

Indicator species

Pierre Legendre
Département de sciences biologiques, Université de Montréal

Beijing, September 2009

Indicator species —

The identification of characteristic or indicator species is traditional in ecology and biogeography. Field studies describing sites or habitats usually mention one or several species that characterize each habitat. There is clearly a need for the identification of characteristic or indicator species in the fields of monitoring, conservation, and management. Because indicator species add ecological meaning to groups of sites discovered by clustering, they provide criteria

- to compare typologies derived from data analysis,
- to identify where to stop dividing clusters into subsets, and
- to point out the main levels in a hierarchical classification of sites.

Indicator species differ from species associations in that they are indicative of particular groups of sites. Good indicator species should be found mostly in a single group of a typology and be present at most of the sites belonging to that group.¹

- Classical problem in community ecology and biogeography:
Species are the best indicators we have for particular environmental conditions.
- In long-term environmental follow-up, conservation, ecological management: researchers are looking for bioindicators of habitat types to preserve or rehabilitate.

“McGeoch & Chown (1998) found the indicator value method important to conservation biology because it is conceptually straightforward and allows researchers to identify bioindicators for any combination of habitat types or areas of interest, e.g. existing conservation areas, or groups of sites based on the outcome of a classification procedure. In addition, it may be used to identify bioindicators for groups of sites classified using the target taxa [...] or using non-target taxa (e.g. insect bioindicators of plant community classifications).”²

¹ Legendre, P. & L. Legendre. 1998. Numerical ecology, 2nd English edition. Elsevier Science BV, Amsterdam.

² McGeoch, M. A. and S. L. Chown. 1998. Scaling up the value of bioindicators. *Trend Ecol. Evol.* 13: 46-47.

The IndVal method was proposed by Dufrière and Legendre (1997).³

$$\text{IndVal}_{\text{Group } k, \text{Species } j} = 100 \times A_{k,j} \times B_{k,j}$$

In that equation, $A_{k,j}$ = Specificity, $B_{k,j}$ = Fidelity

$$\text{IndVal}_{\text{Espèce } j} = \max[\text{IndVal}_{k,j}]$$

TABLE 2. Test case example: abundance of three species in 25 sites divided into five clusters.

Species	Group 1					Group 2					Group 3					Group 4					Group 5				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
A	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	3	3	3	3	3	3	3	3	3	3
B	8	8	8	8	8	4	4	4	4	4	6	6	6	6	6	4	4	2	0	0	0	0	0	0	0
C	18	18	18	18	18	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

	Group 1	Group 2	Group 3	Group 4	Group 5
Species 1 (A)					
A	4/20 = 0.20	5/20 = 0.25	5/20 = 0.25	3/20 = 0.15	3/20 = 0.15
B	5/5 = 1	5/5 = 1	5/5 = 1	5/5 = 1	5/5 = 1
IndVal	20 %	25 %	25 %	15 %	15 %
Species 2 (B)					
A	8/20 = 0.40	4/20 = 0.20	6/20 = 0.30	2/20 = 0.10	0/20 = 0.00
B	5/5 = 1	5/5 = 1	5/5 = 1	3/5 = 0.6	0/5 = 0
IndVal	40 %	20 %	30 %	6 %	0 %
Species 3 (C)					
A	18/20 = 0.90	2/20 = 0.10	0/20 = 0.00	0/20 = 0.00	0/20 = 0.00
B	5/5 = 1	5/5 = 1	0/5 = 0	0/5 = 0	0/5 = 0
IndVal	90 %	10 %	0 %	0 %	0 %

³ Dufrière, M. and P. Legendre. 1997. Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecological Monographs* 67: 345-366.

Indicator species analysis using the R language

IndVal analysis for the Dufrêne & Legendre 1997 data, p. 351, Table 2.

Function 'duleg' of the 'labdsv' library computes *IndVal* analysis.

In the latest versions of the library, the Mac OS X version works fine, but the Windows version produces incorrect p-values.

The new library 'indicpecies' available on <http://www.bio.umontreal.ca/legendre/indexEn.html> and <http://sites.google.com/site/miqueldecaceres/software> computes different indicator species indices including *IndVal*.

Note: The "IndVal.g" index computed by function 'strassoc' of the 'indicpecies' library is the square root of the *IndVal* index of Dufrêne and Legendre (1997).

```
library(labdsv)
?duleg          # IndVal method of Dufrêne and Legendre (1997)
```

1. Analyse Table 2 of Dufrêne & Legendre (1997). Artificial data.

Create the data table as well as the vector describing the groups.

The data table must have a class 'matrix' or 'data.frame'

```
table2 = matrix(
c(4,4,4,4,4,5,5,5,5,5,5,5,5,5,5,3,3,3,3,3,3,3,3,3,8,8,8,8,8,4,4,4,4,4,6,6,6,6,6,4,
4,2,0,0,0,0,0,18,18,18,18,18,2,2,2,2,2,0,0,0,0,0,0,0,0,0,0,0,0,0,25,3)
vec = c(rep(1,5),rep(2,5),rep(3,5),rep(4,5),rep(5,5))
colnames(table2) = c("species_1", "species_2", "species_3")
```

```
# IndVal analysis for the real groups
indval.out = duleg(table2, vec, numitr=10000)
indval.out
```

```
# IndVal analysis for random groups
rand.vec = sample(1:5, 25, replace=TRUE) # Bootstrap sample
indval.rand.out = duleg(table2, rand.vec, numitr=10000)
indval.rand.out
```

Apply a correction for multiple testing to obtain a corrected vector of p-values allowing one to identify the significant indicator species. The theory for correction for multiple testing is described in p. 18 of *Numerical ecology* (1998). Use the following R function:

```
?p.adjust
```

Example:

```
prob.corrected = p.adjust(indval.out$pval, "holm")
```

2. A real example: the hunting spider data of Aart & Smeenk-Enserink (1975)

```
spiders.sp = read.table(file.choose()) # Read the data 'Spiders_28x12_spe.txt'
# Create a classification of the sites based upon the environmental variables
spiders.env = read.table(file.choose()) # Read the file 'Spiders_28x4_env.txt'
# Compute a Gower distance matrix from the environmental data
library(vegan)
spiders.gowerD = vegdist(spiders.env, "gower")
# Hierarchical clustering, Ward method
spiders.cl = hclust(spiders.gowerD, method="ward")
plot(spiders.cl, hang=-1) # Dendrogram
# What is the structure of the object 'spiders.cl' ?
summary(spiders.cl)
# Examine the dendrogram. How many groups does it contain?
# Create a vector 'groups' showing the group to which each site belongs,
# in the original order of entry of the sites.
groups = cutree(spiders.cl, k=3)
# Look for the indicator species
library(labdsv)
indval.out = duleg(spiders.sp, groups, numitr=10000)
# Holm correction on the p-values
prob.corrected = p.adjust(indval.out$pval, "holm")
```