

# A new algorithm for the determination of differential taxa

Tsiripidis, Ioannis<sup>1\*</sup>; Bergmeier, Erwin<sup>2</sup>; Fotiadis, Georgios<sup>3</sup> & Dimopoulos, Panayotis<sup>4</sup>

<sup>1</sup>Department of Botany, School of Biology, Aristotle University of Thessaloniki, GR-54124 Thessaloniki, Greece;

<sup>2</sup>Department of Vegetation and Phytodiversity Analysis, Albrecht von Haller Institute of Plant Sciences, University of Göttingen, Untere Karspüle 2, D-37073 Göttingen, Germany; E-mail [erwin.bergmeier@bio.uni-goettingen.de](mailto:erwin.bergmeier@bio.uni-goettingen.de);

<sup>3</sup>Department of Forestry and Management of Natural Environment, Technological Education Institute of Lamia, GR-36100 Karpenisi, Greece; E-mail [gfotiad@for.auth.gr](mailto:gfotiad@for.auth.gr);

<sup>4</sup>Department of Environmental and Natural Resources Management, University of Ioannina, Seferi 2, GR-30100 Agrinio, Greece; E-mail [pdimopul@cc.uoi.gr](mailto:pdimopul@cc.uoi.gr);

\*Corresponding author; Fax +30 2310 998295; E-mail [tsiripid@bio.auth.gr](mailto:tsiripid@bio.auth.gr)

## Abstract

**Question:** How can we determine differential taxa in a vegetation data set?

**Methods:** The new algorithm presented here uses an intuitive fidelity threshold based on relative constancy differences. It is tested on a simulated and a real data set. The results of the proposed algorithm are discussed in comparison with other methods used for the determination of differential taxa.

**Results:** The new algorithm defines each taxon in each group of relevés as: (1) positively differentiating, (2) positively-negatively differentiating, (3) negatively differentiating, or (4) non-differentiating. Each taxon in a data set may be: (1) positively, positively-negatively or negatively differentiating for each group in the data set, (2) differentiating for some groups and non-differentiating for the remaining groups, or (3) non-differentiating for all groups in the data set.

**Conclusions:** The new algorithm finds the relevé groups that are positively differentiated against other groups that are negatively differentiated. It reveals differentiating structures in the data set and thus makes quantification of the relations among and between different syntaxonomic ranks conceivable. As it distinguishes between different types of differential taxa, it might improve standards of typification in vegetation classification.

**Keywords:** Classification; Community ecology; Constancy; Diagnostic taxa; Fidelity; Floristic gradients; Phytosociology; Syntaxonomy; Vegetation.

## Introduction

Differential taxa occur chiefly in a single vegetation unit (character taxa) or are confined to a few vegetation units (Barkman 1989; Chytrý et al. 2002; van der Maarel 2005). The determination of differential taxa is fundamental in vegetation science, especially in the phytosociological approach to

vegetation classification (Braun-Blanquet 1964; Westhoff & van der Maarel 1978; Kent & Coker 1992; Dierschke 1994; van der Maarel 2005), but also in other fields such as ecology and biogeography (Dufrêne & Legendre 1997).

The designation of a taxon as differentiating for a vegetation unit is based on its fidelity to this vegetation unit. Fidelity is a measure of the degree of preference of a taxon for a vegetation unit (Barkman 1989; Dierschke 1994). Szafer & Pawłowski (1927) described the concept of fidelity by employing various criteria including constancy and cover-abundance differences among taxa. Logical arguments were used to define fidelity classes that, sometimes slightly modified, are treated in almost every phytosociology textbook (Braun-Blanquet 1964; Westhoff & van der Maarel 1978; Dierßen 1990; Dierschke 1994). More specific intuitive definitions of differential taxa based on constancy differences between vegetation units were proposed by Bergmeier et al. (1990) and Willner (2001, 2006), and modified by Dengler (2003) and Dengler et al. (2005). Bruelheide (1995, 2000), Botta-Dukát & Borhidi (1999) and Chytrý et al. (2002) introduced statistical methods for calculating fidelity. Chytrý et al. (2002) reviewed the statistical methods suitable for measuring fidelity with categorical data and concluded that the frequency of species occurrence in vegetation units (constancy) gives more robust fidelity estimation than cover-abundance values, as constancy is less affected by temporal fluctuations and observer bias. Further arguments for defining fidelity on the basis of constancy alone were given by Bergmeier et al. (1990), Bruelheide (2000) and Dengler (2003). Statistical fidelity measures may produce results that contradict those based on intuitive criteria. This is mainly due to the fact that statistical fidelity measures depend on the size of the data set and/or the relative size of the target group

of relevés (Chytrý et al. 2002; Tichý & Chytrý 2006). A method provided by Tichý & Chytrý (2006) makes the results of the fidelity measures that do not depend on the size of the data set also independent of the relative group size.

Although recent years have seen some new developments in ways of measuring fidelity, there is no parallel discussion on the way and the basis on which it is calculated. Barkman (1989) gives an example: species A is present within two alliances *X* (with four associations) and *Y* (with five associations), with constancy classes *X*: III, -, -, I and *Y*: II, IV, III, V, III. According to intuitive criteria, species A should be the differential species for alliance *Y* and for the first association of alliance *X*. Dengler et al. (2005) require that the determination of differential taxa should be performed between syntaxa of the same hierarchical rank and belonging to the same syntaxon of the next hierarchical rank. This criterion corresponds to the intuitive solution of Barkman's (1989) example. However, the problem of determination of differential taxa is much more complicated. Within data sets, even those fulfilling the above-stated criterion of Dengler et al. (2005), one or more gradients of floristic differentiation exist. Any taxon in a data set may respond to these gradients, and its constancy may be differentiated accordingly. So at least some taxa may be ordered along these gradients (these are the possible differential taxa), and subsets or groups representing vegetation units may be ordered on the basis of these taxa. Of course, if taxa respond to more than one gradient, more than one group sequence may emerge. Assuming that there is an interpretable classification in a given data set, various sequences of the groups may be expected. In fact, many different group sequences should be expected if many gradients occur in the data set. This is a common experience among phytosociologists when sorting vegetation tables. They will commonly define one taxon as differential for two or more vegetation units, while another taxon is found as differential for a different but overlapping combination of vegetation units. So the determination of differential taxa should consider not only each vegetation unit separately against all others in a data set, but also all the possible combinations of units against the remaining units or some of the remaining units. For instance, in a data set of *n* vegetation units, every taxon has to be checked as to whether it exclusively differentiates any of the *n* units, as well as whether it differentiates any possible combination of two, three or more (up to *n* - 1) units together. The number of possible combinations can be very large

in data sets that include many vegetation units (e.g. for ten vegetation units the number of possible combinations of two units at one time is 45, that of three units is 120, etc.).

In order to interpret the results of such a classification, a multidimensional approach is obviously required. In phytosociology, multidimensional subdivision of data sets has been used chiefly below association level (Dierschke 1994), with the most relevant gradients being geographical, altitudinal, locally ecological (water supply, nutrient supply, etc.), successional-dynamic, and compliant with the degree of human impact or land-use intensity. Schwabe (1985) provides an example of multidimensional subdivision of an association using the first three of these gradients.

The aim of this paper is to present a new algorithm for the determination of differential taxa in vegetation data sets. This algorithm uses specific constancy differences as a fidelity threshold and provides a basis for multidimensional subdivision of vegetation data sets and quantitative and qualitative syntaxonomy at any level.

### Fidelity Threshold

We used a fidelity threshold based on relative constancy differences (hereafter the word constancy refers to relative constancy), because a main concern in our attempt to establish the new algorithm was to obtain results in harmony with intuition. Being guided by the criteria put forward by Bergmeier et al. (1990) and Szafer & Pawłowski (1927) and after testing on real data, we used the following criterion for the determination of differential taxa: a differential taxon has at least double constancy plus 20 percentage points in the target group than outside the target group. In other words, a taxon differentiates the target group if  $X \geq 2Y + 20$  or  $Y \leq (X/2) - 10$ , with *X* being the constancy of the taxon within the target group, and *Y* that outside the group. We used the above criterion in both possible directions, that is if  $X \geq 2Y + 20$  then the taxon is positively differentiating for the target group, while if  $X \leq (Y/2) - 10$  then the taxon is negatively differentiating for the target group.

### The New Algorithm for the Determination of Differential Taxa

The new algorithm can be employed on synoptic vegetation tables of constancy values that can be

produced by any appropriate kind of classification method. The following steps of the algorithm are repeated for each taxon in the data set. The description below concerns one taxon.

*Step 1:* The algorithm finds and compares the highest and lowest constancy values of the taxon. Only if the highest value is at least twice as high plus 20 percentage points higher than the lowest value can the taxon be differentiating for a group in the data set, and the algorithm proceeds. Otherwise the algorithm stops here.

*Step 2:* The algorithm finds the next lowest constancy values (second, third, etc.) and compares each with the mean of all lower constancy values by applying the following condition (a): The  $y$  (i.e. second, third, etc.) lowest constancy value must be lower than double of the mean of the  $(y - 1)$  lowest constancy values plus 20 percentage points (in the case of the second lowest constancy, the comparison is made with the lowest constancy). If the condition is not true, the algorithm stops calculating the mean values and finds the constancy values that fulfill condition (a).

*Step 3:* The algorithm finds the next highest constancy values (second, third, etc.) and compares each with the mean of all higher constancy values by applying the following condition (b): the  $y$  (second, third, etc.) highest constancy value must be higher than half of the mean of the  $(y - 1)$  highest constancy values minus 10 percentage points (in the case of the second highest constancy, the comparison is made with the highest constancy). As before, if the condition is not true, the algorithm stops calculating the mean values and finds the constancy values that fulfill condition (b).

*Step 4:* The algorithm evaluates the constancy values that fulfill condition (a) by applying the following condition (c): The  $y$  (second, third etc.) lowest constancy value must be lower or equal to half of the mean of the  $(y - 1)$  highest constancy values fulfilling condition (b) minus 10 percentage points (in the case of the second lowest value the comparison is made with the highest constancy value). When the last highest constancy value fulfilling condition (b) is reached the mean value remains stable and equal to the mean of all the constancy values fulfilling condition (b). From the lowest constancy values fulfilling condition (c), plus of course the lowest constancy value, the algorithm calculates their mean value, hereafter called the last mean value of the lowest constancies. If the second lowest constancy value does not fulfill conditions (a) or (c) then the last mean value of the lowest constancies is equal to the lowest constancy value of the taxon.

*Step 5:* The algorithm evaluates the constancy values that fulfill condition (b) by applying the following condition (d): The  $y$  (second, third etc.) highest constancy value must be higher or equal to twice the mean of the  $(y - 1)$  lowest constancy values fulfilling condition (a) plus 20 percentage points (in the case of the second highest value the comparison is made with the lowest constancy value). When the last lowest constancy value fulfilling condition (a) is reached the mean value remains stable and equal to the mean of all the constancy values fulfilling condition (a). From the highest constancy values fulfilling condition (d), plus of course the highest constancy value, the algorithm calculates their mean value, hereafter called the last mean value of the highest constancies. If the second highest constancy value does not fulfill conditions (b) or (d) then the last mean value of the highest constancies is equal to the highest constancy value of the taxon.

*Step 6:* Each constancy value of the taxon is examined with respect to the following two conditions: (I) It must be higher or equal to twice the last mean of the lowest constancy values plus 20 percentage points, and (II) it must be higher than half of the last mean of the highest constancy values minus 10 percentage points. If both conditions are true, then the taxon is designated positively differentiating for the corresponding group. If both conditions are false, then the taxon is designated negatively differentiating for the corresponding group. If the first condition is true but the second is not, then the taxon is designated positively-negatively differentiating (positively differentiating against the group or groups for which the taxon was found negatively differentiating, and negatively differentiating against the group or groups for which the taxon was found positively differentiating). If the first condition is not true and the second is true, then the taxon is designated non-differentiating.

The algorithm rates a taxon for each group: (1) positive differential, (2) positive-negative differential, (3) negative differential, and (4) non-differential. Below we present the calculations for a taxon in a data set with five groups (A-E) in which it has the following constancy values: 0, 10, 30, 80, 100.

*Step 1:* The highest constancy value (100) is higher than  $(2 \times 0) + 20$  (with 0 being the lowest constancy value). So the algorithm proceeds to the next step.

*Step 2:* The second lowest value (10) is lower than  $(2 \times 0) + 20$ . The third lowest value (30) is not lower than  $(2 \times 5) + 20$  (with 5 being the mean of the two lowest constancy values 0 and 10). Therefore, only the two lowest constancy values fulfill condition (a).

*Step 3:* The second highest value (80) is higher than  $(100/2) - 10$  (with 100 being the highest constancy value). The third highest value (30) is not higher than  $(90/2) - 10$  (with 90 being the mean of the two highest constancy values). Therefore, only the two highest constancy values fulfill condition (b).

*Step 4:* The second lowest constancy (10) is lower than  $(100/2) - 10$  (with 100 being the highest constancy value). Therefore, the last mean value of the lowest constancies (0, 10) is equal to 5.

*Step 5:* The second highest constancy (80) is higher than  $(2 \times 0) + 20$  (with 0 being the lowest constancy value) and thus the last mean value of the highest constancies is equal to 90 (the mean value of the two highest constancies).

*Step 6:* The algorithm uses the last mean values found in steps 4 and 5, and applying conditions I and II, it finds that the taxon is negatively differentiating for groups A and B, positively differentiating for groups D and E and positively-negatively differentiating for group C (positively differentiating against groups A and B and negatively differentiating against groups D and E).

**Results**

*Applying the new algorithm to simulated data*

In order to test the algorithm and to present the way in which it functions, we created a table with simulated data (Table 1). This table comprises three groups, and the taxa show the constancy values 0, 25, 50, 75 and 100. Table 1 includes all possible combinations of constancy values among the three groups, except for zero constancy of a taxon in all groups. Combinations generated by the same constancy values but in different order have been omitted. The total number of “taxa” in Table 1 is thus equal to 34. We used two ways of determination of differential taxa, one that follows the proposed algorithm (Part I in Table 1), and a second that compares the constancy values in each group with the average constancy in the remaining groups (Part II in Table 1). The latter is the method commonly used in vegetation studies, especially when the determination of differential taxa is performed by calculating fidelity measures.

*Applying the new algorithm to real data*

The new algorithm has been tested with various data sets. A case example is presented here

**Table 1.** Determination of differential taxa in a simulated data set using the new algorithm (part I), and comparing the constancy value in each group with the average of the constancy values in the remaining two groups (part II). Bold typescript indicates positive differentiation, bold italic typescript indicates positive-negative differentiation, italic typescript negative differentiation, and regular typescript non-differentiation.

Part I			Part II		
A	B	C	A	B	C
<b>25</b>	<i>0</i>	<i>0</i>	<b>25</b>	0	0
<b>50</b>	<i>0</i>	<i>0</i>	<b>50</b>	<i>0</i>	<i>0</i>
<b>75</b>	<i>0</i>	<i>0</i>	<b>75</b>	<i>0</i>	<i>0</i>
<b>100</b>	<i>0</i>	<i>0</i>	<b>100</b>	<i>0</i>	<i>0</i>
<b>25</b>	<b>25</b>	<i>0</i>	25	25	<i>0</i>
<b>50</b>	<b>25</b>	<i>0</i>	<b>50</b>	25	<i>0</i>
<b>75</b>	<b>25</b>	<i>0</i>	<b>75</b>	25	<i>0</i>
<b>100</b>	<b>25</b>	<i>0</i>	<b>100</b>	25	<i>0</i>
<b>50</b>	<b>50</b>	<i>0</i>	50	50	<i>0</i>
<b>75</b>	<b>50</b>	<i>0</i>	<b>75</b>	50	<i>0</i>
<b>100</b>	<b>50</b>	<i>0</i>	<b>100</b>	50	<i>0</i>
<b>75</b>	<b>75</b>	<i>0</i>	75	75	<i>0</i>
<b>100</b>	<b>75</b>	<i>0</i>	<b>100</b>	75	<i>0</i>
<b>100</b>	<b>100</b>	<i>0</i>	100	100	<i>0</i>
25	25	25	25	25	25
50	25	25	50	25	25
<b>75</b>	25	25	<b>75</b>	25	25
<b>100</b>	25	25	<b>100</b>	25	25
50	50	25	50	50	25
<b>75</b>	50	25	<b>75</b>	50	25
<b>100</b>	50	25	<b>100</b>	50	25
<b>75</b>	<b>75</b>	25	75	75	25
<b>100</b>	<b>75</b>	25	100	75	25
<b>100</b>	<b>100</b>	25	100	100	25
50	50	50	50	50	50
75	50	50	75	50	50
100	50	50	100	50	50
75	75	50	75	75	50
100	75	50	100	75	50
100	100	50	100	100	50
75	75	75	75	75	75
100	75	75	100	75	75
100	100	75	100	100	75
100	100	100	100	100	100

using a data set of five vegetation units from *Fagus sylvatica* forests. The data were taken from Tsiripidis et al. (2007a) and concern the units A (acidophytic forests of north-east Greece), B (thermophytic forests of north-east Greece), C (acidophytic forests of central Greece), D (thermophytic forests of central Greece), and E (calcicolous forests of high altitudes in central Greece). The units have been chosen in such a way as to reflect two different gradients of floristic differentiation (ecological and geographical) and one unit (E) relatively distant from the remaining units (for details see Tsiripidis et al. 2007a). The results of the application of the algorithm are presented in Table 2.

**Table 2.** Determination of differential taxa using the new algorithm in a case example. Vegetation units from A to E are equivalent to units 3, 5, 9, 12 and 10 in Tsiripidis et al. (2007a). The nomenclature of taxa corresponds to that in Tsiripidis et al. (2007a). Taxa with constancy values lower than 20% in any column have been omitted. Different typescript formats have the same meaning as in Table 1.

Group name	A	B	C	D	E
<i>Calamagrostis arundinacea</i>	80	28	1	0	0
<i>Vaccinium myrtillus</i>	69	4	22	1	24
<i>Epilobium montanum</i>	52	19	23	17	0
<i>Rubus idaeus</i>	40	4	14	5	0
<i>Pinus sylvestris</i>	33	2	0	0	0
<i>Veronica officinalis</i>	30	7	14	7	0
<i>Symphytum tuberosum</i> ssp. <i>angustifolium</i>	27	5	0	0	0
<i>Hypericum perforatum</i>	23	18	1	5	0
<i>Poa pratensis</i> ssp. <i>angustifolia</i>	22	3	0	0	0
<i>Rosa canina</i>	20	14	0	1	0
<i>Poa nemoralis</i>	74	97	66	77	12
<i>Veronica chamaedrys</i>	24	46	49	80	0
<i>Monotropa hypopitys</i>	47	27	25	14	0
<i>Luzula luzuloides</i>	100	67	23	4	0
<i>Euphorbia amygdaloides</i> ssp. <i>amygdaloides</i>	71	67	0	16	0
<i>Pteridium aquilinum</i>	33	72	20	77	0
<i>Deschampsia flexuosa</i>	60	36	10	2	0
<i>Carex digitata</i>	50	38	0	10	0
<i>Galium mollugo</i> agg.	32	51	6	5	6
<i>Primula veris</i>	28	47	1	8	0
<i>Corylus avellana</i>	21	47	1	8	0
<i>Orthilia secunda</i>	37	2	73	13	82
<i>Rosa pendulina</i>	26	5	0	0	35
<i>Campanula persicifolia</i>	22	72	2	18	0
<i>Digitalis viridiflora</i>	19	30	0	0	0
<i>Hypericum cerastoides</i>	12	30	0	0	0
<i>Prunus avium</i>	12	40	0	3	0
<i>Quercus petraea</i> ssp. <i>medwediewii</i>	10	75	2	18	0
<i>Dactylis glomerata</i>	8	37	0	17	0
<i>Tanacetum corymbosum</i>	8	26	0	0	0
<i>Hieracium olympicum</i>	7	36	2	1	0
<i>Silene atropurpurea</i>	7	28	0	1	0
<i>Fraxinus ornus</i>	3	36	0	20	3
<i>Epilobium lanceolatum</i>	1	27	1	1	0
<i>Chamaecytisus austriacus</i>	1	25	0	8	0
<i>Rosa arvensis</i>	1	30	1	18	0
<i>Crataegus monogyna</i> var. <i>monogyna</i>	1	21	0	0	0
<i>Thymus species</i>	1	20	0	0	0
<i>Melittis melissophyllum</i> ssp. <i>albida</i>	0	32	2	19	0
<i>Luzula forsteri</i>	0	44	20	69	0
<i>Viola alba</i>	0	24	2	44	0
<i>Sorbus torminalis</i>	0	32	0	27	0
<i>Asplenium adiantum-nigrum</i>	1	32	2	23	0
<i>Lathyrus alpestris</i>	0	0	24	57	0
<i>Galium rotundifolium</i>	1	0	85	70	44
<i>Luzula sylvatica</i>	1	0	40	44	94
<i>Festuca drymeja</i>	0	1	25	58	65
<i>Moehringia trinervia</i>	12	15	23	41	0
<i>Lathyrus laxiflorus</i> ssp. <i>laxiflorus</i>	0	8	14	83	0
<i>Cyclamen hederifolium</i>	0	0	6	66	0
<i>Rubus hirtus</i>	22	32	11	55	26
<i>Physospermum cornubiense</i>	0	8	9	54	0
<i>Hedera helix</i>	0	10	1	53	0
<i>Primula acaulis</i>	0	4	5	53	0
<i>Melica uniflora</i>	1	18	5	47	0
<i>Festuca heterophylla</i>	9	20	7	43	0
<i>Campanula spatulata</i>	0	0	8	41	0
<i>Sanicula europaea</i>	2	8	12	40	26
<i>Doronicum orientale</i>	0	4	10	38	0
<i>Crocus veluchensis</i>	0	0	7	34	0

**Table 2.** (Continued).

Group name	A	B	C	D	E
<i>Lathyrus venetus</i>	0	2	1	34	0
<i>Polystichum setiferum</i>	0	6	8	34	0
<i>Ostrya carpinifolia</i>	4	9	0	34	0
<i>Polygonatum odoratum</i>	0	2	2	33	0
<i>Lathyrus niger</i> ssp. <i>niger</i>	0	10	0	33	0
<i>Epipactis helleborine</i>	14	15	13	33	0
<i>Ilex aquifolium</i>	0	0	3	31	0
<i>Galium laconicum</i>	0	0	0	30	0
<i>Trifolium pignanti</i>	0	0	1	30	0
<i>Campanula patula</i> et <i>sparsa</i>	15	13	5	30	0
<i>Castanea sativa</i>	0	11	3	29	0
<i>Silene italica</i> ssp. <i>italica</i>	0	8	2	29	0
<i>Vicia cracca</i>	0	0	2	27	0
<i>Polypodium vulgare</i>	10	6	12	27	0
<i>Satureja vulgaris</i>	4	16	5	27	0
<i>Hieracium bracteolatum</i>	0	0	4	24	0
<i>Cystopteris fragilis</i>	4	5	6	23	0
<i>Cephalanthera longifolia</i>	0	2	5	21	0
<i>Selinum silaifolium</i>	0	0	0	20	0
<i>Galium odoratum</i>	0	0	18	49	56
<i>Daphne laureola</i>	0	0	1	38	38
<i>Abies borisii-regis</i>	15	0	42	23	100
<i>Prenanthes purpurea</i>	21	6	24	11	94
<i>Arenonia agrimonioides</i>	52	33	35	55	91
<i>Cardamine graeca</i>	0	0	0	0	65
<i>Euphorbia heldreichii</i>	0	0	0	7	62
<i>Sorbus aucuparia</i>	7	0	0	0	59
<i>Corallorhiza trifida</i>	20	0	10	2	56
<i>Satureja grandiflora</i>	0	0	12	17	53
<i>Saxifraga rotundifolia</i> ssp. <i>rotundifolia</i>	10	20	5	23	47
<i>Polystichum lonchitis</i>	0	0	0	0	41
<i>Pinus heldreichii</i>	0	0	0	0	38
<i>Sorbus aria</i>	1	1	0	0	38
<i>Euonymus verrucosus</i>	0	0	0	0	35
<i>Cotoneaster nebrodensis</i>	0	0	0	0	29
<i>Polystichum aculeatum</i>	0	0	0	0	29
<i>Cephalanthera rubra</i>	0	2	8	13	26
<i>Solidago virgaurea</i>	2	9	1	9	24
<i>Pyrola chlorantha</i>	3	0	1	0	24
<i>Pyrola minor</i>	7	0	4	0	21
<i>Lilium martagon</i>	2	2	0	3	21
<i>Mycelis muralis</i>	47	67	75	82	71
<i>Potentilla micrantha</i>	44	66	44	72	41
<i>Neottia nidus-avis</i>	33	30	53	43	38
<i>Viola riviniana</i> et <i>reichenbachiana</i>	38	44	34	52	18
<i>Hieracium murorum</i> et spec.	25	38	32	39	50
<i>Fragaria vesca</i>	44	15	13	13	21
<i>Cardamine bulbifera</i>	11	15	14	16	26
<i>Hieracium racemosum</i>	7	21	3	9	18

**Discussion**

*Is the proposed fidelity threshold arbitrary?*

The fidelity threshold used here is based on true-false arguments (Szafer & Pawłowski 1927; Bergmeier et al. 1990) and it gives a certain threshold for values of constancy difference (see also Dengler et al. 2005). This facilitates its application and also makes it more rigorously defined.

A comparison of the proposed threshold with the arguments of Bergmeier et al. (1990) and Szafer & Pawłowski (1927) is given in Table 3. For this

**Table 3.** Comparison of the proposed threshold with those of Bergmeier et al. (1990) and Szafer & Pawłowski (1927). <sup>1</sup>Constancy class II corresponds to the constancy values 30–40% and 20–30% when constancy class in the target group is V and IV, respectively; <sup>2</sup>constancy classes + and *r* correspond to the constancies in the target group 30–40% and 20–30%, respectively; <sup>3</sup>a taxon is differentiating only when the constancy in the target group is 20% and not for lower values; <sup>4</sup>there is an additional criterion that the constancy of the target group must be at least twice as high as in the reference group; <sup>5</sup>Szafer & Pawłowski (1927) distinguished five constancy classes, with constancy class I equivalent to 0–20%.

Constancy class within target group	Upper limit of constancy class within reference group		
	Proposed threshold	Bergmeier et al. (1990)	Szafer & Pawłowski (1927) <sup>5</sup>
V	II <sup>1</sup>	III <sup>4</sup>	III
IV	II <sup>1</sup>	II <sup>4</sup>	II
III	I	I	I
II	+ or <i>r</i> <sup>2</sup>	+	Absent or very rare
I	Absent <sup>3</sup>	<i>r</i>	Absent or very rare
+		Absent	

comparison, we modified the proposed threshold in the form of true-false arguments applied to constancy classes. The arguments of Bergmeier et al. (1990) are easily comparable as both fidelity thresholds use only constancy as a criterion, while the arguments of Szafer & Pawłowski (1927) cannot be compared as simply because the abundance of taxa is also used as a criterion. However, one can distinguish two different cases in the arguments of Szafer & Pawłowski (1927): one in which the mean abundance of a taxon is much higher in the target group than in the reference group, and a second in which the mean abundances do not differ much or do not constitute a criterion. The latter case may be compared with the proposed fidelity threshold. The comparison concerns only fidelity degrees 5 and 4 (“*treu*” and “*fest*” according to Szafer & Pawłowski 1927), as fidelity degree 3 (“*hold*”) does not make use of constancy as a criterion, and fidelity degrees 2 (“*vag*”) and 1 (“*fremd*”) refer to non-differential taxa (see also Dierßen 1990; Dierschke 1994). In spite of minor differences, the correspondence between the proposed algorithm and the arguments of Bergmeier et al. (1990) and Szafer & Pawłowski (1927) is evident (Table 3).

Yet the proposed fidelity threshold is arbitrary, like the other two fidelity thresholds. All the above-mentioned thresholds are based on intuition and are not derived from a mathematical methodology. Statistical fidelity measures (Chytrý et al. 2002) may constitute an alternative for the definition of differential taxa in vegetation classification. However, even these statistical measures are not commonly used as statistical tests but only as indices for which an arbitrary threshold is chosen.

#### *How does the proposed algorithm work?*

The application of the new algorithm reveals three categories of taxa: (1) positively, positively-

negatively or negatively differentiating for any group in the data set, (2) differentiating for some groups and non-differentiating for the remaining groups, and (3) non-differentiating for any group in the data set. Taxa of the first category may be termed full differential taxa as they differentiate one or some groups positively and all others negatively, and thus match the textbook definition of differential taxa (Westhoff & van der Maarel 1978; Dierßen 1990; Dierschke 1994; van der Maarel 2005). Taxa of the third category show more or less equal (low or high) constancy in all groups of the data set; they are not differentiating any group. The second category comprises taxa that may be termed partial differential taxa. These taxa differentiate positively one or more groups against one or more other groups, which are negatively differentiated by the same taxa. These are not conventional differential taxa, but they do carry information on differentiation.

The second method of determination of differential taxa in the simulated data set compares the constancy in one group with the average constancy in the remaining groups. This method completely fails to find positive fully differential taxa for two groups together. In fact, it succeeds in finding as positive differential taxa only those taxa that show high constancy in a group and a small average constancy in the two other groups. To phrase it more generally, the comparison between the constancy in one group and the average constancy of the other groups fails to find a structure in the data set and the possible relations between the vegetation groups.

As the vegetation units in a data set are connected through complex environmental and geographic or other gradients, the relevé groups along each of these gradients may be ordered so as one group will be closer to the second than to the third. This is because the taxa forming the vegetation units respond to the gradients, and the vegetation units may be understood as specific

segments along these gradients (Austin 2005). Thus, the determination of differential taxa can be described as a problem of finding the taxa that respond clearly to the underlying gradients. The small data set of beech forest vegetation units illustrates that there are several relevant gradients. A cluster analysis of this data set separates at the first level the high-altitude calcicolous forests (unit E) from all other units, at the second level the units of north-east Greece (A and B) are separated from those of central Greece (C and D), and at the third and fourth level the two units of central Greece and the two units of north-east Greece, respectively, are separated. This classification may reveal the dominant gradient, but it is evident from Table 2 that many taxa follow other gradients, e.g. *Orthilia secunda*, *Rosa pendulina*, *Luzula forsteri*, *Viola alba*, *Sorbus torminalis*, *Asplenium adiantum-nigrum*, *Galium rotundifolium*, *Luzula sylvatica*, *Festuca drymeja*, *Galium odoratum* and *Daphne laureola*, which show high constancies in two or three vegetation units separated at quite different levels.

Following the method of determination of differential taxa proposed by Dengler et al. (2005), all differential taxa that respond to the dominant gradient, and thus to the primary linear classification hierarchy, will be found, while differential taxa that respond to other gradients will be missed. Our approach to the determination of differential taxa recognizes the fact that the taxa in a data set usually follow more than one gradient, thus reflecting the multidimensional character of variation within vegetation data sets. Quantifying the relations among and between different syntaxonomic ranks becomes easily conceivable.

An important characteristic of the new algorithm is that it distinguishes certain types of differential taxa: positive differential, positive-negative differential, negative differential and non-differential taxa for one or more groups in a data set. Additionally, each taxon can be a full or partial differential taxon or non-differentiating in the whole data set. Different types of differential taxa and their relative numbers combined with vegetation theory (e.g. Bergmeier et al. 1990; Dierschke 1994; Dengler et al. 2005; van der Maarel 2005; Willner 2006) may offer ways towards improved syntaxonomic classifications based on quantitative differentiation by taxa.

A disadvantage of the proposed algorithm is that it is based on constancy values and that it assumes all groups have the same size. However, by downweighting of the fidelity values in relatively small groups (Bruehlheide 2000) rare vegetation units

of potentially high ecological or conservation value might be disregarded (Tsiripidis et al. 2007b). Furthermore, nor would the result match intuitive expectations (Tichý & Chytrý 2006). In our opinion, the acceptance of a taxon as differentiating for a vegetation unit should not depend on the relative size of the vegetation unit. Such a constraint would be as unreasonable as if the acceptance of a taxon in taxonomy was constrained by its population size. This does not go against the principles that a minimum number of records per vegetation unit is required in order to ensure representativeness and to exclude typification of singular vegetation stands (see Recommendation 7A in Weber et al. 2000). When setting 20 as the minimum number of records of a vegetation unit and applying the new algorithm and fidelity threshold, the largest possible *P* value found for two groups between two differentiating constancies (positively and negatively) is equal to 0.053 (one-tailed Fisher's exact test). However, a minimum number of records is a *sine qua non* and not a factor to be involved in the determination of differential taxa and thus of vegetation types.

**Acknowledgements.** We are grateful to Milan Chytrý, Zoltán Botta-Dukát, Lubomír Tichý and Onno van Tongeren for their helpful and valuable suggestions on earlier versions of the manuscript.

## References

- Austin, M.P. 2005. Vegetation and environment: discontinuities and continuities. In: van der Maarel, E. (ed.) *Vegetation ecology*, pp. 52–84. Blackwell, Malden, GB.
- Barkman, J.J. 1989. Fidelity and character-species, a critical evaluation. *Vegetatio* 85: 105–116.
- Bergmeier, E., Härdtle, W., Mierwald, U., Nowak, B. & Pepler, C. 1990. Vorschläge zur syntaxonomischen Arbeitsweise in der Pflanzensoziologie. *Kieler Notizen zur Pflanzenkunde in Schleswig-Holstein und Hamburg* 20: 92–103.
- Botta-Dukát, Z. & Borhidi, A. 1999. New objective method for calculating fidelity. Example: the Illyrian beechwoods. *Annali di Botanica* 57: 73–90.
- Braun-Blanquet, J. 1964. *Pflanzensoziologie. Grundzüge der Vegetationskunde*. 3rd ed. Springer, Berlin, DE.
- Bruehlheide, H. 1995. Die Grünlandgesellschaften des Harzes und ihre Standortbedingungen. Mit einem Beitrag zum Gliederungsprinzip auf der Basis von statistisch ermittelten Artengruppen. *Dissertationes Botanicae* 244: 1–338.
- Bruehlheide, H. 2000. A new measure of fidelity and its application to defining species groups. *Journal of Vegetation Science* 11: 167–178.

- Chytrý, M., Tichý, L., Holt, J. & Botta-Dukát, J. 2002. Determination of diagnostic species with statistical fidelity measures. *Journal of Vegetation Science* 13: 79–90.
- Dengler, J. 2003. Entwicklung und Bewertung neuer Ansätze in der Pflanzensoziologie unter besonderer Berücksichtigung der Vegetationsklassifikation. *Archiv Naturwissenschaftlicher Dissertationen* 14: 1–297.
- Dengler, J., Berg, C. & Jansen, F. 2005. New ideas for modern phytosociological monographs. *Annali di Botanica Nuova Serie* 5: 193–210.
- Dierßen, K. 1990. *Einführung in die Pflanzensoziologie (Vegetationskunde)*. Wissenschaftliche Buchgesellschaft, Darmstadt, DE.
- Dierschke, H. 1994. *Pflanzensoziologie. Grundlagen und Methoden*. Ulmer, Stuttgart, DE.
- Dufrêne, M. & Legendre, P. 1997. Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecological Monographs* 67: 345–366.
- Kent, M. & Coker, P. 1992. *Vegetation description and analysis*. John Wiley & Sons, Chichester, GB.
- Schwabe, A. 1985. Monographie *Alnus incana* - reicher Waldgesellschaften in Europa. Variabilität und Ähnlichkeiten einer azonal verbreiteten Gesellschaftsgruppe. *Phytocoenologia* 13: 197–302.
- Szafer, W. & Pawłowski, B. 1927. Die Pflanzenassoziationen des Tatra-Gebirges. Bemerkungen über die angewandte Arbeitsmethodik. *Bulletin international de l'Académie polonaise des sciences et des lettres, Série B: Sciences Naturelles* 2: 1–12.
- Tichý, L. & Chytrý, M. 2006. Statistical determination of diagnostic species for site groups of unequal size. *Journal of Vegetation Science* 17: 809–818.
- Tsiripidis, I., Bergmeier, E. & Dimopoulos, P. 2007a. Geographical and ecological differentiation in Greek *Fagus* forest vegetation. *Journal of Vegetation Science* 18: 743–750.
- Tsiripidis, I., Karagiannakidou, V., Alifragis, D. & Athanasiadis, N. 2007b. Classification and gradient analysis of the beech forest vegetation of the southern Rodopi (Northeast Greece). *Folia Geobotanica* 42: 249–270.
- van der Maarel, E. 2005. Vegetation ecology – an overview. In: van der Maarel, E. (ed.) *Vegetation ecology*, pp. 1–51. Blackwell, Malden, GB.
- Weber, H.E., Moravec, J. & Theurillat, J.-P. 2000. International Code of Phytosociological Nomenclature. 3rd edition. *Journal of Vegetation Science* 11: 739–768.
- Westhoff, V. & van der Maarel, E. 1978. The Braun-Blanquet approach. In: Whittaker, R.H. (ed.) *Classification of plant communities*, pp. 289–399. W. Junk, The Hague, NL.
- Willner, W. 2001. Assoziationsbegriff und Charakterarten im Zeitalter der numerischen Klassifikation. *Berichte der Reinhold-Tüxen-Gesellschaft* 13: 35–52.
- Willner, W. 2006. The association concept revisited. *Phytocoenologia* 36: 67–76.

Received 4 April 2007;

Accepted 9 March 2008.

Co-ordinating Editor: M. Chytrý.