Ecography

E6718

Jurasinski, G., Jentsch, A., Retzer, V. and Beierkuhnlein, C. 2011. Detecting spatial patterns in species composition with multiple plot similarity coefficients and singularity measures. – Ecography 34: xxx–xxx.

Supplementary material

Appendix 1

Artificial data set creation and evaluation

Creation of data sets with artificial gradient

The performance of the proposed multiple plot similarity coefficient has been tested with its application to artificial data sets. These have been created with the functions ads() and ads.hot() which are part of the package simba for R (download available at CRAN: <http://cran.R-project.org/>). In the following the function and its progress of calculation are explained in detail.

First, an empty species matrix with a specified number of plots and species is constructed. These parameters are either derived from a real species matrix recorded in the field or can be specified as required. Then a gradient vector is applied to the empty matrix in such a way that the probability of occurrence along the gradient increases from 0 to 1 for half of the species and decreases for the other half.

This probability gradient matrix is randomly transferred into a binary gradient matrix, which represents a random presence–absence manifestation of virtual species according to the probabilities laid down in the probability gradient matrix. Thus, a plot with a 0.4 probability of occurrence of a certain species on average will receive four times out of ten a '1' and six times a '0'. The binary gradient matrix marks the positions at which species can potentially occur in order to form the defined gradient. About 50% of all possible occurrences are now marked with '1' as potential occurrences of species and such describe the artificial gradient.

In the next step, a 'realistic' species frequency distribution is applied to the binary gradient matrix. The shape of the species frequency distribution is controlled by the parameter cf (eq. A1, with S_i = ith species and S_0 = frequency of ith species). It is set to a default value of 0.2, resembling a power-law distribution (see Fig. A1.1 for representations of the curve with 160 species and cf = 0.2 compared to the natural curve of the Abisko data). Now presence of each species is randomly sampled from its probable occurrences manifested in the binary gradient matrix according to its 'natural' occurrence in the data set (defined by its frequency).

$$S_0 = \frac{1}{S_i^{cf}} \quad \text{(A1)}$$

Additionally, ubiquitous species which occur on more than 50% of the plots are randomly re-arranged in the final species matrix by setting reord=TRUE in the function ads() and thus add an irregular element which does not follow the gradient.

This procedure of creating artificial data can either be used to create artificial data sets 'from scratch' or to re-arrange real data sets. It very closely mimics the frequency distribution of the species in the real set, total species richness and average species richness per plot. It does not preserve species co-occurrence patterns (compare Miklós and Podani 2004) because such strong constraints are not necessary for our application.

Evaluating the realized gradient

When the creation of the artificial gradient data set is successful, the first axis of an ordination (in this case we used CA) should represent the gradient quite well. We tested whether this is the case by regressing the CA 1st axis scores against the position of the plots on the gradient (as expressed by the gradient vector that has been used for constructing the gradient). There is a very good representation of the gradient position in the CA 1st axis scores (Fig. A1.2) so that we can take the CA 1st axis scores as a valid representation of the realized gradient. The considerable spread across the second axis proves that there is also a random element in the data. In the real data there is much more clumping. However, we still take the first axis scores as representative for the main gradient in species composition to evaluate the results of the multiple plot similarity calculations.

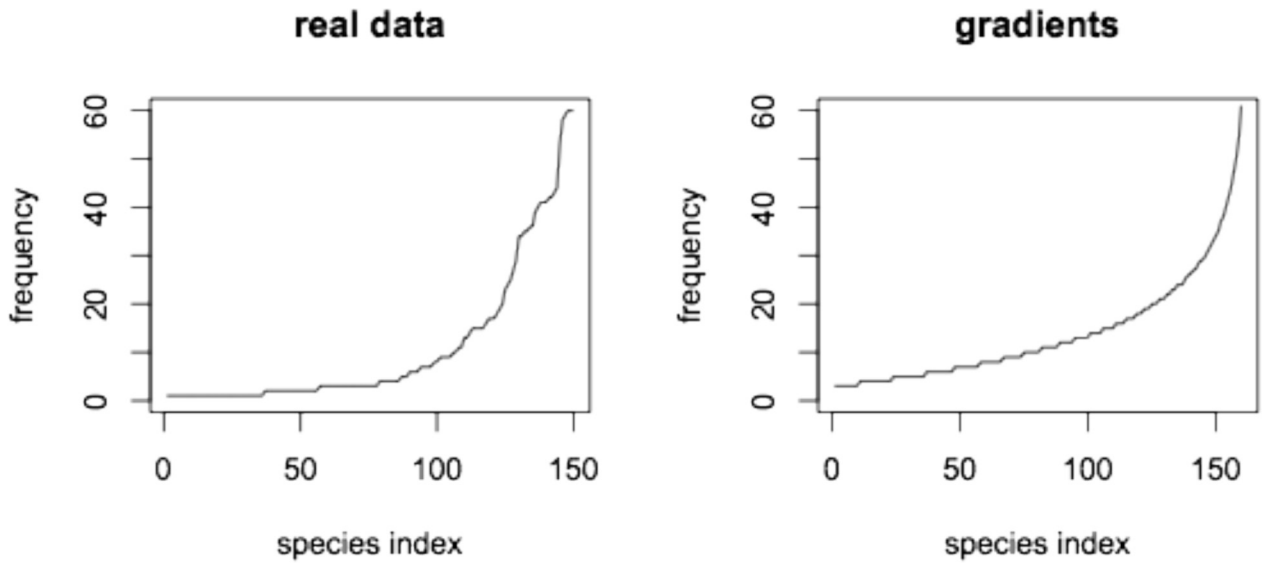

Figure A1.1. Species frequency distributions in the real and all artificial data sets.

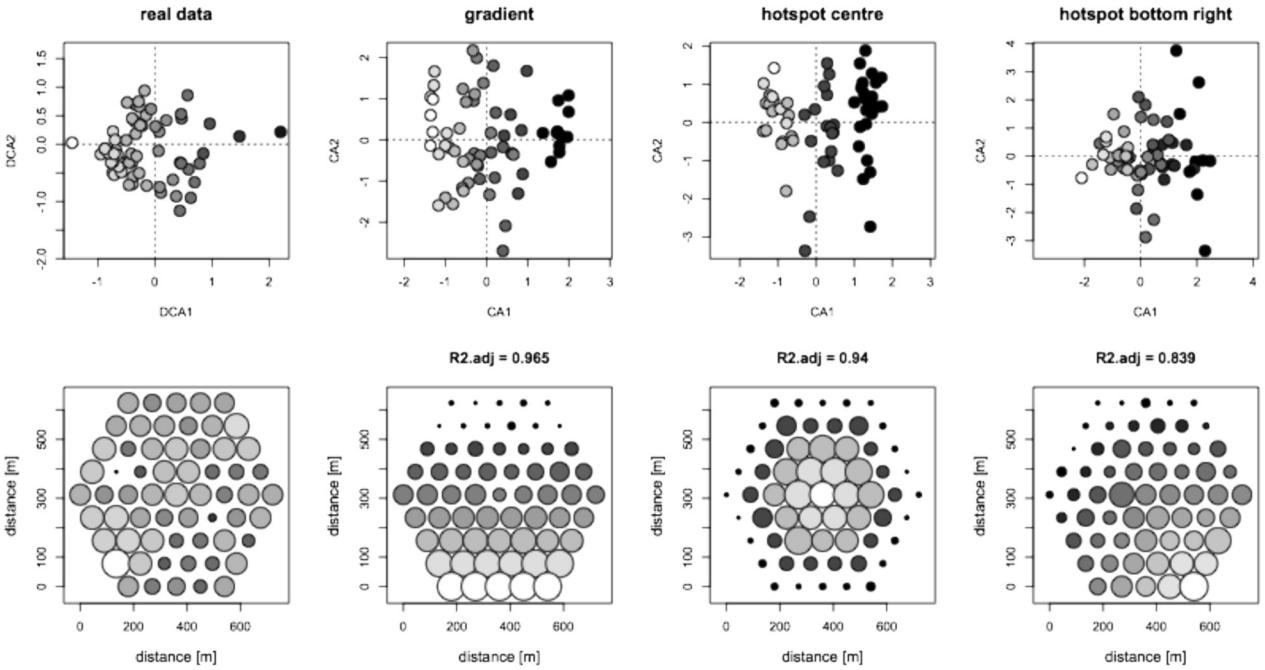

Figure A1.2. The realized gradients. Top: ordination plots of the real (left, DCA because that makes the gradient more apparent) and the artificial data sets (CA). Bottom: maps of the sampling grid including a representation of the applied (greyscale) and the realized position of the plots on the gradient (size of the circles). The R^2_{adj} values for the model fit (realized gradient position (CA 1st axis scores) explained by the applied gradient position from the gradient vector) are given and show, that the gradient representation is very good.

Appendix 2

The real data set from Abisko/Sweden

The data set is derived from a case study in a Tundra ecosystem near Abisko in northern Sweden (Fig. A2.1). Precipitation in the Abisko valley is comparatively low (304 mm yr^{-1}), mean annual temperature is ~0.5°C. The dominant vegetation consists of a mosaic of dwarf shrubs and small fragments of birch forests, mixed with small heath areas. Vegetation data was recorded in the summer of 2004 as abundance data on 61 plots of 5 m radius arranged in an equidistant hexagonal grid (Fig. A2.1, and Jurasinski and Beierkuhnlein (2006) for details on hexagonal sampling grids). The plot centers are 90 m apart. Size and distance to neighbors was determined by a pre-study. For the purpose of testing the new coefficient, the species abundances are transformed to presence/absence data.

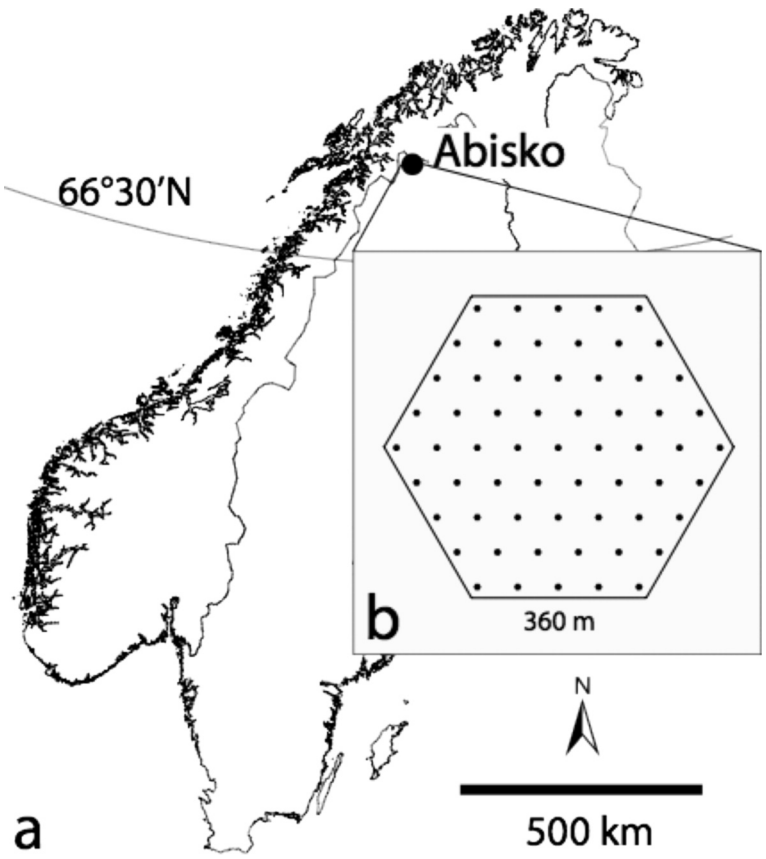

Figure A2.1. (a) Location of Abisko in northern Sweden. (b) Sampling grid established in the summer of 2004. Dots represent plots which were sampled for plant species (incl. mosses). The grid consists of 61 plots with 5 m radius. The plot centers share a distance of 90 m. Size and distance to neighbors was determined during a pre-study in the field.

Appendix 3

The patterns of all tested coefficients

Figure A3.1. Similarity patterns resulting from the application of the tested indices to the five data sets. The first row displays the underlying gradients[1]. The random data set shows no considerable gradient and is therefore not displayed. Subsequently, each row represents the patterns for one multiple plot similarity (solid boxes) or singularity coefficient (dashed boxes)[2]. Bigger dots always indicate higher values of the respective coefficient. Please note, that each subfigure is individually scaled between the minimum and maximum values of the coefficient (these are given in the bottom right corner of each subfigure). The correlation with the main gradient is given in the bottom left corner of each box. The sequence of the indices from top to bottom follows their overall ranking with respect to the performance in detecting the various patterns (see Table 2 caption for details on the ranking). Gray dots display significant values (tested with a permuation procedure against random expectations). The '+'-signs indicate mps or mos values that are higher than it can be expected from random whereas the '–'-signs depict values that are lower than can be expected from random.

[1]) The size of the dots represents the position on a gradient of species composition where the largest size represents one distinct species composition and the smallest represents another distinct species composition (However, there is some random element (Appendix 1). The sizes in between represent the gradual change in species composition. The greyscale depicts the artificial gradient applied (in case of the real data it displays the realized gradient).

[2]) All indices are calculated with a moving window approach. If the distance to a neighbor in the equidistant grid is x, a neighborhood consists of a plot and all other plots up to a distance x apart from that plot. Therefore plots on the borders have less neighbors.

Download <E6718-FigureA3.1.pdf>.