ECo: A new measure evaluating the degree of consistency between environmental factors and spatial arrangement of species assemblages

Giovanni Strona [a,*], Simone Fattorini [b], Simone Montano [c], Davide Seveso [c], Paolo Galli [c], Jesus San-Miguel-Ayanz [a]
a European Commission, Joint Research Centre, Institute for Environment and Sustainability, Via E. Fermi 2749, 21027 Ispra (VA), Italy
b Azorean Biodiversity Group, Departamento de Ciências Agrárias, Universidade dos Açores, Angra do Heroísmo, Terceira, Azores, Portugal
c Department of Biotechnology and Biosciences, University of Milan-Bicocca, 20126 Milan, Italy

* Corresponding author. Tel.: +39 0332 783047.
E-mail address: giovanni.strona@jrc.ec.europa.eu (G. Strona).

Keywords: Biodiversity, Environmental layers, European trees, Fish, Species-area matrix, Terrestrial vertebrates

Abstract

We introduce a measure of Environmental Consistency (ECo), which assesses the probability of reducing homogeneity in the environmental factors within a species' distribution by randomly displacing its occurrences. ECo is computed by applying null model analysis to a species incidence matrix where each locality is associated with a set of environmental values. Environmental homogeneity is measured, for each species, as the average multiparametric distance between any pair of localities where the species occurs. ECo can account for the effect of species interactions and resource availability by using different null models that permit or forbid occurrence displacements altering species local abundance or species prevalence. ECo provides researchers with a flexible statistical framework to address a wide range of ecological and biogeographical issues. We investigated in depth the properties and the potentialities of ECo, showing how it integrates the concepts of Eltonian and Grinnelian niches. We demonstrate that a close relationship exists between niche breadth at species level and environmental consistency of species assemblages. In addition, we provide evidence that ecological consistency is closely related to species range. A software to compute ECo is freely available at http://forest.jrc.ec.europa.eu/download/software/eco.

1. Introduction

Investigating the causes of species distribution is a primary goal in ecology, because it has many applications in both theoretical and applicative fields (Cox and Moore, 2004). Current distributions are determined by the interaction among historical (paleogeographical and paleoecological) causes, inter-specific relationships, and species' environmental needs (Woodward, 1987; Huntley, 1999).

The influence of history on species biogeography can be investigated by searching for non-random patterns in species per site matrices. Nestedness and modularity are typical examples of patterns originating from historical processes (Fortuna et al., 2010). Similarly, significantly high or low species co-occurrence values in species per site matrices may indicate, respectively, positive or negative inter-specific interactions (Stone and Roberts, 1990; Bertness and Callaway, 1994; Gotelli, 2000; Hausdorf and Hennig, 2007).

When environmental variables can be measured at sites of a species' occurrence, Ecological Niche Modelling (ENM) can be used to estimate dimensions of the species' fundamental ecological niche. Then, these estimates can be used to reconstruct species distribution by area suitability (Soberón and Peterson, 2005; Guo and Liu, 2010). The two approaches, i.e. the examination of species per site matrices and ENM are somehow complementary. Yet, the conceptual integration of species/locality analyses and ENM is not straightforward.

Here we describe a new ecological measure that aims at this purpose by providing information about the influence of environmental variables on the spatial arrangement of species assemblages. We called this index ECo, which is the acronym for 'Environmental Consistency', as it measures, in practice, how consistent are the environmental/climatic characteristics of the localities where each species of a species assemblage occurs. This is achieved by combining the valuable information provided by species/locality matrices with environmental/climatic data. Using different null models, ECo makes it possible to discriminate between the effect of environmental factors, and the constraints related to the spatial structure of species distribution (i.e. local species richness, and species prevalence).

ECo can find applications in various ecological fields. For example, as we will show in our case studies, it makes it possible to investigate general ecological patterns such as the potential relationships between ecological specialization and species range size, or how much individual species' niche breadth affect the overall environmental consistency of a species assemblage. Thanks to this latter ability, ECo might be also useful to investigate the relative importance of environmental filtering and species interactions in determining community assembly rules. Moreover, the degree each species contributes to the overall environmental consistency of an assemblage constitutes also a robust measure of how much environmental/climatic variables affect the distribution of that species, and thus can be used to orient and/or justify the use of ecological niche modelling approaches.

ECo may also be useful in conservation biology. For instance, assessing how much environmental/climatic characteristics affect the spatial structure a species assemblage can help identifying species assemblages and localities most susceptible to habitat degradation and climate change.

2. Description of the ECo algorithm

The basic functioning of the ECo algorithm is described in Fig. 1. Species occurrences are first used to compile a matrix of presence/absence of species (rows) per areas (columns). A set of desired environmental (for example, values of elevation, soil pH, salinity, etc.) and/or climatic variables (for example air temperature, precipitation, etc.) is then associated with each area. Each species occurrence (presence in a row) is therefore linked to the values of each variable in the areas (column) in which the species is found. These variables are used to compute distances (for example Euclidean) between any possible pair of occurrence of a given species. Finally, these pairwise values are used to compute an overall (average) distance among all the species occurrences in a given row. If all occurrences are placed in areas (columns) that have identical values for all measured environmental variables, the average distance between species occurrences will be 0, i.e. the ecological consistency for that species (row) will be maximum. Otherwise, the average distance will be higher than 0. Increasing average distances indicate decreasing consistency between the distribution of the considered species (row) and the variables associated with the areas (columns).

We may now imagine to randomly modify the matrix by adding or removing the presence of a species in a given area, i.e. by replacing a 0 in the matrix with a 1 or vice versa (in the following, we will refer to this kind of replacement simply as a "shift"). This may increase, decrease or leave unaltered the average distance computed between the occupied cells of the row (i.e. of the species) involved in the shift. In a matrix where occurrences are placed with a strong consistency with respect to environmental variables, a random shift will tend to increase the average distance. By contrast, in a matrix where the placement of occurrences in columns (i.e. of species in localities) is not or weakly in agreement with the corresponding environmental variables, a random shift would have the same probability of decreasing, increasing or leaving unaltered the overall ecological distance for the row. Thus, we define the Environmental Consistency (ECo) index as the probability that a random shift in a matrix will increase the overall environmental distance of the species involved in the shift: the higher this probability, the higher the ecological consistency.

Let us consider the theoretical case depicted in Fig. 2A. Here a row including the values of an environmental variable (say, for example, a standardized measure of water temperature, T) is associated with a species per area matrix. The first species of the matrix (Sp1) occurs in three localities (A1, A3 and A4). The average distance (D) of the three localities with respect to T will be the mean of the Euclidean distances computed between the T values of the pairs A1–A3, A1–A4, and A3–A4, and thus it will be: (1 + 0 + 1)/3 = 0.67. Now we can imagine to modify the distribution of Sp1. For example we may delete its occurrence in A1 and, simultaneously, we may add a new occurrence in A2 (see Fig. 2B). The average distance (D) will be the mean of the Euclidean distances for T computed between values of the variable measured in the new area pairs A2–A3, A2–A4, A3–A4, i.e. (1 + 2 + 1)/3 = 1.33. As the distance value is increased, this means that the performed shift has reduced the ecological consistency of Sp1 and, in general, of the matrix. To compute ECo, we will simply have to reiterate this procedure several times (ideally performing all possible shifts in the matrix) and calculate the ratio between the number of shifts producing the same effect as above and the number of performed shifts.

In the above example, we performed a shift that did not alter the total number of presences in a row, i.e. we took into account species prevalence (P) but not local species richness (LSR). However, this is not mandatory, and ECo can be calculated under different shift rules to take into account the potential effects of P and LSR in the observed species distribution patterns. The default rule in the ECo Software assumes non randomness in both P and LSR, by allowing only shifts that do not alter matrix row and column totals. However, ECo Software provides three more options with different constraints on row and column totals. The rules work as follows.

(I) Fixed LSR and P (default option): two columns (C_1 and C_2) and two rows (R_1 and R_2) are extracted at random from the matrix. Then cells are moved. Let be X_{11} the cell in R_1 C_1, X_{12} the cell in R_1 C_2, X_{21} the cell in R_2 C_1 and X_{22} the cell in R_2 C_2. Shifting these cells without altering row and column totals will be possible only when X_{11} is equal to X_{22}, X_{21} is equal to X_{12} and X_{11} is different from X_{21}. If the cells do not satisfy this requirement, nothing is done.
Otherwise, the average Euclidean distances for both R_1 and R_2 are computed by averaging the Euclidean distances of each possible pair of presences in the respective row. A pre-shift score (SC_1) is calculated as the mean of these two values. Then the cells are shifted (i.e. X_{11} is shifted with X_{12} and X_{21} is shifted with X_{22}) and a post-shift score (SC_2) is calculated in the same way of SC_1. Finally, cells are moved back to their original position.

(II) Fixed LSR, random P: a column (C_i) and two rows (R_1 and R_2) are extracted from the matrix. Let be X_1 the cell in R_1 C_i, and X_2 the cell in R_2 C_i. If X_1 is equal to X_2, nothing is done. Otherwise, the average Euclidean distances for both R_1 and R_2 are computed by averaging the Euclidean distances of each possible pair of presences in the respective row, and a pre-shift score (SC_1) is computed as the mean of these two values. Then the cells are shifted (i.e. X_1 is shifted with X_2) and a post-shift score (SC_2) is calculated in the same way of SC_1. Finally, cells are moved back to their original position.

(III) Fixed P, random LSR: a row (R_i) and two columns (C_1 and C_2) are extracted from the matrix. Let be X_1 the cell in R_i C_1, and X_2 the cell in R_i C_2. If X_1 is equal to X_2, nothing is done. Otherwise, a pre-shift value (SC_1) is computed by averaging the

Euclidean distances of each possible pair of presences in R_1 . Then the cells are shifted (i.e. X_1 is shifted with X_2) and a post-shift value SC_2 is calculated in the same way of SC_1 . Finally, cells are moved back to their original position.

(IV) Random P and LSR: two cells (X_1 and X_2) are extracted at random from the matrix. If X_1 is equal to X_2 , nothing is done. Otherwise, the average Euclidean distances for both R_1 and R_2 (that are respectively the rows from which X_1 and X_2 were taken) are computed by averaging the Euclidean distances of each possible pair of presences in the respective row, and a pre-shift score (SC_1) is computed as the mean of these two values. Then the cells are shifted (i.e. X_1 is shifted with X_2) and a post-shift value (SC_2) is calculated in the same way of SC_1 . Finally, cells are moved back to their original position.

Ideally, for each rule, all possible combinations of rows and/or columns can be tested by searching for all possible shifts. However, since this can be time consuming for large matrices, ECo Software allows users to indicate a number of row and column pairs to be extracted at random and used in the analysis.

The ECo index is calculated as the ratio between the number of shifts where $SC_2 > SC_1$ and the total number of performed shifts. ECo values higher than 0.5 indicate that shifts are more likely to decrease the ecological consistency, suggesting therefore that the study community is influenced by the selected ecological variables. By contrast, ECo values close to 0.5 indicate that the species is not influenced by the considered environmental variables, as any shift has the same probability of increasing or decreasing the ecological consistency for the species. ECo values lower than 0.5 are unlikely to arise, as they would indicate that each species in the community tends to be systematically present only in areas very different one from another with respect to their environmental features, which would be in contrast with the general ecological laws.

In addition to ECo index, ECo Software returns a measure of the individual contribution of each species to the overall ecological consistency of the matrix under examination. This measure (rECo) is computed as the ratio between the number of shifts increasing the average Euclidean distance between each possible pair of presences

in the species row, and the total number of shifts performed in that row. The closer rECo is to 1, the higher is the contribution of the corresponding species to the overall ecological consistency of the

matrix. Both ECo and rECo evaluate the combined effect of a set of environmental variables on species distributional patterns, without making any assumption (nor drawing any conclusion) on the relative importance of such variables. Yet, an evaluation of the specific effect of a given factor (or a combination of factors) can be evaluated by replicating the analysis with the exclusion (or inclusion) of selected variables.

3. Case studies

3.1. European trees

A species having, on average, narrow tolerance has fewer chances to find a suitable area than species with wider ranges. Consequently, we should expect to find a negative relationship between the overall niche breadth of the species of a particular assemblage, and the corresponding ECo value. To test this expectation, we compared ECo indexes computed for 1000 random matrices of species per site distribution of European trees with the overall niche breadth of the species included in each matrix. A list including almost 1 million occurrence records for 206 tree species across all Europe, at a spatial resolution of 1 km2 , was compiled by merging information from three large datasets. These datasets were provided by the European Forest Data Centre of the European Commission (http://efdac.jrc.ec.europa.eu) and include: (1) presence/absence data from 22 National Forest Inventories, (2) plot information from the so-called Level I and Level II schemes set up by countries for the monitoring of atmospheric pollution on forests in the context of Regulation No. 2152/2003 (Forest Focus, Official Journal of the European Union, 2003, L 324/1), and (3) the Biosoil project (Durrant et al., 2011) in which forest tree biodiversity was sampled in around 3379 plots in Europe. Then, a grid composed of 25 square cells of 0.25◦ × 0.25◦ latitude/longitude was superimposed at random on the tree distribution map. Each grid cell was assumed as an area and the total set of species present in the grid was considered as the species pool. Random grid overlap was repeated to generate a set of species per area matrices of similar size. Only non-empty columns (i.e. grid cells) were included in each resulting matrix. Only matrices including at least 5 species and 5 columns were retained. Climatic variables were obtained from Wordclim (Hijmans et al., 2005). In particular, we used the Current σdataset, which is based on interpolations of observed data representative of the period spanning from 1950 to 2000, at a resolution of 2.5 . We included in our analyses a moderately small subset of the available bioclimatic layers (annual mean temperature, temperature seasonality, temperature annual range, annual precipitation, and precipitation seasonality) in order to keep our case study as simple as possible and to limit potential biases due to variable inter-correlation (Leaché et al., 2009). To account for differences in variable units, we replaced each value (x_i) with its corresponding standardized z-value, computed as $(x_i - \mu x)/\sigma x$ where μx and σx are, respectively, the mean and the standard deviation of the variable under study. For each matrix we computed ECo using the default shifting rule (that preserves row and column totals), to take into account local species richness and species prevalence. All analyses were conducted exhaustively, i.e. performing all possible row and column shifts. Finally, we compared each ECo value of each matrix with a measure of overall 'ecological specialization' of that matrix. This measure was computed as follows. For each species listed in the matrix we computed tolerance limit ranges for the same climatic variables used to calculate ECo. Tolerance limit ranges were calculated as the difference between the minimum and the maximum value of each variable measured in all point records known for that species. Then we averaged these values to obtain, for each species, a score indicating its overall niche breadth. Finally we computed the overall 'ecological specialization' of a given matrix as the reciprocal of the average of the niche breadth scores of its species. Climatic variables and tree distribution point records were obtained from the same sources as above (Hijmans et al., 2005). Again, to cope with the difference in unit measure of climatic variables we used z-values instead of raw values.

In addition, for each matrix, we also computed rECo values for each tree species. Then, we averaged the rECo values of each tree species included in at least one matrix and compared these values with the respective mean value of tolerance limit ranges (calculated as above). As expected ECo values resulted >0.5 (average 0.62 ± 0.10,n = 1000) in most of the random matrices (>95%). We found a significant positive relation between ECo values of the 1000 random matrices and the overall ecological specialization of the included species (Spearman's correlatio coefficientrs = 0.29, $p < 0.0001$, n = 1000, see Fig. 3). Moreover, we found a significant relation between the average rECo values of each tree species included in at least one matrix and the respective mean values of niche breadth values (rs = -0.65, $p < 0.0001$, $n = 204$). The ECo values resulted weakly (but significantly) correlated with longitude (rs = -0.11, $p < 0.005$), and moderately correlated with latitude (rs = -0.34, $p < 0.00001$). These negative relationships make sense, as they reflect the effect of variations in environmental characteristics, such as temperature, precipitation and resource availability, which are important constraints of species distributions. It is well known that in Europe biological diversity tends to decrease northwards as a reflection of historical and ecological factors on species richness and species composition, whereas possible longitudinal trends are less clear (Fattorini and Baselga, 2012; Fattorini and Ulrich, 2012). The patterns emerging from Eco analysis indicate an increase in the relative importance of environmental factors over species interactions on the composition of species assemblages when moving towards higher latitudes, thus suggesting that the latitudinal gradient in biodiversity may reflect the increasing severity of climatic factors on species distribution. We also found a weak correlation between the ECo values and the average Euclidean distance of the selected environmental variables computed between any pair of matrix localities ($r_S = 0.13$, $p < 0.00001$), which indicates that ECo is little affected by the degree of environmental heterogeneity in the study area.

3.2. Terrestrial vertebrates

We performed an extensive analysis on European terrestrial vertebrates using the same procedure as above, with the aim of exploring the behaviour of ECo for taxa with very different characteristics. We performed separated analyses for mammals, reptiles, amphibians and birds. All occurrence records were obtained from GBIF (www.gbif.org). An important premise of ECo computation, is that the scale of study should be calibrated in order to identify meaningful distribution patterns. In other words, it would make little sense to compute ECo on a matrix where each species occurs in all localities. This can be accounted simply by using a scale ensuring that matrix fill is not too high. According to this principle, we used grids of the same size as those of the tree case study except for birds, where the use of such a small scale would lead to a situation of overfilled matrices because of the much higher mobility of birds in respect to that of the other considered vertebrates. Thus bird matrices were compiled by superimposing to the area of study grids composed of 25 cells of 1° × 1° latitude/longitude. In this way we obtained a degree of fill comparable in all sets of matrices, even if that reported for birds was still the highest despite the use of the larger grid (see Table 1).

To keep the analyses simple, we computed ECo using only mean annual temperature (again obtained from WordClim) as environmental variable. Results are reported in Table 1 and in Fig. 4A–D. For mammals, reptiles and birds, the relationships between ECo values and the overall ecological specialization of each species listed in a matrix were very similar to those obtained for the European tree species, while relationships were slightly weaker for the amphibians.

Noteworthy, in several bird matrices ECo values resulted equal to 1, i.e. species were arranged among sites in such a way that there was no possible shift increasing the overall environmental consistency. This peculiar pattern, which is substantially different from those observed in the other matrices, could be the consequence of the lower geographical resolution used for the birds, which could emphasize the differences in temperatures between areas and hence lead to scenarios of high environmental consistency for species assemblages significantly affected by the considered environmental variables. This is, however, in accordance with the well-known dependence of bird distributions on climate, and in particular on temperature, that makes them a suitable group to investigate the ecological effects of climate change (see, for example, Sekercioglu et al., 2012).

3.3. Marine fish

We measured ECo of 50 fish families within an area of high species richness (Carpenter and Springer, 2005), including the Indo–Malay–Philippines archipelago (20S–20N; 90E–150E) using a set of environmental variables known to affect marine organisms, and then we compared the resulting ECo values with several biological characteristics of the respective fish families potentially influencing their occurrence and abundance at each site.

We subdivided the study area into 1° × 1° latitude/longitude cells and, for each fish family, we compiled a binary species per area matrix based on the species distribution on the grid. Fish species distribution was obtained from OBIS (Vanden-Berghe, 2012). Although the biogeographic accuracy of OBIS data has been previously questioned (Mora et al., 2008; Robertson, 2008), OBIS database has grown from around 15 million records in 2008 (the year it was heavily criticized, see Mora et al., 2008; Robertson, 2008), to almost 33 million records in 2012 (Vanden-Berghe, 2012). This has recently promoted its use in several macroecological studies (O'Dor et al., 2010; Tittensor et al., 2010; Webb et al., 2010; Tyler et al., 2012). However, to minimize potential bias related to unequal sampling, we selected the 50 fish families with the highest number of known point records.

To calculate ECo values, we related the following 8 environmental variables to each grid cell (column in the matrix): sea depth (mean and standard deviation), surface temperature (annual mean and standard deviation), bottom temperature (annual mean), surface salinity (annual mean), bottom salinity (annual mean), and primary productivity (annual mean). These environmental variables were obtained from Aquamaps (Kaschner et al., 2008) and were chosen as some of the most important macroecological factors affecting the distribution of marine species (Tyberghein et al., 2011). Finally we transformed raw values of each variable into z-values (see Section 3.1). Please refer to Aquamaps documentation (http://www.aquamaps.org/envtdata/main.php) for further details on the variables. We computed ECo values for each fish family sep-arately. As above, we used the default shifting rule (that preserves row and column totals), in order to take into account local species richness and species dispersal ability. All analyses were conducted exhaustively, i.e. performing all possible row and column shifts, using ECo Software.

For each fish family we took in consideration various sets of characteristics potentially related to ecological specialization: (1) range size; (2) frequency of occurrence; (3) habitat preferences (demersal vs. pelagic, reef associated); (4) migratory behaviour; and (5) life-history traits (maximum body size, growth rate, life span, age at maturity, and trophic level).

Species range sizes were calculated using measures of area of occupancy (AOO) and extent of occurrence (EOO) (Gaston and Fuller, 2009). AOOs were calculated as follows: for each species, we plotted all available point records from OBIS (Vanden-Berghe, 2012) on a global grid of 1° × 1° latitude/longitude and then we counted the number of grid cells where the species was present. EOOs were calculated as the number of 1° × 1° grid cells given by the product of their latitudinal and longitudinal ranges. Latitudinal range (Lat) was calculated as the difference between maximum and minimum latitude of species occurrences. Longitudinal range (Lon) was calculated as the difference between maximum and minimum longitude of species occurrences. Range size of each family was computed as the average of the range sizes of the species present in the corresponding matrix.

In addition, we computed frequency of occurrence for each con-sidered fish species as the ratio between its total number of point records and the number of grid cells where it occurs. This measure might reflect an estimate of local abundance (although we cannot exclude that it might also express differences in study efforts, due to unequal sampling among areas). Again, frequency of occurrence of each family was computed as the average of the frequencies of occurrence of the species present in the corresponding matrix. All other considered fish characteristic was obtained from FishBase (Froese and Pauly, 2013).

Habitat preferences (3 values) for each family were computed as the respective fraction of species in each matrix (i.e. family) falling into one of the following FishBase categories: (1) demersal (including demersal, benthopelagic, and bathydemersal); (2) pelagic (including bathypelagic, pelagic–neritic and pelagic–oceanic); and (3) reef-associated. Similarly, migratory behaviour for each family was estimated as the fraction of matrix species falling into at least one of FishBase migratory category (oceanodromous, anadromous, catadromous, potamodromous, and amphidromous).

Life-history traits (maximum body size, growth rate, life span, age at maturity, and trophic level) for each family were computed as the average of the respective values of each species present in the respective family matrix. Specific details about these ecological parameters can be found in Fishbase online documentation at http://www.fishbase.org/manual/Key%20Facts.htm.

A table reporting the complete results of the ECo analyses for each matrix (including matrix characteristics: rows, columns, fill, shape, and number of occurrences) and the corresponding ecological characteristics (range size, frequency of occurrence, habitat preferences, migratory behaviour, life-history traits, and taxonomic diversity) is provided as supplementary material (Appendix A). As expected, all matrices except one showed ECo values higher than or equal to 0.5, with an average of 0.57 ± 0.06.

Spearman's rank correlation coefficients for the pairwise comparisons between ECo values and the corresponding ecological characteristics are reported in Table 2. ECo values resulted to be significantly correlated ($p < 0.05$) with the respective range size measures (negative correlation for both AOO and EOO, with respectively $r_S = -0.73$ and $r_S = -0.64$), frequency of occurrence (negative correlation, $r_S = -0.51$), fraction of reef-associated species (negative correlation, $r_S = -0.38$), average maximum body size (negative correlation, $r_S = -0.37$), trophic level (negative correlation, $r_S = -0.34$) and fraction of demersal species (positive correlation, $r_S = 0.31$).

4. Differences between null models

In order to test the response of ECo to the use of different null models, we replicated the analysis described in Section 3.1, i.e. the comparison between the ECo index computed for 1000 random matrices of species per site distribution of European trees, and the overall niche breadth of the species included in each matrix, using the less restrictive null model (i.e. the null model IV, random P and LSR, see Section 2) instead of the null model I. We used the same environmental variables used in the previous analysis, and we attempted 10,000 shifts for each matrix. Then we compared the results of this analysis with those obtained in the previous one. Although the two series of ECo values resulted strongly correlated ($r_S = 0.76$, $p < 0.000001$), ECo values obtained using the least restrictive null model were generally higher than those obtained using the most restrictive one (see Fig. 5). This highlights the existence in the structure of examined matrices of peculiar constraints that limit the number of possible shifts reducing the overall environmental consistency. The explanation of this pattern is, however, straightforward. In the most restrictive null model, the only permitted shifts are those involving chequerboard pairs, i.e. submatrices in one of the forms

$$\begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix}$$

or

$$\begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}$$

Chequerboard pairs may result from competitive exclusion, as they indicate situations where a species is present in a locality where another species is missing and vice versa. In a scenario where competitive exclusion matters, and where the absence of a species from a site is therefore not related to local environmental suitability, a shift leading from one chequerboard configuration to another would not have a particular probability of increasing or reducing the overall environmental consistency of the matrix. The fact that in most matrices the null model I led to ECo values lower than the null model IV suggests a prominent role of competitive exclusion in determining species distribution.

To further investigate this aspect empirically, we computed the C-score, i.e. the mean number of chequerboard units per species-pair (Stone and Roberts, 1990), for each tree matrix, and then we compared the C-scores with the ECo values obtained using, alternatively, the null model I and the null model IV. As expected, we found a significant correlation between the C-scores and the corresponding ECo values when the null model I was used ($r_S = 0.14$, $p < 0.000001$), but not the when null model IV was used ($r_S = 0.04$, $p = 0.2$).

5. ECo relationship with matrix features

Since the probability that a shift increases or reduces the overall parametric distance between localities depends only on the environmental variables associated with each locality record, ECo should be in principle independent from matrix properties. However, species local abundance (which affect the fill of a species per area matrix) is often correlated with species range (Strona et al., 2012), and species with wider geographical ranges are likely, as extensively discussed above, to have great environmental tolerance. Consequently, we should expect to find a certain degree of correlation between ECo index and matrix fill naturally emerging from these patterns, and still independent from the statistical principles behind ECo computation. Thus, to test the independence of ECo from matrix properties, we created a theoretical dataset by randomly displacing across Europe the tree species occurrences of the dataset analyzed in Section 3.1, instead of using the original tree species distribution.

We computed ECo values on 1000 random matrices generated from the theoretical tree dataset using the same sampling procedure as described in Section 3.1, using both the most (I) and the least (IV) restrictive null models. Then, we evaluated the relationships between ECo values and number of species, localities, number of occurrences, matrix size (i.e. number of rows × number of columns), and matrix fill (i.e. the ratio between number of occurrences and matrix size).

For both null models, despite the use of random presences, we obtained a good variability of ECo values, with standard deviations equal to 0.19 for the null model I, and 0.14 for the null model IV, which are both higher than the standard deviations observed in the corresponding sets of ECo values computed using the real tree distribution (0.10 and 0.13). This ensures that our comparison is not biased by low variability in the sets of observations.

As expected, for both null models, we found no significant correlation between ECo values and the aforementioned matrix characteristics ($p > 0.1$ in all cases), except for the number of localities, where correlations where significant, but extremely low ($r_S = 0.01$ in both cases, $p < 0.05$).

6. Use of alternative parametric distance measures

We defined ECo as the probability of a random shift in a matrix to increase the overall environmental distance of the species involved in the shift. To compute such probability in the above case studies, we used the Euclidean distance, as it is one of the simplest available distance metrics and offers straightforward interpretations, thus representing an obvious choice in exploratory analyses. However, ECo can be in principle computed using any available multiparametric distance measure. An interesting alternative is represented by Mahalanobis distance, which may offers advantages over the Euclidean one in case of multi-collinearity among environmental variables.

We replicated the analysis described in Section 3.1, computing ECo using the default shifting rule and performing all possible row and column shifts, and using the Mahalanobis distance instead of the Euclidean one. Results obtained using the Mahalanobis distance were very close to those obtained using the Euclidean distance (Fig. 6; regression line equation $y = 0.88x + 0.06$, $R^2 = 0.85$). This indicates that our results are robust towards the use of different distance metrics.

7. Discussion

ECo makes it possible to test the importance of environment in species spatial assemblage taking into account also biotic interaction, local species richness and species prevalence, thus integrating the concepts of Eltonian and Grinnelian niche (Soberón, 2007). Our case studies illustrate some of the potentialities of our approach.

The relation between ECo values measured for the random tree species assemblages and their respective overall degree of specialization provides evidence that our method can confidently assess the effect of environmental variables. Although the measured relationship, with a Spearman's correlation coefficient of −0.31 may seem relatively weak, our results are indeed very consistent with the concept behind ECo, as it emerges from the triangular patterns shown in Figs. 3 and 4. According to these graphs, matrices with a high score of overall ecological specialization, i.e. including mainly species with narrow tolerance limits for the considered variables, can only have high ECo values, while matrices with a low score of overall ecological specialization can have either high or low ECo values. These results make much sense, as they reflect the fact that a species cannot survive outside its tolerance limits. Thus, species with very narrow tolerance limits are necessarily present only in those areas satisfying their requirements, which are therefore likely to be much similar for the considered variables. This minimizes the probability that a random shift could reduce the average distance between the variables. On the other hand, species with larger tolerance for the same environmental variables are less affected by these constraints. This allows them to be distributed with a higher degree of randomness with respect to the environmental variables recorded at each locality, thus having a lower ECo value. In general, however, the level of specialization in a matrix constraints the lower limit of ECo, that is simply another way to state that species distributions cannot contradict species ecology. As regards the contribution of individual species to the overall value of a matrix ECo (which is measured by rECo values), the strong relationship found with the corresponding values of ecological specialization supports the ecological and computational assumptions behind our method. In particular, it demonstrates the existence of a close relation between a species' Grinnelian niche and the probability of a shift to reduce the average distance between the environmental variables of the areas where that species occurs.

All these observations also apply to the case studies of terrestrial vertebrates (see Fig. 4). Amphibians, however, showed a weak relationship. Yet, this is likely due to the fact that, in contrast with mammals and reptiles, amphibians have distributional patterns much more constrained by water availability than temperatures.

The ecological soundness of ECo is demonstrated also by the results of the analyses conducted on marine tropical fish families, which support the general hypothesis that species niche requirements influence range size (Williams et al., 2009) and, in particular, that the species belonging to families with overall high local ecological consistence have narrower geographical distributions in comparison with species belonging to families with overall low local ecological consistence.

Moreover, our analyses indicate that distribution of families with species characterized by high frequency of occurrence (i.e. local abundance) tends to be weakly affected by environmental variables. This sounds much reasonable, considering that organisms that can tolerate wide variation in several factors, tend to be both spatially widespread and locally abundant (Brown, 1984).

The significant positive correlation between the fraction of demersal species per family and ECo values, provides a general support to the high habitat specialization of tropical demersal fish already observed at a finer scale by Fitzpatrick et al. (2012). The negative relationship observed between ECo values and the fraction of reef-associated species per family is much interesting because it corroborates the hypothesis that coral reef fish may be much less specialized than commonly assumed, and that their functional versatility may play a fundamental role in maintaining coral reef biodiversity, in contrast with the traditional assumption that high local species diversity is determined by fine-scale niche partitioning (Bellwood et al., 2006). Finally, given the relevance of several of the selected ecological variables (e.g. temperature and salinity) for fish physiology, the negative relationship between body size and ECo is not surprising, considering that large bodied animals can maintain homeostasis over a wider range of conditions than small bodied ones (Gaston, 1990).

Recent studies demonstrate the existence of a significant relationship between geographical range, body size and local abundance in both marine and freshwater fish at a global scale (Strona et al., 2012). The results of the present analysis not only are highly consistent with this scenario, but also provide further insights into this topic, promoting the hypothesis that species geographical range and local abundance correlate because both are affected by ecological specialization, and that body size correlates with geographical range because it is somehow a measure of species tolerance towards environmental variables.

It should be highlighted that the choice of different null models to compute ECo may lead to different results. In particular the most restrictive null model (see Section 2) takes into account species interactions by focusing on chequerboard pairs, making it possible to disentangle the effect of competition from that of environmental factors.

For the sake of clarity, in the reported case study and, in general, throughout this paper, we deliberately introduced some constraints to the application of ECo. In particular, we used the species as the primary categorical unit to be included in the matrix, and we indicated species area matrices as the necessary input to the analysis. However, ECo can be computed for assemblages of any taxonomic level (populations within species, families within orders, etc.), making it possible to test different ecological, biological or biogeographical hypotheses. For example, when computed for different populations of the same species, ECo could provide valuable information on mechanisms related to intraspecific ecological divergence. Conversely, when applied to higher taxonomic levels, ECo can help disentangling the effect of historical and ecological factors in determining large scale distribution patterns.

In addition, although ECo Software has been developed to be used with species-site matrices, the concepts behind ECo can be easily extended to the more general field of ecological networks (for example it could be used to evaluate how host features affect the distribution of parasites on hosts). Thus, we hope that the immediacy of ECo software will stimulate ecologists to use our approach to ecological consistency in a broad way.

Acknowledgements

We are very grateful to two anonymous referees for their constructive comments on a previous version of this paper. Author contributions: G.S. conceived the idea. G.S., S.F. and J.S.M.A. designed most of the study; S.M., D.S. and P.G. designed part of the study. J.S.M.A. provided data on tree species distribution. G.S. performed most of the analyses and realized the software. S.F. performed some of the analyses. G.S., S.F. and J.S.M.A. wrote the paper. The views expressed are purely those of the writers and may not in any circumstances be regarded as stating an official position of the European Commission.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ecolind. 2014.11.033.

References

Bellwood, D.R., Wainwright, P.C., Fulton, C.J., Hoey, A.S., 2006. Functional versatility supports coral reef biodiversity. Proc. Biol. Sci. 273, 101–107.

Bertness, M.D., Callaway, R., 1994. The role of positive forces in natural communities: a post-cold war perspective. Trends Ecol. Evol. 9, 191–193.

Brown, J.H., 1984. On the relationship between abundance and distribution of species. Am. Nat. 124, 255–279.

Carpenter, K.E., Springer, V.G., 2005. The center of the center of marine shore fish biodiversity: the Philippine Islands. Environ. Biol. Fishes 72, 467–480.

Cox, C.B., Moore, P.D., 2004. Biogeography: An Ecological and Evolutionary Approach. Blackwell Science, Oxford.

Durrant, T., San-Miguel-Ayanz, J., Schulte, E., Suarez Meyer, A., 2011. Evaluation of BioSoil Demonstration Project: Forest Biodiversity. EUR 24777 EN. Publication Office of the European Union, Luxemburg, http://dx.doi.org/10.2788/84823.

Fattorini, S., Ulrich, W., 2012. Spatial distributions of European Tenebrionidae point to multiple postglacial colonization trajectories. Biol. J. Linn. Soc. 105, 318–329.

Fattorini, S., Baselga, A., 2012. Species richness and turnover patterns in European tenebrionid beetles. Insect Conserv. Divers. 5, 331–345.

Fitzpatrick, B.M., Harvey, E.S., Heyward, A.J., Twiggs, E.J., Colquhoun, J., 2012. Habitat specialization in tropical continental shelf demersal fish assemblages. PLoS ONE 7, e39634.

Fortuna, M.A., Stouffer, D.B., Olesen, J.M., Jordano, P., Mouillot, D., Krasnov, B.R., Poulin, R., Bascompte, J., 2010. Nestedness versus modularity in ecological networks: two sides of the same coin? J. Anim. Ecol. 79, 811–817.

Froese, R., Pauly, D., 2013. FishBase. world wide web electronic publication. Available on: http://www.fishbase.org/ (last accessed 04.03.2013).

Gaston, K.J., 1990. Patterns in the geographical ranges of species. Biol. Rev. 65, 105–129.

Gaston, K.J., Fuller, R.A., 2009. The sizes of species geographic ranges. J. Appl. Ecol.46, 1-9

Gotelli, N.J., 2000. Null model analysis of species co-occurrence patterns. Ecology 81, 2606–2621.

Guo, Q., Liu, Y., 2010. ModEco: an integrated software package for ecological niche modeling. Ecography 33, 1–6.

Hausdorf, B., Hennig, C., 2007. Null model tests of clustering of species, negative co-occurrence patterns and nestedness in meta-communities. Oikos 116, 818–828.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high res-olution interpolated climate surfaces for global land areas. Int. J. Climatol. 25, 1965–1978.

Huntley, B., 1999. Species distribution and environmental change. In: Maltby, E., Holdgate, M., Acreman, M., Weir, A. (Eds.), Ecosystem Management: Questions for Science and Society. Royal Holloway Institute for Environmental Research, University of London, Egham, pp. 115–129.

Kaschner, K., Ready, J.S., Agbayani, E., Rius, J., Kesner-Reyes, K., Eastwood, P.D., South, A.B., Kullander, S.O., Rees, T., Close, C.H., Watson, R., Pauly, D., Froese, R., 2008. AquaMaps environmental dataset: Half-Degree Cells Authority File (HCAF), Available on: http://www.aquamaps.org/data (last accessed 04.03.2013).

Leaché, A.D., Koo, M.S., Spencer, C.L., Papenfuss, T.J., Fisher, R.N., McGuire, J.A., 2009. Quantifying ecological, morphological, and genetic variation to delimit species in the coast horned lizard species complex (Phrynosoma). Proc. Natl. Acad. Sci. U.S.A. 106, 12418–12423.

Mora, C., Tittensor, D.P., Myers, R.A., 2008. The completeness of taxonomic inventories for describing the global diversity and distribution of marine fishes. Proc. Biol. Sci. 275, 149–155.

O'Dor, R., Miloslavich, P., Yarincik, K., 2010. Marine biodiversity and biogeography regional comparisons of global volumes, an introduction. PLoS ONE 5, e11871.

Robertson, D.R., 2008. Global biogeographical databases on marine fishes: caveat emptor. Divers. Distrib. 14, 891–892.

Sekercioglu, C., Primack, R., Wormworth, J., 2012. The effects of climate change on tropical birds. Biol. Conserv. 148, 1–18.

Soberón, J., Peterson, A.T., 2005. Interpretation of models of fundamental biological niches and species' distributional areas. Biodivers. Inform. 2, 1–10.

Soberón, J., 2007. Grinnelian and Eltonian niches and geographic distribution of species. Ecol. Lett. 10, 1115–1123.

Stone, L., Roberts, A., 1990. The checkerboard score and species distributions. Oecologia 85, 74–79.

Strona, G., Galli, P., Montano, S., Seveso, D., Fattorini, S., 2012. Global-scale relationships between colonization ability and range size in marine and freshwater fish. PLoS ONE 7 (11), e49465.

Tittensor, D.P., Mora, C., Jetz, W., Lotze, H.K., Ricard, D., Vanden Berghe, E., Worm, B., 2010. Global patterns and predictors of marine biodiversity across taxa. Nature 466, 1098–1101.

Tyberghein, L., Verbruggen, H., Pauly, K., Troupin, C., Mineur, F., De Clerck, O., 2011. Bio-ORACLE: a global environmental dataset for marine species distribution modelling. Global Ecol. Biogeogr. 21, 272–281.

Tyler, E.H.M., Somerfield, P.J., Vanden-Berghe, E., Bremner, J., Jackson, E., Langmead, O., Palomares, M.L.D., Webb, T.J., 2012. Extensive gaps and biases in our knowledge of a well-known fauna: implications for integrating biological traits into macroecology. Global Ecol. Biogeogr. 21, 922–934.

Vanden-Berghe, E., 2012. The ocean biogeographic information system: web pages.
Available on: http://www.iobis.org/ (last accessed 04.03.2013).

Webb, T.J., Vanden Berghe, E., O'Dor, R., 2010. Biodiversity's big wet secret: the global distribution of marine biological records reveals chronic under-exploration of the deep pelagic ocean. PLoS ONE 5, e10223.

Williams, S.E., Williams, Y.M., VanDerWal, J., Isaac, J.L., Shoo, L.P., Johnson, C.N., 2009. Ecological specialization and population size in a biodiversity hotspot: how rare species avoid extinction. Proc. Natl. Acad. Sci. U.S.A. 106, 19737–19741.

Woodward, F.I., 1987. Climate and Plant Distribution. Cambridge University Press.

Table 1

Results of the ECo analysis conducted for European terrestrial vertebrates (see Section 3.1). The table includes average values and standard deviations of matrix features (fill, number of columns, and number of rows), ECo values, and ecological specialization scores × 1000 (E.S.). The table also includes correlation coefficients between ECo values and average ecological specialization of each matrix included in the analysis (Corr1), and correlation coefficients between the average of the values of ecological consistency computed for each species included in at least one examined matrix and the corresponding measures of species' niche breadth (Corr2). All correlations are expressed as Spearman's r_s and are significant ($p < 0.0001$).

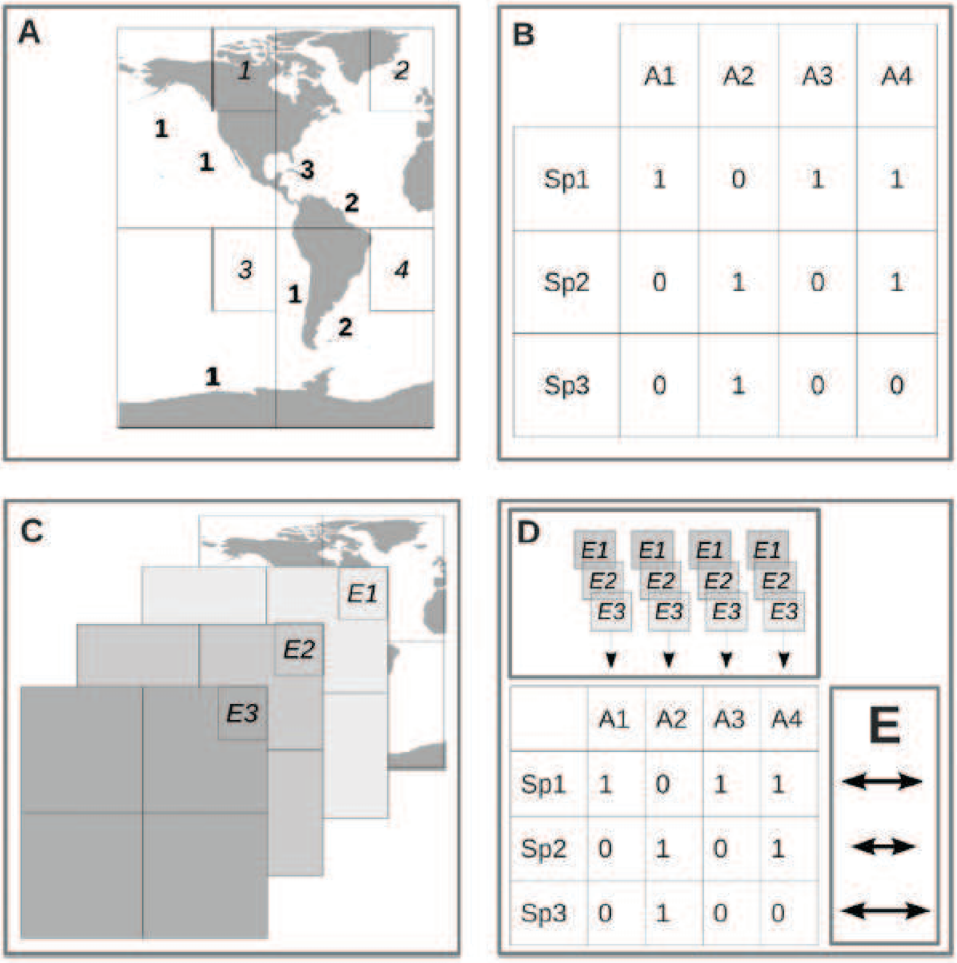
Taxon	Fill	Rows	Columns	ECo	E.S.	Corr1	Corr2
Mammals	0.36 ± 0.18	43.28 ± 23.92	19.60 ± 5.96	0.69 ± 0.14	9.27 ± 1.69	0.55 (n = 1000)	−0.62 (n = 1019)
Reptiles	0.41 ± 0.14	15.77 ± 8.84	18.01 ± 6.48	0.65 ± 0.14	8.30 ± 1.52	0.46 (n = 1000)	−0.74 (n = 372)
Amphibians	0.45 ± 0.14	14.12 ± 4.16	18.34 ± 6.22	0.61 ± 0.12	8.91 ± 0.74	0.14 (n = 1000)	−0.75 (n = 107)
Birds	0.63 ± 0.22	130.27 ± 111.25	24.90 ± 0.94	0.77 ± 0.21	47.95 ± 6.08	0.56 (n = 1000)	−0.22 (n = 565)

Table 2

Spearman's rank correlation coefficients for the pairwise comparisons between ECo values of the 50 marine fish families (see Section 3.3) and their corresponding ecological characteristics. M: fraction of migratory species; D: fraction of demersal species; P: fraction of pelagic species; R: fraction of reef-associated species; L: average maximum body size; K: average growth rate; Y: average life span; Y_m : average age at maturity; T: average trophic level; AOO: area of occupancy; EOO: extent of occurrences; W: frequency of occurrence (i.e. local abundance).

	M	D	P	R	L	K	Y	Y_m	T	AOO	EOO	W
ECo	0.10	0.31^{*}	0.00	-0.38^{**}	-0.37^{**}	0.18	−0.11	−0.02	-0.34^{*}	-0.73^{***}	-0.64^{***}	-0.51^{***}

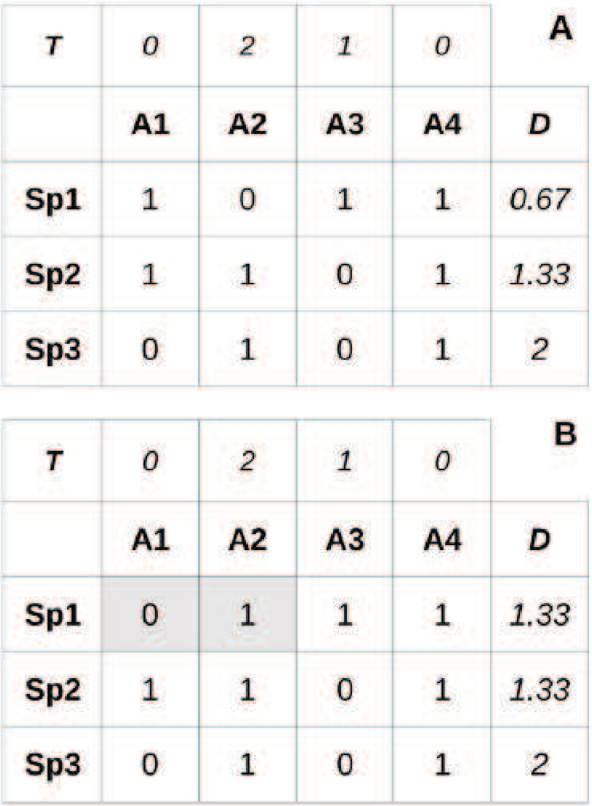
* $p < 0.05$. ** $p < 0.01$. $p < 0.001$.



	A1	A2	A3	A4
Sp1	1	0	1	1
Sp2	0	1	0	1
Sp3	0	1	0	0

	A1	A2	A3	A4
Sp1	1	0	1	1
Sp2	0	1	0	1
Sp3	0	1	0	0

Fig. 1. Functioning of ECo algorithm. Species occurrences (A: species 1–3 are identified by numbers in roman, areas 1–4 are identified by numbers in italic) are used tocompile a binary species (rows) per areas (columns) matrix (B). A set of desired environmental variables is associated with each area (C: layers E1, E2, and E3). Each speciesoccurrence (presence in a row) is therefore linked to the corresponding environmental variables of the areas (column) from which the species is known (D). Using thesevariables it is possible to compute a distance (for example Euclidean) between any possible pair of occurrences in a row, and to use these pairwise values to compute anoverall (average) distance (E) among all the species occurrences in a given row.



A

T	*0*	*2*	*1*	*0*	
	A1	**A2**	**A3**	**A4**	***D***
Sp1	1	0	1	1	*0.67*
Sp2	1	1	0	1	*1.33*
Sp3	0	1	0	1	*2*

B

T	*0*	*2*	*1*	*0*	
	A1	**A2**	**A3**	**A4**	***D***
Sp1	0	1	1	1	*1.33*
Sp2	1	1	0	1	*1.33*
Sp3	0	1	0	1	*2*

Fig. 2. Example of a shift in a theoretical matrix describing the distribution of threespecies (Sp1, Sp2 and Sp3) in four localities (A1, A2, A3, and A4). T represents anenvironmental variable (say, for example, a standardized measure of water temperature). (A, before the shift): Sp1 occurs in three localities (A1, A3 and A4). The averagedistance (D) of the three localities with respect to T is the mean of the Euclidean distances computed between the T values of the pairs A1–A3, A1–A4, and A3–A4 (0.67).(B, after the shift): the distribution of Sp1 has been modified (one of its presenceshas been moved from A1 to A2). The average distance (D) will now be the mean ofthe Euclidean distances for T computed between values of the variable measured inthe new area pairs A2–A3, A2–A4, A3–A4 (1.33). As the distance value is increased,this means that the performed shift has reduced the ecological consistency of Sp1and, in general, of the matrix.

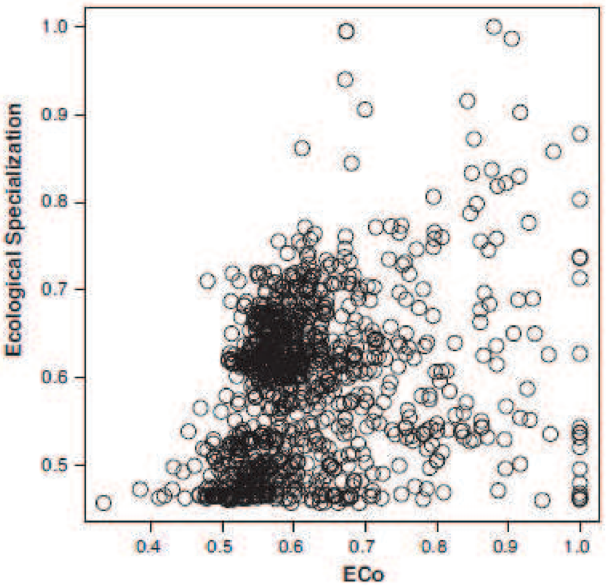


Fig. 3. Relationship between the ECo indexes computed for 1000 random speciesper area matrices of European tree species and their average 'ecological specialization', computed as described in Section 3.1, and standardized using the maximumrecorded value.

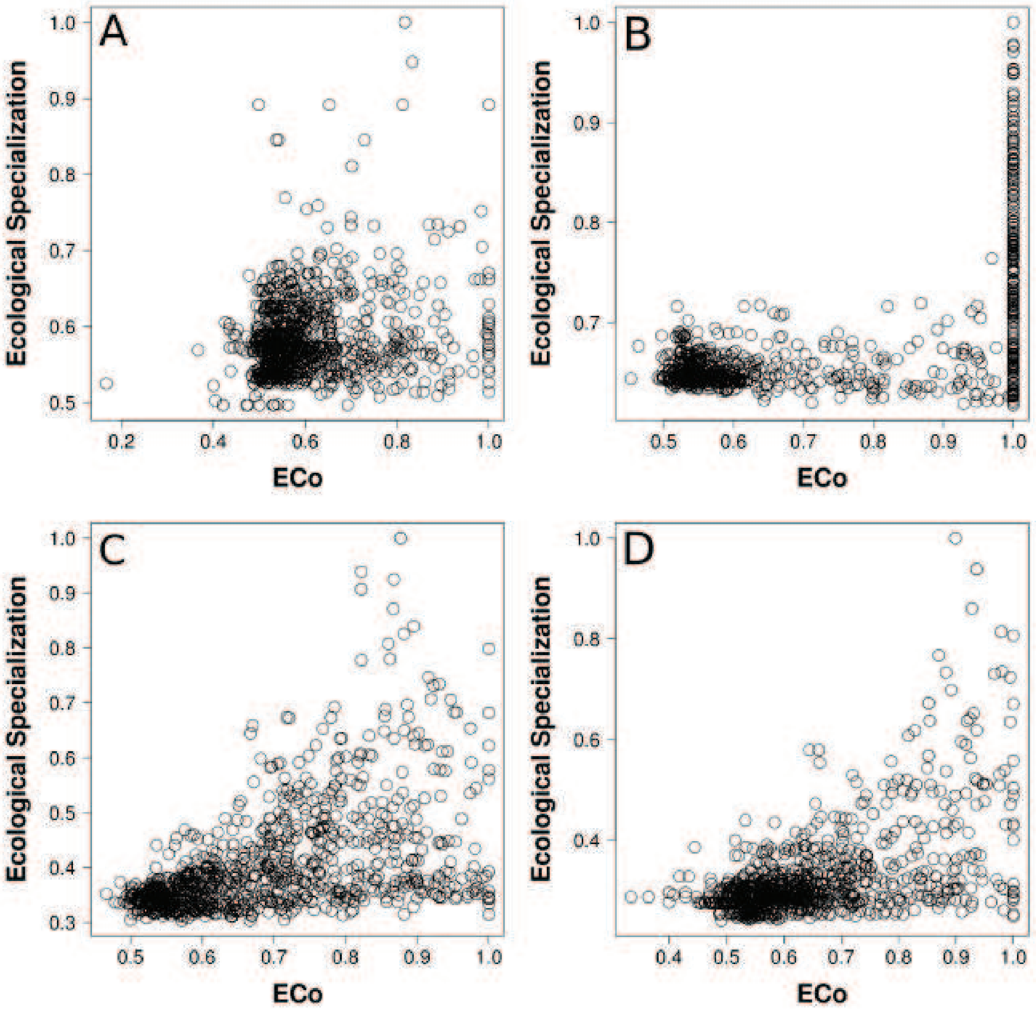


Fig. 4. Relationship between the ECo indexes and average ‘ecological specialization’ for European terrestrial vertebrates, computed as described in Section 3.1, and standard-ized using the respective maximum recorded value. Each plot includes results obtained from 1000 random species per area matrices. A: amphibians; B: birds; C: mammals; D: reptiles. See Section 3.2 for detailed method description.

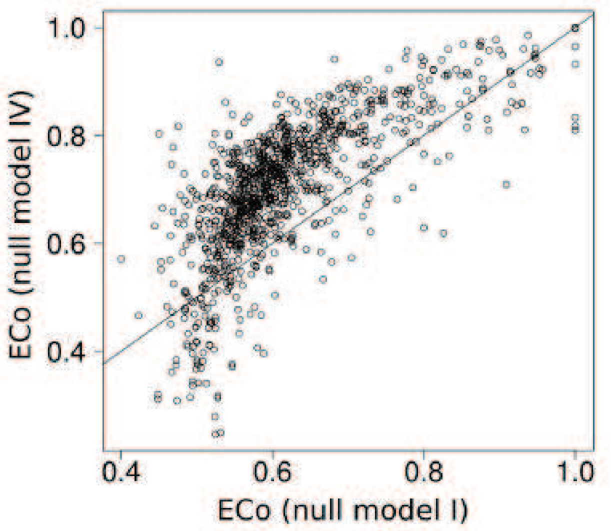


Fig. 5. Comparison between the ECo values computed for 1000 random matrices of species per site distribution of European trees using the most restrictive null model, and the corresponding ECo values computed using the least restrictive one. Values obtained using the least restrictive null model were generally higher than those obtained using the most restrictive one. The continuous line is the line of equality.

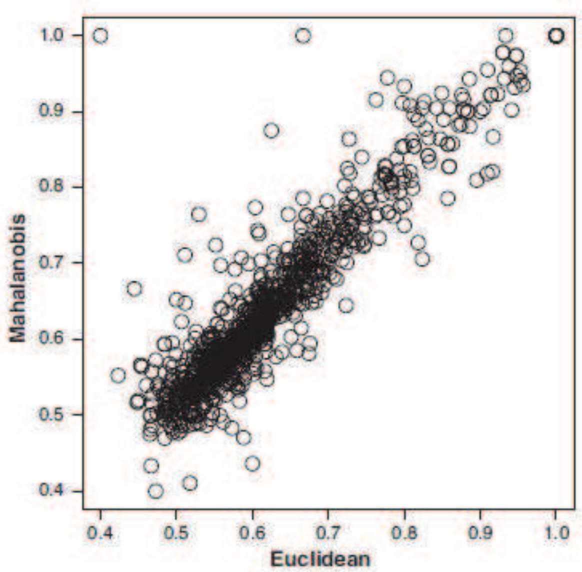


Fig. 6. Relationship between the ECo indexes computed for 1000 random species per area matrices of European tree species using, alternatively, Euclidean and Maha-lanobis multiparametric distances.