Indicator Species Analysis (ISA)

Anabel Dombro

Video seminar: http://tinyurl.com/mv690/indicator/dbRDA
Data: http://tinyurl.com/mv690/seminars/indicator/data

Indicator species are species that preferentially occur in specific conditions. The indicator species analysis uses the relative abundance and relative frequency of occurrence of species in defined groups to narrow a large number of species down to a few species of interest that are the most characteristic species of the defined groups. You can then use the relative occurrence and abundance of these indicator species for the prediction of environmental conditions.

1.1

- Here is an example using a dataset which contains the abundance (% cover) of 28 grassland plant species recorded at 32 locations or “sites” in the dry mixedgrass prairie. The sites are divided into two ecosites: loam and loamy-sand. We can do an indicator species analysis to determine which plant species are indicators of loam or loamy-sand.

```
data = read.csv(“ecosite.csv”)
View(data)
```

- The indicator species analysis requires prior classification of the sites into groups. Here we are interested in loam vs loamy-sand. For other datasets you could obtain a classification using non-hierarchical cluster analysis.

- The indicator species analysis requires a vector of the site groups. The first 16 rows are loam and the next 16 rows are loamy-sand so we will group those.

```
groups = c(rep(1,16), rep(2,16))
groups
```

  So group 1 is loam and group 2 is loamy-sand.

- The second thing the indicator species analysis needs is the community data matrix with sites in rows and species in columns. The species are in columns 3 to 30.

```
data.frame(colnames(data))
matrix = data[ , 3:30]
```

- Then the indicator species analysis is done using the multipatt function in the indicspecies package. The multipatt function is a multi-level pattern analysis that calculates an indicator value for each species in association to the input groups and then finds the group with the highest association to each species. Then the statistical significance of the associations is tested using the indicator value as a test statistic in a permutation test. The option control=how(nperm=999)) determines how many random permutations (affects the precision of the p-value).

```
install.packages("indicspecies")
library(indicspecies)

indval = multipatt(matrix, groups, control = how(nperm = 999))

summary(indval, indvalcomp = TRUE)
```

- The summary shows that group 1 (loam) has PASC.SMI (western wheatgrass) as an indicator species. Group 2 (loamy-sand) has CALA.LON (sandgrass) and EQUI.SPPS (horsetail) as indicator species.

  The indicator species analysis uses an index called Indicator Value (the square root of the indicator value is shown on the generated table under column "stat") which is the product of two components also shown on the table: A and B.

  A is the probability that a site is in the target group given that the species has been found. It is the mean abundance of species in the target group divided by the sum of the mean abundance values over all groups. Horsetail (EQUI.SPPS) has an A value of 1 because it only occurs in the loamy-sand ecosite.

  B is the probability that you will find the species in a site that is in the target site group. B is the relative frequency of occurrence of the species in the target group. Western wheatgrass (PASC.SMI) has a B value of 1 because it occurs in every loam site. Sandgrass (CALA.LON) has a B value of 1 because it occurs in every loamy-sand site. A high B value means that if you do not find the species at a given site, the probability that the site belongs to the target group is low.

  The indicator value is maximum (i.e. 1) when all individuals of a species are found in a single group of sites and when the species occurs in all sites of that group. Western wheatgrass, Sandgrass, and Horsetail seem to be strong indicators of their groups because the square root of their indicator values are all nearly 1. They have high associations to their site groups.

- You can also look at a more comprehensive list of species and group associations, including ones that were not significant, by changing the significance level:

  ```
  summary(indval, indvalcomp = TRUE, alpha = 1)
  ```

  However, some species are missing from the list and are not in either group. This is because the indicator species analysis also looks at the combination of group 1 and group 2, and for some species the indicator value is the highest when the groups are combined. You can look at the object sign to find these species:

```
indval$sign
```

This shows every species. The columns s.1 and s.2 are group 1 and 2. These columns indicate (with ones and zeros) which site group was preferred by the species. The species that correspond the highest to both groups have the number one in both columns. AGRO.SCA (ticklegrass), ARTE.FRI (pasture sage), BOUT.GRA (blue grama grass), CARE.SPP (sedge), HESP.COM (needle and thread grass), KOEL.MAC (junegrass), etc. would not be good indicators for distinguishing between loam and loamy-sand ecosites because they are found similarly at both sites.

1.2

- Here is another short example. The indicator species analysis can be used to identify species that are responding to treatments in an experiment. In this example experiment, access mats, which are 3 x 8 m wooden portable platforms that provide a durable surface and buffer between construction traffic and the ground, were placed onto grassland for either 6 weeks or 24 weeks. 1 year after the access mats were taken off we want to do an indicator species analysis to identify species responding to treatments.

```
dat1 = read.csv("access_mat.csv")
View(dat1)
```

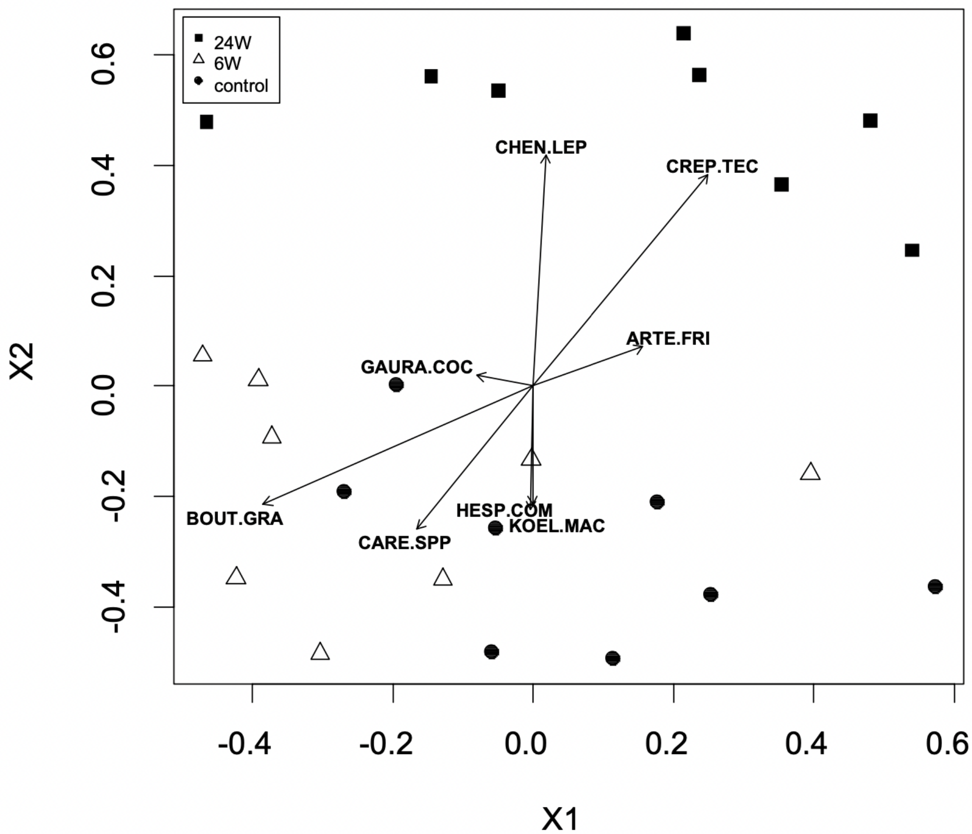
- Grouping by duration there are 3 groups: control (0 weeks), 24 weeks, and 6 weeks.

```
groups = c(rep (1,8), rep (2,8), rep (3,8))
matrix = dat1[ ,4:29]

indval = multipatt(matrix, groups, control = how(nperm = 999))
summary(indval, indvalcomp = TRUE)
```

The summary shows that early-season C3 perennial grass KOEL.MAC (junegrass) is a significant indicator of the control plots (group 1). Annual weeds CREP.TEC (narrowleaf hawksbeard) and CHEN.LEP (narrowleaf goosefoot) are significant indicators of the 24 week treatment (group 2). Native forb GAURA.COC (scarlet gaura) is a significant indicator of the 6 week treatment (group 3). Then there were a few species that had a higher indicator value when groups were combined, like ARTE.FRI (pasture sage) for the control and 24 week, and grasses/grass-likes BOUT.GRA (blue grama grass), CARE.SPP (sedge), and HESP.COM (needle and thread grass) for the control and 6 week.

- You could use the indicator species as vectors overlaid on an NMDS ordination of cover. The indicator species will not always be species with high r^2 because indicator species might not be present in high amounts of cover.



- The original indicator value method by Dufrene and Legendre (1997) only considered individual groups and not combinations of groups. If you want to restrict the analysis to only individual groups without site group combinations, it is possible using duleg = TRUE:

```
indval = multipatt(matrix, groups, duleg=TRUE, control =
how(nperm=999))

summary(indval, indvalcomp = TRUE)
```

- Note: the p-values in the multipatt test are not corrected for multiple testing. It is better to look at the the A and B values.

References:

De Caceres, M. 2020. How to use the indicspecies package (ver 1.7.8). Website: https://cran.r-project.org/web/packages/indicspecies/vignettes/indicspeciesTutorial.pdf

De Caceres, M.D. and Legendre, P. 2009. Associations between species and groups of sites: indices and statistical inference. Ecology, 90(12): 3566-3574.

Dufrene, M. and Legendre, P. 1997. Species assemblages and indicator species: The need for a flexible asymmetrical approach. Ecological Monographs, 67(3): 345-366.

References from video:

Broadbent, T.S., Bork, E.W. and Willms, W.D. 2016. Contingency in defoliation and moisture effects on northern mixedgrass prairie composition and diversity. Rangeland Ecology and Management, 69(4): 292-299.

Dufrene, M. and Legendre, P. 1997. Species assemblages and indicator species: The need for a flexible asymmetrical approach. Ecological Monographs, 67(3): 345-366.

Pyle, L.A. 2018. Influence of management and disturbance history on germinable seed bank composition and legume recruitment in Alberta’s Central Parkland and Dry Mixedgrass prairie. PhD Thesis, University of Alberta, Edmonton, AB.