento.

Es posible que el lector haya notado que la resolución espacial de GlobCover es de 300 metros mientras que la resolución espacial de la imagen Landsat es de 30 metros. Igualmente, que los dos archivos representan dos épocas distintas: Globcover es de 2009 y la imagen en estudio es de 2003. Eso significa que las instrucciones indicadas anteriormente tenían como propósito simplemente ilustrar el procedimiento requerido en el software y, por lo tanto, el resultado obtenido no tiene ninguna validez real.

Como no estamos satisfechos con el resultado, vamos a suponer que conocemos la zona y podemos crear nuestros propios polígonos de entrenamiento. Lo primero que hay que hacer es definir la proyección utilizando GUI Settings en Layer Manager. Luego, hay que ir al Map Display y activar, en el extremo inferior, la opción de usar la proyección definida. Luego, en el extremo derecho del menú, hay que pasar de la opción "2D View" a la opción "Digitize". Enseguida aparecerá la barra de digitalización. En el extremo izquierdo de esa barra, hay una caja con el texto "Select". Allí puede seleccionar la opción de crear un nuevo mapa vectorial. Si así lo hace, aparecerá una nueva ventana en donde se debe indicar el nombre del mapa y el nombre del atributo que va a contener los códigos definidos para cada categoría. Usted puede agregar un nuevo atributo, por ejemplo clase, para ingresar un texto más descriptivo. Luego de crear el nuevo mapa, puede hacer clic en el quinto ícono desde la izquierda (Digitize new area) para empezar a dibujar sus polígonos de entrenamiento.

Para el efecto, se deben dibujar polígonos para las siguientes clases:

1. agua
2. bosque
3. arbusto
4. pasto
5. construccion
6. cultivo
7. vivero
8. nube
9. sombra

Los polígonos de entrenamiento para cada clase se muestran en las figuras 25, 26, 27, 28 y 29 .

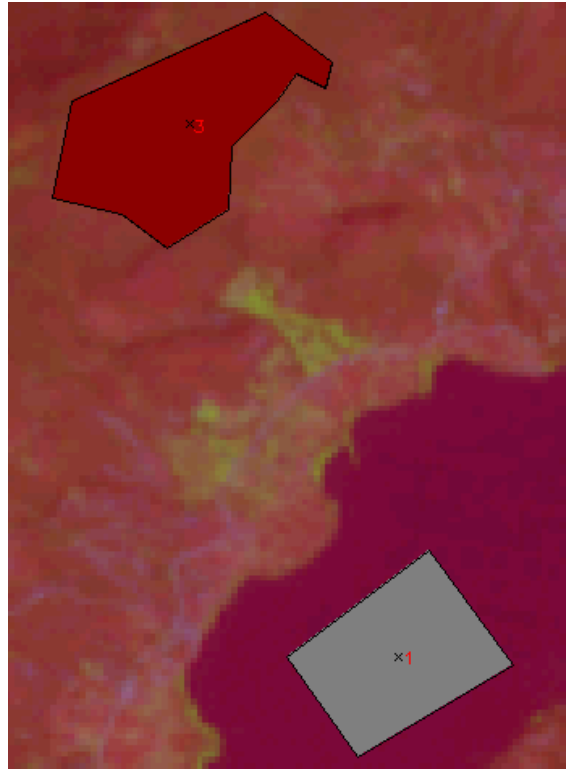


Figura 25. Polígonos de entrenamiento de las clases agua (1) y arbusto (3).



Figura 26. Polígono de entrenamiento de la clase bosque (2).

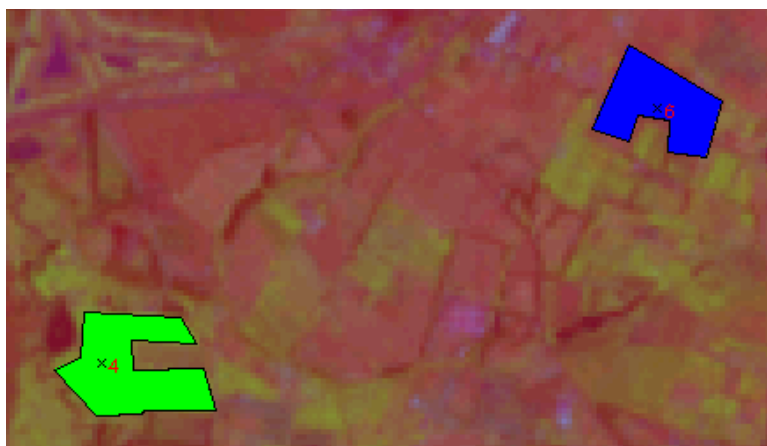


Figura 27. Polígono de entrenamiento de las clases pasto (4) y cultivo (6).

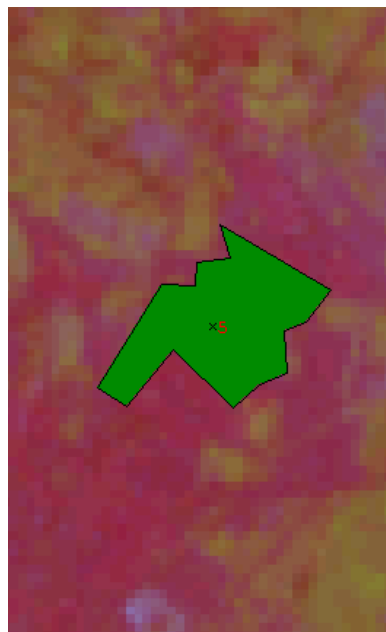
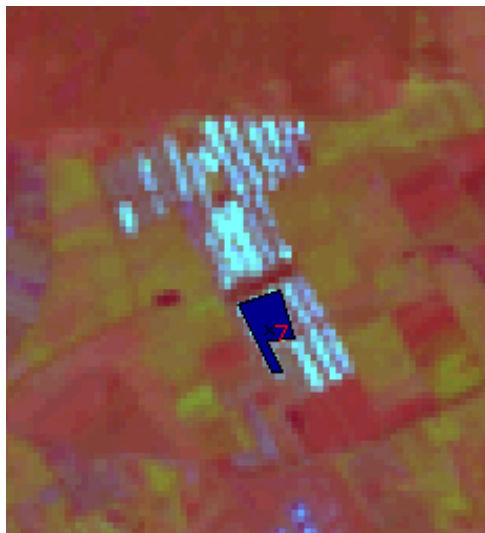


Figura 28. Polígonos de entrenamiento de las clases vivero (7) y construccion (5).

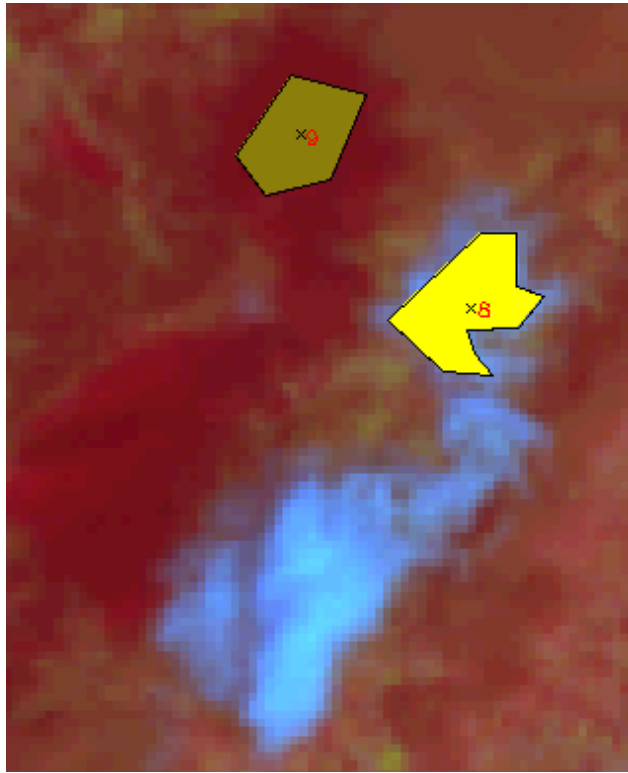


Figura 29. Polígonos de entrenamiento de las clases nube (8) y sombra (9).

Enseguida, hay que convertir el archivo vectorial de entrenamiento en un archivo raster. Para el efecto, se puede usar, en la consola de comandos, el comando *v.to.rast* utilizando instrucciones similares a las siguientes:

```
v.to.rast input="polys@ivan" layer=1 type="point,line,area" output="\rpolys" use="cat" value=1 rows=4096
```

Luego de tener el archivo raster de entrenamiento, se pueden utilizar los comandos *i.gensig* y *i.maxlik* para generar las firmas y realizar la clasificación respectivamente.

La figura 30 muestra los parámetros utilizados para ejecutar el comando *i.gensig*.

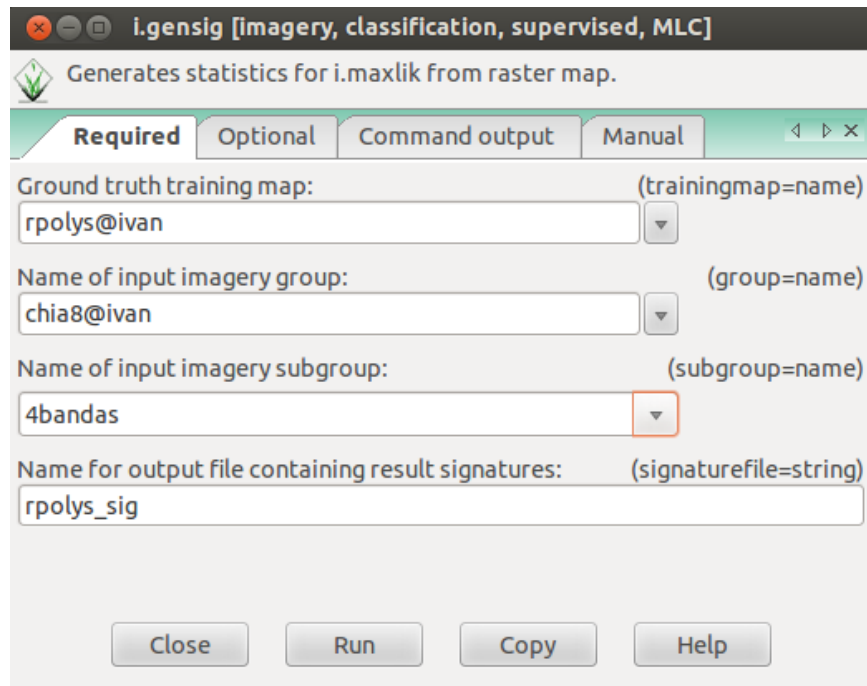


Figura 30. Parámetros utilizados para la creación de las firmas de entrenamiento

La figura 31 muestra los parametros utilizados para ejecutar el comando *i.maxlik*.

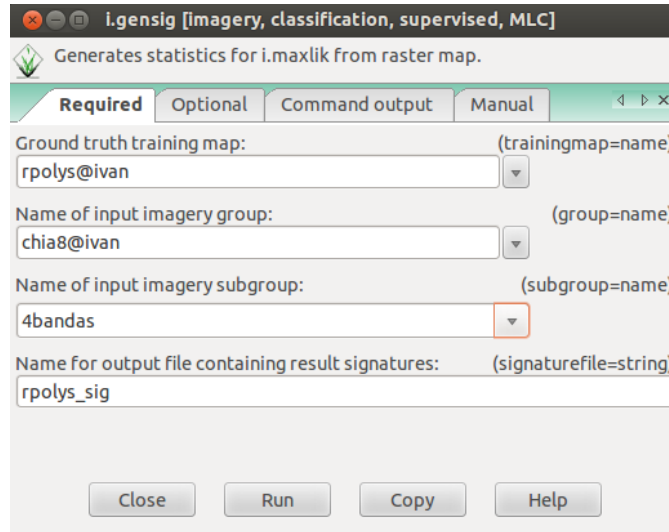


Figura 31. Parametros utilizados para aplicar el algoritmo de maxima probabilidad.

La figura 32 muestra los resultados de la clasificacion. Como se puede ver, los colores asignados a cada clase no parecen muy buenos. Para modificar dichos colores, se requiere crear una tabla de color nueva.

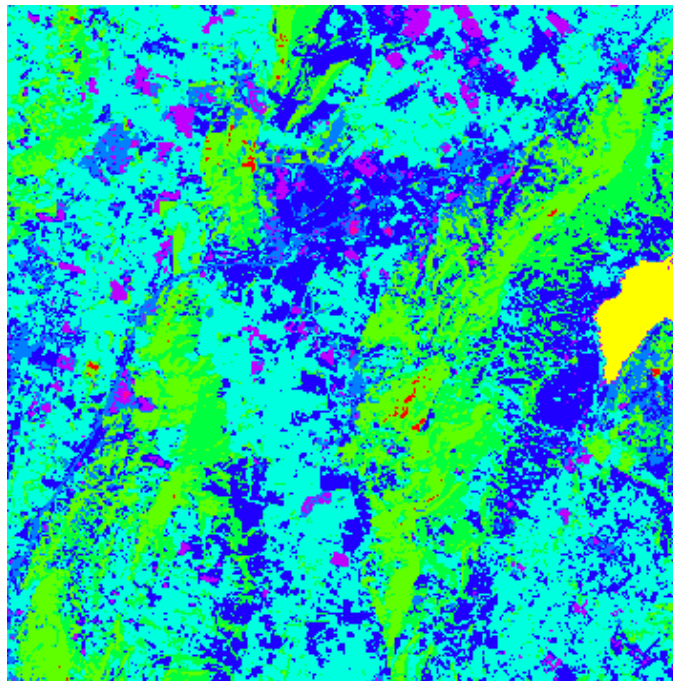


Figura 32. Clasificación digital de la imagen.

La creación de un archivo de tipo texto con las definiciones de los colores deseados para cada clase se puede realizar desde un terminal de Linux. La figura 33 muestra las instrucciones correspondientes. Note que, por ejemplo, a la clase agua (1) se le asigna un color R:G:B igual a 0:0:200, es decir un azul intenso.

```
ivan@ubuntu:~/grasstutoriales/latex$ cat > rules.txt << EOF
> 1 0:0:200
> 2 0:255:0
> 3 0:155:0
> 4 0:75:0
> 5 120:120:0
> 6 0:100:100
> 7 200:200:200
> 8 100:100:0
> 9 50:50:0
> EOF
```

Figura 33. Creación de un archivo con las definiciones de la nueva tabla de color.

La asignación de una tabla de color a un mapa raster específico se puede realizar en GRASS GIS usando el comando *r.colors*. La figura 34 muestra las instrucciones correspondientes para asignar la tabla recién creada al archivo de interés. Los resultados de aplicar esa tabla de color se pueden ver en la figura 35. El lector puede objetar que la tabla de color todavía está defectuosa. Mas aun, puede pensar que hace falta una leyenda de clasificación, una escala y un norte. En todo ello, tiene razón. Por lo tanto, se sugiere que el lector corrija dichos defectos y produzca un mapa de "verdad".

```
r.colors map=chia8_nclas@ivan rules=/home/ivan/grasstutoriales/latex/rules.txt
```

Figura 34. Asignación de una nueva tabla de color al archivo raster que contiene la imagen clasificada.

Usted puede realizar una estimación visual de la calidad de una clasificación mediante la comparación del mapa categórico y de una composición a color de la zona de estudio. Ver, por ejemplo, las ventanas indicadas en la figura 36.

Sin embargo, es conveniente que la evaluación de la calidad se realice mediante una evaluación cuantitativa de la exactitud temática. Para el efecto, GRASS requiere un archivo con muestras de validación (obviamente, este archivo debe ser diferente de las muestras utilizadas para el entrenamiento). Una vez se tenga ese archivo, se puede evaluar la exactitud temática utilizando el módulo *r.kappa* que produce una matriz de error con sus correspondientes indicadores.

El lector puede crear su propio archivo de validación usando el procedimiento que se indicó anteriormente para la creación de un archivo vectorial con las muestras de entrenamiento y su posterior conversión a formato raster. Luego, puede ejecutar la evaluación de exactitud temática.

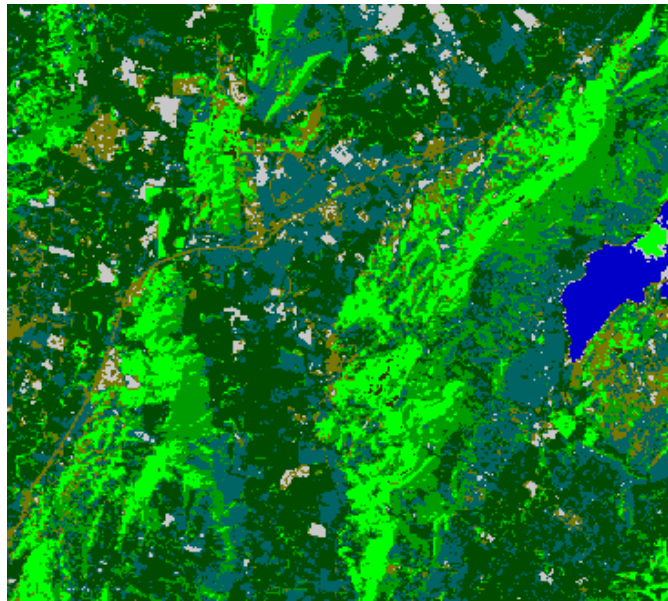


Figura 35. Visualizacion de la imagen clasificada con una nueva tabla de color.

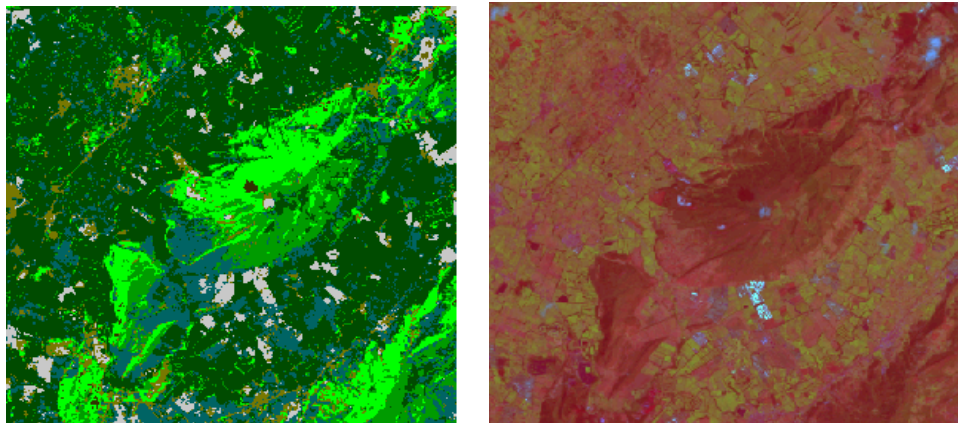


Figura 36. Comparacion entre la imagen clasificada y la imagen original.

Los resultados obtenidos por el autor fueron los siguientes: en la figura 37 se muestran los poligonos de entrenamiento y los poligonos de validacion; en la figura 38 se muestran los parametros de ejecucion del commando *r.kappa*; en la figura 39 se muestran parcialmente los resultados de la evaluacion de la exactitud tematica. Note que se obtuvo un indice Kappa global igual a 0.92.



Figura 37. A la izquierda se encuentran las muestras de entrenamiento. A la derecha las muestras de validacion.

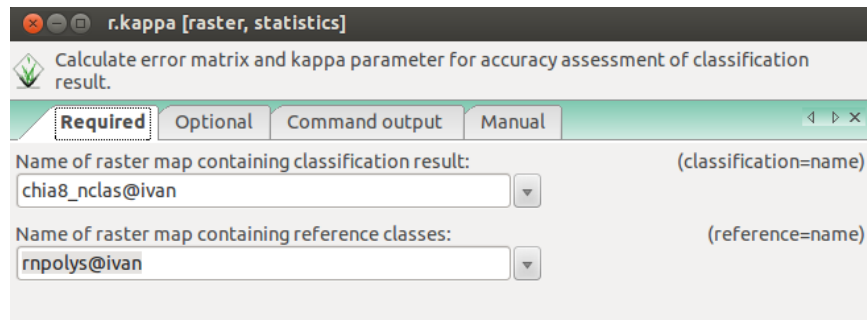


Figura 38. Parametros de entrada para evaluar exactitud tematica.

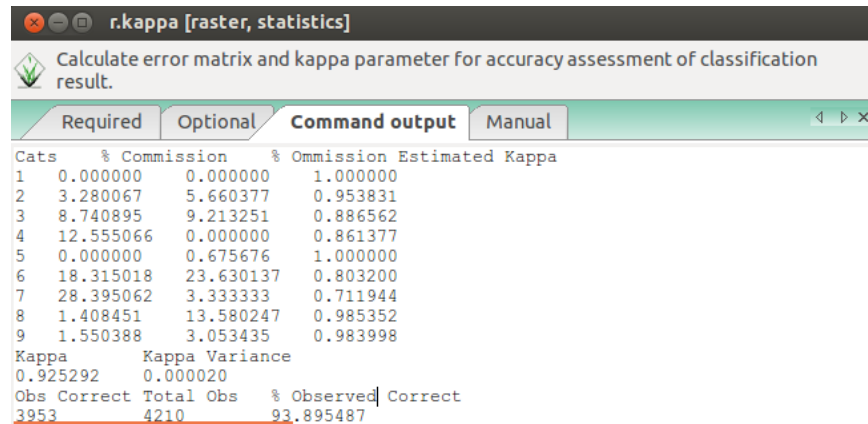


Figura 39. Resultados de evaluación de la exactitud temática.

Si el lector requiere entender a profundidad el tema de exactitud temática puede consultar el siguiente documento: http://www1.usgs.gov/vip/standards/NPSVI_Accuracy_Assessment_Guidelines_ver2.pdf

16.2. Clasificación utilizando el modo interactivo

Enseguida se muestra como realizar el proceso de clasificación supervisada utilizando el modo interactivo.

Para el efecto se deben ejecutar los siguientes pasos (tenga en cuenta que, debido a que este proceso no funcionó en GRASS 7.0, el autor lo realizó utilizando GRASS 6.4.2):

1. en el terminal de GRASS, ingrese la instrucción *d.mon X0* para desplegar un monitor X de ‘viejo estilo’ (este es un visor diferente del visor wxGUI que existe en GRASS 7.0).
2. en el terminal de GRASS, ingrese el comando *i.class* que permite crear las firmas espectrales de cada clase de una manera interactiva. El programa le pregunta, de manera sucesiva, el nombre del mapa raster que va a ser desplegado (ingrese el nombre de una composición a color, por ejemplo *rgb742*), el nombre del grupo de imágenes (ingrese, por ejemplo, *chia8*), el nombre del subgrupo que se utilizará como base (ingrese, por ejemplo, *4bandas*), el nombre el nombre del archivo de salida (ingrese, por ejemplo *4bandas_sup*, y, finalmente, el nombre del archivo con las ‘semillas’ de las firmas (no ingrese nada). El programa despliega luego el mapa raster sobre el cuál el usuario debe ingresar de manera interactiva cada uno de los polígonos de entrenamiento (ver figura 40).
3. en el menú Imagery, seleccione Classify image—Maximum likelihood classification (MLC). En la pestaña Required, indique el nombre del imagery group, por ejemplo *chia8*, el nombre del subgroup, por ejemplo *4bandas*,

el nombre del archivo de firmas, por ejemplo *4bandas_sup*, y el nombre del mapa raster en el cual se van a almacenar los resultados de la clasificación, por ejemplo *chia_sup*. Finalmente, ejecute el proceso correspondiente. Los resultados de la clasificación se muestran en la figura 41)

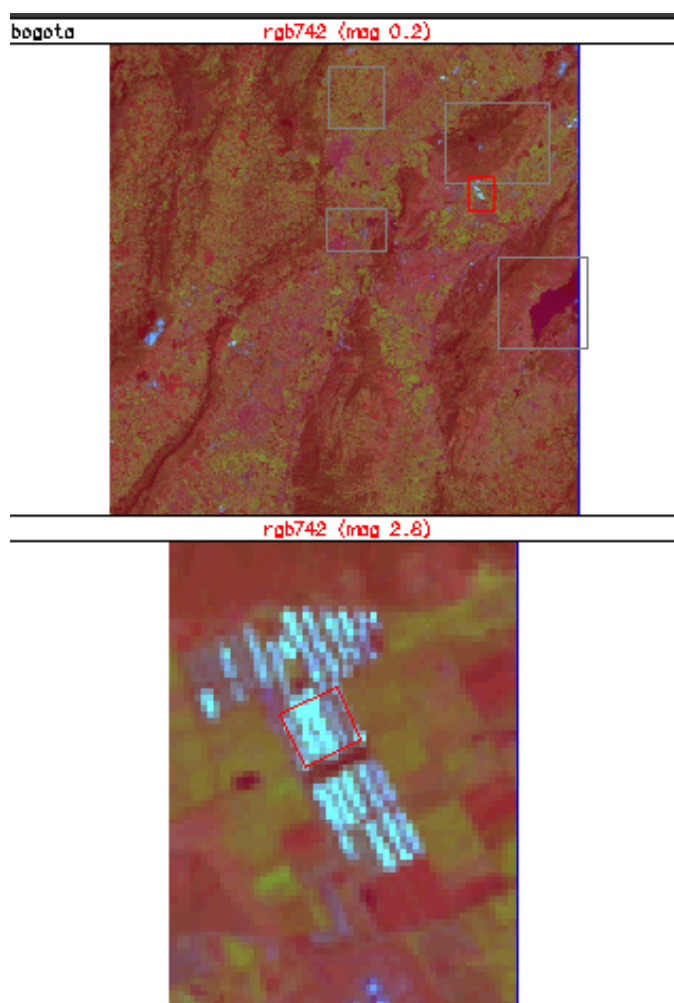


Figura 40. Monitor con la composición a color que sirve de referencia para ingresar los polígonos de entrenamiento.

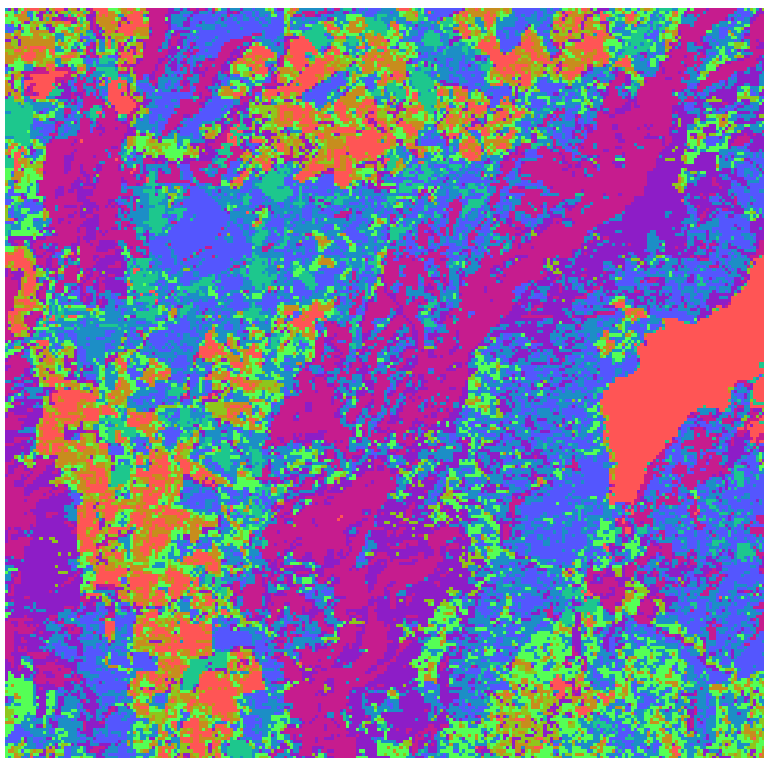


Figura 41. Clasificación de la cobertura del suelo a partir de polígonos dibujados por el usuario.

Nuevamente hay que indicar que es conveniente cambiar la tabla de color para tener una idea mas clara de la calidad de la clasificacion. Igualmente, es necesario evaluar la exactitud temática utilizando el módulo *r.kappa* que produce una matriz de error con sus correspondientes indicadores. Esas dos tareas se le dejan al lector.

17. Integración entre R y GRASS

Las potencialidades del software estadístico *R* pueden ser utilizadas desde una sesión de GRASS. Para poder realizar la integracion entre los dos programas es necesario instalar el paquete base de R, lo mismo que las librerías de R denominadas *sp* y *spgrass6*. Es posible que usted se pregunte cuál es la necesidad de utilizar el programa *GRASS* como intermediario del programa *R*. En opinión del autor, existen dos razones principales: (i) GRASS no tiene suficientes funcionalidades de análisis y modelamiento estadístico, y (ii) R tiene un problema grande de desempeño cuando se requiere procesar datos de gran volumen, como las imágenes de sensores remotos, debido a que dicho programa requiere almacenar los datos en memoria.

La instalación de R y demás librerías requeridas se describe en detalle en la siguiente dirección: http://grass.osgeo.org/wiki/R_statistics

Una vez que haya realizado dicha instalación, se puede ejecutar el programa GRASS y, luego, desde la ventana de comandos en la cual se esté ejecutando dicho programa, se debe iniciar R usando la instrucción *R*. Enseguida, dentro de la sesión de *R*, se debe cargar la librería que permite acceder a las funcionalidades de GRASS usando la siguiente instrucción:

```
library ( spgrass6 ).
```

La importación de un archivo raster existente en GRASS a una sesión de R se realiza utilizando el comando *readRAST6*. Por su parte, la importación de un archivo vector se realiza utilizando el comando *readVECT6*. A manera de ejemplo, enseguida se muestran las instrucciones para importar las siete bandas originales de la imagen de trabajo:

```
chia1 <- readRAST6(" chia_wgs84.1")
chia2 <- readRAST6(" chia_wgs84.2")
chia3 <- readRAST6(" chia_wgs84.3")
chia4 <- readRAST6(" chia_wgs84.4")
chia5 <- readRAST6(" chia_wgs84.5")
chia6 <- readRAST6(" chia_wgs84.6")
chia7 <- readRAST6(" chia_wgs84.7")
```

Si se utiliza el comando *summary(chia1)* se obtienen las características principales del objeto *chia1*, por ejemplo el tipo de objeto (*SpatialGridDataFrame*), las coordenadas mínimas y máximas en las dos dimensiones, su referencia espacial, etc.

Hasta el momento, las siete bandas de la imagen existen de manera independiente. Si se desean ‘agrupar’ las siete bandas se puede utilizar la librería *raster* como se ilustra en las siguientes líneas:

```
#llamado a la libreria requerida
library ( raster )
# conversion de spatial dataframe a raster
rchia1 <- raster ( chia1 )
rchia2 <- raster ( chia2 )
rchia3 <- raster ( chia3 )
rchia4 <- raster ( chia4 )
rchia5 <- raster ( chia5 )
rchia6 <- raster ( chia6 )
rchia7 <- raster ( chia7 )
# remocion de objetos innecesarios
rm ( chia1 , chia2 , chia3 , chia4 , chia5 , chia6 , chia7 )
# creacion de un stack de raster
rchia <- stack ( rchia1 , rchia2 , rchia3 , rchia4 , rchia5 , rchia6 , rchia7 )
```

Si se proporciona la instrucción *rchia* se obtienen las principales características del objeto *rchia*, entre ellas, el tipo de objeto (*RasterStack*), el número de

filas (nrow=1093), el numero de columnas (ncol=1091), el numero de pixeles (ncell=1192463), el numero de bandas (nlayers=7), lo mismo que la resolucio espacial del raster (x=y=0.0002713676).

```
> strm <- readRAST6(c("strm_chia","chia_wgs84.7"), cat=c(TRUE,FALSE),ignore.stder  
r=TRUE,plugin=NULL)
```

Para plotear la banda de elevación en R se usa la siguiente instruccion:

```
> image(strm, attr=1, col=terrain.colors(20))  
> title("Mosaico de elevacion de la zona de estudio")
```

c La figura 42 muestra un ploteo de la banda de elevación en R.

Mosaico de elevacion de la zona de estudio

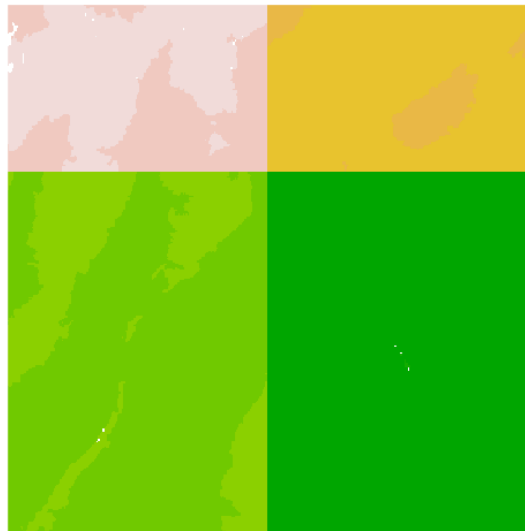


Figura 42. Ploteo de la banda de elevación en R.

Para conocer las características del objeto *strm* se debe usar la instrucción *summary(strm)*. El resultado está desplegado en la figura 44. Observe que *strm* es un objeto de la clase *SpatialGridDataFrame*. Si el lector no está familiarizado con la librería *sp* de R, que permite manejar datos espaciales, se recomienda que revise los conceptos correspondientes en el siguiente enlace cran.r-project.org/web/packages/sp/sp.pdf

```

> summary(strm)
Object of class SpatialGridDataFrame
Coordinates:
      min      max
coords.x1 -74.146239 -73.850177
coords.x2  4.797643  5.094248
Is projected: FALSE
proj4string :
[+proj=longlat +no_defs +a=6378137 +rf=298.257223563
+towgs84=0.000,0.000,0.000]
Number of points: 2
Grid attributes:
  cellcentre.offset  cellsize  cells.dim
1      -74.146103  0.0002713676    1091
2       4.797779  0.0002713676    1093
Data attributes:
  strm_chia  chia_wgs84.7
Min.   : 2533  Min.   : 0.0
1st Qu.: 2773  1st Qu.:127.0
Median : 6560  Median :133.0
Mean   : 7284  Mean   :132.1
3rd Qu.:10701  3rd Qu.:138.0
Max.   :15879  Max.   :168.0
NA's   : 780

```

Figura 44. Características del objeto *strm*.

La librería *sp* proporciona clases y métodos que permiten manejar datos raster y vector desde R.

18. Clasificación de cobertura usando Random Forests

Uno de los métodos más nuevos de clasificación de imágenes es el denominado *Random Forests*. Según *wikipedia* 'Random forest (or random forests) es una combinación de árboles predictores tal que cada árbol depende de los valores de un vector aleatorio probado independientemente y con la misma distribución para cada uno de estos. Es una modificación sustancial de bagging que construye una larga colección de árboles no correlacionados y luego los promedia'.

Las siguientes líneas de código implementan un proceso completo de clasificación de cobertura usando dicho método. El usuario debe abrir una sesión de *R* desde *GRASS* y ejecutar los comandos correspondientes. Algunas líneas están basadas en material desarrollado por el Centro para la Biodiversidad y la Conservación del Museo Americano de Historia Natural.

```
#llamado a las librerias requeridas
library(sp)
library(maptools)
library(rgdal)
library(raster)
library(spgrass6)
library(randomForest)
library(foreign)
library(gstat)
library(mda)
library(vcd)
#lectura de las bandas de la imagen
chia1 <- readRAST6("chia_wgs84.1")
chia2 <- readRAST6("chia_wgs84.2")
chia3 <- readRAST6("chia_wgs84.3")
chia4 <- readRAST6("chia_wgs84.4")
chia5 <- readRAST6("chia_wgs84.5")
chia6 <- readRAST6("chia_wgs84.6")
chia7 <- readRAST6("chia_wgs84.7")
# conversion de spatial dataframe a raster
rchia1 <- raster(chia1)
rchia2 <- raster(chia2)
rchia3 <- raster(chia3)
rchia4 <- raster(chia4)
rchia5 <- raster(chia5)
rchia6 <- raster(chia6)
rchia7 <- raster(chia7)
# remocion de objetos innecesarios
rm(chia1 , chia2 , chia3 , chia4 , chia5 , chia6 , chia7)
# creacion de un stack de raster
rchia <- stack(rchia1 , rchia2 , rchia3 , rchia4 , rchia5 , rchia6 , rchia7)
```

```

# lectura del archivo de poligonos de entrenamiento
vec <- readVECT6(vname="polys",layer=1L, type=NULL, remove.duplicates=TRUE,
ignore.stderr=NULL)
# definicion de sitios de muestreo
p.train <- spsample(vec,500,"random")
# obtencion de clases en los sitios de muestreo
temp <- overlay(vec,p.train)
# creacion de la respuesta del modelo randomForest
response <- temp$cat
# Obtencion de ND de la imagen en cada punto de muestreo
trainvals <- cbind(response, extract(rchia, p.train))
# Ploteo en el espacio espectral
# Definicion de bandas a plotear en espacio espectral
xBand <- 4
yBand <- 7
#
continue <- "c"
while (continue == "c") {
  plotImage <- stack(rchia[[xBand]], rchia[[yBand]])
  #
  cat("Obtencion de valores de pixeles para el ploteo\n\n")
  featurePlotPoints <- sampleRegular(plotImage,100000 )

  # Remocion de valores NA de la tabla trainvals
  featurePlotPoints <- na.omit(featurePlotPoints)

  minBand1 <- min(featurePlotPoints[,1])
  maxBand1 <- max(featurePlotPoints[,1])
  minBand2 <- min(featurePlotPoints[,2])
  maxBand2 <- max(featurePlotPoints[,2])
  rangeBand1 <- maxBand1 - minBand1 + 1
  rangeBand2 <- maxBand2 - minBand2 + 1

  xAxisLabel <- paste("Layer", xBand, sep=" ")
  yAxisLabel <- paste("Layer", yBand, sep=" ")

  plot(featurePlotPoints[,1], featurePlotPoints[,2], col=
"lightgrey", xlab=xAxisLabel, ylab=yAxisLabel)

  uniqueValues <- unique(trainvals[,1])
  for (v in 1:length(uniqueValues)) {
    points(trainvals[which(trainvals[,1]==uniqueValues[v]),
xBand+1],
trainvals[which(trainvals[,1]==uniqueValues[v]), yBand+1], col=v,
pch=20)
  }
}

```

```

    legend(minBand1, maxBand2, col=1:v, pch=20, title="Classes",
legend=as.character(uniqueValues))
#

    continue <- readline(prompt="Entre n para terminar, c para cambiar
bandas o cualquier otra tecla para continuar\n\n")

    if (substr(continue, 1,1) == "n") {
        stop("Proceso terminado por el usuario \n\n", call.=FALSE)
    }
    if (substr(continue, 1,1) == "c") {
        xBand <- as.numeric(readline(prompt="Entre numero de banda
para eje x\n"))
        yBand <- as.numeric(readline(prompt="Entre numero de banda
para eje y\n"))
    }
}
#

```

La figura 45 muestra el despliegue de la imagen de estudio en R usando la tabla de color default. La figura 46 muestra el ploteo espectral de pixeles de entrenamiento en las bandas seleccionadas. Observe que las clases se pueden separar de una manera relativamente facil en estas dos bandas.

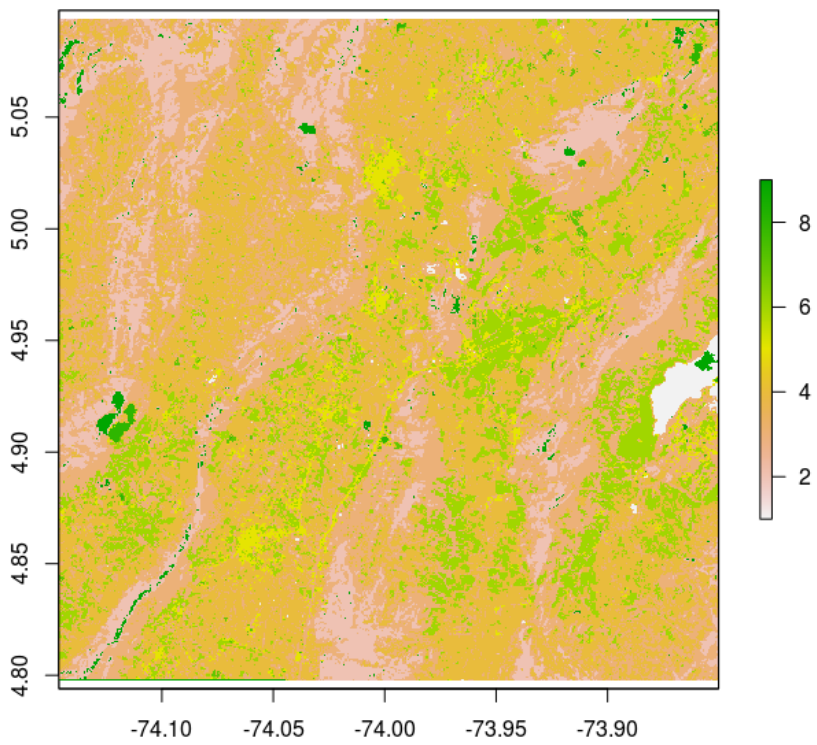


Figura 45. Ploteo de la imagen en R.

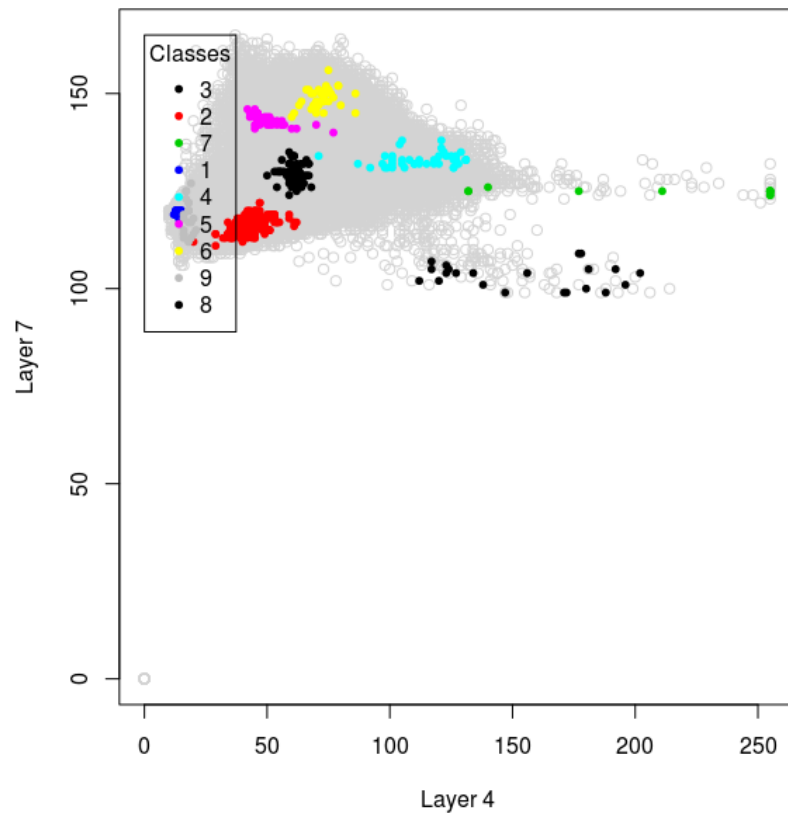


Figura 46. Ploteo espectral de pixeles de entrenamiento en las bandas seleccionadas.

```

# Run Random Forest
cat(" Inicio del proceso random forest\n")
randfor <- randomForest(as.factor(response) ~., data=trainvals ,
importance=TRUE, na.action=na.omit)
#
cat(" Ploteo de la importancia de las variables\n")
varImpPlot(randfor)
#
# Prediction
cat(" Inicio de las predicciones\n")
clas <- predict(rchia , randfor , filename="rchia_rf" , progress='text' ,
format='GTiff' , datatype='INT1U' , type='response' , overwrite=TRUE)
# Plot result
plot(clas)
# Evaluacion de exactitud tematica
# lectura del archivo de poligonos de chequeo
vec2 <- readVECT6(vname="npolys" , layer=1L , type=NULL , remove.duplicates=
TRUE , ignore.stderr=NULL)
# definicion de sitios de chequeo
p.check <- spsample(vec2 , 500 , "random")
# obtencion de clases en los sitios de chequeo
temp <- overlay(vec2 , p.check)
# obtencion de ND de la imagen en cada punto de chequeo
temp$clas2 <- extract(clas , p.check)
# derivacion de la matriz de confusion
conf <- confusion(temp$cat , temp$clas2)
# impresion de la matriz de confusion
conf
# obtencion del valor kappa
k1 = Kappa(conf)
#
k1

```

La figura 45 muestra el despliegue de la imagen de estudio en *R* usando la tabla de color default. La figura 46 muestra el ploteo espectral de pixeles de entrenamiento en las bandas seleccionadas. Observe que las clases se pueden separar de una manera relativamente facil en estas dos bandas.

A estas alturas, el lector se estará preguntando en donde está la imagen clasificada. Revisando el codigo anterior se puede observar que el resultado de la prediccion del modelo RandomForest se almacenó en la variable *clas* y se guardó en la imagen *rchia_rf.tif*. La figura 51 muestra dicha imagen clasificada tal como se despliega en *R*.

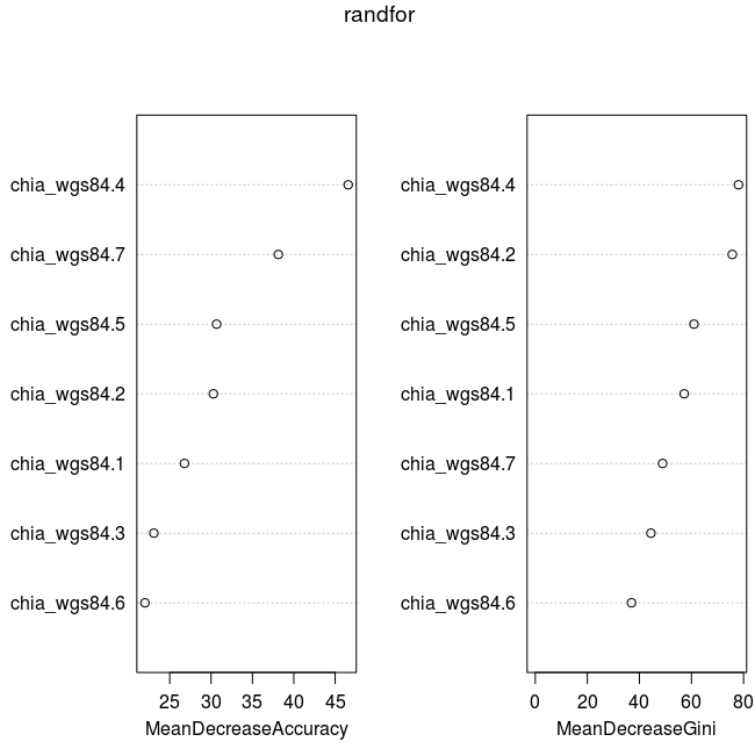


Figura 48. Ploteo de la importancia de las variables.

```

> conf
      true
predicted 1  2  3  4  5  6  7  8  9
1      75  0  0  0  0  0  0  0  0
2      0 132  5  0  0  0  0  0  1
3      0  0 120  0  1  1  0  0  0
4      0  0  0 52  0  0  0  0  0
5      0  0  0  0 40  1  0  0  0
6      0  0  5  3  2 29  0  0  0
7      0  0  0  0  0  0  4  0  0
8      0  0  0  0  0  0  0 15  0
9      0  0  0  0  0  0  0  0 14
attr(,"error")
[1] 0.038

```

Figura 49. Matriz de confusion.

```
> k1 = Kappa(conf)
> k1
           value      ASE
Unweighted 0.953435 0.01047780
Weighted   0.960125 0.05421486
```

Figura 50. Indice kappa de la imagen clasificada.

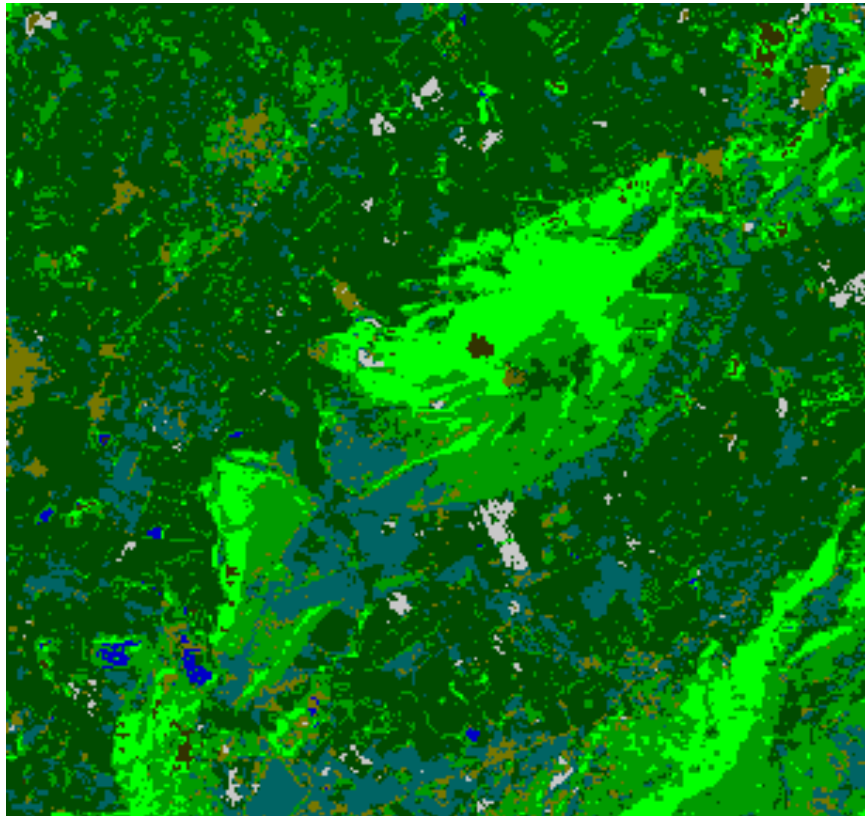


Figura 51. Imagen clasificada usando el metodo RandomForest.

Bueno, esta versión del tutorial termina aquí. Como decía mi abuela, ‘perdonen todo lo malo’. En particular, los problemas de tildes ocasionados por un teclado en inglés y por la decisión de no demorar la publicación de este documento en *GeoTux*. Muchas gracias a Germán Carrillo por esa invaluable labor de promover la difusión de la geoinformática libre!

Espero que el lector se haya motivado a profundizar en estos temas y, ojala, a realizar sus propios experimentos de análisis de imágenes usando *GRASS* y/o *R*. Exitos!