library(vegan)

###NON-HIERARCHICAL CLASSIFICATION----------------------------------

#k-means with climatic data only

clust.env <- kmeans(env.tab.s[,c(9,12)], centers= 3, iter.max = 100)

clust.env

plot(env.tab.s[,c(9,12)], col = clust.env$cluster, pch=16)

points(clust.env$centers, pch = 8, cex = 2, col="blue")

#k-means with all environmental data

clust.env <- cascadeKM(env.tab.s, inf.gr = 2, sup.gr = 10, criterion = "ssi")

clust.env

plot(clust.env)#show optimal number of clusters based on ssi

plot(coord, pch=16, col=clust.env$partition[,1])

#k-means with species data

#transform data using Hellinger transformation

spe.hel <- decostand(spe, "hellinger")

clust.k <- cascadeKM(spe.hel, inf.gr = 2, sup.gr = 10, iter = 100, criterion = "ssi")

plot(clust.k)

plot(coord, pch=16, col=clust.k$partition[,4])

my.cols <- c("blue", "red", "green", "yellow", "orange")

plot(coord, pch=16, col=my.cols[clust.k$partition[,4]])

##HIERARCHICAL CLASSIFICATION--------------------------------------

#single linkage

clust.single <- hclust(beta.b, method='single')

plot (clust.single, main = 'Single linkage')

#complete linkage

clust.complete <- hclust(beta.b, method='complete')

plot (clust.complete, main = 'Complete linkage')

#average linkage

clust.upgma <- hclust(beta.b, method='average')

plot (clust.upgma, main = 'UPGMA')

#is matrix Euclidean?

#install.packages("ade4")

library(ade4)

is.euclid(beta.b)

is.euclid(envdist)

#transform to Euclidean

beta.bc <- cailliez(beta.b, cor.zero = FALSE)

is.euclid(beta.bc)

#what is the constant?

(beta.bc - beta.b)[1]

plot(beta.bc, beta.b)

#ward clustering

clust.ward <- hclust(beta.bc, method='ward.D2')

plot(clust.ward)

###NUMBER OF CLUSTERS---------------------------------------

#for ward clustering

plot(clust.ward)

rect.hclust (clust.ward, k = 3, border="red")

rect.hclust (clust.ward, k = 5, border="blue")

cl.ward <- cutree(clust.ward, k=5)

plot(coord, pch=16, col=my.cols[cl.ward])

#for upgma clustering

cl.upgma <- cutree(clust.upgma, k=5)

plot(forest$X, forest$Y, pch=16, col=my.cols[cl.upgma])

###MANTEL-OPTIMAL NUMBER OF CLUSTERS ACCORDING TO BORCARD ET AL. 2011--------------------------------------------------------------

#define function grpdist

grpdist <- function(X)

{

require(cluster)

gr <- as.data.frame(as.factor(X))

distgr <- daisy(gr)

distgr

}

#define object kt where the results of Mantel correlations will be stored

kt <- data.frame(k=1:10, r=0)

kt

#calculate Mantel correlations

for (i in 2:10)

{

gr <- cutree(clust.ward, k=i) #here paste hclust object

distgr <- grpdist(gr)

mt <- cor(beta.b, distgr, method="pearson") #here paste dissimilarity matrix and distgr matrix

kt[i,2] <- mt

}

kt

k.best <- which.max(kt$r)

#plot optimal number of clusters

plot(kt$k, kt$r, type="h", main="Mantel-optimal number of clusters", xlab="k (number of clusters)",

ylab="Pearson's correlation", las=1, cex.axis=1.2, cex.lab=1.2, cex.main=1.4)

axis(1, k.best, paste("optimum", k.best, sep="\n"), col="red",

font=2, col.axis="red")

points(k.best, max(kt$r), pch=16, col="red", cex=1.2)