

Aplicación del análisis de componentes principales para representar datos usando la información de índices espectrales

| | |
|--------------------------|---|
| Apellidos, nombre | Balaguer Beser, Angel ¹ (abalague@mat.upv.es) Arcos Villacís, María Alicia ² (maar12m@doctor.upv.es) |
| Departamento | ¹ Departamento de Matemática Aplicada ² Departamento de Ingeniería Cartográfica, Geodesia y Fotogrametría |
| Centro | E.T.S.I. Geodésica, Cartográfica y Topográfica |

1 Resumen de las ideas clave

En este artículo se muestra el proceso de obtención de las componentes principales usando el programa Statgraphics Centurion XVIII y el software estadístico R. Para ello se usan datos reales de índices espectrales obtenidos a través de los satélites Sentinel-2, en una zona forestal en la provincia de Valencia (España). Se aplican dos métodos para determinar el número óptimo de componentes principales y se analiza el significado de las componentes seleccionadas. También se describen distintos usos de las componentes principales, viendo las ventajas de sustituir las variables originales por las componentes principales, la cuales son combinación lineal de las primeras y están incorreladas.

2 Objetivos

Después de leer con detenimiento este documento, el lector será capaz de:

- Calcular las componentes principales usando índices espectrales obtenidos con imágenes de satélite.
- Calcular el porcentaje de varianza explicado por las primeras componentes principales y determinar el número óptimo de componentes principales.
- Analizar el significado de las primeras componentes principales.
- Analizar la separación de los datos usando las componentes principales.
- Usar el programa Statgraphics Centurion XVIII y el software R para realizar los cálculos necesarios para alcanzar los objetivos anteriores.

3 Introducción

El análisis de componentes principales (ACP) transforma las variables originales de estudio en otras incorreladas que pueden escribirse como combinación lineal de las primeras. Matemáticamente, dado un conjunto de p variables: X_1, X_2, \dots, X_p , la primera componente principal se define como:

$$Y_1 = t_{1,1}X_1 + t_{2,1}X_2 + \dots + t_{p,1}X_p \quad (1)$$

siendo $t_{j,1}$ los parámetros que se tendrán que estimar. Dichos coeficientes se calculan para que la varianza de la variable Y_1 alcance el máximo valor ($\text{Var}(Y_1)$ máxima) condicionado a que el vector formado por dichos coeficientes, $t_1 = (t_{1,1}, t_{2,1}, \dots, t_{p,1})^T$ tenga norma 1 ($t_1^T t_1 = 1$). Para calcular la componente número k , se busca una dirección de proyección, $t_k = (t_{1,k}, t_{2,k}, \dots, t_{p,k})^T$, tal que:

$$Y_k = t_{1,k}X_1 + t_{2,k}X_2 + \dots + t_{p,k}X_p \quad (2)$$

de modo que sea una variable incorrelada con todas las componentes principales anteriores: $\text{Cov}(Y_k, Y_i) = 0, \forall i = 1, 2, \dots, k-1$, con varianza máxima después de las componentes anteriores ($\text{Var}(Y_k)$ máxima), condicionado a que se cumpla la restricción $t_k^T t_k = 1$. De esta forma, la suma del cuadrado de todos los coeficientes que definen cada componente principal es siempre igual a 1.

Con este procedimiento se pretende que las componentes principales contengan la misma información que las variables originales y eso se consigue al usar combinaciones lineales de las mismas. El hecho de que las componentes principales estén incorreladas indica que las componentes principales no tienen relación lineal entre ellas. Cada componente principal es un nuevo indicador que sirve para ordenar los individuos separando lo máximo posible unos individuos de otros pues la varianza de cada una es la máxima posible. Cuando sea difícil ver el significado de las componentes principales, vamos a identificar cada componente principal con un subconjunto de variables originales, con las que se pierda la menor cantidad de información posible. En los casos en que podamos ver el significado de las componentes principales, cada componente principal será un factor subyacente que no se puede observar por sí solo pero que están detrás de las respuestas.

La mayor aplicación del ACP está centrada en la de reducción de la dimensión del espacio de los datos para simplificar el problema que se estudia. Se pretende identificar un conjunto de variables con menor dimensión que el conjunto de variables originales, con las que se pierda la menor cantidad de información posible. Esta técnica es muy útil cuando se trabaja con datos extraídos con herramientas de Teledetección pues puede aplicarse, por ejemplo, como realce previo a la interpretación visual o como procesamiento anterior a la clasificación [1], incrementando la eficiencia computacional de dichas metodologías al reducir la dimensionalidad de los datos.

Cuando interesa detectar algún aspecto específico de la superficie terrestre pueden utilizarse índices espectrales calculados mediante una transformación de las bandas. Sin embargo, existe una alta correlación entre diferentes índices espectrales y en ocasiones la introducción de varias variables en los modelos de regresión no aporta más información para estimar un nuevo parámetro [2]. Las componentes principales están ordenadas de mayor a menor en relación con el porcentaje de la varianza total de los datos que pueden representar. Ello permite eliminar las componentes que representen un escaso porcentaje de la varianza original y quedarnos con las que tengan mayor variabilidad que además nos sirvan para ordenar los datos.

4 Desarrollo

Para este trabajo se calcularon varios índices espectrales a través de la información de bandas obtenidas de los satélites Copernicus Sentinel-2 A y B, en 135 puntos de una zona forestal de la provincia de Valencia (España), elegidos en una malla regular con una separación de 1 km. La descarga de las bandas se realizó mediante scripts de programación dentro del entorno de Google Earth Engine 14, empleando una ventana de trabajo de 10 m, mientras que el cálculo de los índices espectrales se hizo usando el software R aplicando las fórmulas descritas en la referencia [3]. El producto con el que se trabajó (Nivel 2A) está ortorrectificado y corregido atmosféricamente a la reflectancia de la superficie. Dada la resolución espacial del satélite empleado, se trabajó con cada índice a la máxima resolución posible, remuestreando a 10 m las bandas con resolución de 20 m y 60 m. En cada localización espacial se calculó la mediana los valores obtenidos en todas las fechas disponibles durante 18 meses consecutivos (aproximadamente cada 5 días entre junio de 2020 hasta noviembre de 2021). Los índices obtenidos recogen diferentes características como el verdor de la vegetación (usan la banda verde), la actividad fotosintética (usan la banda roja) o el contenido de agua del suelo y la vegetación (usan bandas SWIR), como se observa en la tabla 1.

| Factor relacionado | Índices Espectrales |
|-------------------------|-----------------------------------|
| Verdor de la vegetación | Vgreen, TCARI, TCARI_OSAVI |
| Actividad fotosintética | NDVI, EVI, VARI, SAVI, OSAVI, RVI |
| Contenido de agua | SLA, NDMI, NMDI, NDWI |

Tabla 1. Tipos de índices espectrales teniendo en cuenta las bandas que intervienen en su definición.

Con Statgraphics Centurion XVIII se puede usar el procedimiento: **Describir + Datos Numéricos + Análisis multivariado**. Luego se elige el gráfico de correlación para obtener coeficientes de correlación de Pearson entre las variables (tabla 2). En dicha tabla 2, en rojo están marcados los pares de variables que tienen un índice de correlación más cercano a 1 (correlación lineal positiva) y en azul los que tienen un índice más cercano a -1 (correlación lineal negativa). Vemos que las variables relacionadas con la actividad fotosintética tienen un índice de correlación entre ellas superior a 0,83. Además, también existe una alta relación entre ellas y tres de las variables (SLA, NDMI y NDWI) relacionadas con el contenido de agua. Las variables relacionadas con el verdor de la vegetación (TCARI y Vgreen) tienen correlación lineal negativa entre ellas, pero guardan poca relación con las variables relacionadas con el contenido de agua. También hemos considerado la variable TCARI_OSAVI definida como el cociente entre esos dos índices espectrales (TCARI/OSAVI) la cual guarda más relación con dichos índices de agua. Por su parte NMDI es un índice de sequía multibanda normalizado que solo alcanza una correlación superior a 0.5 con los otros índices de agua. Correlaciones relativamente altas entre las variables originales es un buen indicador para usar el ACP para reemplazar este conjunto de 13 variables originales por un conjunto más pequeño.

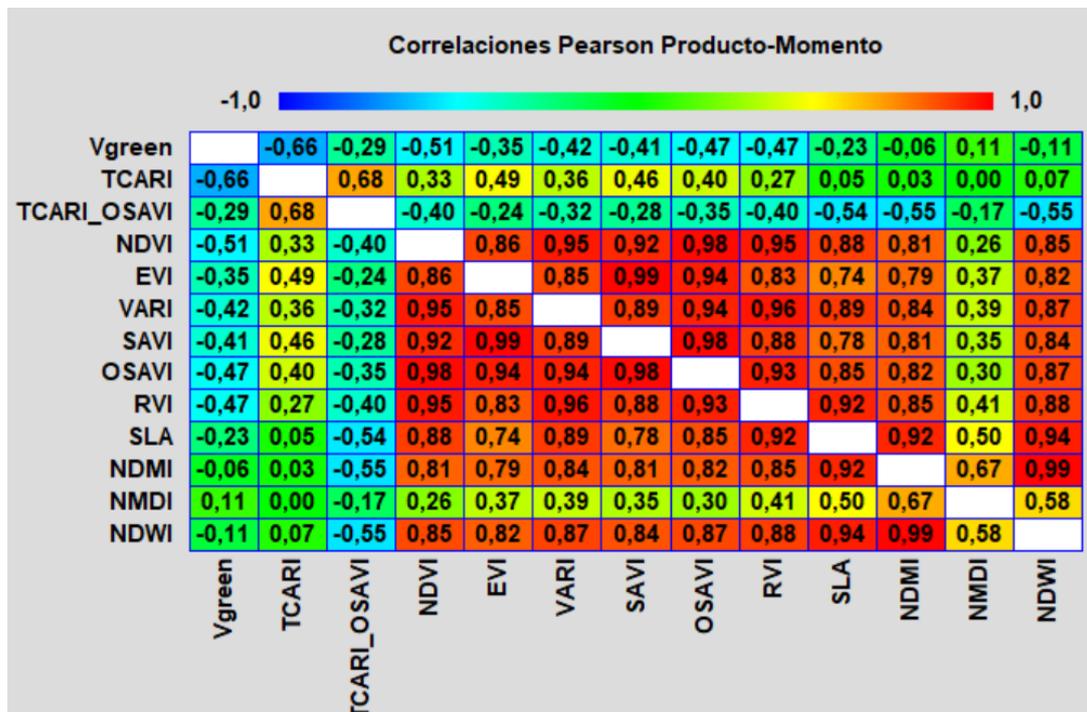


Tabla 2. Valores de los índices de correlación de Pearson entre todas las variables

4.1 Uso de componentes principales normadas.

En el análisis de componentes principales es particularmente importante que las variables sean homogéneas, en el sentido de que ninguna tenga una varianza muy grande en comparación con las restantes. Si esto sucediese siempre identificaríamos la primera componente principal con dicha variable. A menudo las diferencias en las varianzas entre las distintas variables no son más que una consecuencia inmediata de las unidades en que se miden y, por tanto, no es real que unas variables sean más informativas que otras. Por ese motivo conviene mirar el coeficiente de variación que es igual a la desviación típica dividido por la media y mide los dos aspectos anteriores a la vez. Con el Statgraphics Centurion XVIII se puede calcular dicho coeficiente mediante el procedimiento: **Describir + Datos Numéricos + Análisis multivariado**.

La tabla 3 muestra los resultados para cada índice, observando una diferencia importante entre el coeficiente de variación de unos índices a otros. Por tanto, tendremos que trabajar con variables estandarizadas obtenidas tras restar la media y dividir por la desviación estándar a cada variable. Es lo que se conoce por uso de componentes normadas.

| Índice | Vgreen | TCARI | TCARI_OSAVI | NDVI | EVI | VARI | SAVI |
|--------|--------|--------|-------------|---------|--------|---------|--------|
| CV (%) | 18,57% | 25,49% | 28,68% | 23,40% | 19,43% | 325,06% | 18,60% |
| Índice | OSAVI | RVI | SLA | NDMI | NMDI | NDWI | |
| CV (%) | 20,25% | 35,24% | 26,01% | 301,48% | 6,34% | 57,57% | |

Tabla 3: Valores del coeficiente de variación de cada índice espectral expresado en porcentaje.

4.2 Cálculo de las ecuaciones de las componentes principales.

Para calcular las componentes principales con el Statgraphics usaremos el procedimiento: **Describir + Métodos multivariados + Componentes principales**. En nuestro caso, en las opciones marcaremos la opción estandarizar, según lo comentado en el apartado anterior. La tabla 4 muestra los coeficientes de las 4 primeras componentes principales calculadas usando las ecuaciones (1) y (2), teniendo en cuenta las propiedades descritas anteriormente en la introducción. En base a dichas propiedades, ¿existe algún tipo de relación lineal entre dichas componentes principales? En realidad, los coeficientes de dicha tabla 4 corresponden a los vectores propios normalizados asociados a los 4 valores propios de más valor de la matriz de correlaciones de las variables originales.

| | Componente | Componente | Componente | Componente |
|-------------|------------|------------|-------------|------------|
| | 1 | 2 | 3 | 4 |
| Vgreen | -0,128708 | -0,480877 | 0,240414 | 0,635435 |
| TCARI | 0,102449 | 0,579524 | 0,248318 | 0,204693 |
| TCARI_OSAVI | -0,140916 | 0,501404 | 0,409031 | 0,01498 |
| NDVI | 0,326156 | 0,0712829 | -0,220452 | -0,0574053 |
| EVI | 0,311108 | 0,114917 | 0,0684257 | 0,437199 |
| VARI | 0,325517 | 0,0570804 | -0,0397955 | -0,0591107 |
| SAVI | 0,321606 | 0,110313 | -0,00979152 | 0,319424 |
| OSAVI | 0,330218 | 0,0916114 | -0,125802 | 0,139028 |
| RVI | 0,327332 | 0,0223425 | -0,0925745 | -0,2358 |
| SLA | 0,31484 | -0,149561 | -0,028834 | -0,228673 |
| NDMI | 0,311686 | -0,207579 | 0,186716 | 0,0156161 |
| NMDI | 0,160862 | -0,205895 | 0,768071 | -0,351841 |
| NDWI | 0,319662 | -0,176201 | 0,0908213 | 0,0536392 |

Tabla 4. Tabla de los pesos de cada una de las variables en las cuatro primeras componentes principales.

En la tabla 4, podemos observar las variables que tienen más peso en cada una de las componentes principales (tabla de pesos de las componentes). ¿Cuánto vale la suma de todos los coeficientes elevados al cuadrado de cada una de las columnas de la tabla 4? Para contestar a dicha pregunta se pueden revisar las propiedades de los coeficientes de las componentes principales comentadas en la introducción. Si usamos esas columnas para graficar resultados podemos obtener las gráficas de pesos de las componentes que aparecen en la figura 1. La gráfica de arriba a la izquierda de dicha figura es el gráfico de componentes 3D obtenido con Statgraphics Centurion, con el mismo procedimiento usado para obtener la tabla 4. En ella se observa que las variables Vgreen, TCARI, TCARI_OSAVI y NMDI están separadas del resto. Statgraphics también permite obtener los gráficos 2D. Sin embargo, para las gráficas 2D de la figura 1 hemos usado los resultados obtenidos con la función **PCA** de la librería **FactoMineR** del software R, incorporando los datos como un `data.frame` y siguiendo esta sintaxis:

```
df<-as.data.frame(datos)
```

```
respca2<- PCA(X=df, scale.unit=TRUE, ncp=13, graph=TRUE, axes=c(1,2))
```

La gráfica de arriba a la derecha de la figura 1 muestra el gráfico de las componentes 1 y 2.

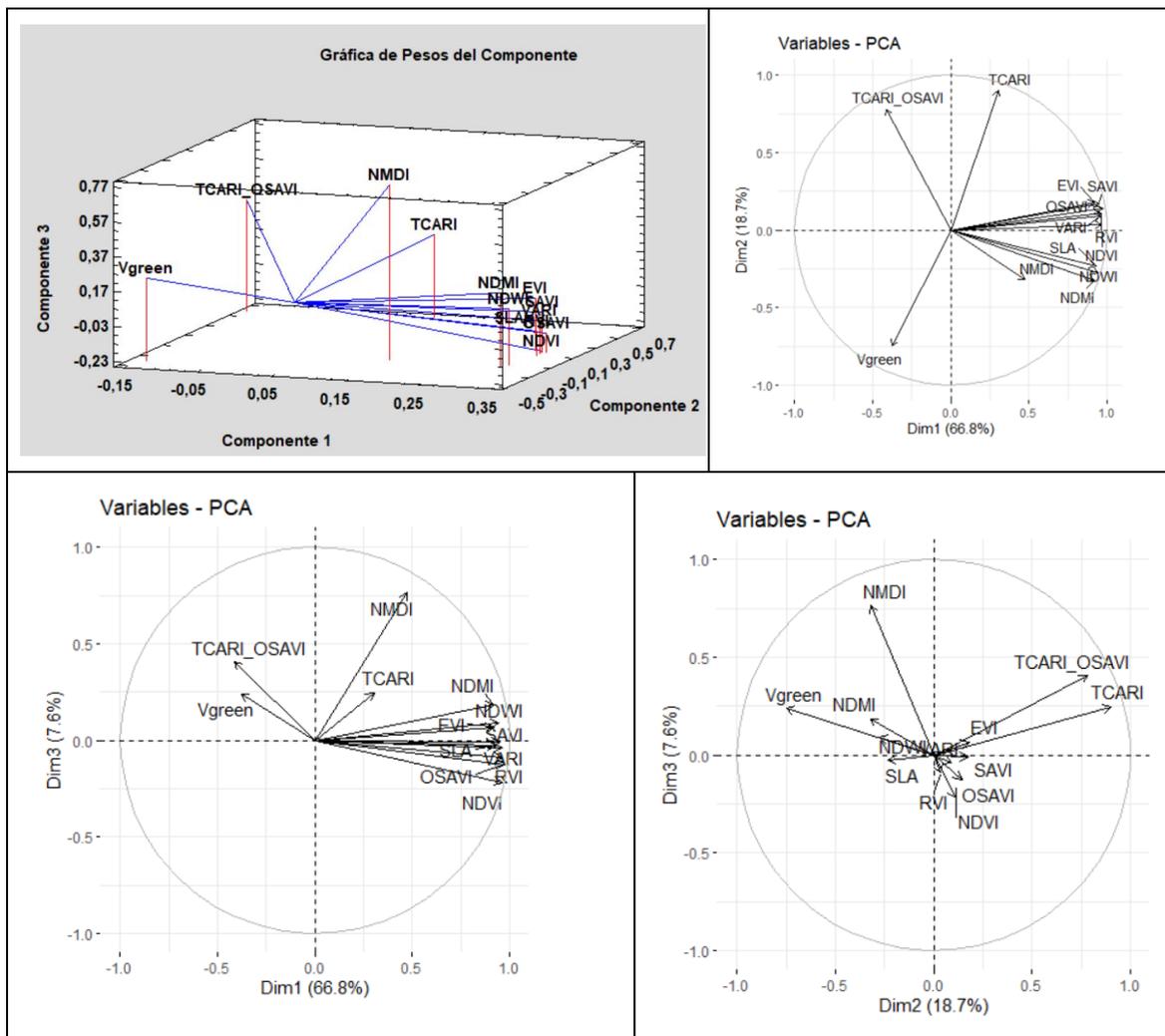


Figura 1. Gráficas que muestran el peso de cada variable en las componentes principales.

En la gráfica de las componentes 1 y 2 se observa que TCARI y OSAVI tienen una dirección parecida, pero sentidos opuestos. Por su parte NMDI tiene un coeficiente más alto en la componente 3, según vemos en las gráficas de la parte de debajo de la figura 1, las cuales han sido obtenidas cambiando la lista de componentes en la opción **axes** de la función **PCA**.

Para ver la contribución de cada variable en la primera componente principal podemos usar la función **fviz_contrib()** de la librería **factoextra** con esta sintaxis:

```
fviz_contrib(respca2,choice="var",axe=1,fill="blue",color="blue")
```

Se puede hacer lo mismo con la componente 2, cambiando el color de la gráfica de esta forma:

```
fviz_contrib(respca2,choice="var",axe=2,fill="pink",color="pink")
```

La figura 2 recoge los resultados obtenidos para las primeras 4 componentes principales. En ella se observa que en la primera componente principal tienen más peso los índices de actividad fotosintética y de agua (salvo el NMDI), en la segunda los índices relacionados con el verdor de la vegetación y en la tercera el índice NMDI junto con TCARI_OSAVI. En la cuarta domina la variable Vgreen. Nos preguntamos, ¿Cuántas componentes principales harán falta para obtener una buena representatividad de las variables originales? A ello dedicaremos el siguiente apartado.

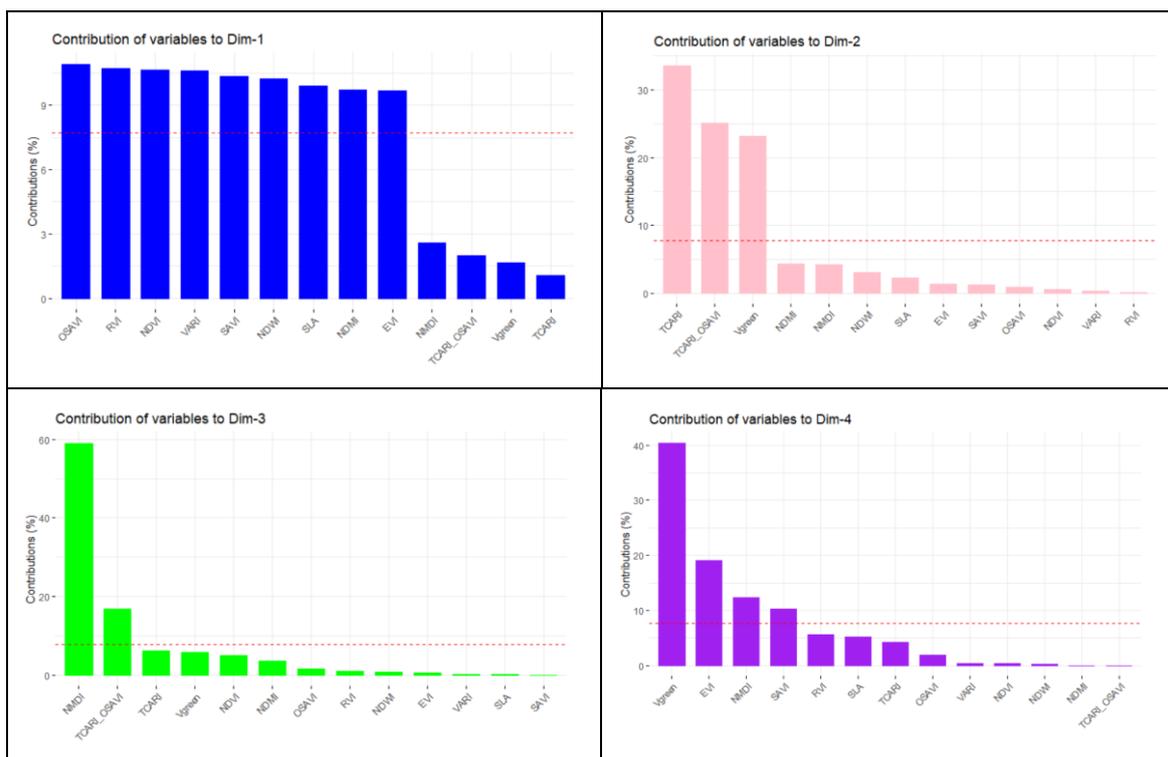


Figura 2. Contribución de cada variable en cada una de las componentes principales.

4.3 Número óptimo de componentes principales.

En el resumen estadístico del procedimiento de Statgraphics usado para obtener la tabla 1 (**Describir + Métodos multivariados + Componentes principales**) se obtiene la tabla 5, en la cual aparecen los valores propios (eigenvalor) de la matriz de correlaciones, ordenados de

menor a mayor, cuyos vectores propios asociados definen los coeficientes descritos en la tabla 4. La varianza de cada componente principal coincide con su valor propio. Al dividir cada valor propio por la suma de todos ellos y multiplicar por 100, se obtiene el porcentaje de varianza explicado por cada componente principal. El porcentaje acumulado para la componente “m” tiene en cuenta el porcentaje de variabilidad explicada por las “m” primeras componentes, sumando los porcentajes de varianza hasta la componente “m”.

| Componente Número | Eigenvalor | Porcentaje de Varianza | Porcentaje Acumulado |
|----------------------|------------|---------------------------|-------------------------|
| 1 | 8,68753 | 66,827 | 66,827 |
| 2 | 2,42521 | 18,655 | 85,483 |
| 3 | 0,986815 | 7,591 | 93,073 |
| 4 | 0,499579 | 3,843 | 96,916 |
| 5 | 0,210049 | 1,616 | 98,532 |
| 6 | 0,0736332 | 0,566 | 99,099 |
| 7 | 0,0463691 | 0,357 | 99,455 |
| 8 | 0,0347922 | 0,268 | 99,723 |
| 9 | 0,0199265 | 0,153 | 99,876 |
| 10 | 0,00967283 | 0,074 | 99,951 |
| 11 | 0,00311394 | 0,024 | 99,975 |
| 12 | 0,00197496 | 0,015 | 99,990 |
| 13 | 0,0013345 | 0,010 | 100,000 |

Tabla 5. Porcentaje de varianza acumulado por las componentes principales.

En la tabla 5 se observa que la primera componente principal explica un porcentaje de variación de los datos superior al 66% y con tres componentes principales se alcanza un porcentaje acumulado superior al 93%. Debemos de tener en cuenta que la variación total de los datos coincide con la suma de esos valores propios descritos en la tabla 5. Por su parte la contribución de la componente 4 contribuye en menos de un 4% a la variación total de los datos y a partir de la componente 6 dicho porcentaje de varianza es inferior al 1%.

Teniendo en cuenta dicho porcentaje acumulado podemos elegir un número de componentes principales que superen el 80% de porcentaje acumulado. Ello nos llevaría a quedarnos con 2 componentes. Sin embargo, también podemos quedarnos con 3 componentes principales pensando que el porcentaje acumulado prácticamente se estabiliza a partir de dicho número y aumentar la dimensión no aporta mucho. Para observar mejor esto podemos usar la sentencia `fviz_eig(respca2)` con la librería **factoextra** del software R, que permite dibujar las varianzas contra el número de dimensiones de las componentes principales (figura 3, izquierda). Por su parte, Statgraphics ofrece el gráfico de sedimentación (figura 3, derecha), donde aparecen los valores propios en el eje Y.

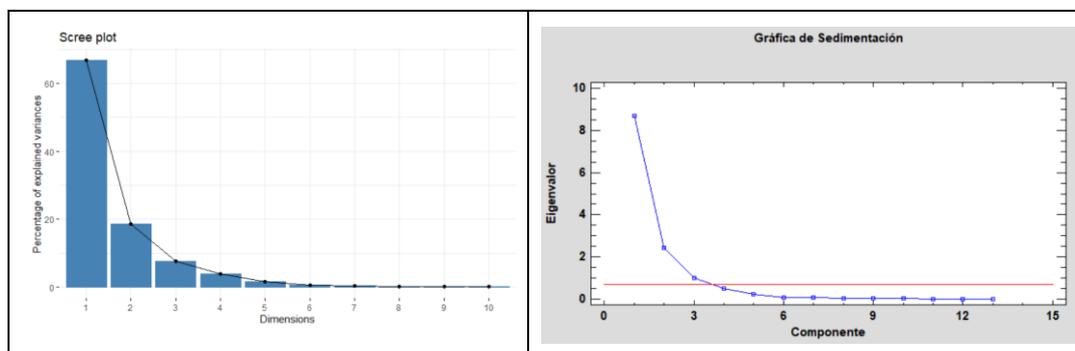


Figura 3. Contribución de componente principal al porcentaje de varianza explicado (izquierda) y comparativa de los valores propios asociados a ellas (derecha).

En el gráfico de sedimentación aparece marcada una línea de color rojo con el valor del eigenvalor mínimo, por encima del cual escogemos las componentes principales. El criterio de Kaiser con correlaciones [4] indica que tenemos que retener las “m” primeras componentes con valores propios superiores a 1. Sin embargo, estudios de Montecarlo prueban que es más correcto el punto de corte con un eigenvalor mínimo igual a 0,7 que es más pequeño que 1 [5], y ese ha sido el valor que ha sido escogido para la figura 3. Con dicho criterio el número de componentes principales elegidas es igual a 3, las cuales describen el 93,073% de la variación total de los datos según hemos visto en la tabla 5. Entonces podremos sustituir las 13 variables originales por las 3 primeras componentes principales, las cuales están incorreladas como se ha visto en la introducción.

Nos preguntamos ahora, ¿para qué podemos usar dichas componentes principales? El hecho que las componentes principales hayan sido calculadas con el criterio de que tengan la máxima varianza implica que sirvan para separar lo máximo posible unos individuos de otros. Y esto es de gran utilidad para el análisis que efectuamos en el siguiente apartado.

4.4 Análisis de la separación de los datos usando las componentes principales.

El procedimiento **Describir + Métodos multivariados + Componentes principales** de Statgraphics ofrece gráficos de dispersión 2D y 3D de los datos de la muestra usando sus coordenadas en las componentes principales. También se pueden elegir los bigráficos 2D que son una representación, en un mismo gráfico, de las filas (individuos) y las columnas (variables) de la matriz de datos. En ellos se cambia la escala de los pesos de las componentes para ponerlo en un mismo gráfico. Este tipo de gráficos sugieren las variables que son responsables de la proximidad entre puntos. La figura 4 muestra el bigráfico creado por las componentes principales 1 y 2. En ella se observa como la componente 1 logra una mayor separación de los puntos que la componente 2. Pichando encima de cada punto de dicho gráfico en Statgraphics se puede ver el número de fila al que pertenece el dato además de sus coordenadas en esas dos componentes principales.

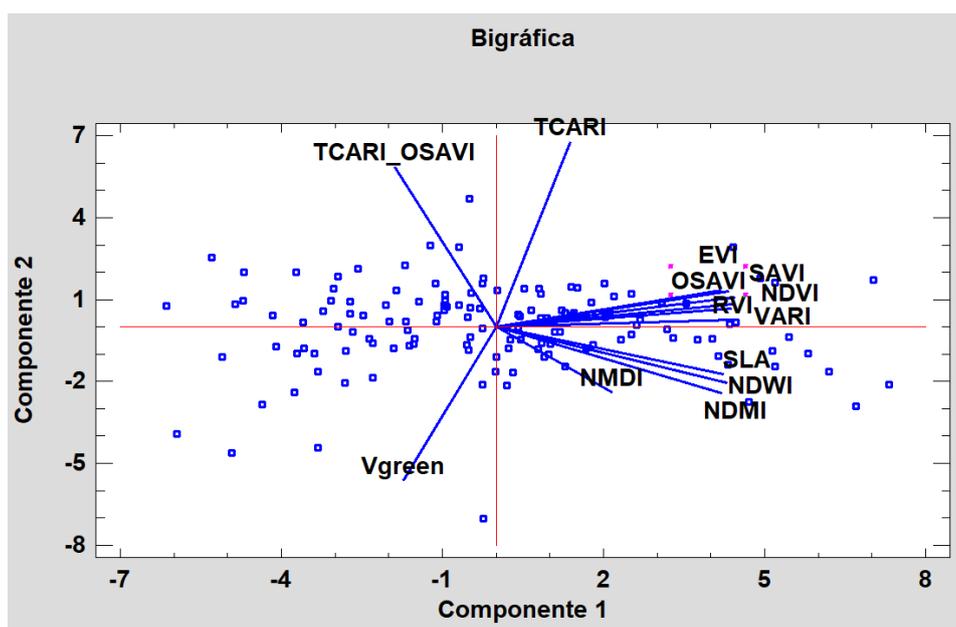


Figura 4. Bigráfico entre la componente principal número 1 y la componente 2.

A la hora de interpretar los valores positivos y negativos se debe tener presente que, aunque tengamos componentes con todos los coeficientes positivos, puede haber datos con valores negativos en una componente pues trabajamos con variables estandarizadas, las cuales toman valores negativos por debajo de la media. De esta forma el punto que se observa en la parte inferior de la gráfica de la figura 4 tiene un valor muy por debajo de la media en el índice TCARI (está en el lado opuesto) y valor por encima de la media en el índice Vgreen (está cerca de esa variable). Ese dato está en la fila 117 de la muestra.

La sentencia `fviz_pca_ind(respca2, axes = c(1, 3))`, usando dicha función de la librería `factoextra` del software R, permite efectuar el gráfico de las coordenadas en las componentes principales 1 y 3, viendo el número de fila de cada dato (figura 5). Así podemos ver la ordenación de los datos mediante esas dos componentes principales.

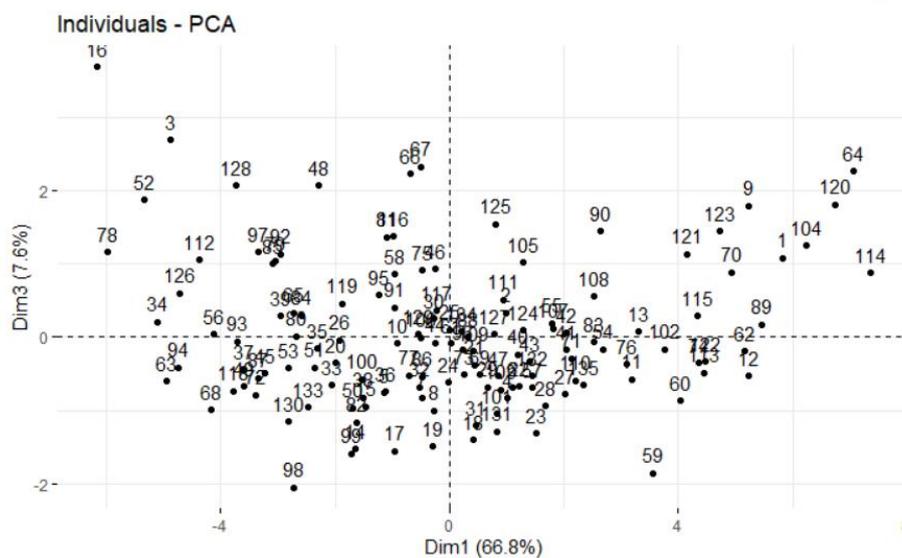


Figura 5. Gráfico de dispersión 2D con las coordenadas de las componentes principales 1 y 3.

Con el programa Statgraphics se pueden guardar las coordenadas de cada dato en las componentes principales en la base de datos y luego dibujar el diagrama de dispersión de esas columnas usando el procedimiento: **Grficar + gráfico de dispersión + gráfico XY** (figura 6).

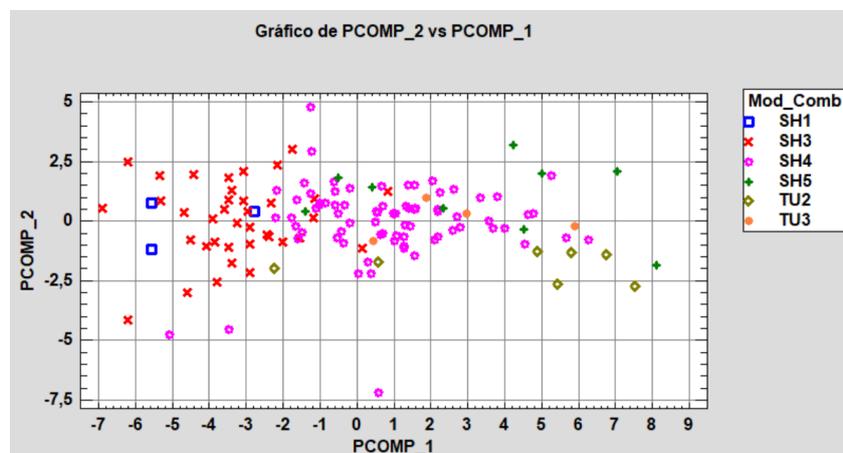


Figura 6. Gráfico de dispersión 2D con las coordenadas de las componentes principales 1-2 y la variable categórica que indica el tipo de vegetación: SH (matorral) y TU (arbolado).

Este tipo de gráfico permite añadir en las opciones de ventana de la gráfica una variable categórica que indica el código de puntos. En el caso de la figura 6 hemos añadido una variable con 6 categorías que indican diferentes tipos de vegetación a las que pertenece cada parcela forestal en las que se han tomado los datos. Las categorías SH son de tipo matorral, las cuales en función del grado de agrupamiento y continuidad llevan los números 1, 3, 4 o 5. Las dos categorías TU son para parcelas de arbolado, y dependiendo del tipo de matorral existente debajo de los árboles se definen los grupos TU2 o TU3 (véase [6]). En la figura 6 se observa que la representación gráfica de datos por medio de las coordenadas en las componentes principales permite separar los datos en función de los tipos de vegetación, observando diferencias entre las categorías SH3, SH4 y TU2, que son las que cuentan con más datos.

5 Conclusiones

En la primera componente principal tienen más peso las variables relacionadas con la actividad fotosintética (índices de vegetación) junto con algunas relacionadas con el contenido de agua como NDWI, NDMI y SLA. Dicha componente es capaz de explicar un 66,8% de la variabilidad total de los datos y por su propia definición es la que separa mejor los datos. El criterio de Kaiser indica que podemos sustituir los 13 índices espectrales originales por las 3 primeras componentes principales, que están incorreladas y con las cuales se alcanza el 93,07% del porcentaje de variabilidad acumulado. En la segunda componente tienen mayor influencia los índices TCARI, TCARI_OSAVI y Vgreen, relacionados con la actividad fotosintética de la vegetación. Y en la tercera componente destaca el índice NMDI que es un índice de sequía, el cual tiene una correlación menor con el resto de las variables originales de nuestra base de datos. Este análisis puede ser usado como paso previo al uso de técnicas de clasificación para asignar una nueva parcela forestal a alguna de las clases de vegetación de matorral o arbolado.

6 Bibliografía

- [1] Ferrero, S., Palacio, M., Campanella, O.R. (2002). Análisis de componentes principales en teledetección. Consideraciones estadísticas para optimizar su interpretación. *Revista de Teledetección*, 17, 43-54. <http://www.aet.org.es/revistas/revista17/AET17-05.pdf>
- [2] Balaguer Beser, A., Ruiz Fernández, L.A. (2021). Selección de un modelo de regresión lineal múltiple para el cálculo de la precipitación media en verano. *Universitat Politècnica de València*. <http://hdl.handle.net/10251/167659>
- [3] Costa-Saura, J.M., Balaguer-Beser, A., Ruiz, L.A., Pardo-Pascual, J.E., Soriano-Sancho, J.L. (2021). Empirical models for spatio-temporal live fuel moisture content estimation in mixed mediterranean vegetation areas using sentinel-2 indices and meteorological data. *Remote Sensing*, 13(18), 3726. <https://doi.org/10.3390/rs13183726>
- [4] González, A.L., Solano, H.L., Tilano, J. (2008). Análisis multivariado aplicando componentes principales al caso de los desplazados. *Revista Científica Ingeniería y Desarrollo*, (23), 119-142. <https://rcientificas.uninorte.edu.co/index.php/ingenieria/article/view/2098>
- [5] Cuadras, C. M. (1996). *Nuevos métodos de análisis multivariante*. CMC Edicions.
- [6] Clave para la identificación de los modelos de combustible de la Comunitat Valenciana. https://agroambient.gva.es/documents/162905929/169203680/Clave+fotogr%C3%A1fica+modelos+combustible_20200430/fd5ae58d-3b3f-4e50-866a-d83544a6f1b2