



Too good to be true: pitfalls of using mean Ellenberg indicator values in vegetation analyses

David Zelený & André P. Schaffers

Keywords

Bio-indication; Circularity of reasoning; Compositional similarity; Environmental calibration; Null models; Ordination analysis; Randomization; Species richness; Weighted averaging

Abbreviations

ANOVA = analysis of variance; CA = correspondence analysis; CCA = canonical correspondence analysis; DCA = detrended correspondence analysis; EIVs = Ellenberg indicator values; TWINSpan = two-way indicator species analysis

Received 2 September 2010

Accepted 14 October 2011

Co-ordinating Editor: Michael Palmer

Zelený, D. (corresponding author, zeleny@sci.muni.cz): Department of Botany and Zoology, Masaryk University, Kotlářská 2, CZ-611 37, Brno, Czech Republic

Schaffers, A.P. (Andre.Schaffers@wur.nl): Department of Environmental Sciences, Nature Conservation and Plant Ecology Group, Wageningen University, PO BOX 47, 6700 AA, Wageningen, The Netherlands

Introduction

Ellenberg indicator values (EIVs; Ellenberg et al. 1992) and their geographic alternatives (e.g. Landolt 1977; Borhidi 1995; Hill et al. 1999; Pignatti 2005; Lawesson et al. 2009) are frequently used by European vegetation scientists as

surrogates for measured environmental variables. Several studies have demonstrated that the calculated mean of EIVs for species present in the vegetation sample are often a good estimate of real environmental conditions, even if this relationship may be limited to certain parts of a given gradient or to a particular vegetation type (for a detailed

Abstract

Question: Mean Ellenberg indicator values (EIVs) inherit information about compositional similarity, because during their calculation species abundances (or presence-absences) are used as weights. Can this *similarity issue* actually be demonstrated, does it bias results of vegetation analyses correlating mean EIVs with other aspects of species composition and how often are biased studies published?

Methods: In order to separate information on compositional similarity possibly present in mean EIVs, a new variable was introduced, calculated as a weighted average of randomized species EIVs. The performance of these *mean randomized EIVs* was compared with that of the *mean real EIVs* on the one hand and *random values* (randomized mean EIVs) on the other. To demonstrate the *similarity issue*, differences between samples were correlated with dissimilarity matrices based on various indices. Next, the three mean EIV variables were tested in canonical correspondence analysis (CCA), detrended correspondence analysis (DCA), analysis of variance (ANOVA) between vegetation clusters, and in regression on species richness. Subsequently, a modified permutation test of significance was proposed, taking the *similarity issue* into account. In addition, an inventory was made of studies published in the *Journal of Vegetation Science* and *Applied Vegetation Science* between 2000 and 2010 likely reporting biased results due to the *similarity issue*.

Results: Using *mean randomized EIVs*, it is shown that compositional similarity is inherited into mean EIVs and most resembles the inter-sample distances in correspondence analysis, which itself is based on iterative weighted averaging. The use of mean EIVs produced biased results in all four analysis types examined: unrealistic (too high) explained variances in CCA, too many significant correlations with ordination axes in DCA, too many significant differences between cluster analysis groups and too high coefficients of determination in regressions on species richness. Modified permutation tests provided ecologically better interpretable results. From 95 studies using Ellenberg indicator values, 36 reported potentially biased results.

Conclusions: No statistical inferences should be made in analyses relating mean EIVs with other variables derived from the species composition as this can produce highly biased results, leading to misinterpretation. Alternatively, a modified permutation test using *mean randomized EIVs* can sometimes be used.

review, see Diekmann 2003). Some issues regarding EIVs have undergone thorough discussion in the scientific literature, such as applicability outside their area of origin (e.g. Gégout & Krizova 2003; Godefroid & Elías 2007), appropriate interpretation in terms of ecological gradients (e.g. Schaffers & Sýkora 2000; Wamelink et al. 2000, 2005) or differences in calculation, such as whether to account for species abundances (e.g. Ewald 2003; Käfer & Witte 2004). Considering the use of EIVs in vegetation analysis, an often recognized problem is the circularity of reasoning when attempting to explain vegetation pattern using EIVs, as these EIVs were themselves derived from empirical experience with vegetation (e.g. Exner et al. 2002). To quantify the risk of this circularity in real analysis is, however, nearly impossible, and one may also argue that the vegetation used as the source of information for the original construction of EIVs usually differs considerably (in space, scale or time) from the vegetation under study, causing the actual effect of this circularity to be low. However, there is yet another aspect of circularity that applies to the use of calculated average species EIVs. The mean for a particular sample is calculated as an average of tabulated species EIVs weighted by species abundance (or presence–absence) data (Ellenberg et al. 1992). This means that when calculated, mean EIVs are derived from two information sources: (1) tabulated species EIVs specifying the ecological behaviour of species along main ecological gradients and (2) the species composition of the vegetation samples for which the mean EIVs are calculated. Due to the first information source, the mean EIVs contain external ecological information, based on thorough expert knowledge about the species ecological requirements (Ellenberg et al. 1992). The use of the second source, however, also has its consequences: calculated mean EIVs inherit information about compositional similarity to other vegetation samples. Hereafter, this will be referred to as the *similarity issue*.

If two vegetation plots have exactly the same species composition, their calculated mean EIVs will indeed be identical. Likewise, mean EIVs of two plots differing in only one or a few species will be very similar. Indeed, compositional similarity corresponds to true ecological similarity between the two plots, but not completely. This is because real communities are structured not only according to the ecological niche theory (Hutchinson 1957), by which the use of mean EIVs is also justified, but also by stochastic processes not related to the ecological conditions of plots (Hubbell 2001). This implies that two plots with identical ecological conditions do not need to have identical species composition and *vice versa*. In case of mean EIVs, however, the more similar the two plots are in terms of species composition, the more similar their calculated mean EIVs will be. This may be illustrated by the apparent relationship between the calculation of mean EIVs and the algorithm of

correspondence analysis (CA; Hill 1973) involving weighted averaging. In CA, the ordination of samples and species along the first ordination axis is derived by iterative repetition of two steps: (1) calculation of sample scores as the abundance weighted average species scores; and (2) calculation of new species scores as the abundance weighted average sample scores calculated in step 1 (the scores are then rescaled and the process is repeated until the scores become stable). If we use species EIVs as the initial species scores, the first step is identical to the calculation of mean EIVs. Here also lies the historical context; Hill (1974) derived CA as an extension of Whittaker's (1960) algorithm of gradient analysis, which was based on the average of estimated species optima along the moisture gradient weighted by species abundances. An interesting property of the CA algorithm is that the result of the iterative process is independent of the choice of the initial species scores (indicator values); these can be set without previous knowledge about species ecology simply as random values.

The main aim of this paper is to quantify the amount of compositional similarity preserved in mean EIVs and to evaluate the influence of this *similarity issue* on analyses that correlate mean EIVs to other aspects of species composition. Examined are four types of analyses commonly employed by vegetation ecologists: constrained ordination analysis, unconstrained ordination analysis, analysis of variance between vegetation clusters and regression on species richness. The performance of proposed modified permutation tests, introduced to solve the problem of biased results, is also tested. Finally, a literature search is performed to estimate how frequently potentially biased results have been published in scientific studies.

Methods

Data sets

Two vegetation data sets were used to illustrate the problem with mean EIVs on real data; the first one large enough to allow for selection of randomly drawn subsamples, and the second containing measured environmental variables for each plot. For the first data set, which will be referred to as the *National database data set*, 1000 forest vegetation samples were randomly selected from the Czech National Phytosociological Database (Chytrý & Rafajová 2003), which was geographically stratified (Knollová et al. 2005). It covers a wide range of forest types and contains plots ranging in size from 100 to 400 m², sampled in the region of the Czech Republic during the last century by various authors using the Braun–Blanquet approach (Westhoff & van der Maarel 1978). The second data set, referred to as the *River valley data set*, consists of 97 forest vegetation plots originating from a local study in the deep valley of the Vltava River near Český Krumlov, Czech

Republic. During the seasons 2001–2003, 150-m² plots evenly distributed along transects following the steep valley slopes were sampled using also the Braun–Blanquet approach. The data set includes measured environmental factors, such as soil pH, aspect, slope, prevailing soil types and soil depth. Given the strong ecological gradients present in the valley, this data set also represents a wide range of ecologically distinct vegetation types (for more detail, see Zelený & Chytrý 2007). In both data sets, bryophytes were excluded, identical species from different layers were merged and data were transformed into presence–absence form. Species Ellenberg values for temperature, continentality, light, moisture, nutrients and soil reaction were taken from Ellenberg et al. (1992).

Randomization of species EIVs

To illustrate the problems caused by the *similarity issue*, the part of the information in mean EIVs originating from external knowledge was separated from the information about compositional similarity. This was achieved by the randomization of species EIVs among the species in the entire table and using these randomized values to calculate mean EIVs (randomization includes also missing values in cases of species without assigned species EIVs). This randomization process removes the external ecological information from mean EIVs but retains the information on compositional similarity inherited in the values. These values can subsequently be used in parallel with mean EIVs calculated from real species EIVs and the difference in performances of these two variables can be attributed to the external ecological information in mean EIVs. Additionally, the performance of both (real and randomized) mean EIVs can be compared to the performance of a random variable, which contains neither ecological nor similarity information. Hence, three calculated variables were used in further analyses: (1) *mean real EIVs*, calculated as a mean of real species EIVs and carrying both external ecological information and information about compositional similarity among plots (Fig. 1a); (2) *mean randomized EIVs*, calculated from species EIVs randomized among species in the table and carrying only information about compositional similarity among plots (Fig. 1b); and (3) *random numbers*, obtained by randomization of *mean randomized EIVs* among plots and carrying no meaningful information at all (Fig. 1c). Note that prior to analysis, species data were transformed into presence–absence data and mean EIVs therefore do not account for species abundances.

Identification of the similarity issue

It could be assumed that the more similar two samples are in terms of their species composition, the more similar they

(a)	EIV-react	Plot 1	Plot 2	Plot 3
Species 1	7	1	0	0
Species 2	8	1	0	0
Species 3	6	1	0	0
Species 4	7	1	1	0
Species 5	5	1	1	0
Species 6	6	0	1	1
Species 7	4	0	1	1
Species 8	5	0	1	1
Species 9	3	0	0	1
Species 10	4	0	0	1
<i>Mean real EIV for soil reaction:</i>	6.6	5.4	4.4	

(b)	EIV-react	Plot 1	Plot 2	Plot 3
Species 1	4	1	0	0
Species 2	6	1	0	0
Species 3	3	1	0	0
Species 4	8	1	1	0
Species 5	5	1	1	0
Species 6	5	0	1	1
Species 7	6	0	1	1
Species 8	7	0	1	1
Species 9	4	0	0	1
Species 10	7	0	0	1
<i>Mean randomized EIV for s. r.:</i>	5.2	6.2	5.8	

(c)	Randomized EIV-react	Plot 1	Plot 2	Plot 3
Species 1	4	1	0	0
Species 2	6	1	0	0
Species 3	3	1	0	0
Species 4	8	1	1	0
Species 5	5	1	1	0
Species 6	5	0	1	1
Species 7	6	0	1	1
Species 8	7	0	1	1
Species 9	4	0	0	1
Species 10	7	0	0	1
<i>Random numbers:</i>		5.8	5.2	6.2

Fig. 1. Schematic representation of the calculation of (a) *mean real EIVs*, (b) *mean randomized EIVs* and (c) *random numbers*, the three alternative variables used in the analyses (see text for details).

will also be in terms of *mean real EIVs* because similarity in species composition reflects similarity in environmental conditions. On the other hand, more similar samples will not have more similar *random numbers*, as these do not carry any information (ecological or otherwise). The question is how is it with *mean randomized EIVs*. These values carry no meaningful external ecological information because the indicator values for individual species were

randomized; so if plots having higher compositional similarity also have more similar *mean randomized EIVs*, this should be attributed to the *similarity issue*.

To investigate this in detail, a data set created of 100 randomly selected vegetation samples from the *National database data set* was used to calculate matrices of compositional dissimilarities among samples based on different distance metrics: (1) Bray–Curtis distance; (2) relative Euclidean (chord) distance; (3) chi-square (χ^2) distance; and (4) sample distance in the space of correspondence analysis (defined by all CA axes). While the first three metrics are commonly used to calculate resemblance matrices among samples, e.g. in cluster analysis, the fourth one was included to investigate how the methodological relationship between the calculation of mean EIVs and the algorithm used in CA (mentioned earlier) project into their correlation. For comparison with these similarity matrices, three matrices were calculated with absolute differences between samples in terms of their: (a) *mean real EIVs* for soil reaction; (b) *mean randomized EIVs* for soil reaction; and (c) *random numbers*. We compared the compositional dissimilarity matrices (1–4) on the one side with the differential matrices (a–c) derived from mean EIVs on the other, using the Mantel statistic for matrix correlation (Legendre & Legendre 1998). All analyses were repeated 100 times, each time on a newly created data set of 100 randomly selected samples and newly randomized EIVs.

Performance of mean EIVs in vegetation analyses

Constrained ordination analysis

To evaluate the bias in performance of mean EIVs as explanatory variables in constrained analysis caused by the *similarity issue*, two analyses were performed: (1) investigation of all six mean EIVs; and (2) comparison of mean EIVs with a measured environmental variable. The ordination method used was canonical correspondence analysis (CCA; ter Braak 1986).

In the first analysis, the explanatory variables in CCA were *mean real EIVs*, *mean randomized EIVs* and *random numbers*, calculated for each of the six Ellenberg indicator values. As a result, $3 \times 6 = 18$ variables were created and used separately as the predictor in a series of 18 CCA analyses. These analyses were performed on a subset of 100 plots, randomly selected from the *National database data set*. To evaluate the stability of the pattern, the whole procedure (i.e. 18 CCA analyses on 100 randomly selected plots) was repeated 100 times, each time using a different subset of 100 randomly selected plots.

In the second analysis, using the *River valley data set*, the variance explained by mean EIVs for soil reaction was compared with that explained by measured soil pH. The explanatory variables used (in turn) were *mean real EIVs*

for soil reaction and its two randomized derivatives (*mean randomized EIVs* for soil reaction and *random numbers*), and the actual and randomized soil pH. The analysis was repeated 100 times, each time using newly randomized variables (*mean randomized EIVs*, *random numbers* and randomized soil pH).

Unconstrained ordination analysis

The *River valley data set* was used to calculate the sample scores along the first three axes of a detrended correspondence analysis (DCA, Hill & Gauch 1980), detrended by 26 segments. These scores were correlated with two variables containing no external ecological information: *mean randomized EIVs* for soil reaction and *random numbers*. Two parameters were studied: the distribution of correlation coefficients (r) and the number of significant ($P < 0.05$) correlations with particular DCA axes. Each correlation was repeated 1000 times, each time with newly generated *mean randomized EIVs* and *random numbers*. A high number of permutations was used to emphasize differences between axes.

Analysis of variance among classification groups

Differences in mean EIVs among groups of samples assembled according to their compositional similarity (e.g. by cluster analysis) can be tested using analysis of variance (ANOVA) or analogous, non-parametric methods. The samples from the *River valley data set* were classified into five groups, using a cluster analysis based on the relative Euclidean (chord) distance and Ward's clustering algorithm (the same procedure was used in Zelený & Chytrý 2007, except that species data were transformed into presence–absence form in the present study). Next, one-way ANOVA was performed on *mean real EIVs* for soil reaction, *mean randomized EIVs* for soil reaction and *random numbers*, respectively, where assignment of samples into groups was used as a fixed factor. These analyses were repeated 100 times, each time with newly generated *mean randomized EIVs* and *random numbers*, and the number of significant results ($P < 0.05$) was counted.

Regression of species richness

The relationship between species richness and mean EIVs representing environmental factors was analysed using regression analysis. A data set of 100 randomly selected vegetation samples from the *National database data set* was used and subjected to linear regression analysis of species richness on *mean real EIVs* for soil reaction, *mean randomized EIVs* for soil reaction and *random numbers*. For each regression analysis, the coefficient of determination (R^2)

was calculated and the significance tested ($P < 0.05$). Each regression was repeated 100 times, each time with a newly selected set of 100 plots and newly generated *mean randomized EIVs* and *random numbers*.

Modified permutation test

Randomization of EIVs among species prior to calculation of the weighted average can be used as the starting point for a modified randomization test, which can be an alternative to the standard tests of significance. Null model permuting species EIVs was recently also implemented by Jansen et al. (2011). A permutation test generally consists of two steps: (1) creation of the null distribution for the test statistic (e.g. correlation coefficient r) that would be expected under the null hypothesis; and (2) comparison of the observed test statistic with this null distribution and estimation of its probability. For a standard permutation test, the null distribution (step 1) is created by calculating the test statistic using randomized data. To account for the *similarity issue*, a modification of this first step is introduced in the sense that the test statistic is calculated not using randomized *mean real EIVs* but using (not randomized) *mean randomized EIVs* instead (i.e. randomizing species EIVs among species in the table instead of randomizing calculated mean EIVs). The performance of this modified permutation test, as compared to the standard one, was evaluated using the *River valley data set* in two analyses: (1) relating mean EIVs to unconstrained ordination axes; and (2) ANOVA of mean EIVs between classification groups (the same five groups as described before).

To relate mean EIVs (or environmental measurements) to ordination axes, the algorithm of Oksanen et al. (2011) was followed, as implemented in the function *envfit* from the R-library *vegan*. Instead of calculating the correlation between an environmental variable and each ordination axis, *envfit* calculates a multiple linear regression of the environmental variable being the dependent variable and site scores on ordination axes being the independent variables. The original function returns a table with normalized regression coefficients, coefficient of determination (R^2) and significance based on the original permutation test. The normalized regression coefficients multiplied by the square root of the coefficient of determination are used to locate the arrowhead of the vector projected onto the ordination diagram. Note that normalized regression coefficients are not correlation coefficients of environmental variables with ordination axes. The coefficient of determination is used as a test statistic, and its null distribution is created by 999 permutations of the environmental variable. The modified permutation test using *mean randomized EIVs* was implemented into the function *envfit.iv* (App. S2). Similarly, the function *summary.aov* using a parametric

ANOVA test was modified into *summary.aov.iv* by implementing the modified permutation test, using an F-value as the test statistic (App. S3). Along with the R scripts of the modified functions, the *River valley data set* is also provided (App. S4).

All analyses and most of the figures were calculated and drawn in the R program, version 2.11 (R Development Core Team 2010, Vienna, Austria).

Literature analysis

To evaluate the frequency of published studies reporting results that are potentially biased due to the *similarity issue*, a literature analysis was conducted. For this, the focus was on studies published between 2000 and 2010 in the *Journal of Vegetation Science* (JVS) and *Applied Vegetation Science* (AVS). The selection of these journals was justified by the fact that in a preliminary search in ISI Web of Knowledge database (<http://apps.isiknowledge.com>) within the subject area of Ecology and Plant Sciences between the years 2000 and 2010, these two journals had the highest number of studies citing the work of Ellenberg et al. (1992) (33 papers in JVS and 18 papers in AVS). Using the full text search engine of the publisher's website (<http://onlinelibrary.wiley.com>), all studies from JVS and AVS between the years 2000 and 2010 containing the word 'Ellenberg' were identified and downloaded. Studies related to mean Ellenberg indicator values were manually selected, and within these, the way in which mean EIVs were used was evaluated. Particularly, the focus was on the use of mean EIVs in constrained ordination, unconstrained ordination, ANOVA between clusters and correlation/regression with species richness. It should be stressed that the aim of the analysis was not to point out 'wrong' studies and to identify the 'sinners' who wrote them, but to document how frequently mean EIVs are used in a way that may yield biased results.

Results

Identification of the similarity issue

For all four distance metrics used, the matrix of compositional dissimilarities correlated best (as expected) with the differences among *mean real EIVs*, while the differences among *random numbers* showed no correlation (Fig. 2). However, differences in *mean randomized EIVs* also correlated to compositional dissimilarity, even though they carry no external ecological information. There were also remarkable differences between the four applied distance metrics: Bray–Curtis and chord distance showed relatively low correlation coefficients compared to χ^2 and CA distances, with CA distance yielding highest correlations.

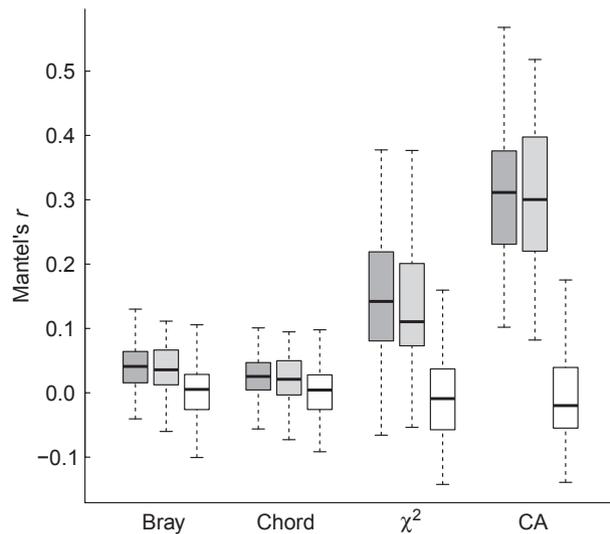


Fig. 2. Mantel correlations between compositional dissimilarity and mean EIVs differences for soil reaction. Presented are the results for: *mean real EIVs* differences (dark-grey boxplots), *mean randomized EIVs* differences (light-grey boxplots) and *random numbers* (white boxplots). The four compositional dissimilarity metrics used were: Bray-Curtis distance (Bray), relative Euclidean distance (Chord), chi-square distance (χ^2) and Euclidean distance between samples in the space of correspondence analysis (CA). Analyses were performed on a random selection of 100 samples from the *National database data set* and were repeated 100 times. Outliers are not shown.

Performance of mean EIVs in vegetation analyses

Constrained ordination analysis (CCA)

In the first CCA analysis (Fig. 3), comparing the variance explained by each of the six mean EIVs with their randomized alternatives, *mean real EIVs* always explained most, while the amounts differed among particular Ellenberg indicator values, with values being highest for moisture and lowest for continentality. In contrast, *mean randomized EIVs* showed variances that were very similar among the six Ellenberg indicator values and these variances were systematically higher than the variances explained by *random numbers*.

In the second CCA analysis, in which the amounts of variance explained by calculated mean EIVs and measured soil pH were compared (Fig. 4), *mean real EIVs* for soil reaction explained more variance (4.4%) than measured soil pH (3.2%), while on average the variances explained by both randomized soil pH and *random numbers* were similar and lower (1.2%). However, comparison of variance explained by *mean real EIVs* (4.4%) and *mean randomized EIVs* (2.4%) revealed that only $4.4 - 2.4 = 2.0\%$ of the variance was related to external ecological information on soil reaction, while $2.4 - 1.2 = 1.2\%$ was caused by the

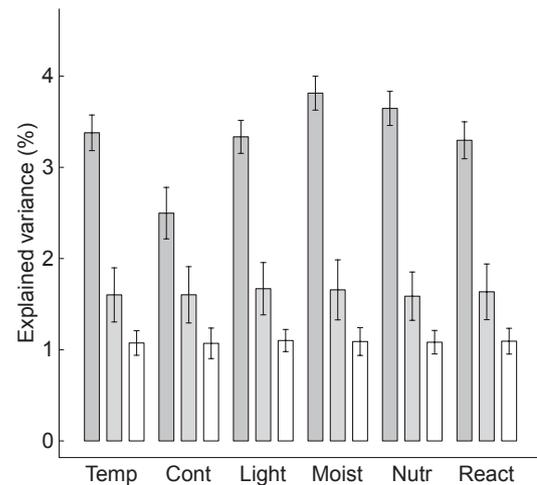


Fig. 3. Variance explained in CCA by *mean real EIVs* (dark-grey columns), *mean randomized EIVs* (light-grey columns) or by *random numbers* (white columns), for all six Ellenberg indicator values. In each case, the average value is calculated from 100 analyses of random subsets selected from the *National database data set*, with error bars indicating standard deviation. Temp, Cont, Light, Moist, Nutr and React, respectively, are abbreviations for mean Ellenberg indicator values for temperature, continentality, light, moisture, nutrients and soil reaction.

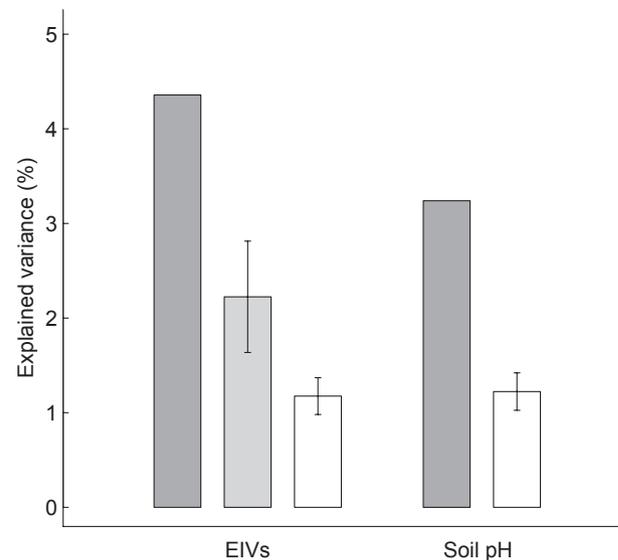


Fig. 4. Variance explained in CCA by mean EIVs for soil reaction and soil pH (measured environmental variable), using the *River valley data set*. Dark-grey bars indicate results for the actual data without any randomization (*mean real EIVs* and real soil pH). White bars indicate variances explained by random variables (*random numbers* and randomized soil pH) and light-grey bar indicates the variance explained by *mean randomized EIVs*. Error bars reflect standard deviations as calculated from 100 new randomizations of the Ellenberg indicator values and pH data.

similarity issue (variance explained by *mean randomized EIVs* minus that explained by *random numbers*). Therefore, even though the *mean real EIVs* explain more variance than

measured soil pH, the part of their explained variance truly related to soil reaction is identical in both cases ($3.2 - 1.2 = 2.0\%$ for measured soil pH and $4.4 - 2.4 = 2.0\%$ for *mean real EIVs*). The fact that they are exactly identical (2.0%) is probably just a coincidence; it can be argued that EIVs for soil reaction reflect content of calcium in the soil more than soil pH (Schaffers & Sýkora 2000).

Unconstrained ordination analysis (DCA)

There were remarkable differences between the performance of *mean randomized EIVs* and *random numbers* when correlated with DCA axes. Correlations with *mean randomized EIVs* (Fig. 5a) showed higher variability of correlation coefficients than correlations with *random numbers* (Fig. 5b). Similar differences could also be found in the number of significant correlations: for *mean randomized EIVs*, 52.4% of 1000 correlations with the first axis were significant, 45.0% of those with the second and 38.6% of those with the third axis. However, for *random numbers*, the proportion of significant results was close to 5% for all

three main axes (4.2%, 4.9% and 4.9%, respectively) corresponding to the 5% significance level used as a threshold.

Analysis of variance among groups

For *mean real EIVs*, the effect of cluster was indeed significant (Fig. 6a). When using *mean randomized EIVs*, 77% of the analyses were still significant at $P < 0.05$ (one example is in Fig. 6b), while for *random numbers*, only 6% were significant. The use of *mean randomized EIVs* thus dramatically increases the probability of getting significant results, even though the values carry no external ecological information.

Regression of species richness

For the 100 random sets of samples, coefficients of determination (R^2) were highest for the regressions of species richness on *mean real EIVs* (Fig. 7) and all 100 were significant at $P < 0.05$. This is in itself not too interesting; the

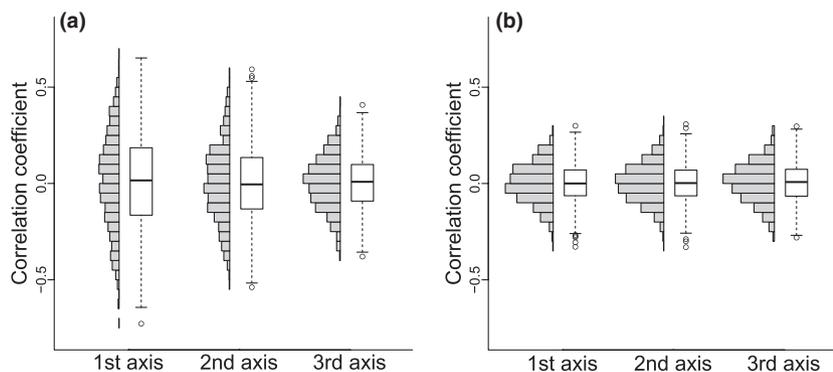


Fig. 5. Correlation coefficients along the first three DCA axes for (a) *mean randomized EIVs* for soil reaction and (b) *random numbers*. Each boxplot with vertical histogram represents the distribution of correlation coefficients from 1000 analyses of the *River valley* data set, each time using newly generated *mean randomized EIVs* and *random numbers*.

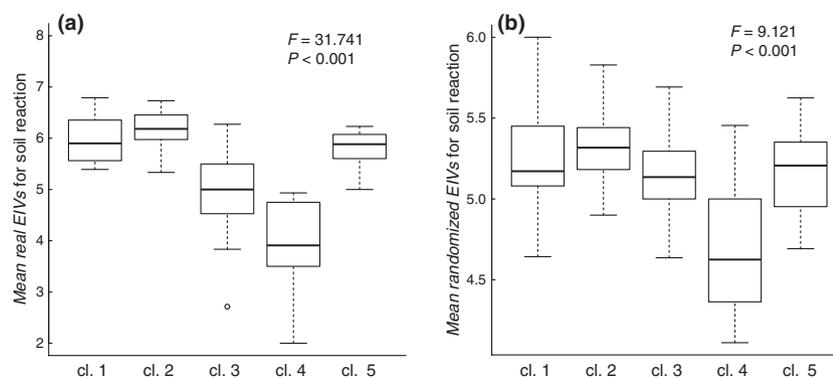


Fig. 6. Differences in site mean EIVs for soil reaction between cluster analysis groups: (a) *mean real EIVs* and (b) *mean randomized EIVs* (one example from the 100 randomizations performed). Even though *mean randomized EIVs* do not contain any ecologically relevant information, the differences among the groups were often significant, as can be seen in the (b) example.

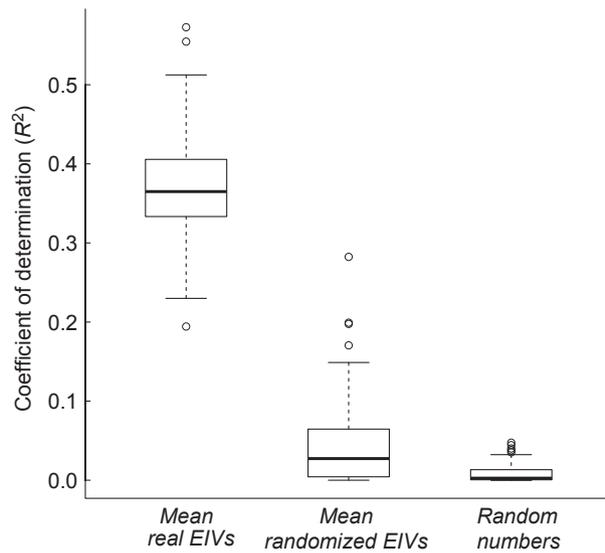


Fig. 7. Coefficients of determination for regression of species richness on mean real EIVs for soil reaction, mean randomized EIVs for soil reaction and random numbers. For each variable, the regression was performed 100 times, each time using a new set of 100 vegetation samples randomly selected from the National database data set.

comparison between mean randomized EIVs and random numbers is more informative. Even though both these variables carry no external ecological information, regressions on mean randomized EIVs often have higher coefficients of determination than regressions on random numbers (Fig. 7), and they are also much more often significant (37% for mean randomized EIVs and 5% for random numbers). This means that even if species richness is regressed on mean EIVs without meaningful external information, the probability of getting a significant result is still high.

Modified permutation test

When the original permutation test was used, all six mean EIVs were significantly related to the first two DCA axes ($P < 0.001$, except continentality with $P < 0.01$; Table 1). With the modified permutation test, however, continentality is no longer significant, and temperature and nutrients are now only significant at $P < 0.05$. In the case of ANOVA of mean EIVs among groups of samples, the parametric test of significance returned all mean EIVs as highly significant (Table 2). Using the modified permutation test, temperature and continentality were no longer significant.

Literature analysis

Ninety-five papers were found to use Ellenberg indicator values, published in JVS and AVS between the years 2000 and 2010 (55 in JVS and 40 in AVS; App. S1). Among

Table 1. Relationship between six mean EIVs and the first two DCA axes – Comparison between the results of the original permutation test and the modified one.

	DCA1	DCA2	R^2	P (orig. test)	P (perm. test)
Light	0.477	0.879	0.600	<0.001	0.004
Temperature	0.350	0.937	0.471	<0.001	0.011
Continentality	0.726	0.688	0.148	0.004	0.452
Moisture	-0.925	0.381	0.897	<0.001	<0.001
Nutrients	-0.998	0.066	0.429	<0.001	0.032
Soil reaction	-0.653	0.757	0.032	<0.001	<0.001

R^2 : goodness of fit (coefficient of determination) from the *envfit* function. P (orig. test) and P (perm. test): significance of the relationship calculated using the original permutation test and modified permutation test, respectively. Significant values ($P < 0.05$) are in bold. DCA1 and DCA2: normalized regression coefficients for the first and second axis of DCA, respectively (see Methods for details).

Table 2. One-way ANOVA for mean EIVs between five groups of samples assembled by cluster analysis.

	F -value	P (param. test)	P (perm. test)
Light	36.242	<0.001	<0.001
Temperature	7.149	<0.001	0.148
Continentality	5.858	<0.001	0.511
Moisture	91.823	<0.001	<0.001
Nutrients	42.982	<0.001	<0.001
Soil reaction	31.741	<0.001	0.014

F -value: the value returned by parametric F -test. P (param. test) and P (perm. test): significance of the analysis, calculated using the parametric F -test and the modified permutation test, respectively. Significant values ($P < 0.05$) are in bold.

them, 36 papers carried analyses potentially biased due to the similarity issue (21 in JVS and 15 in AVS; Table 1 in App. S1). From these, two papers were using mean EIVs as explanatory in CCA, 19 papers correlated mean EIVs with sample scores along ordination axes in DCA or similar (15 of them also tested the significance of this correlation), 16 papers calculated ANOVA among groups of samples and six papers calculated correlation (or regression) of mean EIVs and species richness (all of them tested the significance). In most cases, however, the potential bias was of only minor importance and did not influence interpretation of the results.

Discussion

The results of this study clearly show that information on compositional similarity between samples is inherited into mean EIVs. Differences between mean EIVs are correlated with compositional dissimilarity between samples, even when all external information on species ecology is removed by randomizing the individual species values

before calculating the sample mean (Mantel tests; Fig. 2). The inheritance of compositional information (*similarity issue*) is therefore not caused by the individual Ellenberg species values, but originates in the process of creating a mean value based on the samples' species composition. This is where the problem starts, with clear consequences if this sample mean indicator value is subsequently related to the samples' species composition again (ordination or classification) or to a variable derived thereof (e.g. species richness). This also suggests that even if the sample value is calculated not as the site mean but, for instance, as the mode or median value (Möller 1987), the problem would still persist (although possibly less so) because information is still being used on the samples' species composition. Moreover, the *similarity issue* is not limited only to mean EIVs, but also extends to other variables derived from the species composition using external information about individual species (e.g. proportion or diversity of plant functional traits).

Our study focused in more detail on the performance of mean EIVs in various vegetation analyses. The effect of the *similarity issue* in these analyses has become obvious: stronger relationships and more significant results occur than the ecological information present in the mean EIVs actually allows for.

Concerning the use of mean EIVs as explanatory variables in constrained ordination analysis, we noted that application of this method is rather infrequent in the scientific literature (our analysis revealed two papers), perhaps because the danger of circularity of reasoning (as mentioned in the Introduction) is most obvious here. It appears (Figs 3, 4) that the variance explained by mean EIVs consists of three parts: (1) the part truly originating from the external information on species ecology (the difference between the variance explained by *mean real EIVs* and *mean randomized EIVs*); (2) the part resulting from the *similarity issue* (the difference between the variance explained by *mean randomized EIVs* and *random numbers*); and (3) the part that would be explained by any random variable (the variance explained by *random numbers*). If the focus of the analysis is on the relative differences between variances explained by mean EIVs for various factors (Fig. 3), the results can be meaningful, although the analysis itself is not too appealing. However, if the focus is on the comparison of mean EIVs with measured variables on the same ecological factor (as in the case of mean EIVs for soil reaction and soil pH; Fig. 4), mean EIVs will tend to be a better predictor and the extent of this advantage is given by the part of the explained variance resulting from the *similarity issue*. This may partly explain why mean EIVs are often better predictors than measured environmental variables (Dupré & Diekmann 1998; Smart et al. 2010).

In contrast to constrained ordination, use of mean EIVs to interpret unconstrained ordination is very common (19 papers). From the circularity of reasoning point of view, 'indirect' correlation of mean EIVs with sample scores along unconstrained ordination axes is generally perceived as more correct than 'direct' use of mean EIVs as explanatory variables in constrained ordination. However, this impression is wrong. The sorting of samples along axes reflects compositional similarities (similar samples are closer), and the *similarity issue* may cause a situation where even mean EIVs with no external ecological information (e.g. *mean randomized EIVs*) can have correlation coefficients with ordination axes that are spuriously high (Fig. 5a) and with a high probability of being significant. We might expect mean EIVs that contain no ecologically relevant information to show correlations around zero, as can indeed be seen in Fig. 5b using truly random values. But although *mean randomized EIVs* do not contain actual information on their own, differences between them do. If, for instance, during the randomization many of the species in a sample have by chance received high values (leading to high mean EIVs), samples with similar species composition will also tend to have high values. If these samples are located at the right-hand side of the ordination axis this will lead to a positive correlation with the first DCA axis scores. But if (new randomization) many species from the original sample have received low values by chance, similar samples (close by on the right-hand side of the ordination) will also tend to have low mean EIVs, leading to a negative correlation with the axis score this time. This illustrates that a tendency exists towards stronger correlations (away from zero; either positive or negative) and explains the wide distribution of correlation coefficients in Fig. 5a when compared to the expected distribution in Fig. 5b. This tendency towards stronger correlations does, of course, exist not only when species values are randomly assigned. Information on similarity is inherited in exactly the same way if the assigned species values do have actual meaning (as is the case for true EIVs). Moreover, ordination axes in DCA (and other eigenvalue-based ordinations) are ordered according to decreasing eigenvalues, which correspond to the variance in species composition explained by a particular axis and consequently also to the similarity among samples retained by the sample scores. This is why the problem with biased correlation coefficients and significances increases with the importance of ordination axes, usually being strongest for the first one. As a result, if we project *post hoc* mean EIVs as vectors onto an ordination diagram, they will artificially tend to be more correlated with the ordination axes of higher importance (e.g. more with the first than the second axis). Also, their vectors will tend to be longer than vectors for measured environmental factors due to the

tendency towards higher correlation coefficients explained above.

Comparison of mean EIVs among groups of samples using ANOVA or *t*-test (or analogous non-parametric analyses, e.g. Mann–Whitney *U* test, Kruskal–Wallis test) is fairly common (16 papers). In the literature analysis, three types could be distinguished, differing in how the groups of samples were assembled: (a) according to the species composition (vegetation type), (b) according to the experimental treatment or management and (c) according to the year of sampling. All three types can theoretically yield biased results if used with mean EIVs. This is most obvious for case (a), but also in cases (b) and (c) the groups may show internal similarity, e.g. if an experimental treatment was applied for a period long enough to cause convergence of the species composition (type b) or if different plots experience similar compositional changes in time (type c). In our study, only the first type was tested, i.e. groups assembled according to vegetation type (based on numerical classification). If analysed by ANOVA, differences and significances tended to be biased to more optimistic values. Even if mean EIVs contained no external ecological information (as in the case of *mean randomized EIVs* in Fig. 6b), differences among groups were often highly significant. To what degree the *similarity issue* biases ANOVA results mainly depends on the distance measure chosen for clustering, because particular distances differ in how they reflect the similarity inherited by mean EIVs (e.g. Bray–Curtis dissimilarity reflects this much less than chi-square distance; Fig. 2). An extreme case is probably TWINSPAN (Hill 1979), which assembles the samples into groups according to their position along the main ordination axis of CA. As mentioned earlier, the weighted averaging algorithm of mean EIVs is closely related to the algorithm of CA, which means that the similarity among samples assembled into groups and the similarity inherited into mean EIVs are of a similar kind (see Fig. 2).

The relationship between mean EIVs and species richness is also occasionally subject to statistical inference (six papers). The reason why the differences in species richness would reflect compositional similarity among samples is not necessarily trivial. A possible explanation is that local species richness (number of species in a sample) is influenced by the size of the species pool (Zobel 1997), and the size of the species pool varies among vegetation types (or habitats; e.g. Sádlo et al. 2007). Thus, for a data set covering a wide range of environmental factors and/or including different vegetation types, similar plots will be of similar vegetation type, they will have a similar size of the species pool and hence also similar species richness. This also means that the strength of the relationship between species richness and compositional similarity among plots is

specific for each data set and perhaps increases with the data set's compositional heterogeneity. The *National database data set*, which was used for the analysis (Fig. 7), is relatively heterogeneous, covering wide range of ecologically very different forest vegetation types. This is perhaps why a relatively high number of regressions (37%) between *mean randomized EIVs* and species richness was significant.

Other methods that may be influenced by the *similarity issue* but were not analysed in this paper are classification and regression trees (CART; Breiman et al. 1984), which have recently experienced a boom in ecological literature (e.g. De'ath & Fabricius 2000). The problem of biased performance of mean EIVs as predictors may occur when the dependent variable is derived from the species composition, which is the case for species richness in regression trees or the grouping of samples into groups in classification trees. Preliminary trials with these methods show that if mean EIVs are used together with other (measured) environmental variables, mean EIVs perform better, being placed more close to the root of the tree. Such biased performance may also be a result of the *similarity issue*, which favours mean EIVs against the other (measured) factors. If only mean EIVs are used as predictors, this problem does not occur.

Our study does not of course prove that mean EIVs will *always* perform better than measured variables, only that a tendency exists for mean EIVs to outperform actual measurements. Nevertheless, there may also exist true ecological reasons for mean EIVs to perform better than measurements. EIVs could be considered a form of bio-indication, and the plants used as indicators can be supposed to 'measure' exactly what is ecologically relevant. Actual measurements on the other hand depend on arbitrary choices (depth of soil sampling, elements measured, chemicals used for extraction, etc.). In addition, plants integrate over all relevant factors (e.g. soil moisture content, water tension and air humidity, different nutrient forms, etc.). They also integrate over time, while actual measurements often only provide snapshots. EIVs may thus be truly superior to measured variables in many cases. But in order to settle such issues we need unbiased comparisons accounting for the *similarity issue*.

Modified permutation tests may be a partial solution for analyses where a test of significance is required but standard tests yield biased results due to the *similarity issue*. The modified test changes the null hypothesis from 'there is no statistical relationship between variable X and mean EIVs' into 'variable X is not related to the information based on external ecological data in mean EIVs', i.e. accounting for the relationship caused by the *similarity issue*. Its result makes more ecological sense. For example, given the small spatial scale of the *River valley data set*, the originally significant effect of continentality when regressed on DCA axes

(Table 1) or when tested for differences among vegetation types (Table 2) was difficult to interpret. After the modified permutation test was applied, it appeared that continentality (and also other factors) should not have been considered significant. In the past, significant relationships for unlikely Ellenberg factors were sometimes explained by their interrelated nature. For example, Chytrý et al. (2009) used this argument to explain why the plots in a fertilization experiment significantly differed in mean EIVs for continentality, even if this was not experimentally manipulated.

Practical considerations

Ellenberg indicator values are a highly valuable tool for the interpretation of vegetation patterns. To ensure that such interpretations are free of bias, several practical recommendations for the use of mean EIVs in analyses of vegetation data can be based on the results of this study.

1. In constrained ordination analysis (e.g. RDA and CCA), mean EIVs should not be used as explanatory variables together with measured environmental variables because mean EIVs will tend to explain more variance than the measured variables.
2. In unconstrained ordination (e.g. DCA), correlations with ordination axes offer a quick and simple interpretation aid, but should not be tested because the probability of obtaining significant results is spuriously high. Furthermore, also here it should be taken into consideration that mean EIVs will tend to show higher correlations than measured environmental variables. Preferably, separate diagrams should be made for mean EIVs and measured environmental factors. Also, it should be kept in mind that mean EIVs tend to be correlated more strongly to ordination axes of higher importance.
3. Among clusters derived from species composition, differences in mean EIVs should not be tested (e.g. by ANOVA, *t*-test or non-parametric alternatives). A modified permutation test, as proposed in this study, may be used instead.
4. Variables derived from the species composition (e.g. species richness and the proportion of species traits) should not be correlated to mean EIVs unless the modified permutation test is used.
5. In classification and regression trees with dependent variables directly derived from species composition (e.g. species richness in the case of regression tree or cluster assignment according to similarity in the case of classification tree), mean EIVs and measured environmental variables should not be used as possible predictors at the same time. Mean EIVs will tend to be better predictors than measured environmental factors, occurring at higher levels

of the tree hierarchy. Additionally, the variance explained by such a tree will be too optimistic.

Conclusion

Three main points can be made from this study:

1. Mean EIVs always inherit information on compositional similarity.
2. Using mean EIVs in vegetation analyses (or in other analyses utilizing aspects of species composition) leads to stronger relationships and more significant results than is actually warranted by the external ecological information present in these values.
3. Mean EIVs will have an inappropriate advantage in vegetation analyses compared to actual environmental measurements.

Acknowledgements

We would like to thank colleagues from the Department of Botany and Zoology, Masaryk University Brno, for their discussion, namely Milan Chytrý, Ching-Feng Li, Lubomír Tichý and Jan Roleček. The ideas were also improved after discussions with several participants of the 52th Symposium of the International Association of Vegetation Science held in Jun 2009 in Crete. Our thanks go also to Jörg Ewald, Martin Diekmann and Mike Palmer for valuable comments during the review process. D.Z. was supported by the Ministry of Education of the Czech Republic (MSM 0021622416) and by the Czech Science Foundation (505/11/0732).

References

- Borhidi, A. 1995. Social behaviour types, the naturalness and relative ecological indicator values of the higher plants in the Hungarian flora. *Acta Botanica Hungarica* 39: 97–181.
- Breiman, L., Friedman, J.H., Olshen, R.A. & Stone, C.G. 1984. *Classification and regression trees*. Wadsworth International Group, Belmont, CA, US.
- Chytrý, M. & Rafajová, M. 2003. Czech National Phytosociological Database: basic statistics of the available vegetation-plot data. *Preslia* 75: 1–15.
- Chytrý, M., Hejcman, M., Hennekens, S.M. & Schellberg, J. 2009. Changes in vegetation types and Ellenberg indicator values after 65 years of fertilizer application in the Rengen Grassland Experiment, Germany. *Applied Vegetation Science* 12: 167–176.
- De'ath, G. & Fabricius, K.E. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* 81: 3178–3192.

- Diekmann, M. 2003. Species indicator values as an important tool in applied plant ecology – a review. *Basic and Applied Ecology* 4: 493–506.
- Dupré, C. & Diekmann, M. 1998. Prediction of occurrence of vascular plants in deciduous forests of south Sweden by means of Ellenberg indicator values. *Applied Vegetation Science* 1: 139–150.
- Ellenberg, H., Weber, H.E., Dull, R., Wirth, V., Werner, W. & Paulissen, D. 1992. Zeigerwerte von Pflanzen in Mitteleuropa. *Scripta Geobotanica* 18: 1–248.
- Ewald, J. 2003. The sensitivity of Ellenberg indicator values to the completeness of vegetation relevés. *Basic and Applied Ecology* 4: 507–513.
- Exner, A., Willner, W. & Grabherr, G. 2002. *Picea abies* and *Abies alba* forests of the Austrian Alps: numerical classification and ordination. *Folia Geobotanica* 37: 383–402.
- Gégout, J.C. & Krizova, E. 2003. Comparison of indicator values of forest understory plant species in Western Carpathians (Slovakia) and Vosges Mountains (France). *Forest Ecology and Management* 182: 1–11.
- Godefroid, S. & Elías, D.D. 2007. Can Ellenberg's indicator values for Mediterranean plants be used outside their region of definition? *Journal of Biogeography* 34: 62–68.
- Hill, M.O. 1973. Reciprocal averaging: an eigenvector method of ordination. *Journal of Ecology* 61: 237–249.
- Hill, M.O. 1974. Correspondence analysis: a neglected multivariate method. *Applied Statistics* 23: 340–354.
- Hill, M.O. 1979. *TWINSPAN – A FORTRAN program for arranging multivariate data in an ordered two-way table by classification of the individuals and attributes*. Ecology and Systematics, Cornell University, Ithaca, NY, US.
- Hill, M.O. & Gauch, J.G. 1980. Detrended correspondence analysis: an improved ordination technique. *Vegetatio* 42: 47–58.
- Hill, M.O., Mountford, J.O., Roy, D.B. & Bunce, R.G.H. 1999. *Ellenberg's indicator values for British plants*. Institute of Terrestrial Ecology, Huntingdon, UK.
- Hubbell, S.P. 2001. *The unified neutral theory of biodiversity and biogeography*. Princeton University Press, Princeton, NJ, US.
- Hutchinson, G.E. 1957. Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology* 22: 415–427.
- Jansen, F., Ewald, J. & Zerbe, S. 2011. Ecological preferences of alien plant species in north-eastern Germany. *Biological Invasions* 13: 2691–2701.
- Käfer, J. & Witte, J.P.M. 2004. Cover-weighted averaging of indicator values in vegetation analyses. *Journal of Vegetation Science* 15: 647–652.
- Knollová, I., Chytrý, M., Tichý, L. & Hájek, O. 2005. Stratified re-sampling of phytosociological databases: some strategies for obtaining more representative data sets for classification studies. *Journal of Vegetation Science* 16: 479–486.
- Landolt, E. 1977. Ökologische Zeigerwerte zur Schweizer Flora. *Veröffentlichungen des Geobotanischen Institutes der Eidgenössischen Technischen Hochschule, Stiftung Rübel, Zürich* 64: 1–208.
- Lawesson, J.E., Fosaa, A.M. & Olsen, E. 2009. Calibration of Ellenberg indicator values for Faroe islands. *Applied Vegetation Science* 6: 53–62.
- Legendre, P. & Legendre, L. 1998. *Numerical ecology*. 2nd English ed Elsevier, Amsterdam, NL.
- Möller, H. 1987. Wege zur Ansprache der aktuellen Bodenazidität auf der Basis der Reaktionszahlen von Ellenberg ohne arithmetisches Mitteln dieser Werte. *Tuexenia* 7: 499–505.
- Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H. & Wagner, H. 2011. *Vegan: Community Ecology Package*. R package version 1.17-6. Available at: <http://cran.r-project.org/web/packages/vegan/index.html>.
- Pignatti, S. 2005. Valori di bioindicazione delle piante vascolari della flora d'Italia. *Braun-Blanquetia* 39: 1–97.
- Sádlo, J., Chytrý, M. & Pyšek, P. 2007. Regional species pools of vascular plants in habitats of the Czech Republic. *Preslia* 79: 303–321.
- Schaffers, A.P. & Sýkora, K.V. 2000. Reliability of Ellenberg indicator values for moisture, nitrogen and soil reaction: comparison with field measurements. *Journal of Vegetation Science* 11: 225–244.
- Smart, S.M., Scott, W.A., Whitaker, J., Hill, M.O., Roy, D.B., Critchley, C.N., Marini, L., Evans, C., Emmett, B.A., Rowe, E.C., Crowe, A., Le Duc, M. & Marrs, R.H. 2010. Empirical realised niche models for British higher and lower plants – development and preliminary testing. *Journal of Vegetation Science* 21: 643–656.
- ter Braak, C.J.F. 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology* 67: 1167–1179.
- Wamelink, G.W.W., Joosten, V., van Dobben, H.F. & Berendse, F. 2000. Reliability of Ellenberg indicator values for moisture, nitrogen and soil reaction: a comparison with field measurements. *Journal of Vegetation Science* 11: 225–244.
- Wamelink, G.W.W., Goedhart, P.W., van Dobben, H.F. & Berendse, F. 2005. Plant species as predictors of soil pH: replacing expert judgment with measurements. *Journal of Vegetation Science* 16: 461–470.
- Westhoff, V. & van der Maarel, E. 1978. The Braun-Blanquet approach. In: Whittaker, R.H., ed *Classification of plant communities*, pp. 287–399. Junk, The Hague, NL.
- Whittaker, R.H. 1960. Vegetation of the Siskiyou mountains, Oregon and California. *Ecological Monographs* 30: 279–338.
- Zelený, D. & Chytrý, M. 2007. Environmental control of vegetation pattern in deep river valleys of the Bohemian Massif. *Preslia* 79: 205–222.
- Zobel, M. 1997. The relative of species pools in determining plant species richness: an alternative explanation of species coexistence? *Trends in Ecology and Evolution* 12: 266–269.

Supporting Information

Additional supporting information may be found in the online version of this article:

Appendix S1. Review of papers published in *Journal of Vegetation Science* and *Applied Vegetation Science* from 2000 to 2010, which report analysis results using mean EIVs potentially biased due to the similarity issue.

Appendix S2. R function `envfit.iv` with modified test of significance of relationship between mean EIVs and sample scores along axes of unconstrained ordinations.

Appendix S3. R function `summary.aov.iv` calculating one-way ANOVA among groups, using modified permutation test.

Appendix S4. River valley data set, containing two files: `vltava-spe.csv` with matrix of sample species data, and `vltava-spec.eiv.csv` with matrix of species Ellenberg indicator values. The data set is intended for use with `envfit.iv` and `summary.aov.iv` functions.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.