

Classifying vegetation stands in terms of structure and composition

Miquel De Cáceres¹

¹Centre Tecnològic Forestal de Catalunya, Solsona, Catalonia, Spain

November 7, 2013

Contents

1	Introduction	1
2	Example vegetation data	1
3	Cumulative abundance profiles	2
4	Dissimilarities in structure and composition	4
5	Classification of vegetation stands	6

1 Introduction

The aim of this tutorial is to show how to implement the method to calculate dissimilarities between communities presented in [De Cáceres et al., 2013]. The new method allows incorporating both the structure and the composition of the community in the dissimilarity measurement. The functions needed to carry out computations have been included in package `vegclust` so we start by loading the package:

```
> library(vegclust)
```

2 Example vegetation data

In order to illustrate the method we will use a stratified vegetation data set containing data from 96 stands. The data was obtained to investigate patterns vegetation recovery three years after a wildfire. Data were collected in 2012 by Miquel De Cáceres and Albert Petit in Horta de Sant Joan

(Catalonia, Spain). The R object is of class `stratifiedvegdata` (actually a list).

```
> data(medreg)
> class(medreg)
```

```
[1] "list"                "stratifiedvegdata"
```

```
> length(medreg)
```

```
[1] 96
```

The dataset contains 96 stands (list elements), each of them a data.frame where rows correspond to broad plant functional groups (Pine trees, Oak trees, Tall shrubs and small trees, Scrubs and small shrubs and Grass) and columns correspond to vegetation strata (1 to 7). The upper heights of the vegetation strata are the following (in cm.) vector:

```
> strataUp = c(20,50,100,300,600,1200,2400)
```

And the width (range of heights) of each stratum is:

```
> strataWidths = c(20,30,50,200,300,600,1200)
```

Species abundance values are percentage cover values estimated using cover classes:

```
> medreg[[1]]
```

	1	2	3	4	5	6	7
Pine trees	0.0	0.0	0	0	0	0	0
Quercus trees	12.5	25.0	0	0	0	0	0
Tall shrubs and small trees	0.0	62.5	25	0	0	0	0
Scrubs and small shrubs	12.5	37.5	0	0	0	0	0
Grass	50.0	0.0	0	0	0	0	0

The data is read as follows. Shrubs reaching stratum 3 (50 - 100 cm) had a cumulative cover of 25%, while shrubs reaching only stratum 2 (20 - 50 cm) had a cumulative cover of 62.5%. Thus, the observers grouped plants according to their height and functional group, and estimated the cover for those groups of plants.

3 Cumulative abundance profiles

The cumulative abundance profile (CAP) is a function that takes a value of *size* as input (here the *size* is a vegetation stratum) and returns the cumulative abundance of organisms (here the cumulative cover value) whose

size is *equal to* or *larger than* the input value. In our case, the CAP function is the cumulative cover of plants reaching the current stratum or higher strata. The calculations are made using by calling the function `CAP()` of `vegclust`:

```
> medreg.CAP <- CAP(medreg)
```

Note that a different CAP function is calculated for each functional group. The structure of the resulting R object is very similar to the stratified data:

```
> class(medreg.CAP)
```

```
[1] "list" "CAP"
```

```
> length(medreg.CAP)
```

```
[1] 96
```

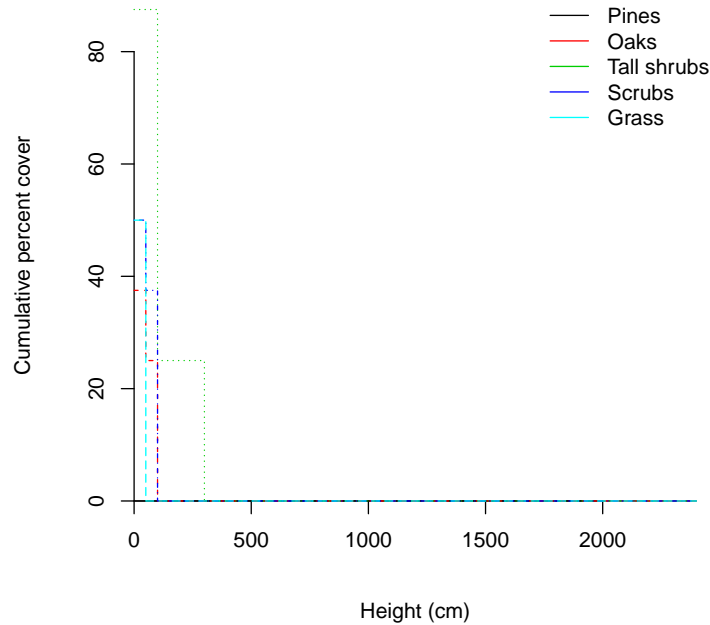
If we inspect the first element of the list, we can see the difference between the original data and the cumulative abundance profile.

```
> medreg.CAP[[1]]
```

	1	2	3	4	5	6	7
Pine trees	0.0	0.0	0	0	0	0	0
Quercus trees	37.5	25.0	0	0	0	0	0
Tall shrubs and small trees	87.5	87.5	25	0	0	0	0
Scrubs and small shrubs	50.0	37.5	0	0	0	0	0
Grass	50.0	0.0	0	0	0	0	0

Moreover, it is possible to graphically display the CAP of a given stand (it may become difficult to interpret when the number of species is large). For example, we can display the CAP for each functional group of the first stand (`plots="1"`):

```
> plot(medreg.CAP, plots="1", sizes=strataUp, xlab="Height (cm)",
+       ylab="Cumulative percent cover")
> legend("topright", col=1:5, lty=1,
+       legend=c("Pines", "Oaks", "Tall shrubs", "Scrubs", "Grass"),
+       bty="n")
```



In this case the vegetation is a short but dense shrubland. Note that in the plot we used `strataUp` to set the x-axis, so that real heights are adequately represented.

4 Dissimilarities in structure and composition

Although CAPs can be used to graphically display the structure and composition of vegetation stands, the whole point of defining the CAP function was to allow comparisons between stands. We can calculate dissimilarities for all pairs of stands, thus obtaining a square and symmetric matrix with dissimilarity values:

```
> medreg.D = vegdiststruct(medreg.CAP, method="bray",
+                           classWidths=strataWidths)
```

In the above sentence we forced strata to have different weight, according to the range of heights that each stratum occupies. There are different alternatives with respect to the dissimilarity index. In our case we chose the CAP generalization of Bray-Curtis [De Cáceres et al., 2013]. If we want to know, for example, the dissimilarity between stands '1' and '2' we simply write:

```
> as.matrix(medreg.D)[1,2]
```

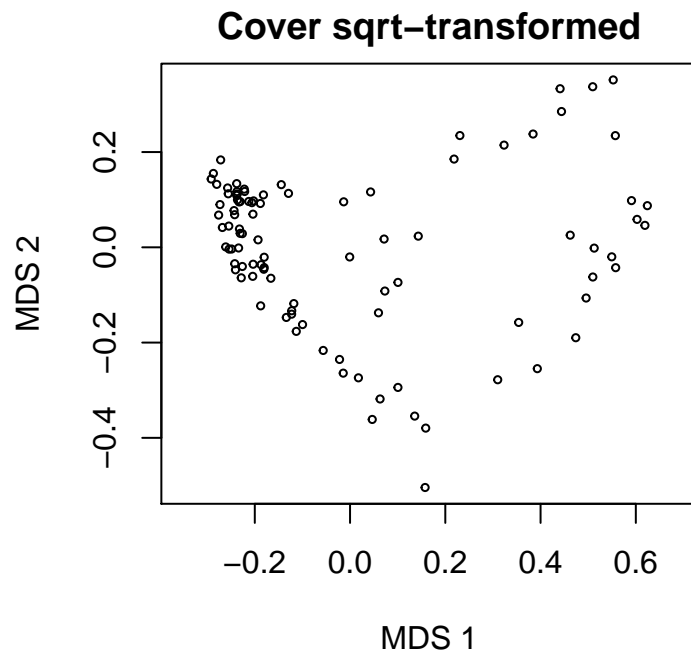
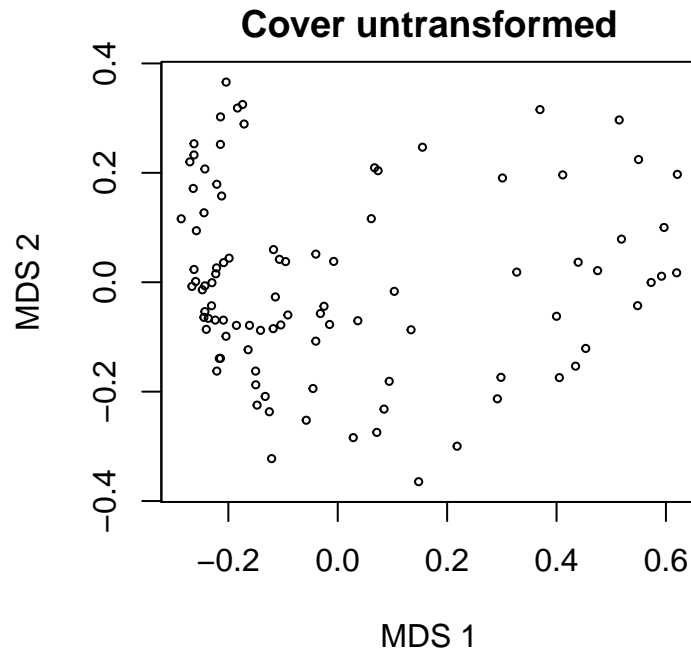
[1] 0.2713178

When calculating dissimilarities it is possible to transform the CAP values in order to prevent large abundance values to have an undue influence in the analysis. In our case we choose to take the square root of cumulative cover values:

```
> medreg.Dsqrt = vegdiststruct(medreg.CAP, method="bray",  
+                               classWidths=strataWidths, transform="sqrt")
```

We can use metric multidimensional scaling to represent the distances between stands obtained in both cases:

```
> par(mfrow=c(2,1), mar=c(4,5,2,1))  
> X<-cmdscale(medreg.D, k=2)  
> plot(X, xlab="MDS 1", ylab="MDS 2", asp=1,  
+       main="Cover untransformed", cex=0.5)  
> Xsqrt<-cmdscale(medreg.Dsqrt, k=2)  
> plot(Xsqrt, xlab="MDS 1", ylab="MDS 2", asp=1,  
+       main="Cover sqrt-transformed", cex=0.5)
```



Note that the differences between the two ordination plots are remarkable.

5 Classification of vegetation stands

In this section we use the square-root transformed dissimilarities between vegetation stands to obtain a classification of the stands in terms of their

structure and composition. If you are not familiarized with non-hierarchical clustering, you can read the tutorial about `vegclust` package. We start by setting the number of clusters to be found (`nclusters`) and the size of clusters (`dnoise`, a parameter used to leave stands that are far from all group prototypes unclassified):

```
> nclusters = 6
> dnoise = 0.40
```

We call function `vegclust` using the clustering method `HNCdd`, which indicates (a) hard clustering, (b) medoids as prototypes, and (c) noise clustering (i.e. excluding outliers in a special class called *noise class*):

```
> vc<-vegclustdist(medreg.Dsqrt, mobileMemb = nclusters,
+                 method="HNCdd", dnoise=dnoise, nstart=100)
```

With `nstart=100` we indicate that the algorithm should be run 100 times starting from random seeds. This is advisable in order to maximize the chance of having suboptimal solutions. The prototypes identified by the algorithm are the following medoids (these are indices of stands in `medreg`):

```
> medoids<-vc$mobileCenters
> print(medoids)
```

```
[1] 20 38 60 24 92 71
```

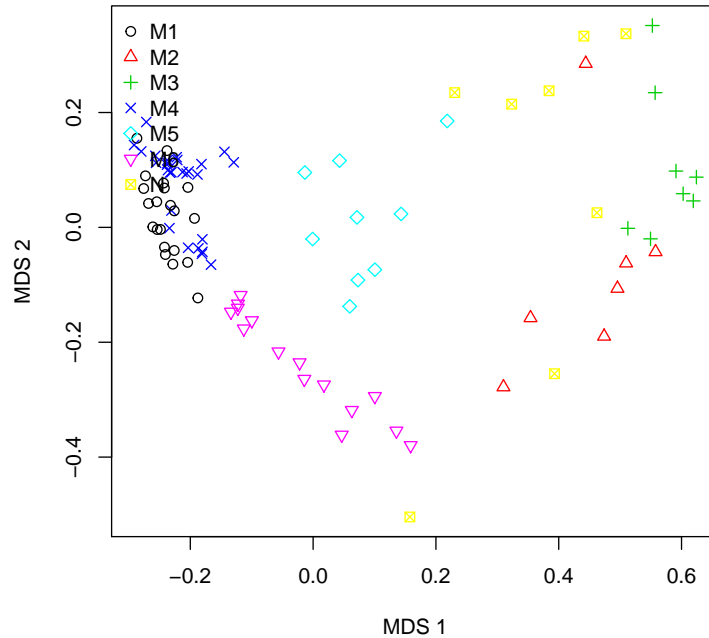
The number of stands belonging to each cluster can be found using:

```
> cluster<-defuzzify(vc)$cluster
> table(cluster)
```

```
cluster
M1 M2 M3 M4 M5 M6  N
21  7  8 28  9 15  8
```

Note that, because of the model chosen (and with the parameter `dnoise`), there are a number of stands that are left unclassified (i.e. those assigned to class 'N'). A useful way to display the results of the cluster analysis is by showing the stand memberships to clusters in the ordination:

```
> clNum = as.numeric(as.factor(cluster))
> plot(Xsqrt, xlab="MDS 1", ylab="MDS 2",
+      pch=clNum, col=clNum)
> legend("topleft", col=1:(nclusters+1), pch=1:(nclusters+1),
+      legend=levels(as.factor(cluster)), bty="n")
```



While the stands belonging to *true* clusters are more or less close, those that are assigned to the *noise* can be far apart, because the only fact that makes them be in the same class is their lack of membership for *true* clusters.

To facilitate the interpretation of the clusters we can calculate the average cumulative abundance profiles:

```
> CAPm = CAPmeans(medreg.CAP, as.factor(cluster))
> names(CAPm)

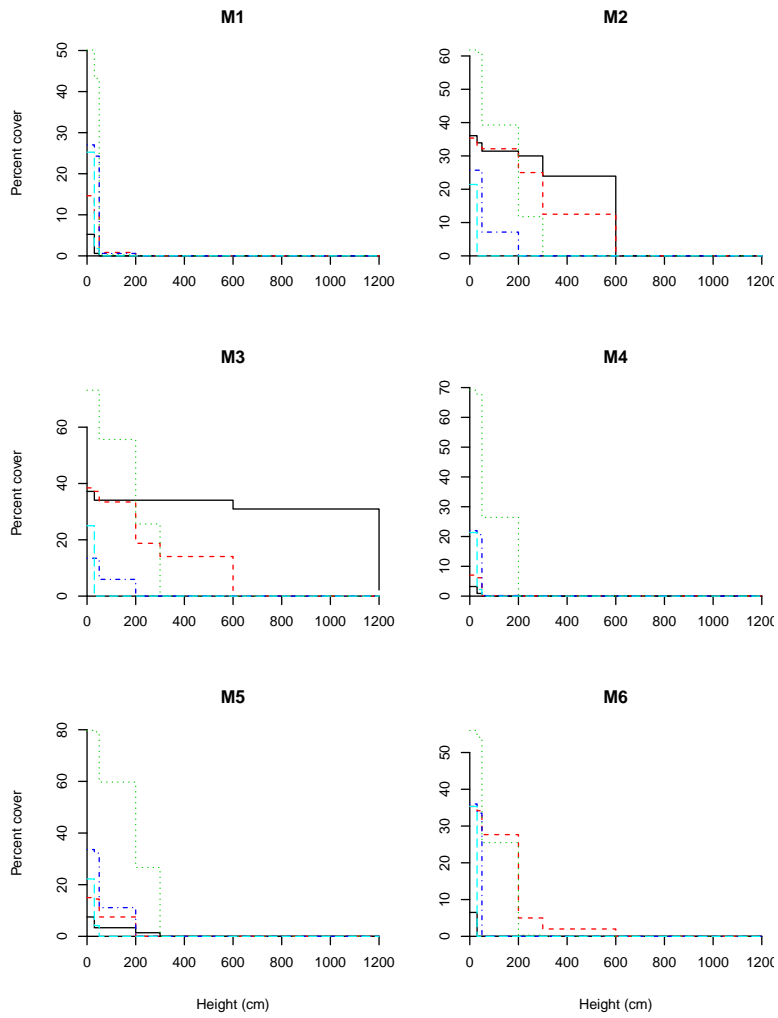
[1] "M1" "M2" "M3" "M4" "M5" "M6" "N"
```

For example, we can inspect the structure and composition of group M4:

```
> round(CAPm$M4, dig=1)

              1    2    3 4 5 6 7
Pine trees    3.2  0.9  0.0 0 0 0 0
Quercus trees  7.1  6.2  0.2 0 0 0 0
Tall shrubs and small trees 69.1 67.6 26.4 0 0 0 0
Scrubs and small shrubs    22.0 20.6  0.2 0 0 0 0
Grass           21.3  2.2  0.0 0 0 0 0
```

The following displays graphically the CAPs of all six groups of vegetation stands (code not shown):



References

Miquel De Cáceres, Pierre Legendre, and Fangliang He. Dissimilarity measurements and the structure of plant communities. *Methods in Ecology and Evolution*, 2013.