

bord-ER — COMPUTER PROGRAM FOR PERFORMING MOVING SPLIT-WINDOW ANALYSIS

L. Körmöczi and Cs. Varró

*Körmöczi, L. and Varró, Cs. (2015): **bord-ER** — computer program for performing moving split-window analysis. — Tiscia 40, 25-30*

Abstract. Vegetation ecologists and soils scientists frequently use moving split-window analysis for partitioning transect data, and reveal boundaries. Few existing softwares supported this research activity, but non of them was equipped with statistical significance test. A new software was developed to mitigate this shortcoming. We offer a user-friendly tool with five distance/dissimilarity functions and two types of random reference. This paper is a short description of the program.

Key words: boundary analysis, C-program, statistical test, user manual.

L. Körmöczi, Cs. Varró, Department of Ecology, University of Szeged, H-6726 Szeged, Közép fasor 52.

Introduction

Moving split-window (MSW) analysis is frequently used in soil and vegetation pattern analysis. It is very efficient in revealing discontinuities (boundaries) in one dimensional data sets. User friendly software, however, is not available that may compute the confidence limits for significance test. We intended to fill this gap, and provide an easy to use application for MSW

Softwares for MSW-computations

The computation of dissimilarity metrics in the MSW analysis is not a big challenge, it can be performed even with a spreadsheet software. This works well with a single window width, but dealing with multiple window widths and handling randomized data are rather difficult in such a way. The PASSaGE software also contains the computation of a dissimilarity profile (Rosenberg and Anderson, 2011). This software package is declared as “a free, integrated, easy-to-use software package for performing spatial analysis and statistics on biological and other data”. Though it may compute dissimilarities for several window sizes the statistical significance is missing. Computations can be done in the R environment. For example, the

software of Rossiter (2013) enables to carry out MSW analyses using multiple window widths and it offers two dissimilarity indices. However, it does not offer any randomizations.

We developed earlier our own software in the statistical language R (version 2.10.1, www.R-project.org). This is equipped with random references, and computes upper confidence limits but the runtime is rather long (Erdős et al. 2014b). Therefore the new version of the software has been written in C, and compiled with QT Creator.

Installation

The software is freely downloadable from <http://expbio.bio.u-szeged.hu/ecology/kormoczi/bordER/index.html>

The installation of **bord-ER** is easy: Just download the file 'bordER.exe' and put it anywhere on your harddisk. Double-clicking the file will start the program. Windows will consider this a breach of security, and will ask if you trust the software provider. If you want to use the program, you will have to answer yes.

The lack of “formal” Windows installation is intentional, and allows installation without administrator privileges.

bord-ER is a stand-alone program to perform moving-split window analysis with various dissimilarity indices. It gives confidence limits for statistical significance of deviation from random distribution.

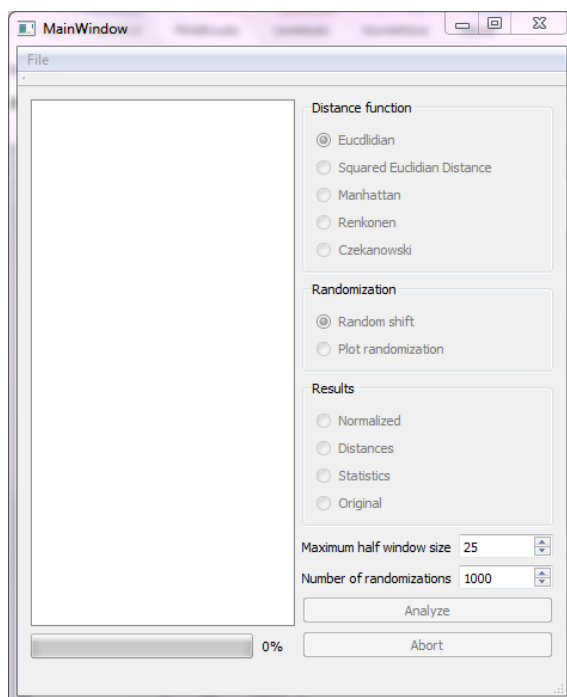
The procedure known as Moving Split-Window Analysis is designed to reveal discontinuities in one-dimensional sequential data set on the basis of dissimilarity values of neighbourhoods. Sequential data set may be any regular sequence of univariate or multivariate data vectors. In ecology it is often used on regular samples of transect data. Increasing window width results in a kind of spatial scaling. Significance test of this statistic is based on the random distribution of attribute values.

Input data structure

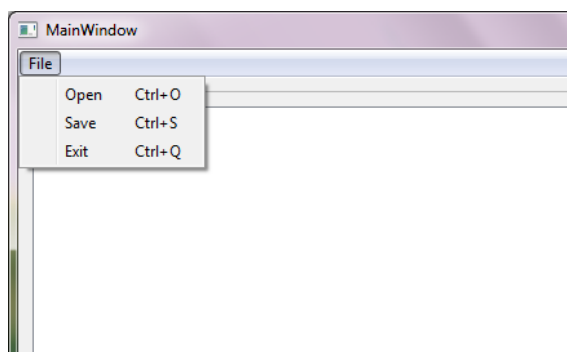
Input data must be arranged in an ASCII file (as that comes from EXCEL). Tab, comma, semicolon or space are accepted as field delimiter. Data should be either integer or real numbers. Only decimal point (.) are accepted as decimal separators. Sequential data vectors are arranged in rows: rows are the sequential sampling units, and columns are the variables. No row or column headers are used. Sample data structure can be found in `sample.txt` that is also available at the homepage of the software.

Performing the analysis

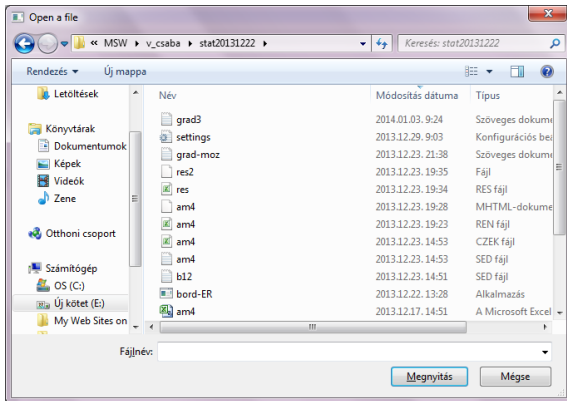
Double-clicking the program name opens the following window:



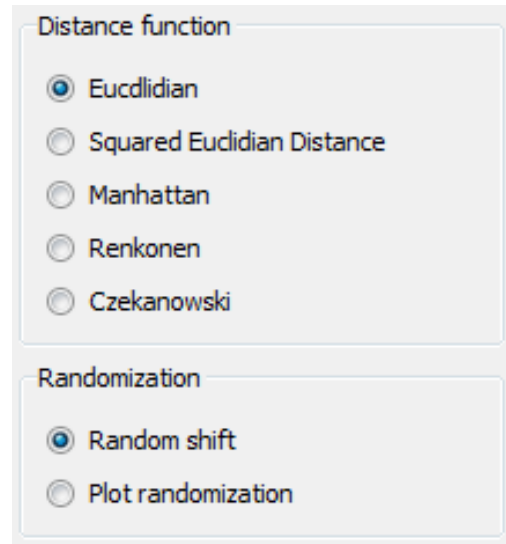
“File” dropdown list contains “Open”, “Save” and “Exit” commands.



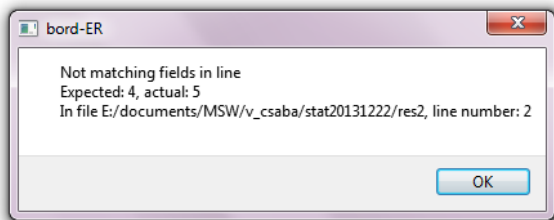
Clicking “Open” you can search for the desired file in the usual browser window. You can open only matching files,



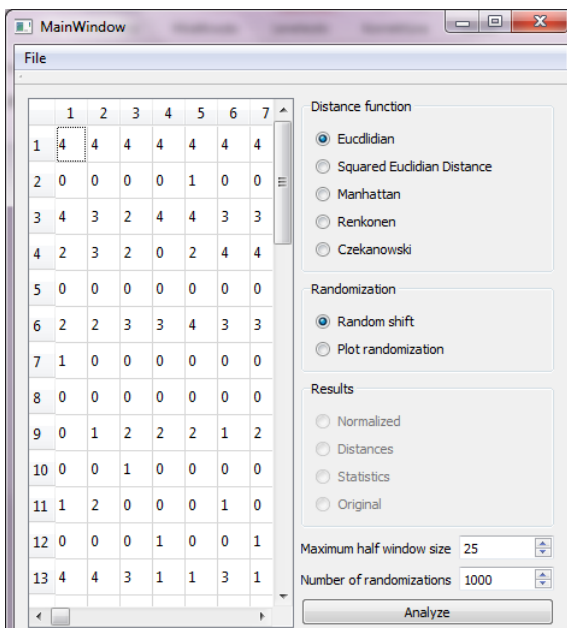
One of the five distance/dissimilarity functions can be chosen at the right panel. Two types of randomization is implemented in the program, but the use of Random shift is suggested.



otherwise an error message appears.



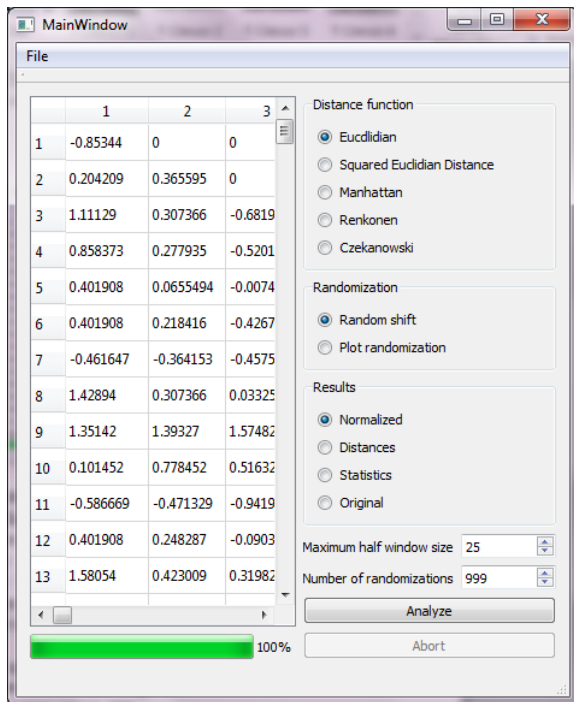
If an appropriate file is open, the raw data appear in the MainWindow in tabular form, and the following options can be chosen:



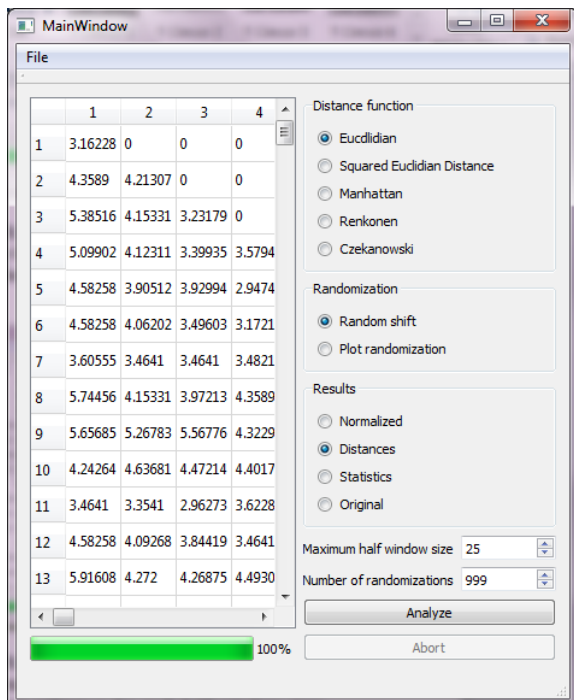
You may set the Maximum half-window size (by default: 25) and the number of Random iterations (by default: 999). The last set of these values is saved in settings.ini file, and is used in the next run of the program.

Clicking the Analyze button starts the analysis with actual settings; a process indicator shows the state of analysis. The resulted values can be checked in three windows:

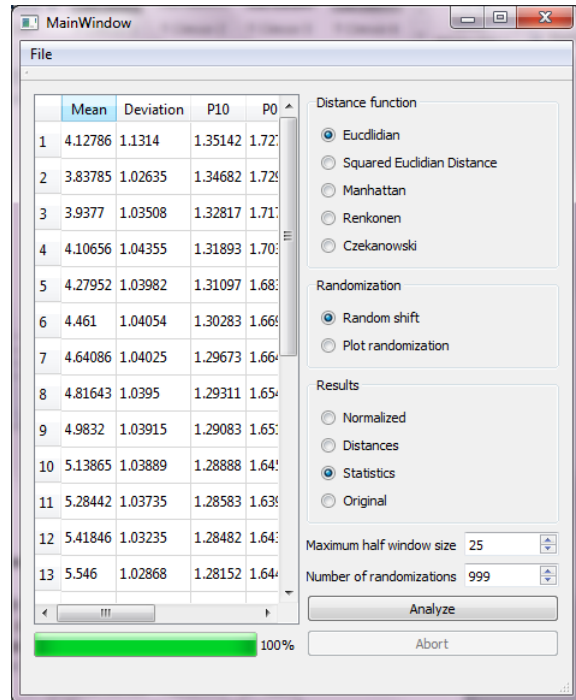
Results: Normalized option window shows the Z-scores of actual running for each half-window size. Those data are for drawing dissimilarity profiles.



Results: Distances option window shows the dissimilarity/distance values for each half-window sizes in the scale of the dissimilarity function applied.



Results: Statistics window displays the overall mean and standard deviation of dissimilarity values of the random reference for each half-window sizes. Also 10 %, 5 %, 2.5 % and 1 % confidence limits are listed.

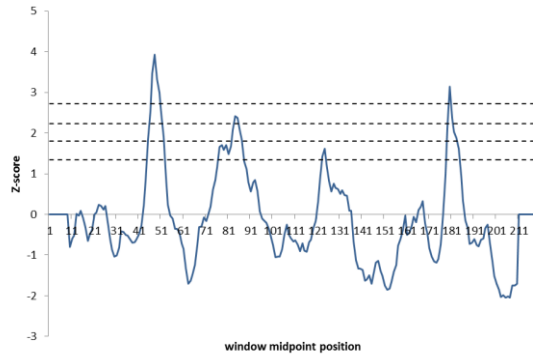


File: Save option saves all of the results in a tab delimited file of specified name. (Important! In the present state the software does not save the results in a directory the name of which contains special characters!) Statistics, original dissimilarity values and Z-scores are saved. Header of the data sequence contains the details of the analysis.

Moving split window analysis
 Maximum window size range: 25
 Randomization count: 999
 Row number:48
 Species number: 220
 Comparative function: Euclidean distance
 Random reference with: Random shift of species

Saved file should be opened in Excel or other spreadsheet application, and dissimilarity profiles should be drawn as appropriate. Dissimilarity profiles are drawn from Normalized (Z-score) data, either for each window size or an average Z-score over several window sizes. Averaging should be done in the spreadsheet application. Level of

significance can be drawn on the basis of confidence limits as is shown below.



Computation

On the basis of Cornelius and Reynolds (1991), the description of MSW-technique is the following (with some comments and improvement). Consider a data series of $j = 1, 2, 3, \dots, m$ ordered positions with $i = 1, 2, 3, \dots, n$ measurement variables at each position. Mostly belt transects consisting of contiguous or evenly placed quadrats are used in community composition measurements. x_{ij} is the abundance of the i th variable at the j th position along the series. Consider a window of width q that brackets sequential positions along the series. q is a non-zero, even integer expressed in the measure of the sampling units, $q < m$. A window of width q is split into two equal halves, designated by W_A and W_B . A series of length m will have $m-q+1$ sequential overlapping windows of width q , where the location of each window midpoint is $r + 0.5$, where $r = q/2, q/2 + 1, q/2 + 2, \dots, m-q/2$. For each window midpoint location each variable i is averaged in each window half. A dissimilarity/distance index can be calculated between each of the $m-q+1$ pairs of average vectors. Any dissimilarity or distance index can be used.

The following marks are used in the index descriptions. $D_{j,j+1,k}$ is the dissimilarity value at the $j+0.5$ -th window midpoint position, for half-window size k ; n is the number of variables; m is the length of data series; $j=k, k+1, \dots, m-k, k=1, 2, \dots, m/2$; k is the half-window width; $x_{i,j,k}$ is the abundance of i -th variable at the j -th position along the data series averaged for the half-window size k .

Euclidean distance:

$$EUC_{j,j+1,k} = \sqrt{\sum_{i=1}^n (x_{i,j,k} - x_{i,j+1,k})^2}$$

Squared Euclidean distance:

$$SED_{j,j+1,k} = \sum_{i=1}^n (x_{i,j,k} - x_{i,j+1,k})^2$$

Manhattan metric:

$$M_{j,j+1,k} = \sum_{i=1}^n |x_{i,j,k} - x_{i,j+1,k}|$$

Complement of Renkonen index:

$$DREN_{j,j+1,k} = 1 - \sum_{i=1}^n \min \left\{ \frac{x_{i,j,k}}{\sum_i x_{i,j,k}}, \frac{x_{i,j+1,k}}{\sum_i x_{i,j+1,k}} \right\}$$

Complement of Czekanowski index:

$$DCZ_{j,j+1,k} = 1 - \frac{2 \sum_{i=1}^n \min(x_{i,j,k}, x_{i,j+1,k})}{\sum_{i=1}^n (x_{i,j,k} + x_{i,j+1,k})}$$

Expected mean dissimilarity/distance and standard deviation (SD) can be calculated from dissimilarity/distance values that are revealed from n -times randomized data with the same procedure as that in the previous paragraph. Expected mean and standard deviation for a given window width are calculated as the mean and SD of all data. Any kind of randomization can be used but we tested complete randomization and two restricted randomizations: random repositioning the plots and random shift of species.

Overall mean dissimilarity and SD are used in Z-transformation according to Cornelius and Reynolds (1991):

$$Z_{j,j+1,k} = \frac{D_{j,j+1,k} - \overline{D_{exp,k}}}{SD_{exp,k}}$$

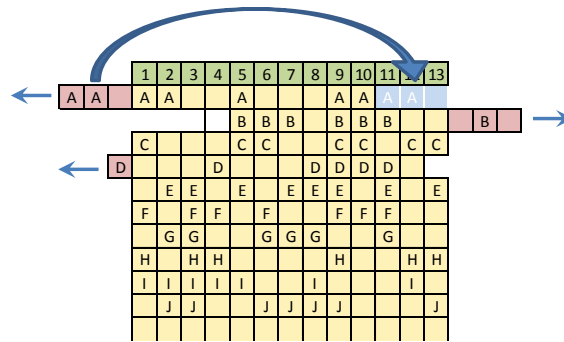
where $D_{exp,k}$ is the overall mean and $SD_{exp,k}$ is the standard deviation of dissimilarities from random reference for half-window width k .

Randomization

Reference distribution for statistical tests is given by permutation procedure. Two kinds of randomization are implemented in the programme: an unrestricted randomization – plot randomization, and a restricted one – random shift (c.f. Erdős et al, 2014a).

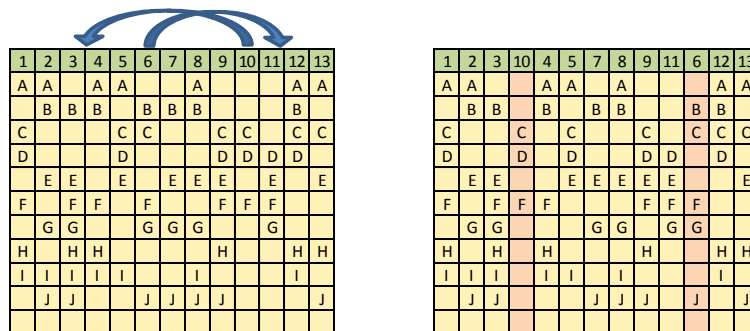
Random shift

During random shift, data belonging to the same variable are randomly shifted along the transect. Values shifted beyond the end of the transect are wrapped back on to the opposite end.



Plot randomization

Each plots is randomly repositioned along the transect.



References

- Cornelius, J.M., Reynolds, J.F. (1991): On determining the statistical significance of discontinuities with ordered ecological data. – *Ecology* 72: 2057-2070.
- Erdős, L., Bátor, Z., Tölgyesi, Cs., Körmöczi, L. (2014a): The Moving Split Window (MSW) analysis in vegetation science – an overview. - *Applied Ecology And Environmental Research* 12: 787-805.

- Erdős, L., Zalatnai, M., Bátor, Z., Körmöczi, L. (2014b): Transitions between community complexes: a case study analysing gradients through mountain ridges in South Hungary. – *Acta Botanica Croatica* 73: 63-77.
- Rosenberg, M.S., Anderson, C.D. (2011): PASSAGE: Pattern Analysis, Spatial Statistics and Geographic Exegesis. Version 2. – *Methods in Ecology and Evolution* 2: 229-232.
- Rossiter, D. G. (2013): Technical note: Optimal partitioning of soil transects with R (http://www.css.cornell.edu/faculty/dgr2/teach/R/R_OptPart.pdf)