

# Indicator species

Pierre Legendre  
Département de sciences biologiques, Université de Montréal

*September 2009, February 2010, October 2014*

## **Indicator species —**

- 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.

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 the most interesting clustering levels in a dendrogram.

“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).” (Legendre & Legendre 2012, p. 397).

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.

## **References**

- Legendre, P. & L. Legendre. 2012. *Numerical ecology, 3rd English edition*. Elsevier Science BV, Amsterdam.
- McGeoch, M. A. & S. L. Chown. 1998. Scaling up the value of bioindicators. *Trend Ecol. Evol.* 13: 46-47.

The IndVal method was proposed by Dufrêne & Legendre (1997).<sup>1</sup>

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

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
<b>Species 1 (A)</b>					
$A_{k,j}$	4/20 = 0.20	5/20 = 0.25	5/20 = 0.25	3/20 = 0.15	3/20 = 0.15
$B_{k,j}$	5/5 = 1	5/5 = 1	5/5 = 1	5/5 = 1	5/5 = 1
IndVal	20 %	25 %	25 %	15 %	15 %
<b>Species 2 (B)</b>					
$A_{k,j}$	8/20 = 0.40	4/20 = 0.20	6/20 = 0.30	2/20 = 0.10	0/20 = 0.00
$B_{k,j}$	5/5 = 1	5/5 = 1	5/5 = 1	3/5 = 0.6	0/5 = 0
IndVal	40 %	20 %	30 %	6 %	0 %
<b>Species 3 (C)</b>					
$A_{k,j}$	18/20 = 0.90	2/20 = 0.10	0/20 = 0.00	0/20 = 0.00	0/20 = 0.00
$B_{k,j}$	5/5 = 1	5/5 = 1	0/5 = 0	0/5 = 0	0/5 = 0
IndVal	90 %	10 %	0 %	0 %	0 %

<sup>1</sup> Dufrêne, 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 `indval()` of the `{labdsv}` package computes *IndVal* analysis.

```
library(labdsv)
?indval
```

### 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(4,4,4,4,4,5,5,5,5,5,5,5,5,5,5,3,3,3,3,3,3,3,3,3,3,3,8,8,8,8,8,4,4,4,4,4,6,6,6,6,6,6,4,
4,2,0,0,0,0,0,0,0,18,18,18,18,18,2,2,2,2,2,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0),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 = indval(table2, vec, numitr=10000)
```

```
indval.out
```

# *IndVal* analysis for random groups

```
rand.vec = sample(1:5, 25, replace=TRUE) # Bootstrap sample
```

```
indval.rand.out = indval(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)          # Example: division into k = 3 groups
# Look for the indicator species
library(labdsv)
indval.out = indval(spiders.sp, groups, numitr=10000)
# Holm correction on the p-values
prob.corrected = p.adjust(indval.out$pval, "holm")
```

### 3. An R package R: {indicspecies}

A new R package called ‘indicspecies’ is available on CRAN and at the following address:

<https://sites.google.com/site/miqueldecaceres/software>.

It allows the computation of different indices including IndVal. Note that the “IndVal.g” index computed by function `strassoc()` of the {indicspecies} package is the square root of the IndVal index of the {labdsv} package and of the Dufrêne Legendre (1997) paper.

The {indicspecies} package is a companion to the following paper published in December 2009:

De Cáceres, M., and P. Legendre. 2009. Associations between species and groups of sites: indices and statistical inference. *Ecology* 90: 3566-3574.

The following recommendations are found on p. 3573 of the paper. They were written to guide the choice of users between IndVal-type indices (‘IndVal’ and ‘IndVal.g’) and correlation-type indices (‘r’, ‘r.g’, ‘cos’ and ‘cos.g’):

“Although both the correlation and indicator value approaches can be successful in many applications, they have been developed with slightly different purposes. (1) For determining the ecological preference of a given species among a set of alternative site groups, the correlation approach is probably more useful than the indicator value approach, because the former naturally allows the detection of negative preferences. [...] (2) For assessing species predictive values (e.g., for field determination of community types or ecological monitoring), one should preferably compute indicator values (McGeoch and Chown 1998), because they are decomposable into components that are interpretable as probabilities. [...]”

The most recent paper on the IndVal method for indicator species is:

De Cáceres, M., P. Legendre and M. Moretti. 2010. Improving indicator species analysis by combining groups of sites. *Oikos* 119: 1674-1684.

These two papers are available in PDF on the page <http://adn.biol.umontreal.ca/~numericecology/Reprints/>.

=====