NMDS to Study Dead Wood Fungi Communities in Parks of New Jersey

Christopher Zambell, Maria Shumskaya. Kean University, NJ, USA.

**Student handout**

**Introduction**

Dead wood is an essential component of any forest ecosystem. It protects soil against erosion, contributes to soil quality with massive organic and mineral inputs, improves water retention and creates multiple habitats for plants, animals and fungi. Enzymatic digestion of tough woody polymers such as lignin and cellulose is primarily performed by saproxylic fungi and bacteria, whose actions are complemented with additional mechanical disintegration by invertebrates[[1]](#footnote-1).

Dead wood fungi are the species that initialize ecological succession processes that sustain forest biodiversity, which makes them invaluable forest organisms. The decomposition pathways vary greatly among the tree species, surrounding biotopes, the landscape matrix, forest history, spore rain and numerous other factors. Different types of saproxylic fungi demonstrate a gradient of enzymatic abilities, specific to substrates they digest such as hardwood or softwood.

In this exercise, we will perform multivariate statistical analysis called NMDS (non-metric multidimensional scaling) ordination of the dead wood fungal communities collected at different parks and reservations across New Jersey, USA. The collection sites were chosen so as they cover both northern parts of the state with hardwood or mixed forests, and southern parts of the state with softwood forests (Pine Barrens).

Ordination analysis orders objects in such way that similar objects are near each other and dissimilar objects are far away; once we perform this analysis on the dataset of dead wood fungi, we will be able find differences between species communities, if there are any, and map them in space.NMDS is a way of plotting sampling sites and species on a 2D or 3D graph in such a way that rank order of differences between sites is retained. In other words, similar communities will be located closer to each other on the plot, and dissimilar communities will be further from each other.

We will use R package vegan to perform this analysis. Software used was: R / RStudio; packages: dplyr and vegan.

The dataset used in this module is an adapted version from the full occurrence dataset produced by a group of student researchers and faculty of Kean University, NJ, USA[[2]](#footnote-2). The full dataset is available at the Global Biodiversity Information Facility portal:

https://www.gbif.org/dataset/d1d59f9f-e130-4715-bc9b-8a25e7b62e89

Occurrence datasets provide evidence of the occurrence of a species (or other taxon) at a place on a specified date. They contain general locality information to support fine-scale analysis and mapping of species distributions. The column names in this dataset are written using Darwin Core (DwC) term names: <http://rs.tdwg.org/dwc/terms/>.

You will be provided a CSV (comma delimited) file created from this dataset with species as columns, and collection sites as rows. Species data is either Presence/Absence (0/1). The description of all DwC term names from the file is described in the table below:

|  |  |
| --- | --- |
| **DwC term** | **Explanation** |
| basisOfRecord | Specific nature of the data record (e.g. human observation or preserved specimen) |
| eventDate | Date the event occurred |
| scientificName | Name of the species (or other taxon) observed |
| kingdom | The full scientific name of the kingdom in which the taxon is classified |
| countryCode | The standard code for the country in which the Location occurs (US for the USA) |
| stateProvince | The name of the next smaller administrative region than country (state, province, canton, department, region, etc.) in which the Location occurs |
| locality | The specific description of the place. |
| habitat | A category or description of the habitat in which the Event occurred. |
| decimalLatitude | The geographic latitude (in decimal degrees, using the spatial reference system given in geodeticDatum) of the geographic center of a Location. |
| decimalLongitude | The geographic longitude (in decimal degrees, using the spatial reference system given in geodeticDatum) of the geographic center of a Location. |
| geodeticDatum | The ellipsoid, geodetic datum, or spatial reference system (SRS) upon which the geographic coordinates given in decimalLatitude and decimalLongitude as based. |
| recordedBy | A person, group, or organization responsible for recording the original Occurrence. |

**Preliminary question:** do you think there will be a correlation between types of communities and latitude of locations where the sampling was performed? Formulate a hypothesis.

**Procedure**

Open NMDS\_script\_students.R in R studio. Import the dataset deadwood.csv. Follow the steps in the script.

**Glossary:**

**Stress** is a statistic of goodness of fit in NMDS.

**Shepard plot**: plot with the ordination distance on y axis and community dissimilarities on x axis.

**Ordination distance**: how far apart the points on the NMDS graph.

**Community dissimilarities:** how dissimilar are the communities of species.

Basically, if these variables fit well, that means we can successfully visualize how the communities differ from each other.

**Hull**: a binding line, a boundary

**Please answer the questions:**

1. How many columns are in the dataset deadwood?
2. View the dataset deadwood, how many fungal species are there? (Hint: subtract non-species columns from the total number of columns).
3. How many rows are in the dataset deadwood?
4. How many localities (field sites) are in the dataset? (Hint: check the rows).
5. How many dimensions were used? What was the stress level?
6. Is the stress level acceptable?
7. How is the non-metric fit (R squared)? This is what we are interested in. Do the blue points follow the line closely?
8. What do you think cex command does?
9. Do the pine (green hull) and the hardwood forest (brown hull) overlap?
10. Do either of the factors (latitude, longitude) have a significant p value (Pr)? What does it tell us about the correlation between the factor and NMDS axes?
11. Which type of communities tend to be found with higher latitudes? Which are found at lower latitudes?
12. Write the names for the three species you selected.
13. Take a screenshot of your complete zoomed in NMDS (or save as TIFF) and insert here:

1. Fill in the table:

|  |  |  |
| --- | --- | --- |
| **Species name** | **Parks (sites) found** | **Hardwood or pine barrens?** |
|  |  |  |
|  |  |  |
|  |  |  |

1. Can you spot a relationship between where the species is found and the position of the red cross representing this species in your NMDS graph? Please explain. Provide one example using a species from the Table above.
2. What can you say about the specificity of dead wood fungal communities? Is there a difference between communities in hardwood vs pine forests?
1. Bobiec A, Gutowski JM, Zub K, Pawlaczyk P, Laudenslayer WF (2005) The afterlife of a tree. WWF Poland. [↑](#footnote-ref-1)
2. Shumskaya M, Zambell C, Mishra S, Bell E, Blue S, Yearwood-Marut J, Marut W, Vindas-Cruz A, Jennings A, Hylton N, Burghardt J (2019). Survey of saproxylic fungi across parks of New Jersey. Kean University. Occurrence dataset https://doi.org/10.15468/ngpb5m accessed via GBIF.org on 2019-07-26. [↑](#footnote-ref-2)