

Ecoinformática

Reto final

Producto 2

Carlos Martínez Núñez

Caracterización de poblaciones de robledal en función de variables biofísicas. Se realizará una clasificación de los distintos subtipos o poblaciones de ecosistemas teniendo en cuenta variables climáticas, edáficas y de funcionamiento ecosistémico. El producto final será un mapa o una tabla que muestre a qué grupo pertenece cada píxel del mapa de distribución del robledal en Sierra Nevada. Es decir, se indicarán los distintos “tipos” de robledal en función de las variables biofísicas utilizadas en la clasificación.

Flujo de trabajo:

- 1.-Obtenemos un archivo con las variables que utilizaremos para realizar la clasificación, así como sus lugares asociados (para luego realizar un mapa).
- 2.-Realizamos grupos teniendo en cuenta todas las variables. Usaremos el método de las k medias. El número de grupos los puede definir el usuario como quiera.
- 3.- Se obtienen los valores salidos del cluster y se añaden a una tabla con las posiciones “x” e “y” que sitúan a cada registro en el espacio.
- 4.- Se representa en un mapa los distintos puntos, con diferente color en función de a qué grupo pertenece cada punto.
- 5.-Se realiza un análisis para concluir qué número de grupos es el más apropiado seleccionar y qué sentido ecológico tiene.

Ejercicio detallado e implementado en R:

- ✓ Primero cargamos el archivo con las variables ambientales y su situación geográfica.

```
datos<-read.csv("robles_ecoinfo.csv", header = T, sep = ",", dec=".")
```

- ✓ Eliminamos las variables "x" e "y" del dataframe. Estas variables nos servirán luego para hacer el mapa, pero no para hacer los clusters.

```
variables<- subset(datos, select=-c(x,y))
```

- ✓ Creamos una variable que definirá el número de grupos a separar. Esto permitirá cambiar fácilmente el número de grupos que deseamos establecer.

```
n_cluster<-3
```

- ✓ Con la función "kmeans" hacemos los grupos teniendo en cuenta las variables del dataframe "variables". Definimos el número de grupos "n_cluster" y el número de iteraciones máximas para separar los grupos "iter.max". No se predefinió ningún número de centroides inicial "nstart".

```
cluster<-kmeans(variables,n_cluster, iter.max=200)
```

- ✓ Observamos la estructura del objeto resultante, y comprobamos el tercer número del primer componente de la lista (para probar que funciona correctamente).

```
str(cluster)

## List of 9
## $ cluster      : Named int [1:2451] 1 1 1 1 3 3 1 1 3 1 ...
## ..- attr(*, "names")= chr [1:2451] "1" "2" "3" "4" ...
## $ centers      : num [1:3, 1:31] 5.05 5.36 5.05 228.53 177 ...
## ..- attr(*, "dimnames")=List of 2
## .. ..$ : chr [1:3] "1" "2" "3"
## .. ..$ : chr [1:31] "hidro_tci" "hidro_acum" "tp_elev" "tp_pend"
## ...
## $ totss       : num 1.87e+10
## $ withinss    : num [1:3] 1.98e+09 1.44e+09 2.23e+09
## $ tot.withinss: num 5.66e+09
## $ betweenss   : num 1.3e+10
## $ size        : int [1:3] 640 1111 700
## $ iter        : int 2
## $ ifault      : int 0
## - attr(*, "class")= chr "kmeans"

cluster[[1]][3]
## 1
```

- ✓ Observamos el tamaño de los distintos clusters (esto nos servirá para las conclusiones).

```
cluster$size
```

```
## [1] 640-1111-700
```

- ✓ Realizamos un subset de los datos iniciales en el que sólo seleccionamos las variables "x" e "y"

```
resultado<-subset(datos,select=c(x,y))
head(resultado)
```

```
##      x      y
## 1 456550 4099050
## 2 456550 4098650
## 3 456650 4101550
## 4 456650 4101450
## 5 456650 4101350
## 6 456650 4101250
```

- ✓ A este dataframe le añadimos los datos del cluster al que pertenece cada posición.

```

resultado<-cbind(resultado, cluster[[1]])
head(resultado)

##           x           y cluster[[1]]
## 1 456550 4099050           1
## 2 456550 4098650           1
## 3 456650 4101550           1
## 4 456650 4101450           1
## 5 456650 4101350           3
## 6 456650 4101250           3

str(resultado)

## 'data.frame':    2451 obs. of  3 variables:
##  $ x          : int  456550 456550 456650 456650 456650 456650
456650 456650 456650 456750 ...
##  $ y          : int  4099050 4098650 4101550 4101450 4101350
4101250 4099150 4098650 4098550 4101550 ...
##  $ cluster[[1]]: int   1 1 1 1 3 3 1 1 3 1 ...

```

- ✓ Cambiamos el nombre de la columna del cluster.

```

colnames(resultado)[3]<- "cluster"
head(resultado)

```

```

##           x           y cluster
## 1 456550 4099050           1
## 2 456550 4098650           1
## 3 456650 4101550           1
## 4 456650 4101450           1
## 5 456650 4101350           3
## 6 456650 4101250           3

```

- ✓ Cargamos los paquetes necesarios para realizar los mapas con los distintos grupos.

```

library(sp)
library(rgdal)
library(classInt)
library(RColorBrewer)

```

- ✓ Definimos las coordenadas con las variables x e y.
- ✓ Establecemos el sistema de coordenadas WGS84
- ✓ Definimos el número de colores de la paleta "Spectral"
- ✓ Queremos el mapa en pdf (nombre, altura, anchura, color del fondo).
- ✓ Con dev.off, cerramos los mapas (de R y pdf).

```

coordinates(resultado) =~x+y
proj4string(resultado)=CRS("+init=epsg:23030")
plotclr <- rev(brewer.pal(n_cluster, "Spectral"))
pdf("Mapa robledal 3 grupos.pdf", height=8,width=10,bg="lightblue1")
plot(resultado, col=plotclr[resultado$cluster], pch=19, cex = .6, main
= "Mapa de 3 grupos de roble")
dev.off()

```

Interpretación de los resultados:

Se realizaron diferentes mapas (con 2, 3, 4, 5 y 7 grupos) para analizar qué número de clusters tenía más sentido ecológico.

Como puede observarse en los mapas:

2 grupos: Diferenciar únicamente dos zonas con todas las variables tenidas en cuenta parece ser demasiado simple. Se obtiene un mapa con muy poco detalle y no parece el adecuado.

3 grupos: Tamaños: 640-1111-700.

Parece una clasificación bastante simple, pero adecuada. Se aprecian parches bastante bien definidos en base a las variables ambientales tenidas en cuenta. Si cada grupo (y color en el mapa) es un “tipo de robledal”, se puede ver que existe bastante aislamiento entre parches y “tipos de robledal”.

4. grupos: Tamaños: 532-468-953-498.

Estos grupos no tienen demasiado sentido ecológico. Se encuentran manchas de la misma población en zonas demasiado alejadas, sin presencia en las manchas intermedias. Esto podría suceder, pero no es lo más probable. Teniendo en cuenta el principio de máxima verosimilitud esta clasificación no parece muy buena.

5 grupos: Tamaños: 727-441-488-372-423.

Esta clusterización parece bastante adecuada. Refleja una complejidad mayor que la clasificación en 3 grupos. Grupos de tamaños homogéneos. Muestra parches más heterogéneos, lo cual significaría que los distintos tipos de robledal se encuentran espacialmente asociados.

7 grupos: Tamaños: 544-350-16-414-294-408-425.

Este mapa muestra demasiada heterogeneidad. Observando el tamaño de los distintos grupos se ve que hay algún grupo tan pequeño (16) que no debería tenerse en cuenta.

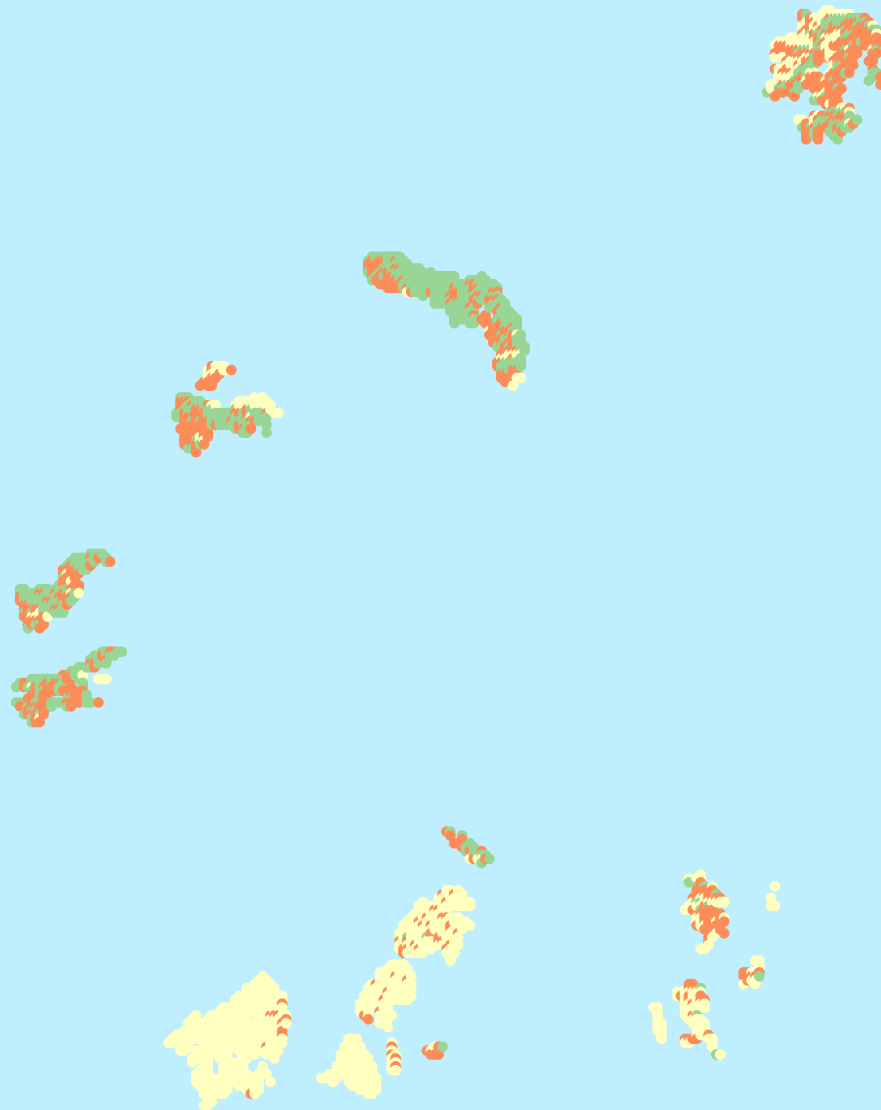
No permite la identificación espacial de tipos de robledal. Los clusters no presentan aislamiento y no se diferencian “tipos” espacialmente definidos. Según este mapa todos los tipos de robledal se encontrarían “entremezclados”.

Una posible aplicación de este tipo de estudios, sería la identificación de poblaciones con la finalidad de favorecer la heterogeneidad genética. Si una población de robledal está genéticamente deprimida, por ejemplo, se podría identificar otra población de la misma especie pero con alelos diferentes. Repoblando la zona de la población endogámica con algunos individuos de otras poblaciones sería una buena solución.

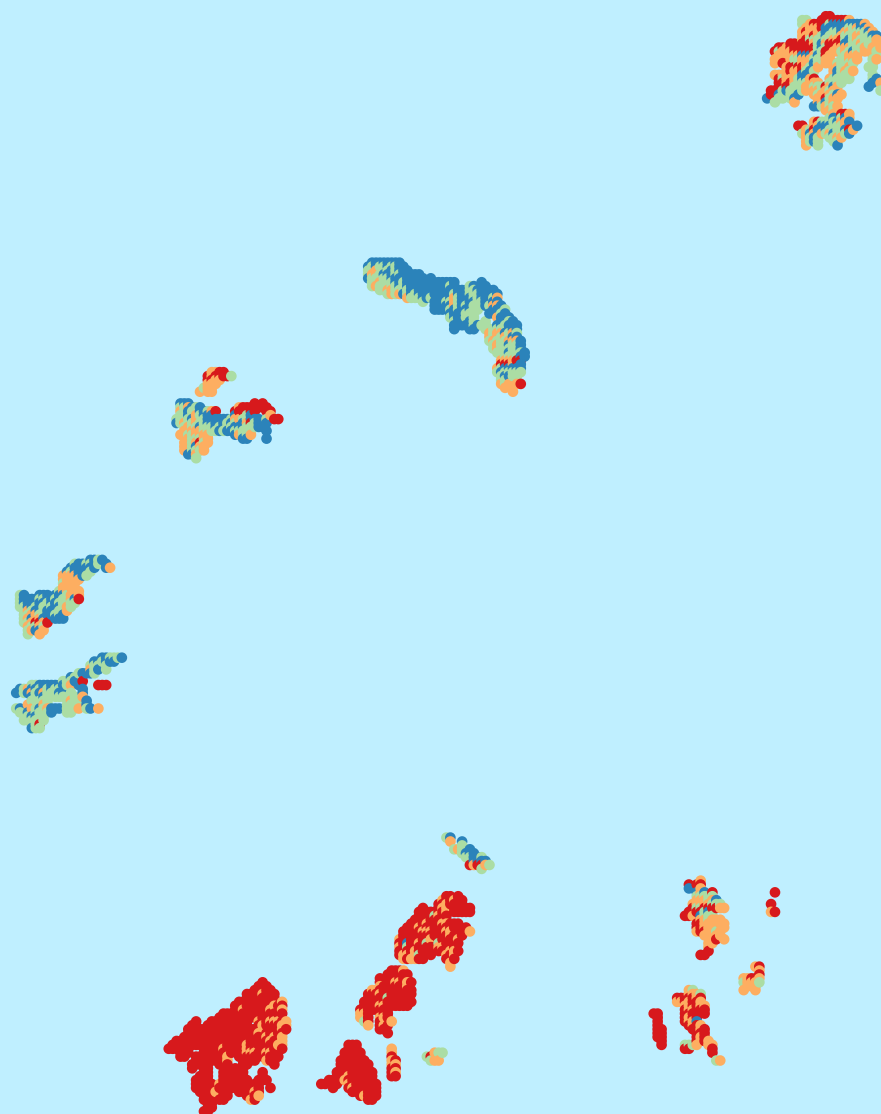
Mapa de 2 grupos de roble



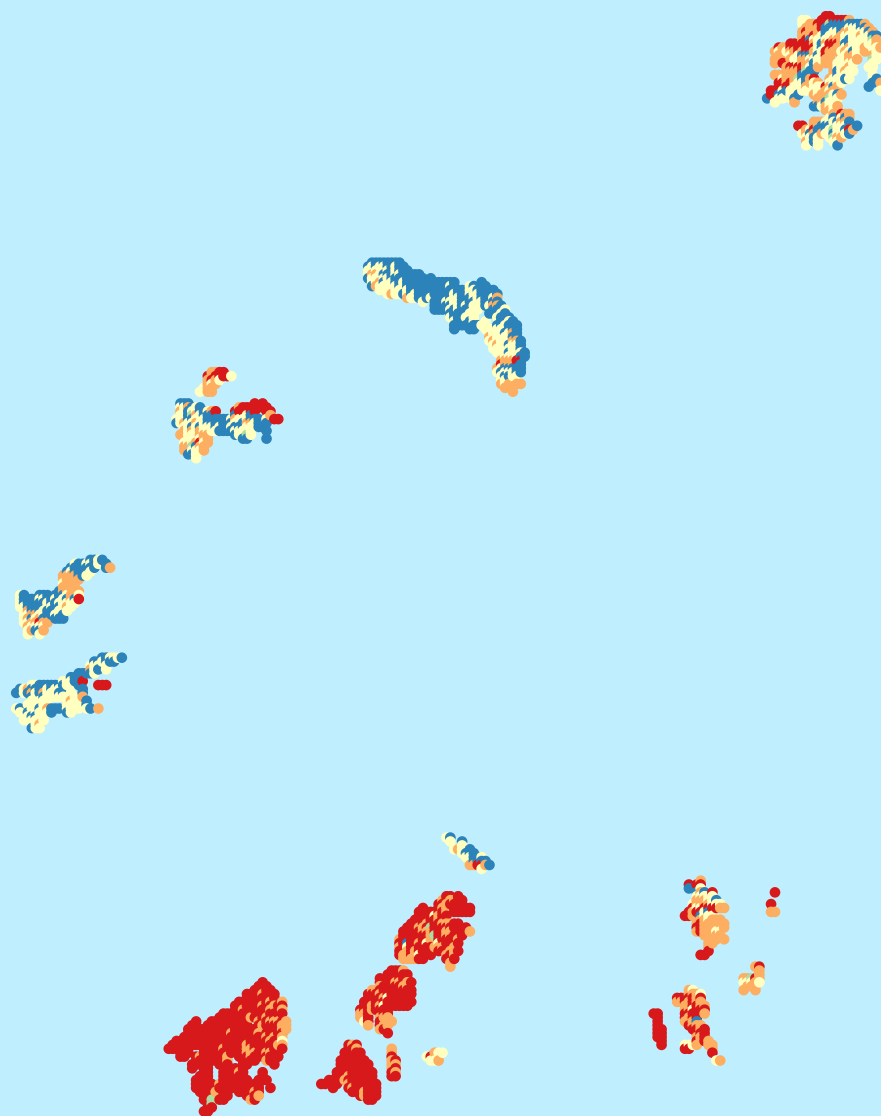
Mapa de 3 grupos de roble



Mapa de 4 grupos de roble



Mapa de 5 grupos de roble



Mapa de 7 grupos de roble

