Iteratio: Calculating environmental indicator values for species and relevés.

Holtland, W. J., ter Braak, C. J. F., & Schouten, M. G. C.

Holtland, W. J., ter Braak, C. J. F., & Schouten, M. G. C. (2010). Iteratio: Calculating environmental indicator values for species and relevés. Applied Vegetation Science, 13(3), 369-377. https://doi.org/10.1111/j.1654-109X.2009.01069.x

Iteratio: calculating environmental indicator values for species and relevés

W.J. Holtland[1], C.J.F. ter Braak[2] and M.G.C. Schouten[3]

This is the pre-peer reviewed version of
Holtland W.J., ter Braak C.J.F., Schouten M.G.C. (2010) Iteratio: Calculating environmental indicator values for species and relevés. Applied Vegetation Science 13:369-377. DOI: 10.1111/j.1654-109X.2009.01069.x

[1]Staatsbosbeheer, Box 1300, 3970 BH, Driebergen, the Netherlands, E-mail j.holtland@staatsbosbeheer.nl
[2]Biometris, Wageningen University and Research Centre, Box 100, 6700 AC Wageningen, the Netherlands
[3]Nature Conservation and Plant Ecology Group, Wageningen University and Research Centre, Box 47, 6700 AA Wageningen, the Netherlands

Abstract
Question:
Is it possible to translate vegetation maps into reliable thematic maps of site conditions?
Method
This paper presents a new method, called *Iteratio*, by which a coherent spatial overview of specific environmental conditions can be obtained from a comprehensive vegetation survey of a specific area. Iteratio is a database application which calculates environmental indicator values for vegetation samples (relevés) on the basis of known indicator values of a limited number of plant species. The outcome is then linked to a digitalized vegetation map (map of plant communities) which results in a spatial overview of site conditions.
Iteratio requires the indicator values of a minimum of 10-20% of the species occurring. The indicator species are given a relative weight according to their amplitudes: species with a narrow range are weighted stronger, species with a broad range are weighted weaker.
Conclusion
The method presented here enables a coherent assessment of site conditions on the basis of a vegetation survey and the indicator values of a limited number of plant species.

Keywords: site conditions, indicator species, vegetation samples , weighted averaging

Introduction

Indirect assessment of site conditions by means of indicator species has become common usage in nature conservation management. Indicator species are determined by studying the correlation between the species and classes of sites (habitat types) and they provide qualitative expressions of environmental niches. The quantitative expression of the species' niche is known as the species indicator value. Assessment of site conditions on the basis of indicator species often remains rather fragmentary as the most useful indicator species have narrow niches and, therefore, limited distributions. In this paper we will therefore use all occurring species, but we will require knowledge of the environmental preference of only a limited number of species. The environmental preference will be expressed as by indicator values of species.

This paper presents a new method, called *Iteratio*, by which a coherent spatial overview of specific environmental conditions can be obtained. *Iteratio* is a database application which

calculates environmental indicator values for vegetation samples (relevés) on the basis of known indicator values of a limited number of plant species. The outcome is then linked to a digitalized vegetation map (map of plant communities) which results in a spatial overview of site conditions.

Description of the method

Required data input

Iteratio proceeds from a comprehensive vegetation survey of a specific area . In the survey local plant communities are mapped and for each plant community one or more vegetation samples (relevés) are taken. The relevés represent the plant species occurring in each community and their abundance. Abundance may cover percentage or any other appropriate measure of degree of presence or merely presence-absence (1/0). As data input *Iteratio* uses the relevés, the digital map of the plant communities and indicator values and weights for some of the species occurring. (see figure 1)

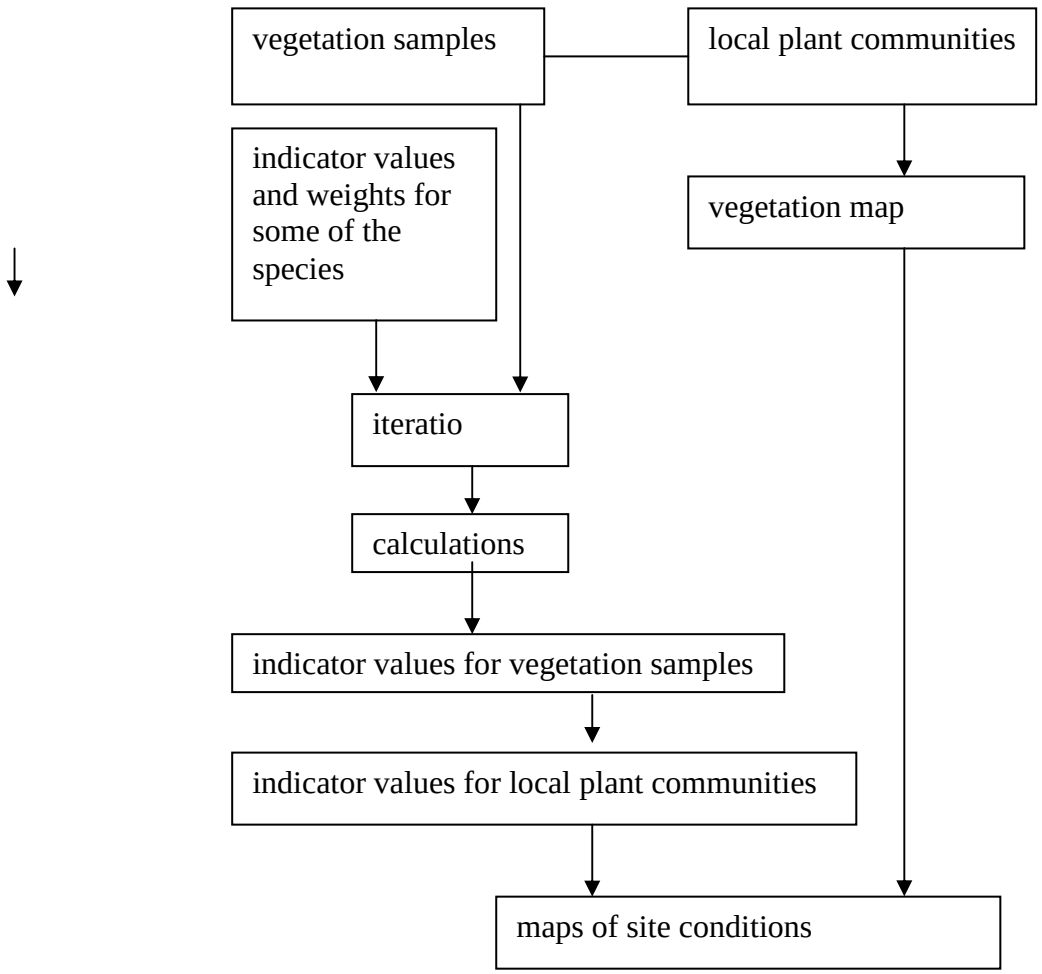

Figure 1.Flowdiagram of the Iteratio method

In order to relate the relevés to environmental variables, *Iteratio* uses species indicator values. Various systems of (plant) indicator species and indicator values have been presented in the

past. Widely used is the Ellenberg system (Ellenberg, 1974). Disadvantages of this system are that it does not provide amplitudes but single ordinal values and that it ignores the fact that in different types of ecosystems or in different climatic regions, the ecological amplitude of a plant species may vary. In recent times, a system of indicator species was developed in the Netherlands (Jalink *et al*., 1996-2007) which is largely based on actual field measurements and which provides ranges in pH, in moisture conditions and trophic conditions at which a selected number of species occur within the context of specific plant communities and of the landscape types in which these communities are represented. Within the ranges the system also indicates the species'optimal occurrence and in the procedure presented in this paper, single values representing these optimal occurrences are used..The different species are given a relative weight according to their amplitudes: species with a narrow range are weighted stronger, species with a broad range are weighted weaker (a weighting method is provided with the programme).

When *Iteratio* is applied to the vegetation data of a study site, it requires the indicator values of a minimum of 10-20% of the species occurring (see below). The program gives the most reliable results if this set contains some species with a narrow ecological amplitude as well as some species with a broad range.

Calculations

The relevés are read into a database application - if needed - after transformation of the species abundance values. If vegetation analysis is according to the Braun-Blanquet method, cover/abundance degrees could be transformed to numerical values on a ratio scale such as the van der Maarel scale (1979). Known species indicator values (and species weights) are then imported. Next, an iterative calculation process is started (see Appendix). In the first run, *Iteratio* calculates per relevé a weighted average on the basis of the known indicator values of the species occurring in the relevé, their cover/abundance degrees and their given weights. In a second run, the program calculates for each species with unknown indicator value a weighted average on the basis of the averages calculated for those relevés in which the species occurs as well as the species cover/abundance degrees. These calculated values are then used as additional indicator values in a next iteration which calculates a new weighted average per relevé. The known species indicator values which formed the basis of the first run, are kept constant during all iterations. After a certain number of iterations, usually between 20-40, the weighted averages for species and relevés remain constant. At this stage, all the species have acquired stable relative positions within the environmental range represented by all the relevés. The program then calculates two indices which characterize the amplitude (niche breadth) of each species: a numerical value and a standard deviation. The program presents the amplitudes also graphically.

Thematic maps of site conditions

The indicator values that have been obtained for the relevés together with the local maps of plant communities can now be used to construct spatial overviews of specific environmental variables. Where a plant community is represented by more than one relevé, an average indicator value is calculated for the community type. By means of GIS these values are then transposed to the digitalized vegetation map which results in thematic maps of site conditions.

Site conditions for pH in 1997 and spring water level in 1991 and 1997

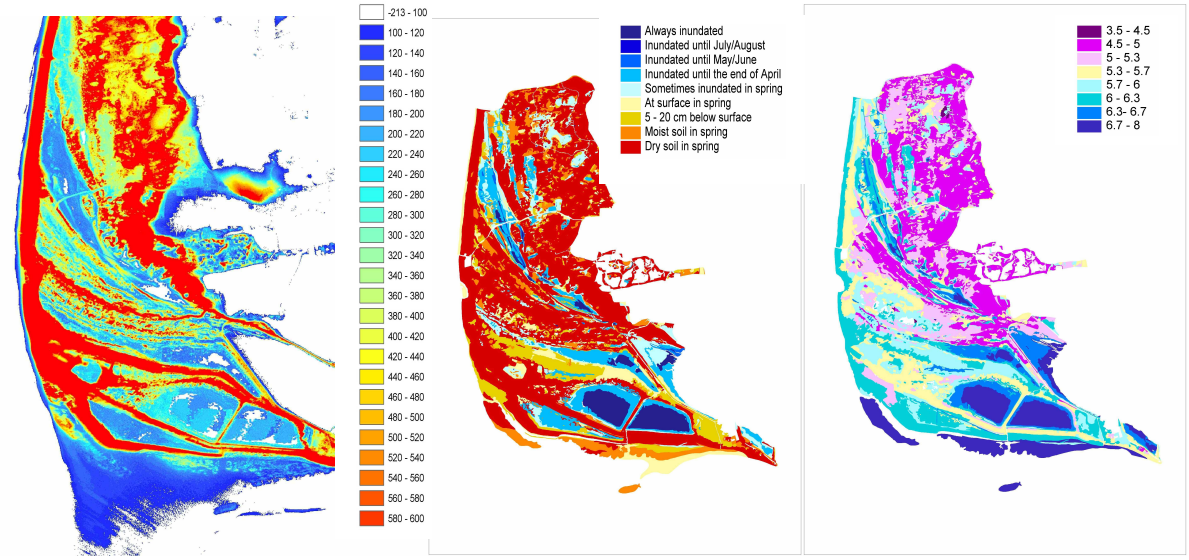

Fig.2a Elevationmap cm's above sealevel Fig.2b Spring water level 1997 Fig.2c pH 1997

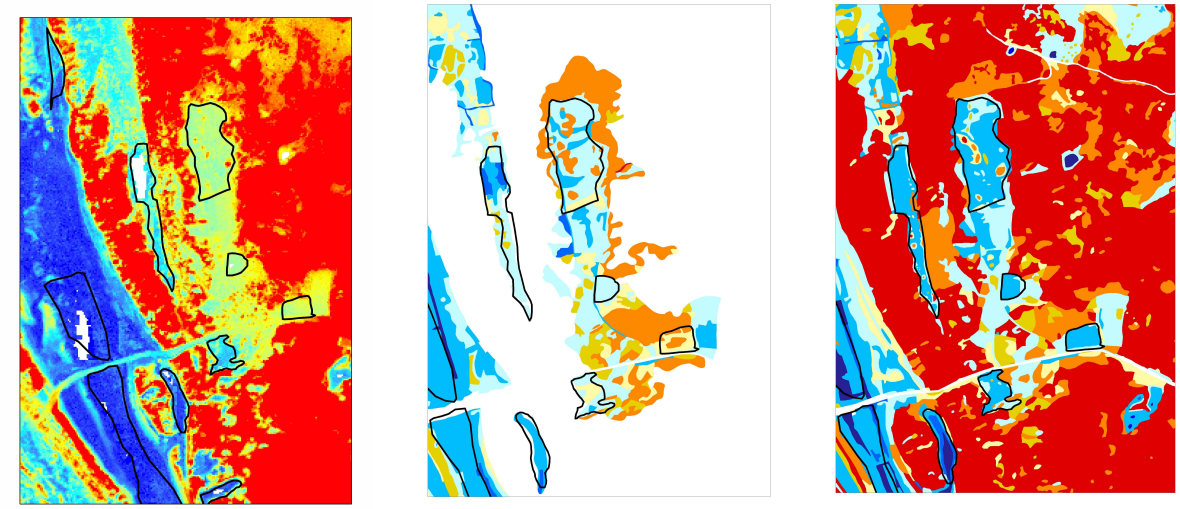

Fig.2d Detail elevationmap, sod-stripped areas (legend see 2a)

Fig.2e Spring water level 1991 (legend see fig.2b)

Fig.2f Spring water level 1997 (legend see fig.2b)

Example of the method

Figure 2 provides an example of the method for a coastal sand dune area on the island of Texel in the northwest of The Netherlands. Figure 2a shows the surface levels of the dune system (derived from a detailed digital database of surface levels in The Netherlands). In 1991, the vegetation of a number of dune slacks in this area was mapped in preparation of an ecological restoration project. The mapping was done on the basis of some 190 relevés recorded following the Braun-Blanquet method. The restoration project was carried out in 1993. Several of the dune slacks were sod-stripped and at the same time ground water extraction from the sand dunes came to a halt after an agreement was reached with the local drinking water company. In 1997, a vegetation map of the entire dune system was produced on the basis of some 450 Braun-Blanquet relevés. *Iteratio* was applied using the relevés and

maps made in 1991 and 1997 and the species indicator values as presented in the system by Jalink *et al*. (1996-2007). Mean water tables in spring and the pH of the substrate as derived from the 1997 vegetation map by the *Iteratio* program are presented in Figures 2b and 2c.. The north-eastern part of the area consists of old and stabilized sand dunes, whereas the southwest shows coastal accretion and here new sand dunes are still being formed. The age of the sand dunes is clearly reflected in Figure 2c: the oldest part of the system has decalcified whereas the younger sections are still very alkaline. Figures 2a and 2c also show that elevation also has a bearing on the soil pH. The more elevated sections of the dune systems decalcify more readily than the lower parts which are under the influence of calcareous ground water.

Figures 2d-2f show the section of the dune system in which restoration measures were carried out. The areas marked in black were sod-stripped in 1993. A comparison of the spring water levels in 1991 and 1997 (both derived from the vegetation maps by means of *Iteratio*) shows that in the dune slacks and particularly in the sod-stripped sections, the period of inundation in spring has increased (Figures 2e and 2f). Similar maps of pH and trophic level produced by means of *Iteratio* (not shown) reveal that in the sod-stripped areas pH has increased and the nutrient availability decreased.

Stability of the method

In this section we investigate the stability of the method in relation to the percentage of known indicator values. We take the gradient trophic level as an example. For this gradient ~30% (64 out of 209) of the species have a known indicator value. (The other two gradients have fewer species with known indicator value). By treating randomly some of these as unknown we generated new data sets with 1, 2, 5, 10, 20 % of the species having known indicator value. We applied *Iteratio* to each new data set. Table 1 shows that the correlation of the obtained indicator values with those obtained from the original data set is slightly negative when only 1 or 2% of the indicator values is assumed known, whereas the correlation is above 0.7 when at least 10% of the indicator values is known. Similar correlations were obtained on the basis of the site scores of the algorithm. We conclude that the meaning of the gradient is reasonably retained when 10-20% of the species has a known indicator value.

Table 1. The influence of the percentage species with known indicator value (% known) on the final *Iteratio* indicator values, expressed as the correlation of the *Iteratio* indicator values with those obtained from the original data set (with 30.6 % of the indicator values known). The value shown are the mean and standard deviation (sd) based on 50 independent replicates.

% known	correlation	sd
1	-0.07	0.00
2	-0.01	0.04
5	0.27	0.18
10	0.70	0.02
20	0.82	0.05

Discussion

In this paper a simple iterative weighted averaging algorithm is proposed to order species and sites along gradients defined by known indicator values of a, possibly small, number of species. The indicator values quantify here the ecological preferences of the species. We now indicate how this algorithm relates to previous methods.

Clausman (1980) also attempted to improve indicator values from existing ones by an iterative procedure. The procedure calculated site scores from the initial indicator values, new indicator values from the site scores, then new site scores form the new indicator values, and so on until stabilization. ter Braak & Gremmen (1987) pointed out that this procedure is essentially an ordination method, much like correspondence analysis (Hill, 1973, 1974;ter Braak & Prentice 1988), and that, by consequence, the original meaning of the indicator values might get lost. Our iterative algorithm differs from this procedure in that some indicator values remain fixed, namely the known indicator values of the species that define the gradient. The fixation of indicator values of at least two species prohibits the iterative algorithm to converge to the first axis of correspondence analysis. This is the mathematical minimum. From our stability analysis (Table 1) we advise to fix the indicator values of many more than two species (*e.g.* 10-20%) in order to be certain that the meaning of the gradient is retained in the analysis.

Conceptually, the work by ter Braak & Gremmen (1987) is closer. In their test of the internal consistency of Ellenberg's indicator values for moisture, they fixed all indicator values, except the one of the species under test. However, because only a single indicator value was not fixed, ter Braak & Gremmen (1987) did not do more than one iteration of the weighted averaging algorithm. ter Braak and Gremmen (1987) attempted to improve upon weighted averaging, so as to avoid some of its drawbacks (ter Braak & Looman (1986; ter Braak and Looman 1986). They calculated indicator values as the centroid of the so-called response histogram and, in a more advanced approach, used maximum likelihood using Gaussian response curves. Either improvement might be incorporated in our iterative algorithm. We did not do so for two reasons: (1) it destroys the simplicity of the methods and (2) the improvements are not unequivocal. Specifically, the improvements treat absences as informative whereas unembellished weighted averaging neglects absences and thus treat absences as uninformative. Because a species may be absent for many reasons - "it may be nibbled by a sheep" (Hill, 1973) - we prefer to treat them as uninformative and thus to stick to unembellished weighted averaging.

Our simple iterative algorithm is less closely related to correspondence analysis (alias reciprocal averaging) than one might think. The reason is that the iterative two-way weighted averaging algorithm of correspondence analysis contains a rescaling step - such as setting the range of the newly obtained scores to 100 or setting the variance to unity (Ter Braak, 1995)- which is missing in our algorithm. This observation led us to consider a modification that is closer to correspondence analysis (see Appendix). As correspondence analysis, the modified algorithm is very sensitive to outliers, in particular to sites with few species which themselves occur in few other sites in the data set, as shown in the Appendix for the example data. We concluded that the simple algorithm is to be preferred over the modified algorithm.

Iteratio is based on weighted averaging, that is, indicator values of species that are present in a stand are averaged whereby species abundance is used as a weight. This method treats the species abundances as ratio-scaled and the (known) indicator values as interval-scaled. Often

abundance values, such as from the Braun-Blanquet method, and known indicator values, such from the Ellenberg system, are on an ordinal scales. Podani (2005) warned against uncritical use of ordinal data in multivariate exploratory analysis. This raises the question whether such values can be used in *Iteratio*. In our view, they can, at least after appropriate transformation. The situation is not very different from when we would have had available ratio-scaled abundance values and interval-scaled indicator values. In that 'ideal' case we would also consider transformation, and the choice of the transformation would influence the result. The question is thus which transformation is appropriate. Ter Braak and Barendregt (1986) showed good performance of the method of weighted averaging for Poisson distributed counts. A transformation is thus appropriate if it makes the data after transformation look more like Poisson distributed. More informally, weighted averaging treats the abundance value as a degree of presence. So we should stay close to presence-absence. If more than a single species is present, no single species should dominate the result. The van der Maarel (1979) scale achieves this more or less in the Braun-Blanquet system, and a logarithmic transformation is often useful for cover percentage data. What about the scale of the known indicator values? These values should be centroids of the species response curves with respect to the underlying environmental variable and, for best performance, the modes of symmetric species response curves (Ter Braak and Barendregt, 1986). So this is the target for transformation. Ter Braak and Gremmen (1987) discuss the issue for Ellenberg's moisture indicator values and concluded that no transformation was required.

As Iteratio orders species and vegetation samples along gradients, the method is. best suited for sites with an internal variation in environmental conditions and, consequently, a variation in plant cover. The basic assumption underlying *Iteratio* is that the vegetation of a specific site is a reflection of the environmental conditions, which implies that most plant species have reached the habitats suited to them and that a competitive balance has established. Therefore, the method is not suited for sites in which environmental conditions have recently changed or the vegetation has recently been disturbed.

The hydrological regimen of habitats can be expressed in a single range as was done by Ellenberg (1974) in his system of indicator species. However, this regimen represents a complex of variables, each of which has a different ecological impact, such as mean annual water table, mean spring or summer water table, water table fluctuation or duration of inundation period. Research into indicator species (Jalink *et al*., 1996-2007) and the testing of *Iteratio* showed that the indirect method of assessing site conditions gives reliable results for some but not all of the hydrological variables. For many plant species, presence and cover/abundance degree appear to be highly indicative of spring water tables and of the duration of a inundation period. Correlations with summer water tables appeared to be less clear (data not shown); in the dry season physical properties of the substrate gain in importance with respect to water availability for plants as they determine water retention capacity of the soil as well as capillary rise. Therefore, in using plant species as indicators of hydrological conditions it is important to identify which specific hydrological variables are concerned.

Applications

The method presented here enables a coherent assessment of site conditions on the basis of a vegetation survey, the indicator values of a limited number of plant species and a limited set of field measurements of abiotic variables. Such assessments are of importance for ecosystem management and for monitoring programs and they also constitute a sound basis for

ecological restoration projects. As shown in the example given in Figure 2, they can also be used to evaluate the effects of restoration measures.

Iteratio uses the total species composition of relevés (both in qualitative and in quantitative terms) which provides a more adequate reflection of prevailing environmental conditions than just a number of selected indicator species would. The program has been tested for some 50 nature reserves in The Netherlands with a total area of about 40,000 ha. The outcome in terms of maps of pH, hydrological conditions and nutrient availability appears to be reliable as it is in agreement with available actual field measurements and the patterns shown are consistent with surface level maps. Nevertheless it is recommended to carry out actual field measurements in a few locations within the study area (preferably linked to the locations of relevés) in order to validate the maps. Comparison with surface level maps and ground truth are required in order to detect possible inconsistencies which then need to be addressed.

Iteratio was tested with a limited number of species for which indicator values had been established through actual field measurements. In the iteration process, for all species present in an area surveyed a relative position along environmental gradients is determined. Consequently insight was gained into habitat requirements of an increasing number of species. This implies that by this method indicator values can be assigned to new species without (expensive) field measurements. Moreover, *Iteratio* helps in identifying which species in a specific location are best suited for continuous monitoring programs.

A Dutch version of the computer program for general use is in preparation; an international (English) version is under consideration.

Appendix

The simple iterative algorithm, its modification and a comparison

Notation. Let $\mathbf{Y}=\{y_{ik}\}[i = 1 \ldots n;\ k= 1 \ldots p]$ be the $n \times p$ matrix containing the abundances of each of p species at each of n sites. Note that in this notation each species is a column, whereas in a spreadsheet it may be a row. If species are weighted differentially by weight w_k multiply y_{ik} by w_k. We assume here that this multiplication has already been done. By denoting summation across an index by a +, the site (row) weights are y_{i+} and the species weights are y_{+k}. We assume without loss of generality that the species with known indicator values are in the first m columns of $\mathbf{Y}$. We denote the known indicator values by α_k ($k = 1 \ldots m$), the unknown species scores by u_k ($k = (m + 1) \ldots p$) and the unknown site scores by $\underline{x_i}$ ($i = 1 \ldots n$).
Simple iterative algorithm. The simple iterative algorithm starts from arbitrary site scores $\{x_i\}$ and then iterates:

Step 1. Calculate species scores by a weighted average of the site scores, *i.e.*

$$u_k = \sum_{i=1}^{n} y_{ik} x_i / y_{+k} .$$

Step 2. Set the species scores of the species with known indicator equal to the known indicator value, *i.e.*

$u_k = \alpha_k$ for $1 \le k \le m$.

Step 3. Calculate site scores by a weighted average of the species scores, *i.e.*

$$x_i = \sum_{k=1}^{p} y_{ik} u_k / y_{i+}.$$

Step 4. Stop on convergence, that is, if the newly obtained site scores differ at most ε (e.g. $\varepsilon = 10^{-12} \times$ the range of the new site scores) from those in the previous iteration, else goto step 1.

In the main text the simple algorithm starts with step 2 and a fixed, but arbitrary choice for the scores of the remaining species (e.g. equal to there index *k*) and then iterates as above. The resulting scores either way are the same.

Modified iteration algorithm. In the modified algorithm the newly calculated indicator values of the species with known indicator values are regressed on these known values and then replaced by the fitted values of this regression. Thereafter the variance of the newly obtained indicator values is set to unity, as in correspondence analysis. The modified algorithm maximizes the ratio of the variance of the site scores and the variance of the species scores (*i.e.* the indicator values), subject to the constraint that the species scores of the species with known indicator values are linearly related to these known values. As in correspondence analysis and its canonical form (Ter Braak 1987) the variances and the regression are all weighted by the species totals.

The modified algorithm maximizes the variance ratio (*VR*)

$$VR = \sum_{i=1}^{n} y_{i+}(x_i - \bar{x})^2 / \sum_{k=1}^{p} y_{+k}(u_k - \bar{u})^2,$$

subject to the constraint that $u_k = a + b\alpha_k$ for $1 \le k \le m$, where α_k are the fixed indicator values and

$$\bar{x} = \sum_{i=1}^{n} y_{i+} x_i / y_{++} \text{ and } \bar{u} = \sum_{k=1}^{p} y_{+k} u_k / y_{++}.$$

This optimization problem can be rewritten in the form of that of canonical correspondence analysis with species and sites interchanged. The predictor variables of length *p* of this canonical correspondence analysis are $\mathbf{z}_1 = (\alpha_1, \ldots, \alpha_m, 0, \ldots, 0)$ and *p*-*m* predictor variables $\mathbf{z}_2, \ldots, \mathbf{z}_{p-m+1}$, one for each of the non-fixed species. For the j^{th} species, the predictor variable ($\mathbf{z}_{j-m+1}$) is all zeros except at position *j* where a 1 is inserted ($j = m+1, \ldots, p$). The solution of this optimization problem can thus be obtained by any algorithm that does canonical correspondence analysis. One such algorithm is the two-way weighted averaging algorithm with an extra weighted multiple regression step (ter Braak, 1986, 1987). Because of the structure of the predictor variables this extra weighted multiple regression step can be simplified to a simple weighted regression step. The resulting modified iterative algorithm then becomes:

Step 1. Calculate species scores by a weighted average of the site scores, *i.e.*

$$u_k = \sum_{i=1}^{n} y_{ik} x_i / y_{+k} .$$

Step 2. Calculate a weighted regression of $\mathbf{u} = (u_1, \ldots u_m)'$ on $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p,)'$ and set the species scores of the species with known indicator values equal to the fitted values of this regression and standardize all species scores to unit variance , as detailed in Steps 2a-c , *i.e.*

Step 2a. Calculate the regression coefficient, *i.e.*

$$b = \sum_{k=1}^{m} y_{+k} u_k (\alpha_k - \bar{\alpha}) / \sum_{k=1}^{m} y_{+k} (\alpha_k - \bar{\alpha})^2 \text{ with } \bar{\alpha} = \sum_{k=1}^{m} y_{+k} \alpha_k / \sum_{k=1}^{m} y_{+k} .$$

Step 2b. Calculate the fitted values and assign to the new species scores for the species with known indicator values, *i.e.*

$$u_k = \bar{u}_0 + b(\alpha_k - \bar{\alpha}) \text{ with } \bar{u}_0 = \sum_{j=1}^{m} y_{+j} \alpha_j / \sum_{j=1}^{m} y_{+j} \text{ for } 1 \le k \le m .$$

Step 2c. Standardize the species scores to zero mean and unit variance,

$$u_k \leftarrow (u_k - \bar{u}) / s_u \text{ with } s_u^2 = \sum_{k=1}^{p} y_{+k} (u_k - \bar{u})^2 / y_{++} .$$

Step 3. Calculate site scores by a weighted average of the species scores, *i.e.*

$$x_i = \sum_{k=1}^{p} y_{ik} u_k / y_{i+} .$$

Step 4. Goto step 1 if the newly obtained site scores differ from those in the previous iteration, else stop.

Without the Steps 2a and 2b, the modified algorithm is the same as the two-way weighted averaging algorithm of correspondence analysis (Ter Braak and Prentice, 1988) and the maximized variance ratio would be the first eigenvalue of correspondence analysis. For the modified algorithm to differ from correspondence analysis, at least three species must have known indicator values. (With just two species fixed, the regression part would always fit perfectly and therefore Steps 2a and 2b would not affect the solution).

Comparison.

Both the simple and modified algorithm converge to a unique solution. A data set is connected when each species can be linked to any other species via direct links of intermediate species. There is a direct link between two species when they co-occur in at least one site. To give a counter example, a data set with two sets of species such that a species of one set never occurs with a species of the other set is not connected.

As correspondence analysis, the modified algorithm turns out to be very sensitive to outliers, in particular to sites with few species which themselves occur in few other sites in the data

set. This sensitivity showed up clearly in the case study data set, which indeed contained two sites with each one species that did not occur elsewhere in the data. For the original data set the modified algorithm gave identical answers for the three different gradients pH, spring water level (swl), trophic level (tro); it set apart these two sites and these two species and gave the species with known indicator values (and all other species) equal scores; the regression coefficient in the modified algorithm was indeed zero. After this observation, we cleaned the data set so that each site contained more than *a* species and each species occurred in more than *a* sites. We used both $a = 1$ and $a = 5$, denote the resulting data sets by D1 and D5 and applied the simple and modified algorithm to both data sets. D1 and D5 contained 209 and 129 samples and 209 and 97 species, respectively.

Table A1. Variance ratio's (*VR*) of the simple iterative algorithm and the modified iterative algorithm for data sets D1 and D5, the weighted correlation of the species scores of the simple algorithm with those of the modified algorithm, and the final regression coefficient (*b*) of the modified algorithm. VR is the ratio of the variance of the site scores and the variance of the species scores.

	D1	D5	D1	D5	D1	D5
	pH		swl		tro	
VR simple alg.	0.84	0.77	0.88	0.84	0.84	0.77
VR modif. alg.	0.95	0.84	0.95	0.87	0.94	0.83
Correlation	0.06	0.82	0.06	0.96	0.11	0.76
b modif. alg	0.03	1.17	0.01	0.31	0.04	0.66

Table A2. Weighted correlation between the species scores obtained from data set D1 and the species scores obtained from data set D5 for the species occurring in D5.

	pH	swl	tro
simple alg	0.98	1.00	0.98
modif alg	0.76	0.55	0.80

Table A1 the variance ratios obtained with the simple and modified algorithm on both data sets. The variance ratio in the modified algorithm is by definition always higher than in the simple algorithm (because the modified algorithm maximizes VR), but the difference is small in both data sets. However, the correlation between the final species scores (indicator values) of both algorithm is very low (<0.13) in D1 and high (>0.76) in D5. This shows that the algorithms may result in very different solution if there are sites with few species and species with few occurrences, and that their solutions are much more alike if there are no such sites or species. The last row in Table A1 shows by the size of the regression coefficient *b* that the known indicator values are much more spread out in the final solution in D5 than in D1. This is an indication that the modified algorithm failed to retain the meaning of the original indicator values in D1 and retained, or at least better retained, the meaning in D5.

The high correlations in the first row of Table A2 show that the simple algorithm yields about the same solution in both data sets whereas the low correlations in the second row show that the modified algorithm yields rather different solutions. The correlations among the site scores give a similar result.

We concluded from this that the simple algorithm is much more stable than the modified algorithm and that, at least for these data, the simple algorithm is to be preferred over the modified algorithm. We remark that the modified algorithm may have virtue for data in which the variance ratio is much smaller than 1.

References

Clausman, P. H. M. A. 1980. Ecologische interpretatie van vegetatieopnamen m.b.v. een computer. WLO-Mededelingen **7**:92-98.

Ellenberg, H. 1974. Zeigerwerte der Gefässpflanzen Mitteleuropas. Scripta Geobot. Göttingen 9: 1-97.

Hartog, P.S., F.H. Everts, A.P. Grootjans et al. 1991. Vegetatiekartering van de Mokslootvallei (Texel). Everts&De Vries e.a. Opdrachtgever Staatsbosbeheer

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 **3**:340-354.

Jalink, M.H. *et al.* 1996-2007. Indicatorsoorten voor verdroging, verzuring en eutrofiëring. 10 vols. Staatsbosbeheer, Driebergen, The Netherlands.

Podani, J. (2005) Multivariate exploratory analysis of ordinal data in ecology: Pitfalls, problems and solutions. *Journal of Vegetation Science*, **16**, 497-510.

Pranger, D.P. and F.H. Everts. 1998. Vegettiekartering Duinen-Zuid Texel. Everts&De Vries e.a. Opdrachtgever Staatsbosbeheer

ter Braak, C. J. F. 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology **67**:1167-1179.

ter Braak, C. J. F. 1987. The analysis of vegetation-environment relationships by canonical correspondence analysis. Vegetatio **69**:69-77.

ter Braak, C. J. F. 1995. Ordination. Pages 91-173 *in* R. H. G. Jongman, C. J. F. ter Braak, and O. F. R. van Tongeren, editors. Data analysis in community and landscape ecology. Cambridge University Press, Cambridge.

ter Braak, C. J. F. and Barendregt, L. G. (1986) Weighted averaging of species indicator values: its efficiency in environmental calibration. *Mathematical Biosciences*, **78**, 57-72.

ter Braak, C. J. F., and N. J. M. Gremmen. 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., and C. W. N. Looman. 1986. Weighted averaging, logistic regression and the Gaussian response model. Vegetatio **65**:3-11.

ter Braak, C. J. F., and I. C. Prentice. 1988. A theory of gradient analysis. Advances in ecological research **18**:271-317.

Van der Maarel, E. 1979. Transformation of cover-abundance values in phytosociology and its effects on community similarity. Vegetatio **39**:97–114