Program for biplot species scores

User's notes

Pierre Legendre July 1999

After distance-based redundancy analysis (db-RDA, Legendre and Anderson, 1999) using Bray-Curtis (or other) distance, species scores may be obtained by computing correlations between the species and the fitted site scores, also called "site scores that are linear combinations of environmental variables" or "sample scores that are linear combinations of environmental variables".

The correlations have to be weighted as follows before being used to draw the species as arrows in biplots:

$$SpeciesScore(jk) = r(jk)*s(j)/s(k)$$

where r(jk) = correlation between species j and fitted site score vector k,
s(j) is the standard deviation of species j,
and s(k) is the standard deviation of fitted site score vector k.

In standard RDA, the lengths of the species score vectors are 1 in distance biplots and they are equal to the square roots of the corresponding canonical eigenvalues in correlation biplots (Legendre and Legendre 1998), just as in PCA.

Different programs may scale the site scores in different ways, however, as can be seen by comparing the three examples below. This may result in species scores that are larger than the original ones by a constant, which does not change the interpretation of biplots. The RDA scalings implemented in different programs are described in Legendre and Legendre (1998, pp. 585-586).

This program offers the option of scaling the species scores as in standard RDA, or as in program CANOCO versions 3 or 4. Compared to standard RDA, described in Legendre & Legendre (1998), the standard deviations of the site scores differ by a constant in CANOCO, so that the site scores also differ by the following constants:

n / sqrt(totinert)	in CANOCO 3.10
sqrt(n*p) / sqrt(totinert)	in CANOCO 4

where n is the sumber of objects (e.g., sites), p is the number of variables (species), and 'totinert' is the total inertia in the species matrix.

Input file 1: Species data text file. Rectangular table of species presence-absence or frequency data where the rows correspond to objects (e.g., sites) and the columns to species. There are no row (= site) or column (= species) identifiers. Data are separated by spaces or tabs. It is recommended to add a carriage return at the end of the last row of data.

Input file 2: Text file of "fitted site scores", or "sample scores that are linear combinations of environmental variables", where rows correspond to objects (e.g., sites) and columns to canonical eigenvalues. This table is copied from the output file of CANOCO or similar program into an ASCII (i.e., text) file. There are no row (= site) or column (= species) identifiers. Data are separated by spaces or tabs. The program checks that the number of rows is the same in input files 1 and 2. Add a carriage return at the end of the last row of site scores.

Output file: Table of species scores for biplots. The rows correspond to the species and the columns to the canonical eigenvalues. This file may be added to the Species Scores and Site Scores (or Fitted Site Scores) tables to produce biplots. An automatic biplot drawing procedure is available in *The R Package* (Casgrain and Legendre, 1999).

The examples that follow were all scaled to obtain distance biplots. Program SPECIESSCORES can also compute biplot species scores from "fitted site scores" scaled to obtain correlation biplots. In all cases, users of CANOCO should request scalings for covariance-based scores in CANOCO 3.1 (negative-number options), or without post-transformation in CANOCO 4.

Example 1

Coral reef fish data from Table 11.3 of Legendre and Legendre (1998, p. 590). There are 10 sites (rows) and 6 species (columns).

Input file 1: Species file (text file; no row or column identifiers).

```
1   0    0    0    0    0
0   0    0    0    0    0
0   1    0    0    0    0
11  4    0    0    8    1
11  5   17    7    0    0
9   6    0    0    6    2
9   7   13   10    0    0
7   8    0    0    4    3
7   9   10   13    0    0
5  10    0    0    2    4
```

Input file 2: fitted site scores from program RDACCA, which can be obtained from the WWWeb site <http://www.fas.umontreal.ca/BIOL/legendre>. Scaling 1 was used here to obtain a distance biplot. The columns correspond to the three canonical eigenvalues. (Text file; no row or column identifiers).

```
 -6.79498   5.49498  -2.24897
 -6.96197   5.91719  -0.63774
 -7.12895   6.33941   0.97349
 -3.55205  -6.52301  -4.39356
 12.69996   0.24686  -3.17159
 -3.88603  -5.67858  -1.17109
 12.36599   1.09129   0.05088
 -4.22000  -4.83415   2.05138
 12.03201   1.93572   3.27335
 -4.55398  -3.98972   5.27384
```

Output file: Species scores for biplots, computed by program SPECIESSCORES. The rows correspond to the species and the columns to the canonical eigenvalues. This table is identical to the table of species scores computed by program RDACCA and reported in Table 11.4 of Legendre and Legendre (1998, p. 591).

```
  0.30127  -0.64624  -0.39939
  0.20038  -0.47265   0.74458
  0.74098   0.16813  -0.25689
  0.55013   0.16841   0.26114
 -0.11588  -0.50594  -0.29319
 -0.06292  -0.21535   0.25679
```

Example 2

Input file 1: Same species file as above.

Input file 2: fitted site scores from program CANOCO 3.1. Scaling –1 was used to obtain a distance biplot with covariance-based scores. The columns correspond to the three canonical eigenvalues. (Text file; no row or column identifiers).

```
 -0.6741    -0.5452    -0.2231
 -0.6907    -0.5870    -0.0633
 -0.7073    -0.6289     0.0966
 -0.3524     0.6471    -0.4359
  1.2600    -0.0245    -0.3147
 -0.3855     0.5634    -0.1162
  1.2268    -0.1083     0.0050
 -0.4187     0.4796     0.2035
  1.1937    -0.1920     0.3247
 -0.4518     0.3958     0.5232
```

Output file: Species scores for biplots. The rows correspond to species and the columns to canonical eigenvalues.

```
  0.95271    2.04359   -1.26323
  0.63365    1.49471    2.35440
  2.34316   -0.53170   -0.81282
  1.73965   -0.53257    0.82547
 -0.36643    1.59994   -0.92714
 -0.19897    0.68102    0.81207
```

Example 3

Input file 1: Same species file as above.

Input file 2: fitted site scores from program CANOCO 4. Biplot scores emphasizing inter-sample distances, without post-transformation (scaling –1), were computed to obtain a distance biplot. The columns correspond to the three canonical eigenvalues. (Text file).

```
 -0.6741    -0.5452    -0.2231
 -0.6907    -0.5870    -0.0633
 -0.7073    -0.6289     0.0966
 -0.3524     0.6471    -0.4359
  1.2600    -0.0245    -0.3147
 -0.3855     0.5634    -0.1162
  1.2268    -0.1083     0.0050
 -0.4187     0.4796     0.2035
  1.1937    -0.1920     0.3247
 -0.4518     0.3958     0.5232
```

Output file: Species scores for biplots. The rows correspond to species and the columns to canonical eigenvalues.

```
  0.73797    1.58296   -0.97849
  0.49082    1.15779    1.82371
  1.81501   -0.41186   -0.62961
  1.34753   -0.41253    0.63941
 -0.28384    1.23931   -0.71816
 -0.15412    0.52751    0.62902
```

Program distribution

A computer program to compute the biplot species scores is available from the following WWWeb site:

• FORTRAN source code (SpeciesScores.f) and compiled versions, written by P. Legendre: <http://www.fas.umontreal.ca/biol/legendre/>

References

Casgrain, P. and P. Legendre. 1999. *The R Package for multivariate and spatial analysis, version 4.0 – User's manual.* Département de sciences biologiques, Université de Montréal. Available from the WWWeb sites <http://alize.ere.umontreal.ca/~casgrain/R/> and <http://www.fas.umontreal.ca/BIOL/legendre>.

Legendre, P., and M. J. Anderson. 1999. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs* 69: 1-24.

Legendre, P. and E. Gallagher. Ecologically meaningful transformations for ordination biplots of species data. *Ecology* (submitted).