

- Flujo neto de CO₂
- Radiación fotosintéticamente activa absorbida por la planta
- Productividad neta de la vegetación
- Cantidad de lluvia recibida por la vegetación
- Dinámica fenológica
- Evapotranspiración potencial

Otra ventaja es que el índice varía entre -1 y 1 lo que facilita su interpretación, puede señalarse como umbral crítico para cubiertas vegetales un valor de NDVI de 0.1 y para la vegetación densa de 0.5. Cualquier valor negativo implica ausencia de vegetación por lo que debieran reclasificarse a cero.

Una de las aplicaciones más habituales es la elaboración de series temporales de índice de vegetación para estudios de desertificación.

Se han propuesto distintas variantes para mejorar este índice. El objetivo fundamental es eliminar la influencia que el suelo puede tener sobre las reflectividades en el rojo y el infrarrojo y que pueda enmascarar las de la vegetación (lo que resulta especialmente útil en medios semiáridos). Entre las formulaciones alternativas destaca el Índice de Vegetación Ajustado al Suelo (SAVI):

$$SAVI = \frac{\rho_{ir} - \rho_r}{\rho_{ir} + \rho_r + L}(1 + L) \quad (10.3)$$

donde L es la constante que ajusta la línea vegetación-suelo al origen. Si se representa en un gráfico el conjunto de píxeles situando en abscisas ρ_r y en ordenadas ρ_{ir} la línea de suelo es la base del conjunto de píxeles, aparece como una línea recta, de tendencia ligeramente ascendente, que si se prolonga hasta cortar el eje de ordenadas lo hace en el valor de L (figura 8.2).

Otros índices, utilizando otras bandas, permitirían analizar otros aspectos como puede ser la litología, por ejemplo:

- TM_5/TM_7 para detectar rocas alteradas con alto contenido en arcilla, mica y roca sulfatada
- TM_5/TM_4 para aislar cubiertas vegetales de la roca desnuda
- TM_3/TM_1 para detectar materiales con alto contenido en limonita

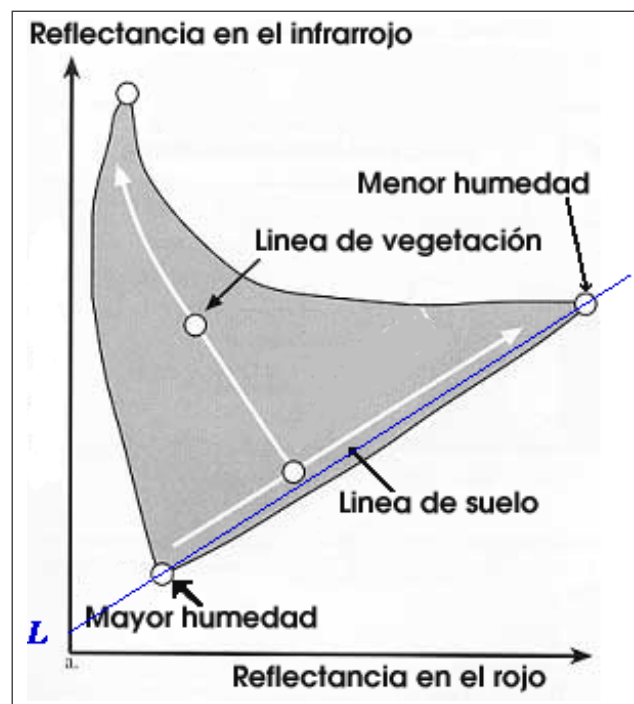


Figura 10.1: Línea de suelo y obtención del parámetro L de SAVI

10.1.2 Transformación Tasseled Cap

El objetivo de esta transformación es obtener unas nuevas bandas por combinación lineal de las originales para realzar determinadas características. Supone pasar de las 6 bandas de landsat (todas excepto la del infrarrojo térmico) a 3 bandas con un claro sentido físico que representan el brillo de la imagen (albedo), el verdor y la humedad. Esta combinación lineal implica una rotación de los ejes definidos por las bandas de forma que los ejes resultantes se corresponden con la línea de suelo (humedad), la perpendicular a esta o sea la línea de vegetación (verdor) y un tercer eje relacionado con el álbedo de la imagen (brillo).

Estas nuevas bandas se obtienen mediante las ecuaciones 8.4 a 8.6, los coeficientes C_{kb} , C_{kv} y C_{kh} de estas ecuaciones aparecen en la tabla 10.1.

$$brillo = \sum_{k=1}^6 C_{kb} \rho_k \quad (10.4)$$

$$verdor = \sum_{k=1}^6 C_{kv} \rho_k \quad (10.5)$$

$$humedad = \sum_{k=1}^6 C_{kh} \rho_k \quad (10.6)$$

El **brillo** refleja los cambios en la reflectividad total de la escena, se trata de una suma ponderada de la reflectividad de todas las bandas excepto el térmico.

El **verdor** muestra el contraste entre las bandas del visible y el infrarrojo próximo, como se puede ver en la tabla 10.1 la banda 4 interviene positivamente mientras que las tres bandas del visible lo hacen negativamente, especialmente la banda 3. Por tanto el resultado es similar a un índice de vegetación. Permite por tanto llevar a cabo análisis de la densidad y estado de salud de la vegetación.

La **humedad** se relaciona con el contenido de agua de la vegetación y el suelo. Se obtiene contrastando la reflectividad en las bandas 5 y 7 en las que el agua absorbe la radiación (con coeficientes de ponderación negativos) respecto al resto de las bandas menos afectadas por la humedad (coeficiente bajos y positivos). También se ha encontrado una relación entre este componente y la madurez de la vegetación. El componente de humedad ha sido también correlacionado con la madurez de las formaciones vegetales.

10.1.3 Con GRASS

Los índices de vegetación y similares pueden programarse fácilmente con **r.mapcalc**.

La transformación tasseled cap también puede programarse, sin embargo GRASS dispone de un módulo específico para ello, **i.tasscap**, de manejo muy sencillo.

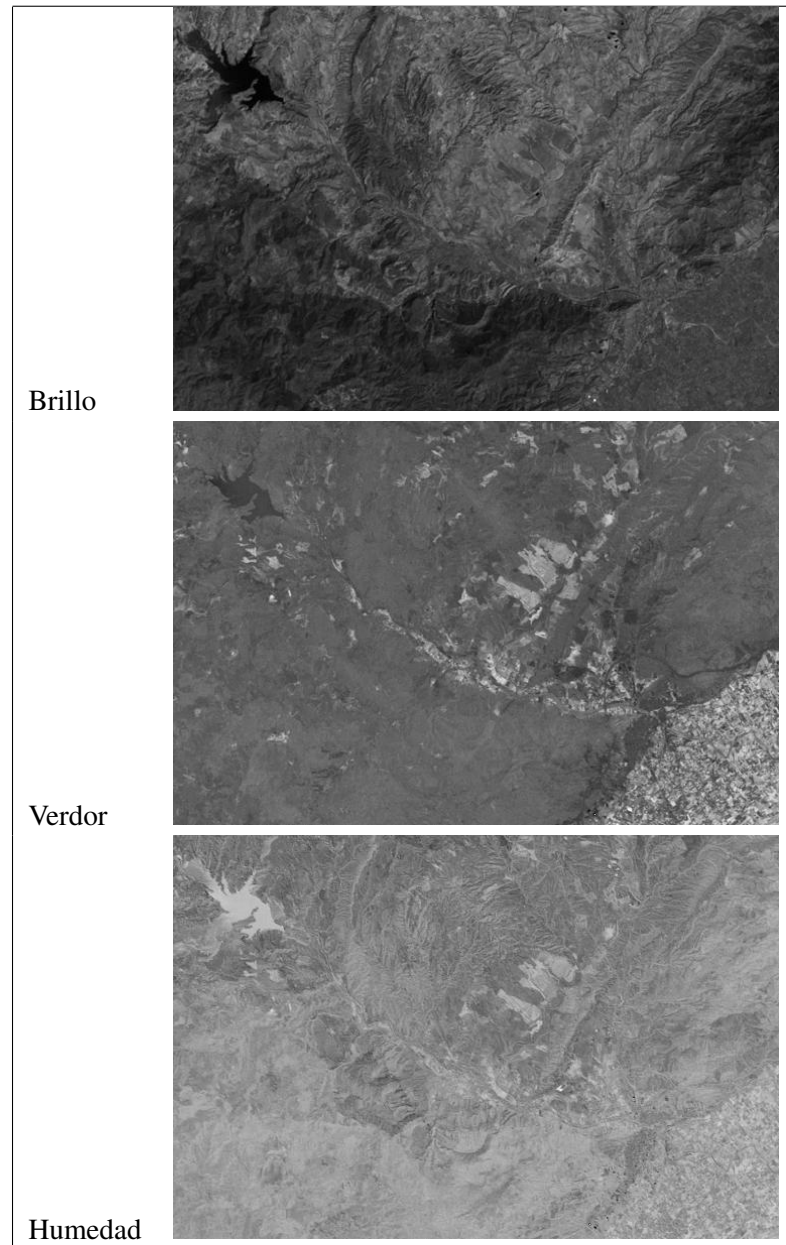


Figura 10.2: Efecto de la transformación tasseled cap

Componente	Brillo	Verdor	Humedad
TM1	0.3037	-0.2848	0.1509
TM2	0.2793	-0.2435	0.1973
TM3	0.4743	-0.5436	0.3279
TM4	0.5585	0.7243	0.3406
TM5	0.5082	0.0840	-0.712
TM7	0.1863	-0.1800	-0.4572

Tabla 10.1: Coeficientes de la transformación tasseled cap

10.1.4 Ejercicios

- Calcula el NDVI y el SAVI, compara los resultados.
- Calcula otros índices y trata de interpretar los resultados
- Utiliza la transformación tasseled cap
- Crea una composición de color con los tres mapas resultantes y determina el significado de los diferentes colores que aparecen.
- Prueba a hacer la composición con los resultados de la transformación reescalados (**r.rescale**) a valores entre 0 y 255 y con una paleta de niveles de gris ecualizados. Analiza las diferencias.

10.2 Análisis de Componentes Principales

Uno de los problemas que aparecen cuando se utilizan técnicas estadísticas para analizar una población es la alta correlación entre algunas de las variables utilizadas. Si el valor absoluto del coeficiente de correlación entre dos variables es muy próximo a 1, ambas variables se sitúan sobre una línea recta con lo que conociendo una se puede conocer la segunda con un margen de error mínimo, por tanto esta segunda variable es redundante e incluirla en un análisis estadístico apenas aporta información nueva (figura 10.3).

Debe por tanto distinguirse entre el número de variables disponibles y la dimensionalidad intrínseca de la muestra de datos; en el ejemplo anterior hay dos variables pero la dimensionalidad es uno.

La siguiente ecuación muestra como calcular el coeficiente de correlación entre las variables x e y (r_{xy}) a partir de la covarianza (COV_{xy}) que se calcula como la media de los productos de las desviaciones de ambas variables respecto a sus medias (m_x y m_y).

$$r_{xy} = \frac{COV_{xy}}{\sigma_x \sigma_y} \quad (10.7)$$

$$COV_{xy} = \frac{\sum_{i=1}^N (x_i - m_x)(y_i - m_y)}{N} \quad (10.8)$$

En el caso de la teledetección, dos bandas situadas muy cerca en el espectro electromagnético tendrán una elevada correlación. Así, en el caso de landsat TM, las tres bandas del visible (muy próximas) aparecen muy correlacionadas, las bandas del infrarrojo cercano (más distantes) aparecen algo menos correlacionadas, y el infrarrojo térmico muy incorrelacionado (en definitiva se trata de una banda de naturaleza completamente diferente). Por tanto la dimensionalidad de una imagen de satélite es inferior al número de bandas.

La ciencia estadística ha desarrollado un conjunto de técnicas para manejar conjuntos de variables correlacionadas entre sí, la más utilizada en teledetección de estas técnicas de análisis multivariable es el Análisis de Componentes Principales (ACP). La justificación teórica del mismo es que, a partir del conjunto de variables puede obtenerse un conjunto de componentes, como combinación lineal de las variables. Estas componentes se caracterizan por:

- Contienen la misma información que las variables originales
- Se conoce el porcentaje de la varianza original que absorbe cada componente
- Los componentes se ordenan de acuerdo con el porcentaje de varianza absorbida

De este modo si se tiene, por ejemplo, n variables pueden seleccionarse los primeros $p < n$ componentes, tal que absorban un porcentaje elevado y fijado de antemano de la varianza, y que además son incorrelacionados. Por tanto podrán aplicarse más fácilmente en algunos análisis estadísticos como la clasificación. El valor de p daría de este modo la dimensionalidad de la imagen.

En el ejemplo de la figura 10.3 un ACP sería equivalente a rotar los datos de manera que la línea de regresión apareciera vertical. De este modo se obtendría un primer componente (en ordenadas) que absorbe la práctica totalidad de la varianza y un segundo componente que absorbe un porcentaje de varianza muy escaso. Por tanto se habrían resumido dos variables en un solo componente.

El resultado de un ACP realizado por cualquier programa de estadística incluye:

- El valor propio (*eigenvalue*, λ_i) de cada componente, que indica la varianza absorbida por el mismo. El porcentaje de varianza absorbida por el componente i se calcula como:

$$V_i = 100 \frac{\lambda_i}{\sum_{j=1}^n n\lambda_j} \quad (10.9)$$

- Una matriz de vectores propios (*eigenvectors*, a) que permiten calcular los valores de cada componente (C_i) a partir de las variables originales (X_j):

$$C_i = \sum_{j=1}^n a_{ij} X_j \quad (10.10)$$

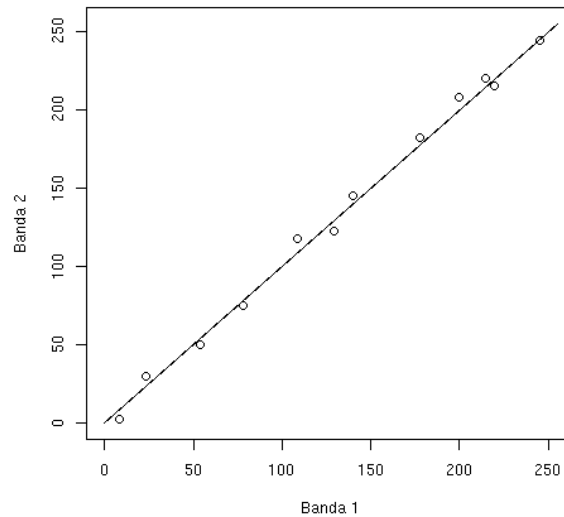


Figura 10.3: Alta correlación entre bandas

Estos valores y vectores propios pueden obtenerse a partir de la matriz de correlaciones o de la matriz de covarianzas. En el primer caso el efecto producido es una estandarización de los datos que resulta adecuada cuando estos no están medidos en las mismas unidades. En el caso de la teledetección deberá utilizarse la matriz de correlaciones cuando se trabaje con ND y la de covarianzas cuando se trabaja con reflectividades. Los módulos de ACP de un programa de SIG deberán proporcionarnos:

- Los mapas de los valores de los componentes, puesto que la varianza de los primeros componentes es mayor que las de los últimos los primeros componentes aparecerán más contrastados que los últimos.
- Los valores propios para determinar el porcentaje de varianza explicada por cada componente.
- La matriz de vectores propios que nos permitirá determinar cual es el peso de cada variable en cada componente. Aunque el ACP es una técnica estadística sin significado físico, puede permitir establecer hipótesis de tipo físico analizando la matriz de vectores propios.

Si la dimensionalidad de la imagen es 3, es decir los tres primeros componentes absorben la mayor parte de la varianza, pueden representarse mediante una combinación de color que permitiría tener en una sola representación gráfica prácticamente toda la información disponible en la imagen.

En ocasiones se nos plantea la posibilidad de reescalar los componentes a valores de 0-255 para facilitar la visualización y su utilización con otras técnicas estandarizadas de teledetección. Hay que tener en cuenta que esta

opción tiene el efecto de igualar el rango de variación de todos los componentes con lo que en la visualización no se apreciarán las diferencias en la varianza absorbida por cada componente.

Las imágenes de los componentes se interpretan bien en blanco y negro. En las figuras 10.4 y 10.5 aparecen los mapas que se obtienen tras un ACP.

Existe la tendencia a considerar a la primera componente como la más relevante, por ser la que mayor varianza absorbe, sin embargo se trata de la varianza general más o menos compartida por todas las bandas y que se relaciona con la reflectividad general de cada pixel. Más interesantes pueden ser el segundo o incluso el tercer componente ya que muestran aquello que resulta distintivo, respecto a la reflectividad general, en cada uno de los píxeles. Por ejemplo en la figura 8.5 el segundo componente parece denotar la vegetación y el tercero la presencia de agua; esta relación se observa con mucha frecuencia en los resultados del ACP en teledetección.

Como se puede comprobar existe una gran relación entre la transformación tasseled cap y el ACP. En el primer caso la matriz de vectores propios vendría predefinida (la matriz de coeficientes de ponderación de la tabla 10.1 mientras que la matriz de la ACP es propia para cada imagen. Sin embargo la tasseled cap tiene una interpretabilidad física directa mientras que en el ACP hay que deducirla de los vectores propios.

Las últimas componentes tienen además cierta relevancia cuando la muestra de datos procede de dos o más poblaciones diferentes con características diferentes. En el caso de la imagen de satélite anteriormente presentada, el Embalse de Puentes constituye un caso completamente distinto, en cuanto a respuesta espectral, que el resto de la imagen de modo que continúa apareciendo incluso en los mapas de los componentes más altos. Una posibilidad sería eliminar el efecto de estos “cuerpos extraños” mediante una máscara.

10.2.1 Con GRASS

El módulo oficial de GRASS para hacer un Análisis de Componentes Principales es **i.pca**. Este módulo tiene el inconveniente de no presentar el vector de valores propios que es importante para interpretar los resultados. Tamoco permite realizar el ACP con la matriz de correlaciones en lugar de con la de covarianzas.

Una opción sería crear un script de R que lea la imagen y utilice la función **princomp** de R

```
# Cargar librerías

library(spgrass6)

# Creación del data frame

d=readRAST6(c("b1"));b1=d$b1
d=readRAST6(c("b2"));b2=d$b2
d=readRAST6(c("b3"));b3=d$b3
d=readRAST6(c("b4"));b4=d$b4
```

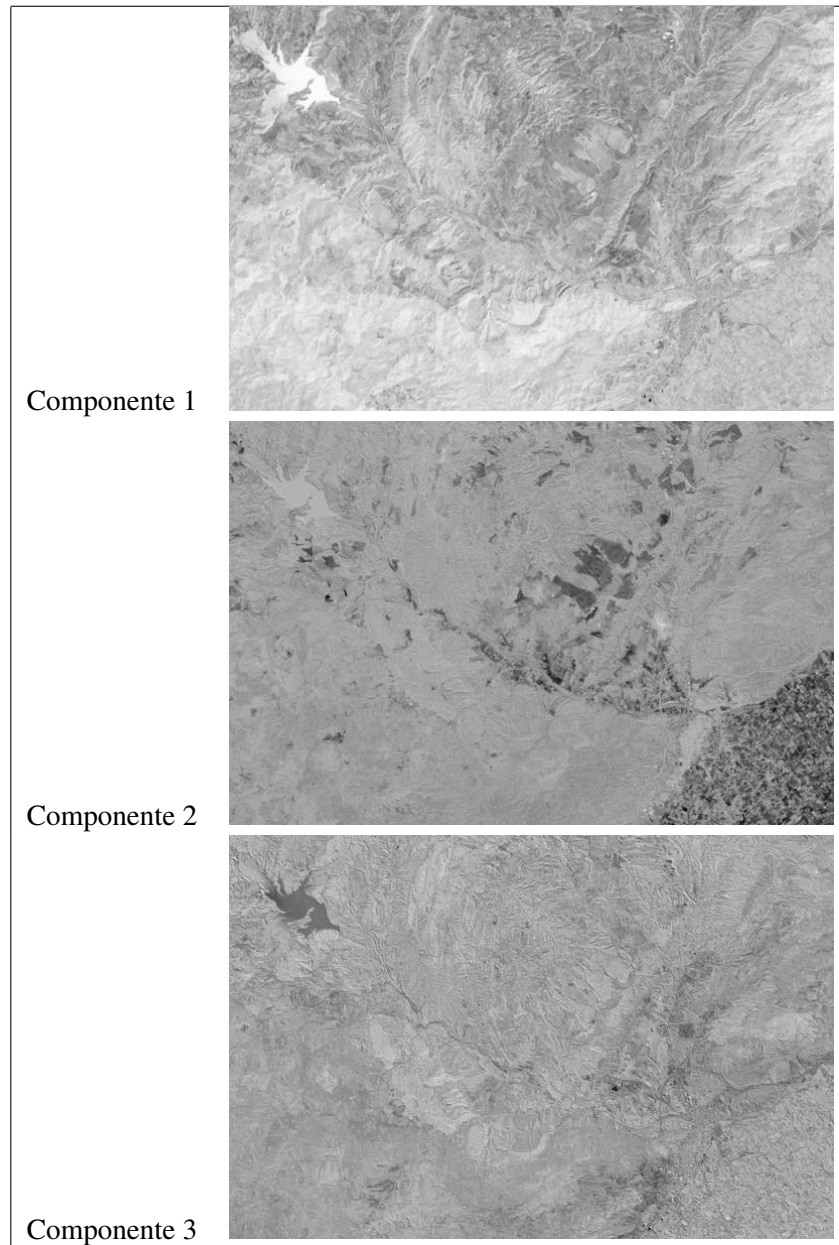



Figura 10.4: Resultados de un Análisis de componentes principales (1)

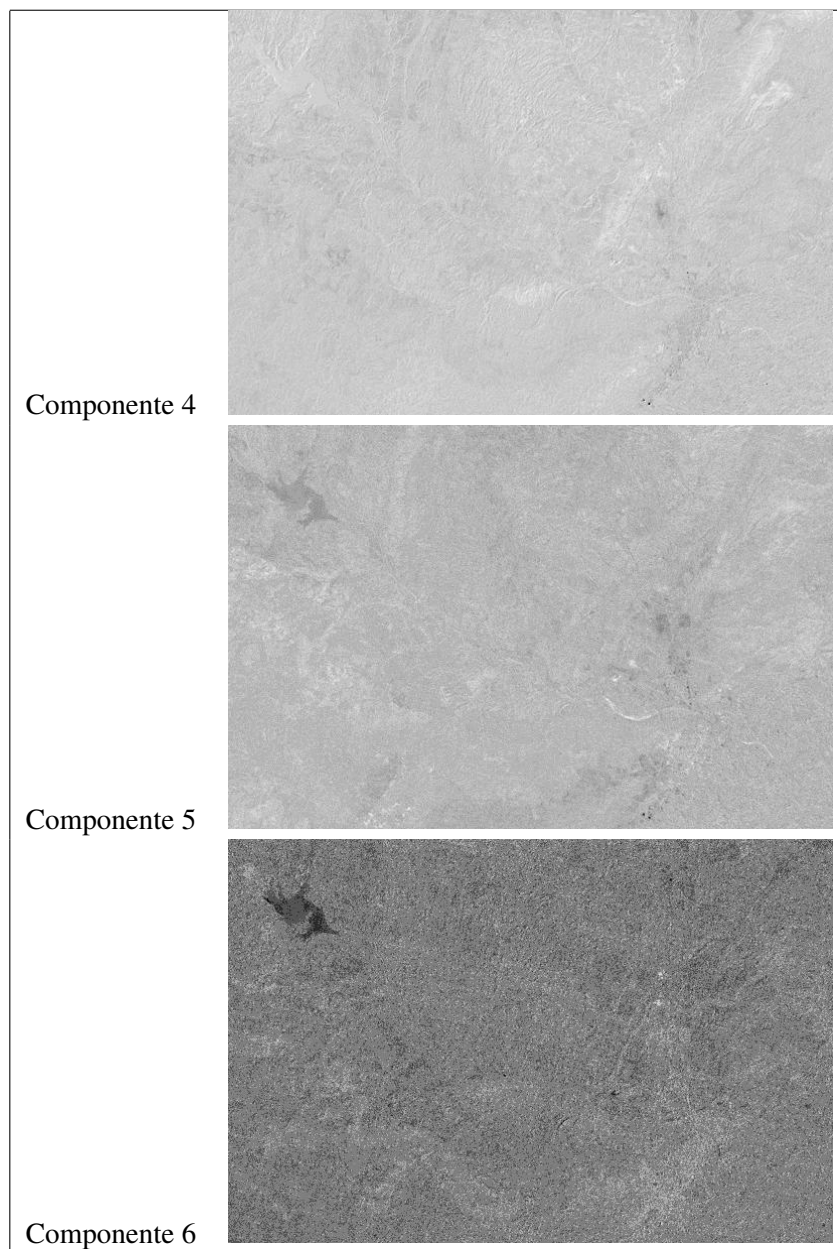


Figura 10.5: Resultados de un Análisis de componentes principales (2)

```

d=readRAST6(c("b5"));b5=d$b5
d=readRAST6(c("b7"));b7=d$b7
imagen=data.frame(b1,b2,b3,b4,b5,b7)

# Análisis de Componentes principales

pc=princomp(imagen)
summary(pc,loadings=T)

# Utiliza los pesos del ACP como coeficientes para generar
# los componentes con r.mapcalc

n=6
k=rep(NA,n*n);dim(k)=c(n,n)
for (i in 1:n){
  for (j in 1:n){
    k[j,i]=pc$loadings[j,i]
  }
  orden=sprintf("r.mapcalc 'pc.%d= %g*b1 + %g*b2 + %g*b3 + %g*b4 + \
                %g*b5 + %g*b7'", i,k[1,i], k[2,i],\
                k[3,i], k[4,i], k[5,i], k[6,i])

  cat(orden,"\n")
  system(orden)
}

```

10.2.2 Ejercicios

- Haz un análisis de componentes principales de tu imagen.
- ¿Con cuantos componentes te quedarías?
- Compara los pesos resultantes del ACP con los coeficientes de la transformación *tasseled cap*
- ¿Puedes asimilar los componentes resultantes a alguna variable física?
- En ese caso trata de hacer una composición de color razonable y analiza los resultados.

Tema 11

Clasificación

En el tema anterior se ha visto como pueden obtenerse medidas cuantitativas de una serie de factores que condicionan la respuesta espectral de los píxeles. Este enfoque responde a la idea de que existe un continuo y una transición suave entre los diferentes valores de reflectividad. La clasificación, por el contrario, responde a la idea de que pueden establecerse límites precisos entre las respuestas espectrales de unos píxeles y otros, siendo estos límites los que definen las clases.

Un ejemplo sencillo sería pensar en un experimento consistente en medir el peso y la altura del conjunto de alumnos de la Universidad de Murcia y representar los valores medidos en un gráfico con el peso en abcisas y la altura en ordenadas. ¿Aparecerían todos los puntos dispersos de manera que hubiese una transición suave entre las diferentes complejiones, o bien los puntos se aglomerarían en áreas del gráfico definiendo clases (elfos, orcos, enanos, hobbits y hombres)? En definitiva el problema es similar a la dicotomía entre objetos y superficies que aparece en el trabajo con SIG.

11.1 Introducción a las técnicas de clasificación

La elaboración de una clasificación del área de estudio en una serie de clases relativas a litología, tipos de vegetación, usos del suelo, etc., es uno de los objetivos fundamentales en teledetección.

La clasificación en teledetección es, por tanto, un caso particular del problema general de clasificar N individuos procedentes de una muestra en un conjunto de $M < N$ clases en función de una serie de K variables (X_1, X_2, \dots, X_K). Este problema se resuelve en dos fases:

1. Determinación del número de clases y de las propiedades de estas en relación a las K variables (sección 9.2)
2. Asignar cada uno de los N individuos a una de las M clases utilizando una regla de decisión basada en las propiedades de los individuos y las clases en relación a las K variables (sección 9.3 y posteriores).

En teledetección el conjunto de variables está compuesto por los ND o la reflectividad en cada una de las bandas. Sin embargo, además de esta información espectral puede utilizarse información textural e información contextual.

La información textural hace referencia a la caracterización de la vecindad de un píxel, especialmente en relación a la variabilidad de la respuesta espectral en distancias cortas. Se definen así una serie de variables (reflectividad media, varianza, autocorrelación, semivariograma etc.) que tratan de cuantificar algunas de las propiedades cualitativas relacionadas con la textura que se estudian en fotointerpretación.

El contexto por su parte hace referencia a la necesaria coherencia en cuanto a la clasificación de píxeles vecinos y en relación a las variables ambientales que caracterizan la celdilla que estamos tratando. No tendría, por ejemplo, mucho sentido clasificar un píxel como cultivo de invernadero al lado de píxeles clasificados como roquedo de alta montaña, ni tampoco localizar áreas de regadío en lo alto de Sierra Espuña.

La incorporación de información textural y contextual tiene, sin embargo, el inconveniente de introducir variables que no suelen seguir una distribución gaussiana con lo que algunos métodos de clasificación no funcionan correctamente.

11.1.1 Interpretación geométrica de la clasificación

Uno de los modos más sencillos de entender la clasificación cuando se basa en medidas cuantitativas es pensar en el conjunto de variables que caracterizan a individuos y clases como un conjunto de ejes que definen un espacio de variables multidimensional. Cada uno de los elementos a clasificar se representa como un punto en dicho espacio.

La figura 11.1 muestra un ejemplo con una imagen de satélite y en un espacio de variables simplificado a las bandas 1 (azul), 3 (rojo) y 4 (infrarrojo cercano) de landsat TM. Cabe pensar que todos los píxeles correspondientes a una misma clase se agruparan en regiones concretas del espacio de variables, si esto es así las clases son fácilmente separables por cualquier procedimiento y el resultado será bueno. Si no es así se producirán errores independientemente del método de clasificación escogido.

A partir de este momento, por simplificar, se presentará un ejemplo en el que se trabaja con las bandas 3 y 4 de landsat (rojo e infrarrojo cercano) para clasificar un conjunto de píxeles (figura 11.2). En este ejemplo los píxeles que hay que clasificar aparecen en color negro, mientras que de algunos píxeles conocemos su clase (clasificación supervisada) y aparecen en el color correspondiente a dicha clase.

11.2 Generación de clases

La determinación de las clases puede hacerse básicamente mediante dos criterios:

1. **Clasificación supervisada.** Se parte de un conjunto de clases conocido *a priori*. Estas clases deben caracterizarse en función del conjunto de variables mediante la medición de las mismas en individuos cuya pertenencia a una de las clases no presente dudas (**áreas de entrenamiento**).

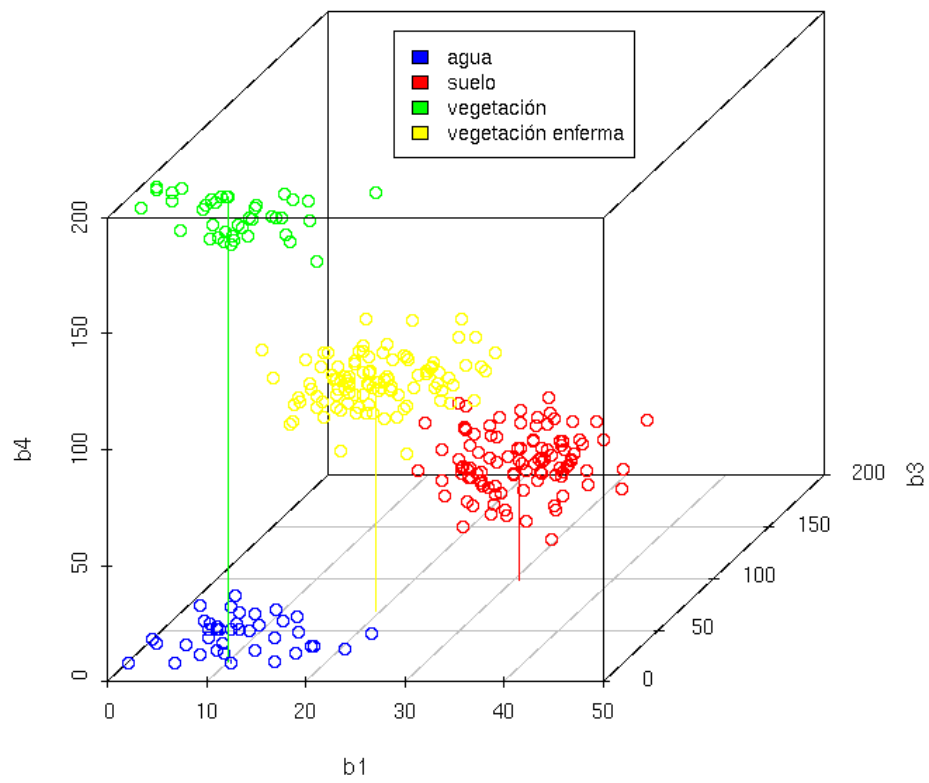


Figura 11.1: Definición de clases en el espacio de variables