

SDR-abunSimplex

User's Manual

by J. Podani

Introduction

The WINDOWS application **SDR-abunSimplex** has been written to calculate the Similarity-Abundance difference-Abundance replacement simplex as proposed by Podani, Ricotta and Schmera (2013). Full description of the procedure is given in this *Ecological Complexity* paper, and is not reproduced here.

Input files – a sample run

In general, any data matrix containing abundance data can be analyzed. The program accepts data with species as rows or as columns (i.e., sites as columns or rows, respectively); the program prompts the user to specify which arrangement is actually given.

In the example, we analyze the random data set as given in Fig. 2, bottom, of Podani et al. (2013). The file format is as follows:

```
Random
10 5
1 2 1 0 3
0 1 0 0 4
4 0 4 1 1
2 1 0 0 0
3 2 1 4 0
2 2 0 0 1
0 4 2 0 0
1 5 0 3 2
0 0 0 1 0
0 6 0 3 2
```

The data file follows the SYNTAX 2000 raw data format. The first line in each file is a title, the second line contains the number of rows (n) and the number of columns (m) of the matrix. Then follow n lines, each corresponding to a species (or site). If necessary, each row in the matrix can be broken into several lines. It is important, however, that every new row in the matrix begins in a new line in the file. At the end, please enter a linefeed character. In the example above, we have 10 species and 5 sites, but the program accepts the same matrix in

transposed form as well. The input file must be present in the same directory as the program.

Sample run and output

The routine is a stand-alone application. After double-clicking the **SDR-abunSimplex** icon, a dialog window appears on the monitor. Suppose that the input file is `RANDOM.TXT`. First, this filename has to be specified by the user and then we specify whether sites are in columns (0) or rows (1). Then, the program outputs the simplex results in percentages, first for the 2D simplex (S, D and R) and then for the three 1D simplices (R+D, S+R and S+D). The latter, the nestedness simplex is calculated in two different ways, as described in the paper (i.e., removing only the anti-nestedness fraction or the anti-nestedness AND the richness identity fractions from S+D). Finally, the program lists three default filenames for the detailed output.

```
ENTER INPUT FILENAME
random.txt
0=analysis between cols, 1=analysis between rows
0
```

PERCENTAGES		
S	SIMILARITY	22.6973
R	ABUND. REPLACEMENT	51.4662
D	ABUND. DIFFERENCE	25.8365
R+D	BETA DIVERSITY	77.3027
S+R	ABUND. AGREEMENT	74.1635
ANTI-NESTEDNESS FRACTION		.000000
S+D-ANTI NESTEDNESS		48.5338
SIMPLE GRADIENT FRACTION		1.81818
S+D-ANTI-GRAD NESTEDNESS		46.7156

```
FULL SET OF SIMPLEX SCORES SAVED IN FILE resu.ord FOR NONHIER
FULL SET OF COORDINATES SAVED IN FILE simplex.dat
PAIRWISE COEFFICIENTS SAVED IN FILE COEFF.ord
10 PAIRS TAKE 9 POSITIONS
REDUCED SET OF SIMPLEX SCORES SAVED IN REDUC.ORD FOR NONHIER
STOP
```

In the output file `resu.ord`, we find the 2D simplex scores which can be directly displayed as a ternary plot using the **Ternary plot** graphics option in the **NonHier** module of the SYN-TAX 2000 package (Podani 2001). In this file true simplex scores are given for each pair of sites (starting with pair 1-2, then 1-

3, 2-3, 1-4, 2-4, 3-4, etc.), rather than percentages. The output sequence by columns is R, D and S

```
Random
10 3
.414      .345      .241
.267      .333      .400
.700      .050      .250
.818      .000      .182
.296      .556      .148
.296      .407      .296
.357      .357      .286
.667      .222      .111
.632      .263      .105
.700      .050      .250
.000      22.697    51.466    25.836    77.303    74.164
48.534    .000
```

The file `simplex.dat` contains the coordinates of points for reproducing the point scatter using any graphics program that displays 2D scatter diagrams. The last three points specify the corners of the triangle.

```
Random
13 2
.634      .507
.754      .327
.849      .858
.836      1.002
.419      .363
.629      .363
.657      .438
.629      .817
.595      .774
.849      .858
.707      1.225
0.000      .000
1.414      .000
```

File `reduc.ord` contains simplex scores only for those points that do not overlap. This file may be useful to display the ternary plot, with reduced memory use, when there are too many overlapping points. In the present small example, the number of points is reduced from 10 to 9 – i.e. two points overlap in the diagram. The output sequence by columns is R, D and S

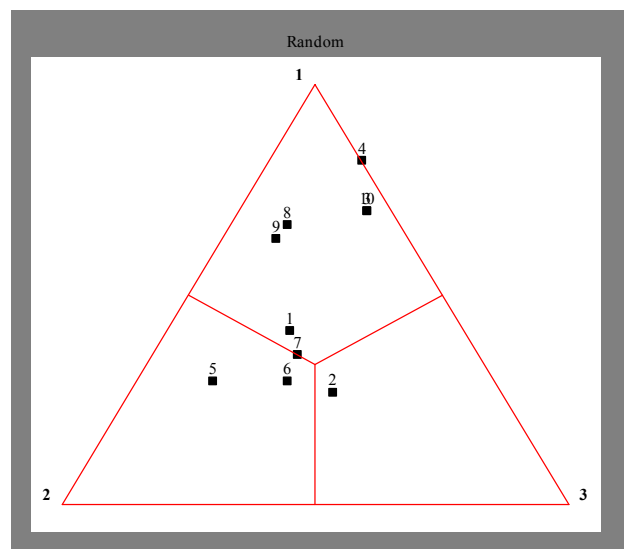
```
Random
9 3
.413793    .344828    .241379
.266667    .333333    .400000
.700000    5.000000E-02    .250000
.818182    .000000    .181818
.296296    .555556    .148148
```

.296296	.407407	.296296
.357143	.357143	.285714
.666667	.222222	.111111
.631579	.263158	.105263

Note that file COEFF.ORD is reserved for technical purposes.

The Ternary plot

Using **NonHier** (Podani 2001), we get the following graphics result, when displaying the data from either `resu.ord` or `reduc.ord`



In this diagram, “3” corresponds to the S corner, “2” corresponds to the D corner and “1” corresponds to the R corner. Each pair of sites is represented by a square symbol.

IMPORTANT!

It can happen (very rarely with actual data) that some column in the input files (either `resu` or `reduc`) contains the same values (for example, zeros). In this case, the graphic routine cannot display the plot. The solution is to edit the file such that some variance is added to the constant values. For example, if the file is

Nested abundance pattern

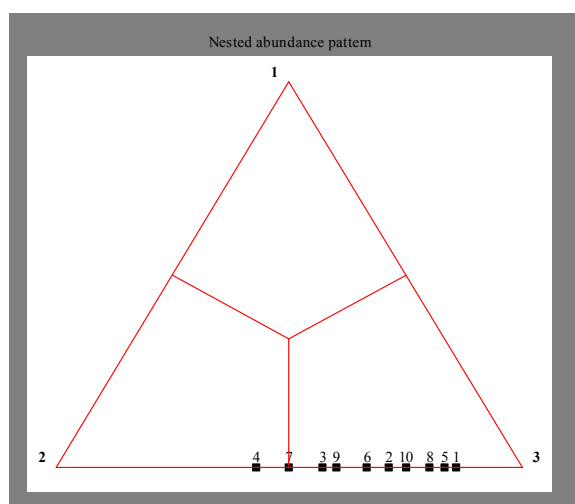
10	3	
.000000	.142857	.857143
.000000	.285714	.714286
.000000	.428571	.571429
.000000	.571429	.428571
.000000	.166667	.833333
.000000	.333333	.666667
.000000	.500000	.500000
.000000	.200000	.800000

```
.000000 .400000 .600000
.000000 .250000 .750000
```

which is the case for the nested pattern in the small data on p. 60 in the paper, then the file should be modified a little bit, for example,

```
Nested abundance pattern
10 3
.0005 .142857 .857143
.000000 .285714 .714286
.000000 .428571 .571429
.000000 .571429 .428571
.000000 .166667 .833333
.000000 .333333 .666667
.000000 .500000 .500000
.000000 .200000 .800000
.000000 .400000 .600000
.0002 .250000 .750000
```

and then the plot will be displayed correctly



References

- Podani, J. 2001. SYN-TAX 2000. Computer programs for data analysis in ecology and systematics. User's Manual. – Scientia.
- Podani, J., C. Ricotta and D. Schmera. 2013. A general framework for analyzing beta diversity, nestedness and related community-level phenomena based on abundance data. *Ecological Complexity*, **15**: 52–61.