**Turnover**

- **a FORTRAN program**
- **for analysis of species associations**

**Version 1.1**

Werner Ulrich

Nicolaus Copernicus University in Toruń

Department of Animal Ecology

Gagarina 9, 87-100 Toruń; Poland

e-mail: ulrichw @ umk.pl


---

**1. Introduction**

The study of associations among species and sites has long been in the focus of ecology. Most previous work centred on patterns of negative co-occurrence and nestedness (inspired by the competition paradigm) and developed a plethora of metrics to assess species segregation, aggregation and the existence of nested subset patterns (reviewed in Gotelli 2000, 2001, Ulrich and Gotelli 2007a, b, Ulrich et al. 2009). Along with the interest in species associations the question of statistical testing has been an issue. This raises the question of what is the appropriate null hypothesis to test and observed pattern against (Gotelli 2000, 2001, Ulrich et al. 2009). Till now no consensus has been reached and many different null hypotheses are used often without clear justification.

Recently species spatial and temporal turnover came into focus (Leibold and Mikkelson 2002, Baselga 2007, 2010, Presley et al. 2010, Ulrich and Gotell 2010). These papers dealt with the problem whether and how it is possible to disentangle patterns species segregation into turnover, competition and richness components. Unfortunately previous papers suffer from a too narrow view on species association and appropriate metrics (Ulrich and Gotelli 2010).

The raw data for species association analysis are typically in the form of a binary presence-absence matrix where rows are species and columns are sites or samples, and the entries indicate the absence (0) or presence (1) of a species. In interaction or foodweb matrices, both rows and columns may represent species, and the entries represent the absence (0) or presence (1) of an interaction link. An abundance matrix contains rows as species, columns as sites or samples, and entries representing the population size (which may be 0) of a particular species in a particular site. Presence-absence matrices can always be constructed from abundance matrices, but not vice-versa. The *C.txt* file on the next page is a typical of an abundance matrix. The task is to find patterns of species and/or site associations within such matrices.

Table 1 presents an admittedly subjective view the most often used metrics to assess aggregation and segregation together with some limitations with respect to data types and appropriate null models. Note that methods that are based on total counts of
joint species co-occurrences and/or joint absences are generally incompatible with null models that retain marginal totals. Note also that many metrics depend in some way on the ordering of matrix rows and/or columns. Only the C-score, the abundance C-score (CA), the abundance aggregation score (AA), and the clumping score are independent of matrix sorting and simultaneously assess patterns in rows and columns.

2. Metrics

Squared Nearest Neighbour

As a first measure of aggregation of occurrences within the matrix Turnover uses one of the members of the Clark and Evans (1951) family of nearest neighbour (NND) measures. It is calculated as the normalized sum of all squared nearest neighbour occurrences in the matrix.

$$\text{NND} = \sum_{i=1}^{\text{species}} \sum_{j=1}^{\text{sites}} \frac{d_{ij}^2}{\text{fill}}$$

where fill denotes matrix fill and $d_{ij}$ is the minimum distance to nearest occurrence for any non-empty cell in the matrix. It is therefore a metric for presence absence matrices. Small values of NND denote an aggregated pattern. The metric depends on the ordering of species and sites.

Block quadrate variance

The Block Quadrate Variance (Hill 1973) is the average variance between the sums of entries of all adjacent 2x2 quadratic blocks in the matrix.

$$\text{Block} = \frac{1}{8L} \sum_{i=1}^{\text{species}-3} \sum_{j=1}^{\text{sites}-3} \left( x_{ij} - x_{i+1,j+1} \right)^2 + \left( x_{i+1,j} - x_{i+1,j+1} \right)^2 + \left( x_{i,j+1} - x_{i+1,j+1} \right)^2$$

where $x_{ij}$ denotes the sum of entries of the 2x2 submatrix $(x_{ij}, x_{i+1,j}, x_{i,j+1}, x_{i+1,j+1})$ and $L=3(\text{species}-3)(\text{columns}-3)$ is the total number of summands. Turnover uses the square root of block to avoid too large values in very large matrices. $\text{Block}$ is then compared with the respective average value of the null model distribution. The metric works with presence—absence and abundance matrices. High values of $\text{Block}$ denote aggregation. The metric depends on the ordering of species and sites.

Join-count statistics

The join-count statistics is similar to the nearest neighbour method and counts for all non-empty cells in the matrix all occurrences within a 3x3 quadratic around the focal non-empty cell (not counting the focal cell).

$$\text{JoinOcc} = \frac{1}{\sum_{i=1}^{\text{species}} \sum_{j=1}^{\text{sites}} x_{ij}^2} \sum_{i=2}^{\text{species}-1} \sum_{j=2}^{\text{sites}-1} \sum_{l=1}^{i-1} \sum_{m=1}^{j-1} x_{il} x_{lj}$$

where $x_{ij}$ and $x_{ijl}$ denote abundances (presence—absences) in cells $(i,j)$ and $(i1,j1)$ and the denominator is the total abundance in the matrix. The metric is applicable to abundance and presence—absence matrix-
The C-score is a normalized count \( NCS \) of all 2x2 submatrices of the form \( \{1,0\}, \{0,1\} \). It is originally calculated from (Stone and Roberts 1990).

### Table 1: Important metrics for assessing species segregation aggregation. The metrics are described section 2.

PA: presence—absence data; A: Abundance data. Fixed-fixed: Null models that retain marginal counts of presences. Whole matrix: metrics that assess patterns for the whole matrix (their values wouldn’t change after matrix transpose). Species: Metrics that compare patterns for species (matrix transpose give the metric for site comparisons only). \( NODF, BR \) and \( T \) measure a special case of aggregation, a nested subset pattern and are described in detail in Ulrich et al. (2009). \( Turnover \) does not include these metrics.

<table>
<thead>
<tr>
<th></th>
<th>PA null models</th>
<th>A null models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole matrix</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Independent of matrix sorting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C-score</td>
<td>PA</td>
<td>All</td>
</tr>
<tr>
<td>Clumping score</td>
<td>A</td>
<td>All</td>
</tr>
<tr>
<td>NestPairs</td>
<td>A and PA</td>
<td>All</td>
</tr>
<tr>
<td>Togetherness</td>
<td>A and PA</td>
<td>All</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Species only</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simpson dissimilarity</td>
<td>No fixed-fixed</td>
<td></td>
</tr>
<tr>
<td>Soerensen dissimilarity</td>
<td>No fixed-fixed</td>
<td></td>
</tr>
<tr>
<td>Morisita</td>
<td>All</td>
<td>All</td>
</tr>
<tr>
<td>Chao</td>
<td>All</td>
<td>All</td>
</tr>
<tr>
<td>Other joint occurrence/absence metrics</td>
<td>No fixed-fixed</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dependent of matrix sorting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whole matrix</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Segregated</td>
<td>PA</td>
<td>All</td>
</tr>
<tr>
<td>Aggregated</td>
<td>A and PA</td>
<td>All</td>
</tr>
<tr>
<td>Aggregated nested</td>
<td>A and PA</td>
<td>All</td>
</tr>
<tr>
<td>Other distance based metrics</td>
<td></td>
<td>All</td>
</tr>
<tr>
<td>NND</td>
<td>A</td>
<td>All</td>
</tr>
<tr>
<td>Block</td>
<td>A and PA</td>
<td>All</td>
</tr>
<tr>
<td>Join-coint</td>
<td>A and PA</td>
<td>All</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Species only</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Embedded absences</td>
<td>PA</td>
<td>All</td>
</tr>
<tr>
<td>r²</td>
<td>PA</td>
<td>All</td>
</tr>
<tr>
<td>C_{Turn}</td>
<td>PA</td>
<td>All</td>
</tr>
<tr>
<td>C_{Segr}</td>
<td>PA</td>
<td>All</td>
</tr>
<tr>
<td>Nodf_{T}/r²</td>
<td>PA</td>
<td>All</td>
</tr>
<tr>
<td>Morisita</td>
<td>BR_{T}</td>
<td>All</td>
</tr>
</tbody>
</table>
Turnover

The togetherness score (Tog) of Stone and Roberts (1992) is a complement of the C-score and defined from “2x2 submatrices of the form \[ \{1,0\}, \{1,0\} \] The metric Tog is a count of the total number of these submatrices. It is standardized with regard to matrix size by

\[
Tog = \frac{4 \sum_{i,j} (1 \ldots 0)}{\text{Species}(\text{Species} - 1)\text{Sites}(\text{Sites} - 1)}
\]

Tog can range from 0.0 to 1.0, with high values of Tog indicating positive species association.

Morisita index

The Morisita index is calculated as proposed by Hoagland and Collins (1997):

\[
\text{Morisita} = \sum_{i=1}^{\text{Sites}} \frac{n_i (n_i - 1)}{N (N - 1)}
\]

where \( n_i \) denotes the number of starting and stopping boundaries at site \( i \). \( N \) is the total number of starting and stopping boundaries and equals therefore two times the number of species. High index values denote clustered boundaries of species occurrence.

Extended Morisita metric

Chao et al. (2008) extended the Morisita index of similarity for two communities to a matrix-wide metric for \( n \) communities of the form

\[
\text{Extended Morisita metric}
\]
\[
Chao = \frac{\sum_{i=1}^{\text{Species}} \left( \frac{\sum_{j=1}^{\text{Sites}} p_{ij}}{\text{Sites}} \right)^2 - \sum_{j=1}^{\text{Sites}} (p_{ij})^2}{(\text{Sites} - 1) \sum_{j=1}^{\text{Sites}} (p_{ij})^2}
\]

where \( p_{ij} \) is the relative abundance of species \( i \) in site \( j \). Low values of \( Chao \) indicate dissimilarity of species relative abundance distribution among sites, which can be interpreted as a measure of negative covariance in relative abundances. The metric is independent of the ordering of the matrix.

**Spearman correlation of cell positions**

In a matrix ordinate by reciprocal averaging, the rank correlation between row and column positions of all non-empty cells is a commonly used metric how well the non-empty cells are placed along the matrix diagonal. *Turnover* uses the respective coefficient of determination \( r^2 \) as test statistic and compares it with \( r^2 \) of the null distribution. High coefficients of determination indicate species turnover.

**Embedded absences**

Preseley et al (2010) proposed a count of the number of embedded absences in the ordinate matrix as a metric of coherence of species range sizes (Leibold and Mikkelson 2002). In Fig. 1 these are all zero cells in each row that are on the left and on the right side embedded by non-empty cells. In Fig. 1 there are 17 such embedded absences. Low numbers of embedded absences with respect to the null distribution indicate matrix wide coherent range sizes.

**Clumped occurrences**

A simple metric for aggregation that is independent of matrix order is a count of the total number of 2x2 submatrices of the form \({\{1,1\},\{1,1\}}\). The metric is therefore the aggregation equivalent of the C-score. As in the case of the C-score *Turnover* calculates Clumping in the normalized form divided by the total number of 2x2 submatrices [species(species-1) sites(sites-1)/4].

**TS index**

*Turnover* calculates also the metric TS as the quotient of \( C_{\text{ant}}/(C-\text{score}-C_{\text{turn}}) \) as an additional metric of co-occurrence. This metric tries to separate checkerboard segregation measured by \( (C-\text{score}-C_{\text{turn}}) \) from vicariant segregation as measured by \( C_{\text{turn}} \). Low values of TS with respect to the null distribution indicate a tendency towards checkerboard segregation.

**NODF**

As a measure of the degree of nestedness, *Turnover* calculates the NODF metric (Almeida-Neto et al. 2008). To compare the metric with \( R^2 \) *Turnover* rescales the original NODF into the range of 0 (perfect antinestedness) to 1 (perfect nestedness) by \( \text{NODF} = \text{NODF}/100 \).

**\( \text{NODF}_T \)**

After seriation a nested matrix concentrates occurrences in the lower left or upper right corner of the matrix in a similar way as after sorting according to marginal totals. Hence we can use NODF as before but using the matrix turned around by 45°. The new metric \( \text{NODF}_T \) again ranges from 0 (perfect antinestedness) to 1 (perfect nestedness).

**NR\(^2\) index**

The NR\(^2\) index tries to catch the gradient between a nested and a turnover segregated matrix and is calculated from the quotient \( NR^2 = \text{NODF}_T / r^2 \). High values of \( NR^2 \) indicate a tendency for nestedness.

**BR and BR\(^T\)**

The discrepancy metric BR (Brualdi and Sanderson 1999) calculates the minimum number of occurrences that have to be shifted in rows (to the left) or columns (upwards) to get a perfectly nested matrix. *Turnover* uses BR in its standardized form \( BR = \text{total count of shifts} / \text{matrix fill} \). Low values of BR indicate nestedness. As for NODF \( BR_T \) is the BR metric for seriated matrices.

**3. Program description**

The present FORTRAN 95 software *Turnover*
calculates the abovementioned metrics for abundance data as well as for presence-absence data. The program accepts standard space delimited text file matrices (cf. the above example file), with sites in columns and species in rows (the Ecosim format, Gotelli and Entsminger 2006). Species and site names must not contain spaces. Multiple analyses using many matrices are possible, and need an additional text file containing the file names as shown below. The first line of this file has to be a comment line.

Statistical inference is done from a null model approach (Gotelli 2000, 2001, Ulrich and Gotelli 2007a, b). The user can choose between five null models commonly used with presence-absence matrices (the fixed row – fixed column constraint model, the equiprobable row – fixed column constraint model, the fixed row – equiprobable column constraint model, and the equiprobable row/column null model, cf. Gotelli 2000 and Ulrich et al. 2009 for detailed descriptions). The fifth null model resamples the matrix that is proportional to the observed row and column marginal totals until the observed number of incidences is achieved (Ulrich and Gotelli 2012). In the case of the fixed row – fixed column constraint model randomization is done with the independent swap algorithm (Gotelli 2000) using 10nm swaps for each random matrix, where n and m are the numbers of rows and columns, respectively. This number of swaps ex-
Turnover exceeds even for moderately sized matrices (80 species, 40 sites) the number of 30,000 swaps recommended by Hausdorf and Hennig (2007) for large matrices.

For the study of matrices that contain abundance data, we use four recently developed randomization algorithms (Ulrich and Gotelli, Ecology in press). These models start with a cleared matrix and place individuals into the matrix cells using three different constraints:

- **rc** assigns individuals to matrix cells proportional to observed row and column abundance totals until, for each row and column, total abundances are reached.
- **aa** resamples the matrix proportional to observed row and column abundance totals until the matrix-wide number of individuals is reached. In a few cases, this algorithm generates matrices with empty rows (species) or columns (sites).
- **is** resamples the matrix proportional to observed row and column abundance totals until, for each column and each row, total numbers of incidences are reached.
- **ss** resamples the matrix proportional to observed row and column abundance totals until the matrix-wide number of species is reached. In a few cases, this algorithm generates matrices with empty columns (sites).

Observed metrics are compared to the distribution of the scores obtained from the null model matrices (the number of matrices is user-defined, with a default of 100). Statistical inference is done from the lower and upper 95% confidence limits.

The program output contains summary data about the matrix (numbers of species, sites, matrix fill and numbers of occurrences). For each metric observed and expected raw scores are given, as well as the standard deviations of the expected scores and the $Z$-transformed scores ($Z = (\text{observed score} - \text{expected score})/\text{StdDev}_{\text{exp}}$). Also given are the skewness and the upper and lower 95% confidence limits of the null distribution.

**4. Program run**

*Turnover* first asks for the input file. Give it with extension (example: file.txt). In the case of multiple runs a carriage return results in the question for the name of the file that contains the matrix file names for multiple analysis (cf. the example above). All of the
files have to be in the same directory.

Next the program asks for the names of the output and matrix file names. Both are shown in the next two Figures. Carriage returns assign the default names Output.txt and Matrix.txt.

The program proposes five presence absence and three abundance based null models as described in Gotelli (2000), Ulrich and Gotelli (2007a) and above. The abundance based null models can also be used with the presence—absence metrics.

The default number of randomized matrices is 1000 but should be higher particularly at small matrix size. Note that at large matrix size high numbers of null matrices might cause extraordinary long computing times. In the above example of matrix of size 47x17 took approximately 3 min to be processed on a standard Pentium Dual Core. However at larger matrix size null distributions tend to become normally distributed and fewer null matrices are needed to get sufficiently precise confidence levels.

5. Output

Turnover returns two output files. The first, Matrix.txt, contains the initial sorted (if chosen) matrix with column and row occurrence and abundance totals together with some basic information about the matrix. It also prints the last randomized matrix.

The second file, Output.txt, gives species richness (numbers of rows), number of sites (column numbers), total numbers of non-empty cells (occurrences), and matrix fill. The output contains the observed value, the expected value obtained from the null model, the respective standard deviation and the Z-transform:

\[ Z = \frac{Obs - Exp}{StDev_{Exp}} \]

PZ(H0) denotes the one-sided probability level of the null hypothesis of no difference between observed and expected metric value based on the assumption of a normally distributed Z. Lastly the program gives the skewness and the lower (L95%CL) and upper 95% confidence limits (U95%CL) of the null model distribution. P(H0) denotes the one-sided probability level of the null hypotheses as estimated from the null distribution. At larger matrix size null distributions are approximately normally distributed and PZ (H0) and P(H0) are similar. At small matrix size and too few numbers of null matrices PZ(H0) and P(H0) might differ. The minimal probability level of P(H0) is given by the inverse of the number of null matrices.

6. Citing Turnover

Turnover is freeware but nevertheless if you use Turnover in scientific work you should cite Turnover as follows:

7. System requirements

*Turnover* is written in FORTRAN 95, has been compiled under a 64 bit architecture, and runs under Windows 7, XP, and Vista. The present version is only limited by the computer’s memory.

8. Acknowledgements

Development of this program was supported by grants from the Polish Science Committee (KBN, 3 P04F 034 22, KBN 2 P04F 039 29).

10. References


