

Using the Niche Programs in R

Shirley Pledger and Shane Geange

16 February 2010

1 Introduction

This file documents the running of the **R** niche overlap program found in **NicheProg.txt**.

See the Research Report (Pledger and Geange 2009) for the theory and statistical methods used:

<http://msor.victoria.ac.nz/Main/ResearchReportSeries>

The user should have the following files in the same directory:
two program files `NicheProgA.txt` and `NicheProgB.txt`,
two data files `ExampleA.csv` and `ExampleB.csv`,
two availability files `ExampleAhabitatAvail.csv` and `ExampleBhabitatAvail.csv`
and one file of R functions, `NicheFunctions.txt`.

There are two types of analysis, one comparing the niches of two or more species (or taxa) at one site (Example A), and the other comparing the niches of one species (or taxon) over several sites (Example B).

2 Example A

In this example, different species are compared for niche overlaps within a single site. The basic data file `ExampleA.csv` has its first two columns labelled `id` (for individual) and `species`. Subsequent columns are the variables of different types which were measured.

Since one of the variables, `habitat`, is of the “resource selection” type, we must also supply a matching availability file `ExampleAhabitatAvail.csv`. There must be one availability file for each resource-selection type of variable.

The variables may be habitat where found, food eaten, morphological columns, etc. They may refer to resource usage, or be categories, counts, or continuous data. If continuous, they may be measurements or ratios of measurement. The possible types of variables are:

"cat"	=	categorical, but not resource selection
"bin"	=	binary
"cts"	=	continuous, use raw data (no transformation)
"meas"	=	measurement, continuous positive, take logs
"pcent"	=	percentage data, bounds at 0 and 100, use logits
"propn"	=	proportion data, bounds at 0 and 1, use logits
"count"	=	count
"rsel"	=	resource selection, categorical

The variable types need to be specified while running the program.

The commands in the program file `NicheProgA.txt` are to be copied and pasted into R, using the directory which contains the program and data files. At some stages in the program file, you need to type in appropriate details, e.g. specifying the types of variables.

When supplying your own data, make the first two columns `id` = individual and `species` = species or some other taxonomic group. The data should have the column names across the top, and thereafter one row per individual.

For any resource type of variable, there must be an associated file giving the availabilities. The first row should be the names of the choices, and the second row is either the percentages or the proportions of the different choices. In Example A, there is one resource selection variable, called `habitat`, with three choices, `grass`, `forest` and `rock`. Within the file, the choices are sorted alphabetically, and a check is done to be sure the choices named in the availability file match those in the appropriate column of the main data file.

There are comments throughout the program to explain what is being calculated at each stage.

3 Example B

In this example, one species is compared for its niche choice over two or more sites. Is there substantial overlap, or is the species occupying different niches at the different sites?

For any resource-selection type of variable, it is necessary to allow for different availabilities when comparing over the different sites. Also, the comparisons may only be done for resource choices which are available at all the sites. If one choice is unavailable at one site, we cannot know if it would have been chosen, if available.

We allow the data set to include all the information collected, but the program restricts the analysis to choices available at all sites.

In `ExampleB.csv` there is one resource selection variable, `habitat`, with four choices, `rock`, `grass`, `sand` and `forest`. However, the availability file `ExampleBhabitatAvail.csv` shows that `sand` is not available at Site 4. Hence the analysis will discard all individuals observed on `sand`.

If we wish to include `sand` as a choice, we would need to supply data only for Sites 1, 2 and 3.

4 Overview of program

The program does these steps:

1. Read in the data, store the species or site names etc.
2. Calculate the electivity score (Manly's alpha) for any resource usage variable.
3. Calculate niche overlaps.
4. Do null model (permutation) tests.

5 Analysing your own data

To analyse your own data, copy the layout styles in the examples. Data may be set up using Excel, then converted to `.csv` for use with R.

For an analysis of type A, ensure the first two columns are named `id` and `species`.

Extra analyses of type A may be run, comparing pairs of species. If all possible pairs are compared, an adjustment for multiple testing (e.g. the sequential Bonferroni) should be used. To do a pairwise analysis, simply provide a data set with only the two species included.

For an analysis of type B, ensure the first two columns are `site` and `id`.

Type B analysis may be run with different data sets. In example B, the analysis uses all four sites, but excludes any observations with `sand` as the chosen habitat, since `sand` is not an available choice at all sites. An alternative analysis could be to exclude Site 4 and run the analysis including `sand`, which was available at all the other sites.

Again, pairwise comparisons between sites may be done, by supplying data from only two sites at a time. Adjustment for multiple testing is necessary.