Ecological Indicators 36 (2014) 441–446

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Original article

Using statistical tests on relative ecological indicator values to compare vegetation units – Different approaches and weighting methods

Csaba Tölgyesi*, Zoltán Bátori, László Erdős

Department of Ecology, University of Szeged, 52 Közép fasor, H-6726 Szeged, Hungary

ARTICLE INFO

Article history:
Received 14 December 2012
Received in revised form 31 July 2013
Accepted 2 September 2013

Keywords:
Relative ecological indicator
Weighted average
ANOVA test
Kruskal–Wallis test

ABSTRACT

Relative ecological indicators are frequently used tools in vegetation analyses. Despite their ordinal nature, it has been shown that average indicator values can characterize an area well, and can provide useful ecological information. Several different averaging methods have been tested against the indicated environmental parameters, but only very slight differences could be found between their reliability. Different statistical tests, including parametric and non-parametric tests, are also often applied on relative ecological indicators. Similarly to the weighting methods, there are several ways to provide source data for the tests from raw indicator values but the possible differences in the reliability of the resulting statistical layouts have never been looked at. In the present study we have chosen the Hungarian adaptation of Ellenberg's indicator for soil moisture as a model system and examined a total of 8 different statistical layouts. Raw indicator values were obtained from vegetation surveys of 16 appropriately chosen sites and were processed in two fundamentally different ways. In the first approach, average indicator values were calculated for each sampling quadrat of the sites and these averages were used as source data for ANOVA tests. The calculation of the averages was carried out in four different ways according to the weighting methods. In the second approach, site specific species lists were compiled using the quadrats of each site and the raw indicator value populations deriving from these lists were analyzed with Kruskal–Wallis tests. Again, four weighting methods were used, but instead of averaging, the indicator value of each species within a site was repeated as many times as its weight required. Finally, the reliability of each method was assessed by comparing the results with the actual soil moisture relations of the sites, determined with physical measurements. According to our results, it can be said that false positive results are rare with any type of the methods but the amount of false negative results varied among the methods considerably. The most reliable method was the Kruskal–Wallis test when performed on frequency weighted raw indicator value populations. This method could best reproduce the original soil moisture relations and could yield the most convincing *p*-values; therefore we can recommend using this method in studies where sets of relative ecological indicator values are intended to be compared with statistical tests.

1. Introduction

Relative ecological indicators express the realized optimum of plant species on ordinal scales defined along environmental gradients (Ewald, 2003). Originally, the system was developed for the flora of Central Europe by Heinz Ellenberg and included the following 7 environmental factors: soil moisture, soil acidity, productivity/nutrients, continentality, soil salt content, temperature and light (Ellenberg, 1952; Ellenberg et al., 1992). The system has been adapted to several regions outside its first definition and has become a wide-spread tool of applied plant ecology, forestry and agriculture (Borhidi, 1993; Diekmann, 2003; Dzwonko, 2001).

The most common applications of relative ecological indicators are to compare the habitat conditions of two or more different areas or to monitor the changes of the vegetation of a permanent plot (Diekmann, 2003; ter Braak and Wiertz, 1994; Tölgyesi and Körmöczi, 2012). Comparing relative indicator values, however, has its difficulties. Owing to the ordinal nature of their scales, several statistical operations cannot be applied to them without further considerations. Möller (1992) recommends the median values of the sites as statistically sound tools for comparisons but several studies have shown that mean indicator values characterize an area well and they can provide useful ecological information (Lengyel et al., 2012; ter Braak and Barendregt, 1986; ter Braak

* Corresponding author. Tel.: +36 304860363.
E-mail address: festuca7@yahoo.com (C. Tölgyesi).

1470-160X/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ecolind.2013.09.002

and Gremmen, 1987). According to Ellenberg et al. (1992) there are three basic ways to calculate mean ecological indicator values. (i) The qualitative method uses only the presence/absence data of the species and results in unweighted averages. (ii) The quantitative method uses the percent cover values of the species as weights and results in weighted averages. (iii) The ordinal method also results in weighted averages but the weights are developed by projecting the percent cover values to an ordinal scale. For example Allen (1992) recommends a 6-grade scale, while van der Maarel (1979) uses a 10-grade scale. There have also been proposals for abundance-independent weighting methods to improve the accuracy of the average values. Schaffers and Sykora (2000) called attention to the general phenomenon that the frequency distributions of indicator values are rather uneven, which creates a tendency for mean values to converge to the value most common in the regional species pool. In practice, this means that the more extreme a value is, the less species belong to it in the flora. Therefore, supplying every indicator value with a weight that appropriately downweights common values and upweights rare values can prevent the average value of an extreme habitat from shifting toward intermediate values.

Surprisingly, apart from some special cases, the correlation between average indicator values and the values of the indicated environmental parameters do not change significantly with any type of the main two weighting methods compared with the unweighted one (Diekmann, 2003; Käfer and Witte, 2004; Klaus et al., 2012). The abundance independent weighting method does not improve the correlation considerably, though it has some beneficial effects such as improving the linearity of mean values along the gradient of the indicated environmental parameter. Therefore, Schaffers and Sykora (2000) recommend the use of this weighting as a standard method, especially when quantitative statements about environmental conditions are to be made.

As it can be seen, the accuracy of different averaging methods is well-studied, but according to our knowledge no study has ever been conducted to examine the reliability of the different statistical layouts used on Ellenberg indicator values. Such tests, however, are widely used in applied vegetation science (Zeleny and Schaffers, 2012). In the literature one can find examples for the use of parametric tests like the *t*-test and the ANOVA test (e.g. Spiegelberger et al., 2006), as well as non-parametric tests like the Mann–Whitney test and the Kruskal–Wallis test (e.g. Zwaenepoel et al., 2006), and in some cases mean indicator values are weighted with species abundance measures (e.g. Roovers et al., 2005) but in other cases they are not (e.g. van Dobben et al., 1999).

In the present study we have chosen a relative ecological indicator, Borhidi's indicator for soil moisture (*F* value), which is the adaptation of Ellenberg's indicator for soil moisture to the Hungarian flora (Borhidi, 1995), and aimed to investigate whether there are differences in the efficiency of different statistical layouts and tried to find the most reliable one for comparing vegetation units. For this purpose we selected 16 appropriately chosen study sites and examined, which statistical layout can best reproduce their humidity relations, previously determined with physical measurements.

2. Materials and methods

2.1. Study sites

The study was carried out on the lowlands of Central Hungary, in the Kiskunság National Park. Considering the purposes, study sites were needed with different water supplies but otherwise with environmental conditions as similar as possible. The sites had to be relatively close to each other to ensure synchronized water supply fluctuations, thus eliminating the need for multiple soil moisture measurements. Areas under severe human influence had to be avoided as it may cause competitive release, making the original indicator values of certain species less usable (Kowarik and Seidling, 1989). The presence of severe disturbance – natural or anthropogenic – would have also been disadvantageous because the vegetation of such areas does not primarily indicate specific soil conditions but reflect the disturbance regime (Briemle, 1997). Using average indicator values on heterogeneous plots can result in misleading results (Diekmann, 2003), therefore special attention was paid to choose study sites with as homogeneous vegetation as possible. To ensure differential water supplies, the sites were chosen so that they were located on different elevations.

Considering the above criteria, eight study sites were chosen in a sand dune range, called Fülöpháza Sand Dunes. Four of them were in hilltop position (dry dune sites, DD1-4, 109–111 m a.s.l.) and four in dune slacks (wet dune sites, WD1-4, 100–101 m a.s.l.). DD sites are covered with sparse xeric vegetation, since their only water source is falling precipitation and their soil has a very poor water holding capacity. The vegetation of the WD sites is denser and taller since they receive some extra water from the adjacent sandhills in the form of leaking moisture at thaw and after rain, and, in addition, they are less exposed to the drying effect of the wind. The water table, however, is still several meters below their deepest points. For a detailed description of the vegetation and the environmental conditions of the Fülöpháza Sand Dunes see Molnár (2003). A set of eight other study sites were chosen in the adjacent Turjánvidék, which is a mosaic of low-lying (92–93 m a.s.l.) wetland and steppe patches. The water table at the wetland sites (wet mosaic sites, WM1-4) is close to the soil surface and the vegetation can be characterized with tall sedge and grass species. The steppe patches (dry mosaic sites, DM1-4) were located 0.5–1.0 m higher than the WM sites and were apparently dryer habitats with shorter vegetation rich in herbaceous plants. More information on the vegetation and the environmental conditions of the Turjánvidék is given by Biró et al. (2007) and Járai-Komlódi (1958).

2.2. Data collection

All field samplings and surveys were carried out in late May 2012. Seven random soil samples were taken from every study site for soil moisture measurements (a total of 112 samples). After removing the litter layer, cylindrical cores were collected from the upper 20 cm of the soil. The cores were analyzed at the University of Szeged, Hungary. The mass of the cores was measured with gravimetry, then they were baked at 90 °C for 2 days and the remaining dry matter was measured again. The difference was the water content, which was then expressed in percents of the original mass. No rain had fallen within 10 days before the samplings, so the soil moisture contents reflected real microclimatic conditions. Vegetation surveys were carried out on 5 (DM and WM sites) or 7 (DD and WD sites) randomly chosen 2 m × 2 m quadrats (a total of 96 quadrats). The DD and WD sites seemed to have less homogeneous vegetation than the WM and WD sites; this is why the larger number of quadrats. Every species in the quadrats was identified and their percent cover was also assessed.

2.3. Data analysis

Average relative soil moisture contents were analyzed with ANOVA, which was followed by Tukey's pairwise comparisons. First, the four groups were tested for within-group differences, which meant four separate tests. If a test detected significant differences, the group was split accordingly and the new groups were used in all subsequent analyses. As a second step, between-group differences were tested to see if the original assumption for the soil

Table 1
The transformation of percent cover values to Allen's ordinal scale.

Cover in percent		Ordinal score
<1%	→	1
1–5%	→	2
6–25%	→	3
26–50%	→	4
51–75%	→	5
76–100%	→	6

moisture regimes of the four habitats was right. In those between-group tests where one of the original groups had to be split due to within-group differences, only that one was included in the test, which was closer to the other group along the soil moisture gradient. As a result, the original 4×4 arrangement of the study sites have been refined, with proved significant differences between neighboring groups.

Indicator value populations of the study sites were treated in two fundamentally different ways. In one set of analyses, hereafter called the first approach, mean indicator values were calculated for every quadrat of every site and after testing them for normality with Shapiro–Wilk tests, they were further analyzed with ANOVA (followed by and Tukey's pairwise comparisons) or two sample *t*-tests, depending on the number sites compared. In the other set of analyses, hereafter called the second approach, a species list was compiled for every study site, based on the occurrences in the quadrates. The raw indicator value populations, developed using these species lists, were then used to characterize the study sites. Thus, the second approach did not require the averaging of indicator values. Since raw indicator value distributions are usually skewed and cannot be transformed due to their ordinal nature, we proceeded to Kruskal–Wallis tests without testing normality. When Kruskal–Wallis tests found significant differences, pairwise Mann–Whitney tests with Bonferroni corrections were performed. The number of tests considered for the corrections was the number of Mann–Whitney tests carried out after each Kruskal–Wallis analysis. When only two sites had to be compared, we applied the Mann–Whitney test without corrections.

During the first approach, analyses were performed on both weighted and unweighted averages. Three types of weights were used with the first approach: (i) raw cover values in percents, (ii) ordinal cover values according to Allen (1992) (Table 1.) and (iii) abundance-independent weights according to Schaffers and Sykora (2000). For the calculation of the abundance-independent weights the following formula was used: $w_k = N/(12 \times n_k)$, with w_k, the weight of the kth F indicator value; N, the total number of species in the Hungarian flora having F values; n_k, the number of species having the kth F value and 12 is the size of the F indicator scale. Abundance-independent weightings were used only alone, no double weightings were performed.

Cover and ordinally weighted averages were calculated for the quadrats of a site with the following general formula: $A_j = \sum_{i=1}^{n}(w_{ij} \times x_i)/\sum_{i=1}^{n} w_{ij}$, with A_j, the average indicator value of quadrat j; w_{ij}, the weight of species i in quadrat j; x_i, the F value of species i; n, the total number of species in the quadrat. The formula for the averages with abundance-independent weightings was similar to the previous one but the weight was defined as follows: $A_j = \sum_{i=1}^{n}(w_{xi} \times x_i)/\sum_{i=1}^{n} w_{xi}$, with w_{xi}, the weight of the indicator value of species i.

Since the Kruskal–Wallis test is a rank test, weightings during the second approach were applied so that the indicator value of a species was repeated in the dataset as many times as its weight required. This also meant that all weights had to be integers. Weightings were carried out with (i) frequency values, with (ii) ordinally transformed average cover values according to Allen (1992) and with (iii) abundance-independent weights. Frequency values were given as the number of quadrats where each species occurred. Average percent cover values included fractions smaller than 0.1, and using a factor to transform them to a positive integer would have resulted in extremely large weights for more abundant species and therefore extremely high increases in the sample size. We considered this effect unrealistic; therefore percent cover values were not used as weights for Kruskal–Wallis tests. To transform abundance-independent weights into integers, all original weights were divided with $min(w_i)$ and were then rounded. Thus, the weight of the most common F values was 1, while the others had bigger integer weights.

In both approaches, the course of the tests performed on the indicator values followed a strict design. Within-group tests were carried out first and they were followed by between-group tests. The between-group tests were carried out on groups next to each other along the humidity gradient. The refined groups were used during the analyses; therefore the number and arrangement of these "basic" comparisons was the same in all the eight statistical layouts. The fact that every refined group differs significantly from their neighbors along the soil humidity gradient implies that they differ from all the other groups as well. However, if a statistical layout cannot find significant difference between two groups, the above statement can no longer be said about these groups. Therefore, in such cases a number of extra tests were carried out to see the extent of the weakness of the statistical layout. The following example illustrates the course of these extra tests: If, based on soil moisture analyses, groups G_1, G_2, G_3, G_4, etc. were known to be along the humidity gradient in this order, but a test could not find significant difference between G_2 and G_3, we tested G_3 against G_1, and G_2 against G_4 as well and so on.

The number of deviations from the expectations (i.e. the number of undetected significant differences and the number of newly found significant differences) was used to evaluate the goodness of the statistical layouts.

Statistical analyses were performed with SPSS 11.5 (SPSS Inc.). The allocation of indicator values to the species was carried out with the SynData software (Horváth, 2006). The level of significance was $p = 0.05$ in all cases.

3. Results

3.1. Soil moisture measurements

The WM3 and WM4 cores were not analyzed for moisture content because they were so saturated that a considerable amount of water flew out from them while they were being removed from the soil. The water table in these sites was only 5–10 cm below the surface and these were considered the wettest sites. Since we did not know whether the WM3 and WM4 sites had different moisture contents, we handled them as if they belonged to the same group. In the case of the other sites, samplings were successful without loss of water and the cores were suitable for measurements. No significant differences were found within the DD and WD sites and between the WM1 and WM2 sites. The DM sites were not uniform but they had the following soil moisture relations: DM1 < DM2 < DM3 = DM4. To elucidate the between-group relations, first the DD sites were compared with the WD sites, and all DD sites were found drier than the WD sites. The WD sites were compared with the DM1 site as well, and all WD sites proved drier. Finally the DM3 and DM4 sites were tested against the WM1 and WM2 sites, and the difference was significant (Fig. 1). Thus, we received seven refined groups along the humidity gradient in the

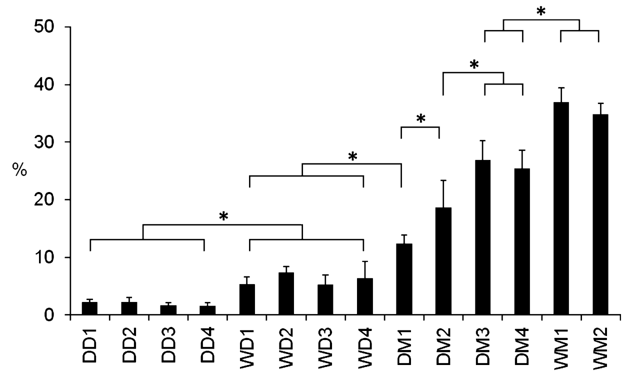

Fig. 1. Average relative soil moisture contents of the study sites based on actual soil moisture measurements. The WM3-4 sites are not included. Error bars indicate standard deviations. $^{*}p < 0.05$.

following order:

$$(\mathrm{DD1}-4) < (\mathrm{WD1}-4) < (\mathrm{DM1}) < (\mathrm{DM2}) < (\mathrm{DM3}, 4) < (\mathrm{WM1}, 2) < (\mathrm{WM3}, 4)$$

Owing to this arrangement, the basic tests of the indicator values included the within group tests of the DD1-4, WD1-4, DM3,4 and WM1,2 groups, whereas the between group tests included the DD1-4 vs. WD1-4, the WD1-4 vs. DM1, the DM1 vs. DM2, the DM2 vs. DM3,4, the DM3,4 vs. WM1,2 and the WM1,2 vs. WM3,4 comparisons. The within group test of the WM3,4 group was not carried out, as it is not confirmed whether the soil moisture contents of the WM3 and WM4 sites are statistically equivalent. The number and arrangement of extra tests depended on the results of these tests.

3.2. ANOVA tests of mean indicator values – the first approach

According to the Shapiro–Wilk tests 59 of the 64 datasets (1 unweighted and 3 weighted datasets for each of the 16 study sites) did not show significant deviation from normal distribution; therefore, considering the robustness of the ANOVA and the need for large sample sizes to gain reliable results with Kruskal–Wallis tests (Khan and Rayner, 2003), ANOVA tests were chosen. Table 2 summarizes the deviations from the expected significances. For exact *p*-values see Tables A.1–A.4 in Supplementary Data Appendix.

With using unweighted data 4 expected significant differences could not be confirmed with the basic tests, and all of these were related to the DM1 site. An unexpected significant difference was found between the WM1 and WM2 sites, increasing the total number of deviations to 5. None of the extra tests had results different from the expectations.

Cover weighted averages performed very poorly, with 12 mistakes in the basic tests and 13 in the extra ones, which makes a total of 25 mistakes. These mistakes did not concentrate to a specific point of the gradient, but were scattered along its entire span.

Transforming the percent cover values into ordinal values improved the performance of the analysis, but 8 expected significances could still not be detected (5 in the basic and 3 in the extra tests). Now the problematic comparisons concentrated to the WD1-4, DM1 and DM2 sites.

For the abundance-independent weighting method, the weights had to be calculated first. The *F* values of 2178 Hungarian plant species were available (Borhidi, 1995), and after developing the *F* value spectrum for the flora using the SynData software (Horváth, 2006), the weights of the values were calculated as described in the 2.4. Data analysis section (Table 3).

When using these weights, only 5 expected significances could not be detected. All of these occurred in the basic tests and were in connection with some WD and DM sites.

3.3. Kruskal–Wallis tests of indicator value populations – the second approach

A summary of deviations from the expected significant differences can be found in Table 4. For exact *p*-values see Tables A.5–A.8 in Supplementary Data Appendix.

Unweighted data performed the worst, with a total of 15 mistakes (13 in the basic and 2 more in the extra tests). These were not localized to a certain part of the spectrum but occurred at various comparisons.

The frequency weighted method was the most efficient of all, with only 2 mistakes. These 2 mistakes were the lack of significance between the DM1 and the WD2,4 sites. This was the only method which was sensitive enough to detect the significant difference between the DM1 and the DM2 sites. The majority of the *p*-values were rather low, and where significant results were not expected, the *p*-values equaled 1 in all but one cases.

The ordinal weighting method resulted in 9 mistakes, affecting various within-group relations. All of these were found with the basic tests.

The performance of the abundance-independent method was moderately good, with a total of 5 mistakes (4 false negative and 1 false positive) in the basic tests of certain DM sites.

4. Discussion

In the present study we attempted to find differences in the efficiency of different statistical layouts designed to compare vegetation units by means of relative ecological indicators. Two different approaches were used during data processing. The first approach, which is widely used in the scientific literature, was based on quadrat specific average indicator values, while the second approach did not require the averaging of any indicator values, which, though accepted, is a mathematically inappropriate operation and has been criticized by some authors (e.g. Dierschke, 1994; Möller, 1992). Instead, the second approach used raw indicator value populations deriving from the species lists of the sites and were weighted by repeating each value as many times as its weight required. The principle of this weighting method is new for vegetation science, and it has an important feature, namely, that it increases the size of the data pool. This can be beneficial when the number of quadrates in the sites is low due to common practical reasons (e.g. when the homogeneous part of a site is small or when the time available for the fieldwork is not sufficient, etc.) but the vegetation is not especially species-poor. In cases like these, the sample size is restricted to the number of quadrats when mean indicator values are used (i.e. in the first approach), which limits the efficiency of the tests used, whereas the sample size of the second approach is less affected.

Since the indicator value set of a quadrat can be considered as a sample from the common data pool of the given study site, quadrat specific averages can theoretically follow normal distribution, while scattering around the average value of the site. When used on the same datasets, parametric tests are more efficient in recognizing significant differences than non-parametric tests (Khan and Rayner, 2003), but they require data with normal distribution. Thus, in studies where the quadrat specific averages severely violate the assumption of normality by pure chance and only non-parametric tests can be used on them, the results will probably be less reliable than what was found with the first approach of the present study.

Table 2
The complete list of study site pairs where the results of the parametric tests (Tukey's tests after the ANOVA tests or *t*-tests) deviated from the expectations.

Unweighted	Cover weighted	Ordinally weighted	Abundance-independent
WD2-4 vs. DM1 DM1 vs. DM2 WM1 vs. WM2[a]	WD1 vs. WD4[a] DD1-4 vs. WD1 WD1-4 vs. DM1-4 WD4 vs. WM1 DM1 vs. DM2 DM2 vs. DM4 WM2 vs. WM3	WD1-4 vs. DM1 WD1-3 vs. DM2 DM1 vs. DM2	WD1,2,4 vs. DM1 DM1 vs. DM2 DM2 vs. DM4

[a] Newly found significance (i.e. false positive). In the rest of the cases significant differences were expected but could not be detected (i.e. false negative).

Table 3
Weights of the indicator values for the abundance-independent weighting method. The weights for both the ANOVA and the Kruskal–Wallis tests are listed.

W values	1	2	3	4	5	6	7	8	9	10	11	12
Number of species	77	223	311	396	344	283	169	138	126	52	33	26
Weights – ANOVA	2.35	0.81	0.58	0.45	0.52	0.64	1.07	1.31	1.44	3.49	5.50	6.98
Weights – Kruskal–Wallis	5	2	1	1	1	1	2	3	3	7	11	14

Based on our results, two main conclusions can be made. Firstly, all but three of the significant differences detected with the various methods were in line with the actual soil moisture relations. This means that regardless of the data processing approach and the weighting method, the chance for a false positive result is very low.

In contrast, the number of false negative results is very different among the methods; therefore their reliability is far from being the same. Thus, unlike in the correlation tests between average indicator values and the actual values of the indicated environmental parameters, it does matter which combination of data processing approach and weighting method is used when the vegetation of two or more sites have to be statistically compared by means of indicator values. According to our results, the Kruskal–Wallis tests on site specific, frequency weighted indicator value populations proved to be the most reliable. This layout resulted in the lowest number of false negative results and, in addition, the reliability of the conclusions based upon the *p*-values was also better than in the case of the other methods, since the *p*-values were very low values where significant differences were expected and $p = 1$ in most cases where no difference was expected. The importance of this feature is advantageous in studies where several study sites are to be compared and therefore the correction factor is high.

The least reliable results were gained with the ANOVA tests of cover weighted averages and the unweighted Kruskal–Wallis tests of site specific indicator value population; therefore the use of these methods should be avoided. Though the ANOVA tests of the ordinally weighted quadrat specific averages and the Kruskal–Wallis tests of the ordinally weighted site specific indicator value populations were better, they were still much less reliable than the frequency weighted method; therefore they cannot be recommended for vegetation analyses either. The ANOVA tests of the unweighted quadrat specific data performed relatively well, therefore this may also be an acceptable option, but none of the above-described theoretical advantages of the second approach (increased data pool size and no need for normal distribution) apply for it, thus, under less ideal circumstances the reliability of the results may drop more rapidly than that of the frequency weighted method. The two abundance-independent weighting methods were intended to improve the resolution of the tests when sites with extreme conditions have to be compared. They were efficient enough to find all expected differences at both ends of the soil moisture gradient, but at middle values they performed a less efficiently than the frequency weighted Kruskal–Wallis tests. Since the frequency weighted tests were also able to detect all expected differences at both the driest and wettest ends of the gradient, the abundance-independent weighting methods had no advantage over the frequency weighted method.

The explanation for the good performance of the frequency weighted method seems to have an ecologically sound explanation, as it does not carry the general problems one may encounter when using cover weighted or unweighted indicator values for vegetation analyses. According to Ellenberg et al. (1992), the drawback of using cover data is that certain generalist species, like clonal grasses, reach high cover values easily and therefore can have disproportionally high weights, while solitary growing, yet strongly indicative species such as some orchids are underweighted, making the cover weighted averages unreliable. In fact, the ordinal method has the same problems, though to a lesser extent as the biggest weight can be only 6-times bigger than the smallest one. The unweighted method, however, allocates the same weight to all species ($w_i = 1$), regardless whether they are characteristic species or just transient ones, accidentally invading the site with a couple of clones or solitary specimens (Ellenberg et al., 1992). When weighting with frequency, however, similar weights are allocated to all characteristic species of the sites, regardless of their cover within the quadrats. On the other hand, transient species will never get high weights, since, by definition, they do not occur in more than a couple of patches. Moreover, the accuracy of cover value estimations is subject to between-observer differences (Sykes et al., 1983)

Table 4
Complete list of study site pairs where significant differences were expected but could not be detected with the non-parametric tests (Kruskal–Wallis and Mann–Whitney tests).

Unweighted	Frequency weighted	Ordinally weighted	Abundance-independent
DD1 vs. WD2 DD4 vs. WD1-4 WD2 vs. DM1 DM1 vs. DM2-4 DM2 vs. DM3,4 WM1,2 vs. WM3,4	WD2,4 vs. DM1	DD4 vs. WD1 WD2,4 vs. DM1 DM1 vs. DM2 DM2 vs. DM3,4 WM1 vs. WM3 WM2 vs. WM3,4	DM1 vs. DM2,4 DM2 vs. DM3,4 DM3 vs. DM4[a]

[a] Newly found significance (i.e. false positive). In the rest of the cases significant differences were expected but could not be detected (i.e. false negative).

but for determining frequency values no subjective estimations are needed.

As a summary, it can be stated that our results along with the theoretical considerations support the use of Kruskal–Wallis tests on site specific, frequency weighted indicator value populations for studies where relative ecological indicators are intended to be used for comparing vegetation units.

Acknowledgements

The authors would like to thank Csaba Biró for his help in choosing the study sites. We are also grateful to Róbert Gallé and Krisztina Boda, who provided valuable suggestions in statistical issues.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ecolind.2013.09.002.

References

Allen, R.B., 1992. RECCE: an inventory method for describing New Zealand's vegetation cover. Forest Res. Inst. Bull. 176, 1–25.
Biró, M., Révész, A., Molnár, Zs., Horváth, F., 2007. Regional habitat pattern of the Danube–Tisza interfluve in Hungary I. The landscape structure and habitat pattern; the fen and alkali vegetation. Acta Bot. Hung. 49, 267–303.
Borhidi, A., 1993. A magyar flóra szociális magatartási típusai, természetességi és relatív ökológiai értékszámai – Social behaviour types of the Hungarian flora, its naturalness and relative ecological indicator values. Janus Pannonius Tudományegyetem, Pécs..
Borhidi, A., 1995. Social behaviour types, the naturalness and relative ecological indicator values of the higher plants in the Hungarian flora. Acta Bot. Hung. 39, 97–181.
Briemle, G., 1997. The applicability of ecological values in Grassland. J. Appl. Bot. – Angew. Bot. 71, 219–228.
Diekmann, M., 2003. Species indicator values as an important tool in applied plant ecology – a review. Basic Appl. Ecol. 4, 493–506.
Dierschke, H., 1994. Ökologische Zeigerwerte. Pflanzensoziologie; Grundlagen und Methoden. Ulmer, Stuttgart, pp. 224–246.
Dzwonko, Z., 2001. Assessment of light and soil conditions in ancient and recent woodlands by Ellenberg indicator values. J. Appl. Ecol. 38, 942–951.
Ellenberg, H., 1952. Landwirtschaftliche Pflanzensoziologie II. Wiesen und Weiden und ihre standortliche Bewertung, Ulmer, Stuttgart.
Ellenberg, H., Weber, H.E., Düll, R., Wirth, V., Werner, W., Paulissen, D., 1992. Zeigerwerte von Pflanzen in Mitteleuropa. Scr. Geobot. 18, 1–258.
Ewald, J., 2003. The sensitivity of Ellenberg indicator values to the completeness of vegetation relevés. Basic Appl. Ecol. 4, 507–513.
Horváth, A., 2006. SYNADATA: szünbotanikai (florisztikai és cönológiai) adatbázis-kezelő és -elemző program. Kitaibelia 11, 55.
Járai-Komlódi, M., 1958. Pflanzengesellschaften in dem Turjángebeit von Ócsa-Dabas. Acta Bot. Acad. Sci. Hung. 4, 63–92.
Käfer, J., Witte, J.-P.M., 2004. Cover-weighted averaging of indicator values in vegetation analyses. J. Veg. Sci. 15, 647–652.
Khan, A., Rayner, G.D., 2003. Robustness to non-normality of common tests for the many-sample location problem. J. Appl. Math. Decis. Sci. 7, 187–206.
Klaus, V.H., Kleinebecker, T., Boch, S., Müller, J., Socher, S., Prati, D., Fischer, M., Hölzel, N., 2012. NIRS meets Ellenberg's indicator values: Prediction of moisture and nitrogen values of agricultural grassland vegetation by means of near-infrared spectral characteristics. Ecol. Indic. 14, 82–86.
Kowarik, I., Seidling, W., 1989. Zeigerwertberechnungen nach Ellenberg – Zu Problemen und Einschränkungen einer sinnvollen Methode. Landschaft und Stadt 21, 132–143.
Lengyel, A., Purger, D., Csiky, J., 2012. Classification of mesic grasslands and their transitions of South Transdanubia (Hungary). Acta Bot. Croat. 71, 31–50.
Molnár, Zs., 2003. Dry sand vegetation of the Kiskunság. TermészetBÚVÁR Alapítvány Kiadó, Budapest.
Möller, H., 1992. Zur Verwendung des Medians bei Zeigerwertberechnungen nach Ellenberg. Tuexenia 12, 25–28.
Roovers, P., Bossuyt, B., Gulinck, H., Hermy, M., 2005. Vegetation recovery on closed paths in temperate deciduous forests. J. Environ. Manage. 74, 273–281.
Schaffers, A.P., Sykora, K.V., 2000. Reliability of Ellenberg indicator values for moisture, nitrogen and soil reaction: a comparison with field measurements. J. Veg. Sci. 11, 225–244.
Spiegelberger, U., Matthies, D., Müller-Schärer, H., Schaffner, U., 2006. Scale-dependent effects of land use on plant species richness of mountain grassland in the European Alps. Ecography 29, 541–548.
Sykes, J.M., Horrill, A.D., Mountford, M.D., 1983. Use of visual cover assessments as quantitative estimators of some British woodland taxa. J. Ecol. 71, 437–450.
ter Braak, C.F.J., Barendregt, L.G., 1986. Weighted averaging of species indicator values: its efficiency in environmental calibration. Math. Biosci. 78, 57–72.
ter Braak, C.F.J., Gremmen, N.J.M., 1987. Ecological amplitudes of plant species and the internal consistency of Ellenberg's indicator values for moisture. Vegetatio 69, 79–87.
ter Braak, C.J.F., Wiertz, J., 1994. On the statistical analysis of vegetation change: a wetland affected by water extraction and soil acidification. J. Veg. Sci. 5, 361–372.
Tölgyesi Cs Körmöczi, L., 2012. Structural changes of a Pannonian grassland plant community in relation to the decrease of water availability. Acta Bot. Hung. 54, 413–431.
van der Maarel, E., 1979. Transformation of cover-abundance values in phytosociology and its effects on community similarity. Vegetatio 39, 97–114.
van Dobben, H.F., ter Braak, C.J.F., Dirkse, G.M., 1999. Undergrowth as a biomonitor for deposition of nitrogen and acidity in pine forest. Forest Ecol. Manag. 114, 83–95.
Zeleny, D., Schaffers, A.P., 2012. Too good to be true: pitfalls of using mean Ellenberg indicator values in vegetation analyses. J. Veg. Sci. 23, 419–431.
Zwaenepoel, A., Roovers, P., Hermy, M., 2006. Motor vehicles as vectors of plant species from road verges in a suburban environment. Basic Appl. Ecol. 7, 83–93.