

[IS-Online-Publication No. 3 | Berlin, September 2002]

nNMDS-Statistic, a Computer-Program for Nonmetric Multidimensional Scaling

by INGO SCHINDLER, Berlin

Part I

File: nNMDS.exe
VisualBasic Program for Windows PCs

Part II

User Guide

Part II

DISCLAIMER

This Visual Basic program is provided without any explicit or implicit warranty of correct functioning. It has been developed for my own research program. To use this program on your computer is your own risk. The author or any other person who distributed this uncommercial program doesn't take any guarantee. If, however, you should have encounter problems with this program, the author will help solve them. Researchers can use this program free for research purposes, but intellectual property and copyright remains the property of Ingo Schindler, Berlin.

Users of the program may refer as follows:

Schindler, Ingo (2002). nNMDS-Statistic, a Computer-Program for Nonmetric Multidimensional Scaling. IS-Online-Public. No. 3.

© 2002 INGO SCHINDLER, 12051 Berlin (Germany)

e-Mail: ingoschindler@web.de

website: www.i-sch.de.vu

1. Description

nNMDS is a computer program that calculates a nonmetric multi-dimensional scaling as a method of nonlinear ordination. It is called nonparametric NMDS because the program transferred the raw data into rank data. This limited the influence of outlier and standardized the variables. The mean city-block metric

$$D_{ij} = (1/n) \sum \text{abs}(X_{ik} - X_{jk})$$

based on the rank data is used to measure the dissimilarity between OTUs. However, the program is also useful for ordination of multivariate data on the basis of any distance matrix. The NMDS itself use only the rank values of the distance matrix.

The nonmetric multi-dimensional scaling is based on the Program NMDS.BAS by Ludwig & Reynolds (1988). This follows the algorithms of KRUSKAL and SHEPARD as modified from a program by ORLOCI & KENKEL, but normalize the configuration to unit size and rotate it to its principal axis. The results from this program may differ slightly from other NMDS programs because of differences in rotation and scale and the use of different algorithms to optimize the rank order (INTERACTIVE BASIC PROGRAM: NMDS.BAS; Ludwig & Reynolds, 1988). The iterations stopped when the value for stress is below 0.001 or when the difference between the successive stress values is below 0.00001. The maximum number of iterations is limited to 200.

2. Procedure

- a) Select data manipulation in the 'adjust' frame and final dimensions in the 'Dimensions' frame.
- b) Select input file (using drive-, directory- and file-box).
- c) Use 'Options' for special applications (see Chapter 5).
- d) Press 'Rank' to start transforming the raw data into rank data.
- e) Press 'Save' if you like to save the rank data.
- f) Press 'Distance Matrix' to calculate the distance matrix
- g) Press 'Save' if you like to save the distance matrix.
- h) Change output file name in the 'Text' box (if necessary).
- i) Press 'nMDS' to calculate the nonmetric multi-dimensional scaling (NMDS).
- j) Press 'Save' to save the results.
- k) Press 'Close' if you finish, or calculate a further nMDS with a new random start configuration (see chapter 5 for more information).

3. Input File

The input file have to be saved as ASCII *.txt, *.cic, *.isv or *.dis. For the necessary data format of raw data see the example (Appendix: Test run) below (data from Costa, 2002). The program allows also the user to load a distance matrix. The distance matrix is in ASCII (*.dis), square (N x N) with a proceeding line for the value of N followed by a number for the dimension of the start configuration (if your data don't include a start configuration use '0' [null]). The program is not checking the input file format and will terminated automatically if the format is not correct.

The limitations for the program are:

Maximum number of variables = 50

Maximum number of observations (OTUs) = 180

Maximum number of groups = 22

4. Output Files

The results are saved in the directory which is shown in the directory box. The final output file (*.txt) is called according to the name in the 'Text' box (see Appendix: Test run for an example). The output files during the procedure (see chapter 2.) are named with the first three letters of the input file as *.rkk or *.dis. For further informations see chapter 5. The final low dimensional data (1D, 2D or 3D) can be imported into any graphics package. Be caution, the program doesn't check if the output file name already exists. Thus it may happen that your old data are deleted.

5. Options

5.1. Start-Configuration

The results of NMDS are strongly depended on the start configuration of the OTUs. If you don't have any idea of the distribution of the OTUs in the low dimensional space, than use the 'Random' start configuration (see Appendix 2). The 'UserData' bottom is only available when your input file is including such data (see Appendix: Test runs). The values for the start configuration data have to be between -1 and +1.

5.2. Automatic Runs

This option is (only) useful with a randomized start configuration. You can check the results and search for the final data with the lowest stress. Or you can use the 'ntsFormat' option to calculate a consense representation (see option 'ntsFormat'). When the limited number of runs (select up to 50) is reached the program is terminated. The name of the output file is added by the the number of automatic runs.

5.3. Jackknife

With the 'Jackknife' option it is possible to calculate "confidence" limits of the final low dimensional representation. In each run an other variable is eliminated and not included in the measure of dissimilarity. This option should be used with 'ntsFormat' to calculate consens representation and the confidence limits (see option 'ntsFormat') or to separate different complexes of variables. The name of the output file is added by a 'J' and the number of variables.

5.4. TernaryData

This option prints TernaryData into the result file. These data can be imports into any graphic package, which is able to draw Ternary – plots.

5.5. ntsFormat

The use of this option print an additional output file with the same name as the result file in *.nts format. This format is used by the program tpsRelw (for 2D) and tpsSmall Version 1.19 (F. James Rohlf, Department of Ecology and Evolution, State University of New York, Stony Brook, 1998) to calculate consens configuration and/or superimposition of jackknife ordinations. The program GRF-ND (Generalized rotational fitting of n-dimensional landmark data) by Dennis E. Slice is also helpful (see Appendix: 2). All these programs are available through the internet (<http://life.bio.sunysb.edu/morph>) © 1999 by F. James Rohlf).

6. Frames

6.1. 'adjust'

Are the data raw measurements (e. g. in mm) use one of the options to scale the data to size free. The use of index data are somewhat problematic (Reist, 1985) and it seems best to use

$$\log(\text{SL}) - \log(\text{Y})$$

[where $\log(\text{SL})$ means the logarithmical Standard length (use this notion as your first variable, when such application is wanted) and $\log(\text{Y})$ any further logarithmical transformed variable] (see Fricke, 1982). If your data are already in percents of standard length (as in the example; see Appendix: Test run), or if you use a more sophisticated size free metric (see Reist, 1985) use the option 'none'.

6.2. 'Dimensions'

Select the final dimensionality 1D, 2D or 3D. Be aware that the stress is not to high. It should be lower than 0.2 for a good ordination.

7. Literature and sources

- Backhaus, K., B. Erichson, W. Plinke & R. Weiber (2000): Multivariate Analysemethoden. Springer, Berlin.
- Costa, W. J. E. M. (2002): The *Austrolebias alexandri* species Group: A taxonomical revision of an annual fish clade (Cyprinodontiformes: Rivulidae) in southern Brazil. *Comun. Mus. Cienc. Tecnol. PUCRS, Ser. Zool.* 15(1): 87-111.
- Cox, T. F. & M. A. A. Cox (2001) Multidimensional Scaling. 2nd Edition. Chapman & Hall, Boca Raton.
- Dunn, G. & B. S. Everitt (1982): An introduction to mathematical taxonomy. Cambridge University Press, Cambridge.
- Fahrmeir, L. & A. Hamerle (1984): Multivariate statistische Verfahren. Walter de Gruyter, Berlin.
- Fricke, R. 1982. Modifizierung und Anwendung von McCunes Körperform-Messsystem für rezente benthische Fische (Pisces). *Braunsch. Naturk. Schr.* 1: 533-559.
- Jardine, N. & R. Sibson (1971): *Mathematical Taxonomy*. Wiley & Sons, New York.
- Lozan, J. L. und H. Kautsch (1998): *Angewandte Statistik für Naturwissenschaftler*. Parey Studientexte 74. 2., überarbeitete und ergänzte Auflage. Parey Buchverlag Hamburg & Berlin.
- Ludwig, John A. and James F. Reynolds. 1988. *Statistical ecology: a primer of methods and computing*. Wiley Press, New York, New York.
- Manly, B. F. J. (1997): *Randomization, Bootstrap and Monte Carlo Methods in Biology*. Second edition. Chapman & Hall, London.
- Mayr, E. 1975. *Grundlagen der zoologischen Systematik*. Parey, Berlin & Hamburg.
- Reist, J. D. (1985): An empirical evaluation of several univariate methods that adjust for size variation in morphometric data. *C. J. Zool.* 63: 1429-1439
- Rohlf, F. J. (1998): tpsSmall, Version 1.19. SUNY at Stony Brook, Stony Brook, New York.

! Be patient, the program is slow, even with small data sets !

Appendix 1: Test run

The data used in this test run are from Costa (2002: Tab. 1-5). The notes in brackets are only for explanation and not part of the data file. Copy the data into editor program and save it. Follow the procedure (Chapter 2). The results and final configuration in your test may differs from the representations of Fig. 1-4, because of a different random start configuration and rotations.

```
[start]
38          [number of OTUs]
15          [number of variables]
2           [dimension of the start configuration]
10         [number of group names]
Alm        [group names]
ALf
IBm
IBf
CYm
CYf
PEm
PEf
AFm
AFF
Body depth [names of variables]
Caudal peduncle depth
Predorsal length
Prepelvic length
Dorsal-fin base
Anal-fin base
Caudal-fin length
Pectoral-fin length
Pelvic-fin length
Head length
Head depth
Head width
Snout length
Lower jaw length
Eye diameter
MDS1       [name of the start configuration]
MDS2       [name of the start configuration]
[data as following format]
1 33.8 14.8 46.7 40.3 44.2 46.6 34 26.5 11.6 24.6 30.1 18 3.8 7.1 9
[first number of each row is the group indicator; the following are the
variables]
1 37 17 46.3 45.3 45.4 47 37.5 31.1 11.5 28.8 30.4 20.3 3.6 7.3 10.1
1 35.5 15.1 45.5 44.1 44.9 46.1 34 26.6 11.8 26.8 30.2 19.3 4.3 7.5 9.5
1 35.9 17 51.4 47 46.2 44.7 35.6 30.3 13.8 30.6 31.4 19.5 4.5 6.7 10.9
2 34.5 15.2 61.7 55.2 31.2 25.2 38.5 27.1 14.2 29.1 29.4 21.1 4.4 7.1 10.3
2 38.6 17 62.6 58.9 32.3 26.6 38.5 26.9 13.9 30.3 32.7 22.6 5 6.7 10.4
2 34.7 14.8 63.9 57.1 30.1 29.2 33.6 24.6 13.3 28.8 28.9 21.3 4.6 6.4 10.2
2 36.9 14.4 60.7 58.3 29 24.9 37.6 28.9 14.7 31.1 30.1 23.2 5 6.6 11.3
3 34 16.1 49.6 47 42.7 41.7 32.2 26.7 12.2 27.2 29.2 18.6 4 7.2 8.9
3 38.6 16.4 49.2 48.7 44.6 39.8 37.7 27.5 13.3 30.5 29.7 20.2 4.1 7.4 9.5
3 35.9 15.7 49.3 45.7 43.4 42.8 31.5 27.7 11.1 28.8 29.7 18.8 4.2 7.6 9.1
3 34.6 14.6 49.8 45.5 42.8 41.8 35 28.4 12 27.9 28.8 20.7 4.5 6.4 9
4 35.2 14.4 61.9 53.3 28.6 25.9 32.8 25 12.5 29 28.6 21.9 3.9 6.4 9.6
4 37.3 15.4 66.9 57.9 27.5 23 38.7 26.8 15.3 32.3 31.1 21.9 4.5 7 10.8
4 34.5 14.2 60.7 53 33 28.1 37.8 25.6 11.4 30.1 27.9 21.5 4.6 6.6 9.9
4 24 14.9 62.3 52.2 28.3 25.6 35.3 22.8 11.8 31 29.2 22.1 4.2 6.7 10.1
5 37.3 14.2 47.7 44.4 47.4 46.6 29.3 24.8 11.1 26.4 29.4 17.6 3.8 6.1 8.9
```

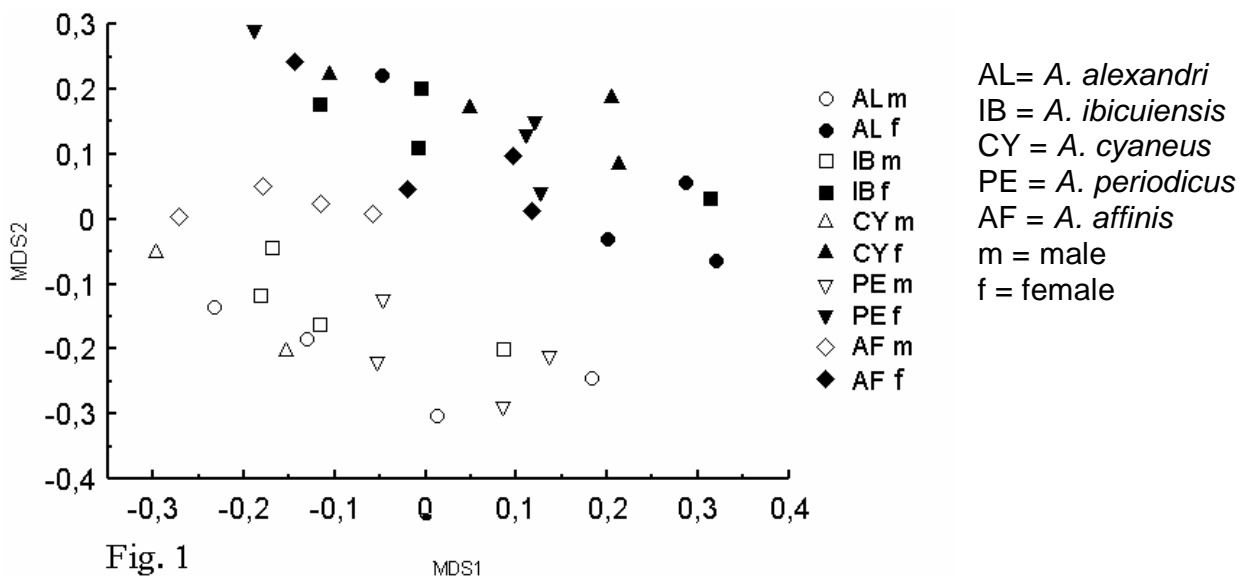
```

5 37 17.1 46.8 45.5 48.1 47.7 30.8 25.3 11.8 27.9 30.6 19.4 3.7 6.5 9.5
6 37.9 15.7 62.2 55.2 28.3 24.6 35.7 23.5 12 31.6 31.5 22.5 4.5 6.7 10.6
6 36.8 14.9 63.1 54.5 30.2 23.2 35.2 25.6 12.4 30.7 31.5 21.2 3.7 6.1 9.9
6 37.2 15.7 59.6 55.2 28.6 24.5 34.7 23.3 10.3 29.6 31.4 22.1 3.9 6.2 8.8
6 39.9 16.8 63 58.1 25.9 22.2 36.1 23.4 12 30.7 31.5 22.8 4.4 5.8 9.5
7 39.3 15.8 48.5 46.2 46.5 45.9 31.4 24.3 10.7 28.9 33.2 20.9 4.2 6.9 8.6
7 41.7 16.3 46.5 47 47.9 47.4 35.1 27.7 11.5 30.4 35.2 21.4 4.2 7 9.6
7 38.2 15.2 48.2 46.1 44.2 41.6 35.1 27.2 10.5 29.4 30.6 18.9 4.4 5.7 9.7
7 38.7 16.2 48.6 47.9 45.5 41 35.6 26.8 11.4 31.3 32.8 19.8 4.3 6.7 10.6
8 37.5 15.4 60.7 57 31.3 25.3 36.2 22.2 11.8 31.1 30.9 22.6 4 5.2 10.3
8 34.4 13.9 61.9 54.3 25.8 25 34.5 22.4 10.4 27.9 27.8 21.3 4.2 5.9 9.9
8 37.6 15.7 57.7 53.8 33.1 24.9 39 25 12.8 30 29.4 21.2 3.9 5.1 11
8 37.3 14.1 62.1 54.4 29.2 26.2 36 26 12.2 30.8 30.8 21.8 4.1 5.2 11.5
9 37.9 14.9 47.7 48.5 44.5 42.2 27.9 20.3 8.9 28.2 29.7 19.3 3.8 6.5 9.3
9 37.5 14.3 51.3 47.7 38.1 37 32.7 22.8 10.9 28.5 30.6 20.1 4.2 6.1 8.9
9 39.9 15.5 51 48.4 42.1 41.4 29.7 23.3 10.8 28.9 30.1 21.4 4.1 6.1 9.9
9 38.5 15.3 53.7 51.5 41.1 38.8 33.2 23.9 10.9 30 31.3 20.9 3.9 6.1 9.9
10 38.4 15.6 60.2 56.4 33 25.2 35.4 23 12.1 30.7 29.8 22 4.3 5.8 10.1
10 36.3 15.3 57.8 52.3 34.6 27.3 36.8 23.6 10.9 30.1 29.1 20.6 4.1 6.4 10.2
10 35.4 14.6 62.8 55.1 25.6 25.5 32.9 22.5 9.3 29.9 29.7 21 4.1 6 9.8
10 37.8 16.1 56 53.6 36.4 29 36.1 23.7 10.7 31.8 30.1 22.3 4.3 5.7 10.6
0 [this 'zero' is to check the data]
-.257268 .005946 [variables of the start configuration]
-.206993 -.032382
-.21978 -.005611
-.04828 -.044681
.148379 .011015
.176074 -.039153
.12446 .01064
.25852 -.016818
-.206727 .004198
-.082876 -.02892
-.196388 -.004471
-.139554 .009911
.04895 .030593
.253174 -.024408
.081309 .008435
.140123 .021604
-.244765 .024274
-.249421 -.003591
.201218 -.016073
.148362 .013046
.068187 .026898
.241396 -.01049
-.191611 -.004303
-.147214 -.039555
-.121678 -.006324
-.056502 -.03658
.183013 -.00022
.076881 .049793
.115005 -.003125
.173231 .001728
-.187725 .02704
-.089741 .021869
-.065404 .011387
-.036278 .006311
.132017 -.002275
.020042 .006836
.070641 .040115
.087222 -.012661
0 [this 'zero' is to check the data]
[end of data]

```

Results of test run:

1. To calculate reasonable start coordinates use the AutomaticRuns option (at least 10 runs with random start configuration). In 2D its easy to produce a consens configuration with the aid of tpsRelw–Program. Figure 1 plots the consens coordinates, which is calculated based on 20 final configurations of randomized starts. The consense coordinates should be included in the data file (see above). The display of coordinates can be used for group discrimination. In this example it is easily recognized that the group of males (open symbols) differ from the group of females (filled symbols). Within the males the *Austrolebias affinis* (AF m) separated from the remain groups of males.



2. Confidence limits can be calculated with the completed data file using the Jackknife option. The consens configuration (Fig. 2) is similar to this of Fig. 1. In Figure 3 the final coordinates of all data by Jackknife procedure are plotted. The cluster of *Austrolebias affinis* (AF m), which was recognized as separate in Fig. 1 and 2, is now braeked by others. However, the multidimensional morphometric differentiation of males and females (sexual dimorphism) is supported. In some species (e. g. *A. affines*) the multidimensional morphological sexual dimorphism is less strong than in others (e. g. *A. alexandri*).

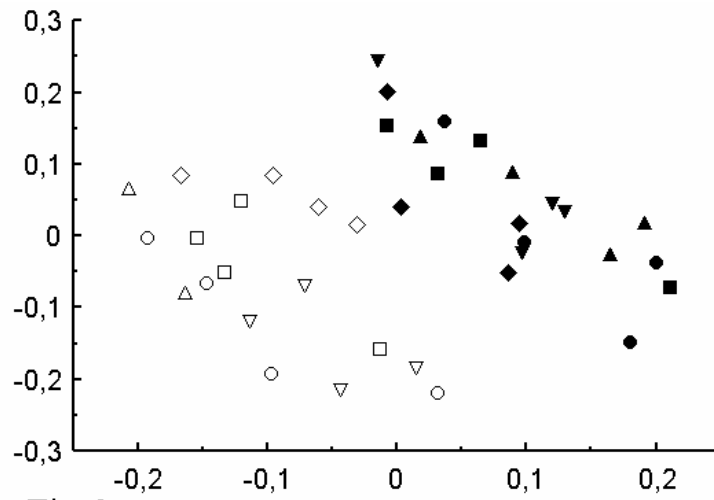


Fig. 2

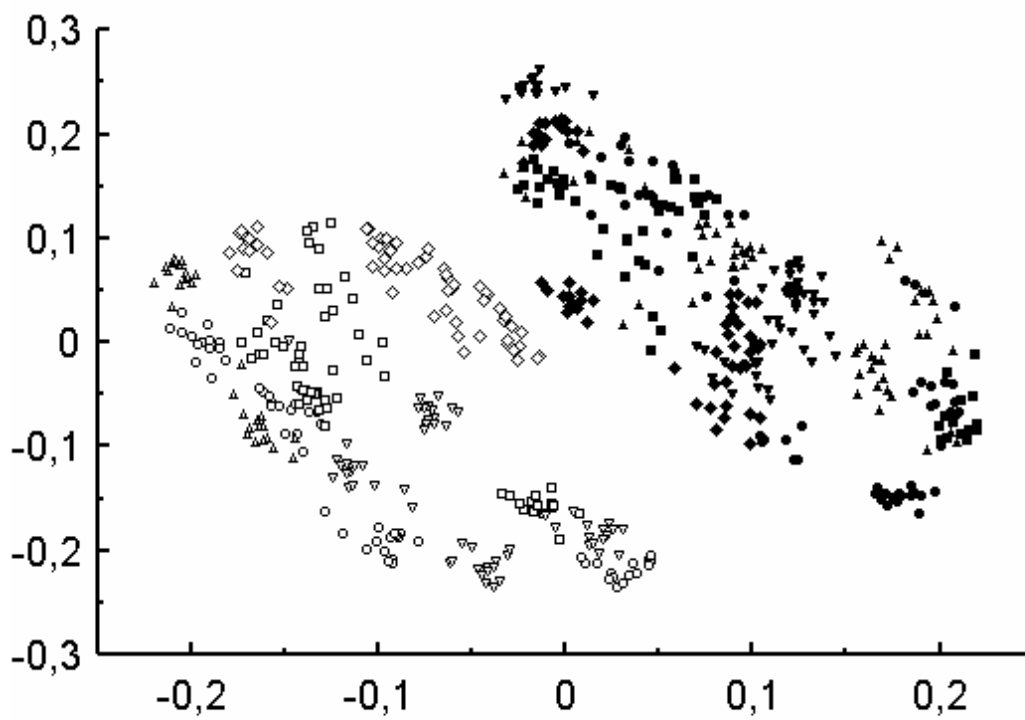


Fig. 3

Appendix 2: Notes

- It is strongly recommended that the user should start with a automatic random application to find an convinced consens initial configuration.
- Be aware that even extreme Datas can be ordinate close to ordinary clusters cause of the rank oriented NMDS. Exclude such OTUs before starting calculation. Use only OTUs with similar appearance.
- This program is mainly developed for small sample size (see limits).
- Use the Jackknife application only with 'UserData' start coordinates. Be sure that the start ordination is sufficient for your application.
- Watch the stress value. It should be less than 0.2 for a responsible final ordination. If not use 3D instead of 2D or 1D.
- If the start configuration is reasonable the results of Jackknife application can be used for character clustering. With the program tpsRelw you can display the RW1 against RW2 (see Fig. 4 for an example) to find characters with a similar or most opposite influence to the final ordination. The example below (Fig. 4) is the result by using the 'Test Run' (see Appendix 1) data.
- Use an own distance matrix (*.dis) if you like to calculate NMDS with an other than city-block-distance (the only supported distance metric by nNMDS.exe).
- Don't test the significant of differences (final coordinates) between groups by meaning of t-Test or by meaning of similar parametric statistic tests. Just observe the configuration with convex clusters.

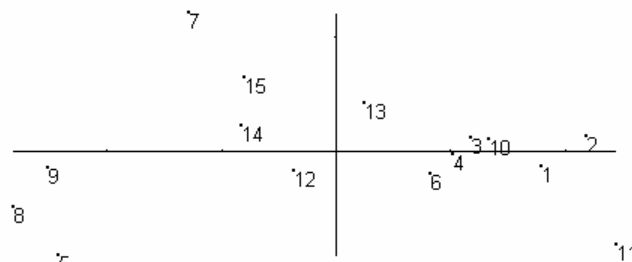


Fig. 4

I hope you find this 'guide' and the program helpful. Feel free to contact me with any suggestion or question. I would try to help to resolve it.
I apologize for my little English.

If you find bugs or mistakes in this program, please contact me immediately!