Journal of Vegetation Science 11: 167-178, 2000
© IAVS; Opulus Press Uppsala. Printed in Sweden

A new measure of fidelity and its application to defining species groups

Bruelheide, Helge

Albrecht-von-Haller-Institute for Plant Sciences, Department of Ecology and Ecosystem Research,
Untere Karspüle 2, D-37073 Göttingen, Germany; Fax +49 551 393556; E-mail hbruelh@gwdg.de

Abstract. The first objective of this paper is to define a new measure of fidelity of a species to a vegetation unit, called *u*. The value of *u* is derived from the approximation of the binomial or the hypergeometric distribution by the normal distribution. It is shown that the properties of *u* meet the requirements for a fidelity measure in vegetation science, i.e. (1) to reflect differences of a species' relative frequency inside a certain vegetation unit and its relative frequency in the remainder of the data set; (2) to increase with increasing size of the data set. Additionally (3), *u* has the property to be dependent on the proportion of the vegetation unit's size to the size of the whole data set.

The second objective is to present a method of how to use the value of *u* for finding species groups in large data bases and for defining vegetation units. A species group is defined by possession of species that show the highest value of *u* among all species in the data set with regard to the vegetation unit defined by this species group. The vegetation unit is defined as comprising all relevés that include a minimum number of the species in the species group. This minimum number is derived statistically in such a way that fewer relevés always belong to a species group than would be expected if the differential species were distributed randomly among the relevés. An iterative algorithm is described for detecting species groups in data bases. Starting with an initial species group, species composition of this group and the vegetation unit defined by this group are mutually optimized. With this algorithm species groups are formed in a data set independently of each other. Subsequently, these species groups can be combined in such a way that they are suited to define commonly known syntaxa *a posteriori*.

Keywords: Binomial distribution; Character species; COCKTAIL; Definition of a vegetation unit; Differential species; European Vegetation Survey; Hypergeometric distribution; Numerical method; Phytosociological data base; Vegetation classification.

Introduction

The aim of this study is to outline a method that uses species groups for numerical vegetation classification. Applications of this method have already been published (e.g. Bruelheide 1995; Bruelheide & Jandt 1995; Jandt 1999, in press), also comparisons to other methods (Bruelheide & Jandt 1997; Bruelheide & Chytrý 2000). The purpose of this paper is to describe the fundamentals of the method. The first part of this article gives a mathematical formulation of fidelity which is the basis for the second part where the extraction of species groups from a data set is described.

Fidelity is one of the most important concepts of the Braun-Blanquet (Zürich-Montpellier) approach. Generally speaking, fidelity is the degree to which a species is concentrated in a given vegetation unit. The fidelity of a species determines whether it can be considered a differential or character species or just a companion or accidental species. A character species can be interpreted as a special case of a differential species: a differential species shows a distinct accumulation of occurrences in one or more vegetation units; whereas a character species should accumulate in only one single vegetation unit (Barkman 1989: 106). The differential and character species combined are the diagnostic species that are used to delimit associations (Braun-Blanquet 1921: 323; Westhoff & van der Maarel 1973: 619; Dierschke 1994: 300).

The concept of fidelity is based on Braun-Blanquet (1918: 10) and was described in detail by Szafer & Pawlowski (1927: 4). For example, species are considered faithful to the highest degree, if they show a relative frequency (= presence degree) of between 61 and 100% inside the vegetation unit (constancy class IV and V; Westhoff & van der Maarel 1973: 647) and not more than 40 % outside this unit (constancy class II). Similarly, four more degrees of fidelity are defined. In addition, differences in the cover degree are considered in Szafer & Pawlowski's definition. This or a slightly modified form of fidelity classes is presented in almost every textbook on phytosociology (Becking 1957: 447; Braun-Blanquet 1964: 95; Westhoff & van der Maarel 1973: 655; Dierschke 1994: 277).

The concept of fidelity has raised fundamental problems because it involves a circular argument: a vegetation unit is defined by differential species (including character species) and at the same time differential species are those that show a preference for this vegetation unit. As Poore stated in 1955 (p. 239), this circularity is not solvable with logical arguments.

Although there is no way out of this dilemma, it might be easier to consider it a problem of optimization. In other words, all differential species defining a vegetation unit should be as faithful as possible. At the same time, only the most faithful differential species are used to define this vegetation unit. In principle, differential species and vegetation units can be optimized stepwise and vice versa.

In order to translate such an optimization into a distinct procedure or algorithm, a continuous measure of fidelity is needed that allows a refined comparison of the quality of a differential species. For this reason, the five classes of fidelity mentioned above are not sufficient. Moreover, the classes of Szafer & Pawlowski harbour some logical deficiencies. For example, they do not consider all possible combinations of differences in constancy classes; this produces serious gaps in the definition.

The need for a new measure of fidelity was already emphasized by Barkman (1989: 110): “What we need, however, is a not too complicated, logical, consistent criterium to define fidelity degrees and one that best expresses our ideas on the optima and role of species in plant communities.” In his proposal for a refinement, Barkman (1989: 115) defines a continuous measure based on the average cover value of a species in the vegetation unit under consideration related to its average cover value in another vegetation unit (TCR, total cover ratio). A disadvantage might be that cover values strongly affect the TCR. The question of whether presence/absence or abundance/cover is more appropriate for a floristically based classification has been the subject of dispute in vegetation science for many decades (Brockmann-Jerosch 1907: 244; Braun-Blanquet 1921: 322 and 1925: 127; Szafer & Pawlowski 1927: 3; Becking 1957: 420; Du Rietz 1936: 587; Whittaker 1973: 387).

There are some arguments for defining fidelity on the basis of presence/absence alone:

1. Cover values are subject to considerable seasonal dynamics and annual fluctuation within plant populations (e.g. Kaiser et al. 1998).
2. Consideration of abundance or cover implies a weighting of different species (Goodall 1973: 584). More importance is attributed to species that specifically occur with higher cover values than those with smaller cover values. Even the choice of cover as a measure of species performance, instead of abundance or biomass, is completely arbitrarily. Furthermore, the decision as to the degree to which the cover is taken into consideration is made arbitrarily (proportional, exponential etc.; van der Maarel 1979).
3. Large-scale vegetation surveys rely on the material of various authors. Since abundance and cover are usually estimated, they show a considerable amount of subjective variation (Tüxen 1972: 171; Lepš & Hadincová 1992). The restriction to presence/absence can minimize these individual deviations.

Fidelity is a relative measure: it compares occurrences of a species inside a vegetation unit with its occurrences in another vegetation unit. In principle, there are two possibilities for comparing the frequency of a differential species in the vegetation unit under consideration:

(1) regarding its frequency in the remainder of the data base, that means in all available relevés not belonging to the unit in question;

(2) regarding its frequency in other vegetation units, which either may be all other known vegetation units or the vegetation unit in which it shows the next highest frequency, often quoted as ‘floristically closest related vegetation unit’ (Becking 1957: 447; Barkman 1989: 109).

In this paper, fidelity is based on the first option: the reference set comprises all relevés in the data set apart from those in the vegetation unit under consideration.
The second option makes use of a species’ frequency in all other vegetation units. This definition exhibits another circularity since all possible vegetation units have to be known beforehand. In their fidelity-based classification approach, Brisse et al. (1995) solved this problem elegantly by defining each possible vegetation unit as being characterized by one of all the species in the French data base. This allows calculation of a species-by-species fidelity matrix, which is further used to define the discriminatory capabilities of plant species. In contrast to the stochastically based fidelity measure described below, Brisse et al. (1995) simply define fidelity as a relation of co-occurrences of species A and B to all occurrences of species A.

Definiton of fidelity

The stochastic basis

It seems appropriate to define fidelity using statistical tools (see Goodall 1973: 582). As suggested by Barkman (1989: 111) the frequency of occurrence of a species in a vegetation unit should be compared with a random distribution. A measure of departure from random distribution has also been used by Goodall (1953), Williams & Lambert (1959, 1960) and Feoli & Orlóci (1979), basing fidelity on the χ^2-value, and by Jancey (1979), suggesting to use the *F*-value of analysis of variance.

The starting point of this chapter is that the affiliation of a relevé to a vegetation unit is known. Likewise, the frequency of all species in the whole data set is known. The observed frequency is used to calculate the number of occurrences of a species to be expected within the vegetation unit if the species is distributed randomly within the data set.

$$\mu = P \cdot n = (N_p \,/\, N) \cdot n \qquad (1)$$

μ = Expected number of occurrences of a distinct species in the vegetation unit;
n = Number of all occurrences of the considered species in the data set;
P = N_p / N = Proportion of the vegetation unit's relevés in all relevés;
N_p = Number of relevés in the vegetation unit;
N = Number of all relevés in the data set.

When the theoretical distribution is known, the probability of all cases can be calculated in which the observed number n_p of occurrences of a species is higher than the expected number μ. In principle, two different distributions come into question.

The *binomial distribution* is appropriate when the chance of a relevé to contain a certain species is considered to be the same as for all other relevés in the vegetation unit. This means that the probability of a species occurring in a relevé is calculated *independently* from other relevés. In this case the probability α of observing n_p occurrences of a certain species or of a larger number than n_p (e.g. a more extreme distribution) in the vegetation unit is described by the cumulative binomial distribution function:

$$\alpha(x \geq n_p) = \sum_{x=n_p}^{N_p} \frac{\binom{n}{x} \cdot \binom{N-n}{N_{p-x}}}{\binom{N}{N_p}} \qquad (2)$$

The cumulative binomial distribution is appropriate as long as the probabilities of species occurrences refer to natural conditions, where the chance of encountering a certain species in the landscape is not influenced by previous encounters and is only dependent on the species' abundance or rarity in the area. In a data base the situation is different. Here the total number of occurrences of a species in the whole data set is limited. When calculating probabilities one has to consider that a species occurrence which has already been allocated to a certain relevé cannot be allocated to another one. With each allocation the chance of encountering the species in the rest of the data base decreases. Consequently, the probability of a species to occur in a relevé is *dependent* on occurrences in other relevés. Sampling from a finite population is described by the cumulative *hypergeometric distribution* function:

$$\alpha(x \geq n_p) = \sum_{x=n_p}^{N_p} \frac{\binom{n}{x} \cdot \binom{N-n}{N_{p-x}}}{\binom{N}{N_p}} \qquad (3)$$

As in the binomial distribution, the resulting α describes the transgression probability of finding n_p or more than n_p occurrences of a limited total number of occurrences of a species in the vegetation unit under consideration.

If the number of all occurrences of a species n exceeds 10, the exact calculation of the binomial distribution becomes more and more laborious. From then on, the normal distribution $\Phi(u)$ can be used for approximating both the binomial and the hypergeometric distribution (Molenaar 1970). The classic normal approximation is:

$$\Phi(u) = \Phi\left(\frac{|n_p - \mu|}{\sqrt{n \cdot P \cdot (1-P)}}\right) \qquad (4)$$

As will be demonstrated below, the argument u of the standard Gaussian distribution is a suited measure for fidelity.

Eq. (4) should be slightly modified for practical work. For small n (number of all occurrences of species) a good approximation is ensured by a term for continuity correction (Bortz et al. 1990; Molenaar 1970). This is performed by subtracting half a unit from the numerator.

The absolute symbols of the term $|n_p - \mu|$ only produce positive values, when the correction term is disregarded. Differences between n_p and μ less than 0.5 are set to 0 in order to prevent negative values of u possibly caused by the correction for continuity.

In order to decide if a species in the vegetation unit under consideration is more frequent or less frequent than expected, u is provided with a positive or negative sign by a SIGNUM term. The sign is negative if a species is more frequent than expected outside the vegetation unit.

$$u = \frac{|n_p - \mu| - 0.5}{\sqrt{n \cdot P \cdot (1-P)}} \cdot \mathrm{SIGNUM}(n_p - \mu)$$

$$\begin{aligned} &\mathrm{SIGNUM}(n_p - \mu) = 1, \text{ if } n_p > \mu + 0.5 \\ &\mathrm{SIGNUM}(n_p - \mu) = 0, \text{ if } n_p = \mu \pm 0.5 \\ &\mathrm{SIGNUM}(n_p - \mu) = -1, \text{ if } n_p < \mu + 0.5 \end{aligned} \qquad (5)$$

It should be noted that, mathematically, u is almost exactly equivalent to the square root of χ^2 (for df = 1 as in a 2 × 2 contingency table). The main reason to use the notation of u as a normal deviate rather than to use the χ^2 notation is that the dependencies on the size of the data set (N) and on the relative size of the vegetation unit (P) can be derived directly from formula (5) (see below).

Goodness of approximation by the normal distribution

The goodness of fit between normal approximation (4) and both the binomial (2) and hypergeometric (3) distributions is shown in Figs. 1 and 2 for various combinations of relative frequency inside and outside of

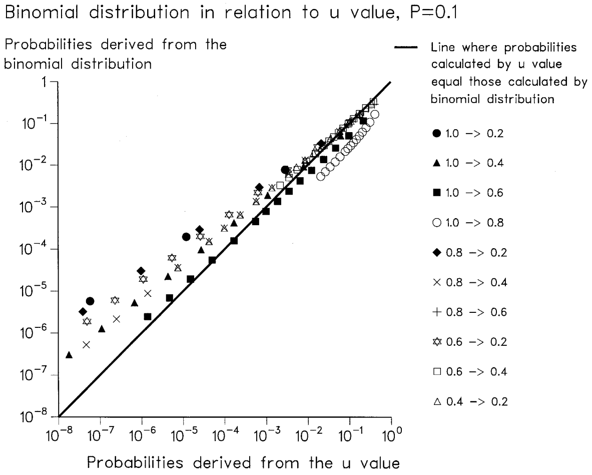

Fig. 1. Goodness of approximation of the binomial distribution by the normal distribution (value of *u*). Symbols refer to different combinations in relative frequency in the vegetation unit and in the rest of the database (e.g. 1.0 -> 0.2: 100 % in the unit, 20 % in the rest of the database). Subsequent points with increasingly smaller probabilities were created by expanding the size of the dataset (1st point: *N* = 50, 2nd *N* = 100, then 150, 200, …450, 500, 600, …900, 1000, 1200, …2000). Note that probabilities for certain relative frequency combinations could not be calculated for high *N* and that such points are missing.

a vegetation unit. The better the approximation, the closer the values are located to the diagonal in the diagram. For most combinations of relative frequencies of occurrence inside the vegetation unit and in the rest of the database (outside), the normal distribution (value of *u*) produces smaller probabilities than the binomial distribution (Fig. 1) but larger probabilities than the hypergeometric distribution (Fig. 2). The approximation by *u* shows large deviations from the binomial distribution when differences in relative frequency inside and outside the vegetation unit are large. In comparison, the approximation to the hypergeometric distribution produces largest deviations for small differences in relative frequency.

Some combinations yield only very poor approximations. It is possible to look for other, more sophisticated approximations, which are also provided in Molenaar (1970). However, Molenaar (1970) emphasizes that there is no optimal approximation for all given values of the parameters and argument. The same approximation may produce large errors near the median and small errors near the tail. Most recommendations refer to accuracy near the customary significance levels. In vegetation data bases, probabilities of species occurrences usually range much beyond any significance level. The typical situations are a high relative frequency inside the vegetation unit (60 to 80%) and a low relative frequency outside (less than 20%), yielding very low probabilities even with low size of the data set *N*. Note that for these frequency combinations, the classic normal approximation is quite satisfactory.

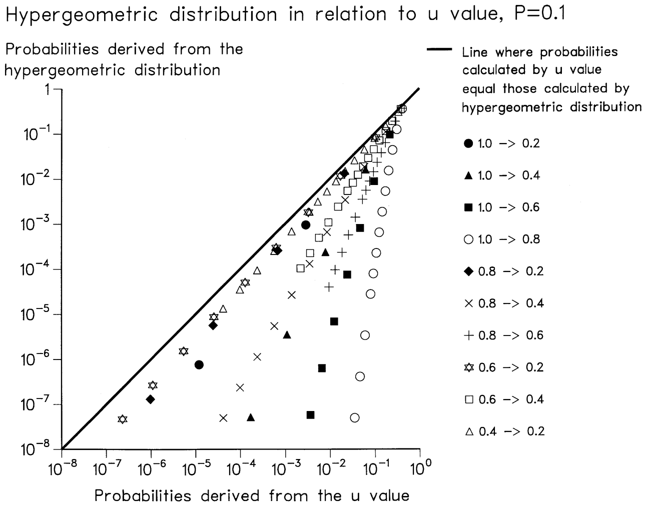

Fig. 2. Goodness of approximation of the hypergeometric distribution by the normal distribution (value of *u*). For details see Fig. 1.

The lack of a general goodness of fit should not be overemphasized because in large data bases the probability values are usually so small that they completely loose their importance as a measure for inferential statistics. However, they may still be used as a descriptive measure. Even very small probabilities allow the comparison of the fidelity of different species with regard to a certain vegetation unit. For this purpose, it is much more essential to use the same approximation for all species and all occuring frequency combinations in the data base than to find best approximations. In this respect, the value of *u* is advantageous because it represents the simplest of all approximations by normal distributions available.

The subsequent chapters do no longer refer to probability values but instead use *u* as the argument in the normal distribution. For practical reasons the values of *u*, which range in large data bases between 0 and 100, are much easier to handle than probability values with negative exponents. If desired, probabilities may be obtained from *u* by integrating the curve of the standard Gaussian distribution between the *x* values 0 and *u*.

u as a function of differences in frequency

To begin with, the absolute level of *u* will be disregarded and the relative differences between graphs in Fig. 3 discussed. The fidelity of a differential species proves to be higher, the more frequently it occurs inside the considered vegetation unit and the rarer it is outside

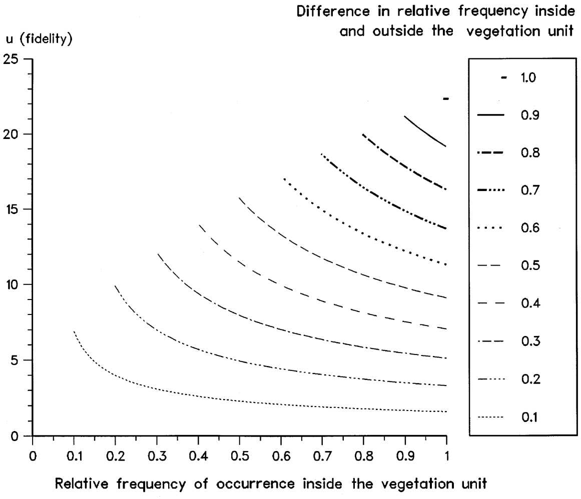

Fig. 3. *u* as a function of differences in relative frequency. The figure shows that fidelity increases, the higher the difference in relative frequency. Values refer to $P = 0.5$ and $N = 1000$.

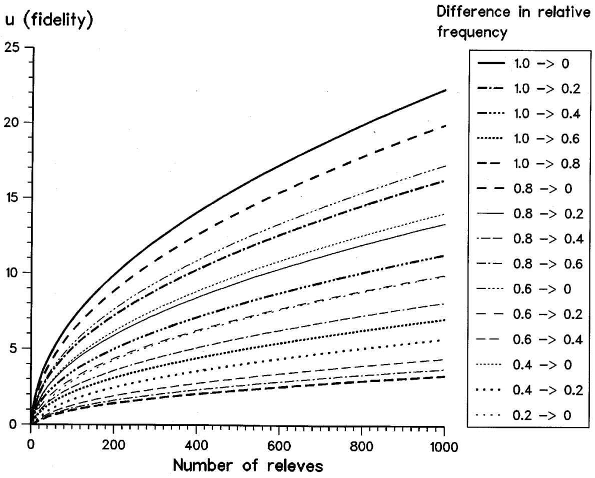

Fig. 4. *u* as a function of the number of relevés *N* in large datasets ($N \leq 1000$, $P = 0.5$). All combinations of relative frequency show increasing fidelity with increasing size of the dataset.

in the rest of the data set. The greater this difference, the higher is the value of *u* (Fig. 3).

For graphs with the same difference in relative frequency, Fig. 3 shows increasing *u* if the species is less frequent in the vegetation unit: this suggests that fidelity increases with rarity, while the difference in relative frequency remains the same. For example, a species with a difference in relative frequency of 0.3 and a degree of relative frequency inside the vegetation unit of 80% (i.e. 50% outside) is far less faithful than a differential species with 40% inside (and 10% outside). This very example was put forward by Barkman (1989: 111) but without statement of the exact figures for fidelity.

Likewise, *u* is appropriate for comparing differences in fidelity. In the example in Fig. 3a species with a difference in relative frequency of 0.4 and an 80%-degree of relative frequency inside the vegetation unit (i.e. 40% outside) is less faithful than a species with a difference of only 0.2 and an occurrence of 20% inside (and absent outside). This example is the confirmation of Lausi & Feoli's (1979) unproved remark that a highly faithful species does not have to be frequent in the considered vegetation unit when it is restricted to this unit. Nevertheless, they only used differential species of high relative frequency in their study about European salt marsh vegetation.

u as a function of the number of relevés in the data set N

Fig. 4 shows *u* for selected combinations of relative frequency inside and outside the vegetation unit depending on the size of the data set. The graphs' sequence essentially reflects traditional criteria for differential species in phytosociology (e.g. Szafer & Pawlowski 1927). Nevertheless, some fundamental differences are obvious:

All graphs ascend with *N*, implying higher fidelity with increasing extent of the data set taken as basis. This behaviour of *u* meets the requirement for considering the size of a data set for its validity (Barkman 1989: 111).

Above $N = 100$, the sequence of the graphs presented in Fig. 4 exhibits no further changes. Consequently, the larger the size of the data set *N*, the less the relation of differences is influenced by *N*. If the size of the data base is extended (by including relevés that have no effect on the distribution of the species under consideration) the species' absolute value of fidelity is raised, but its fidelity in relation to other species is only changed to a much lesser extent.

This change in relative fidelity as a function of *N* is due to the fact that the graphs do not ascend proportionally with respect to one another. Specifically, those graphs whose occurrences are confined to the vegetation unit with no occurrences outside it (1.0->0, 0.8->0 etc.) ascend more steeply than all the others. This means that species with restriction to a specific vegetation unit become more faithful than more scattered species with increasing size of the data base.

In small data sets the situation may be different (Fig. 5). For example, a species with relative frequency 0.8->0.2 is slightly more faithful than a species with 0.4->0 if the size of the data set is less than ca. 35 relevés. In a larger data base the latter becomes more faithful.

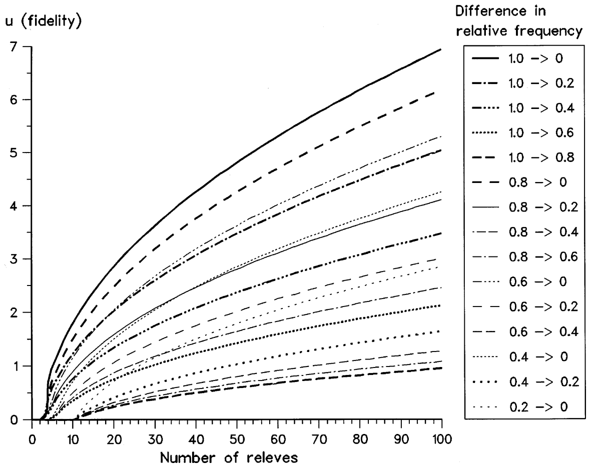

Fig. 5. *u* as a function of the number of relevés *N* in small datasets ($N \leq 100, P = 0.5$). Note the intersection of some of the graphs.

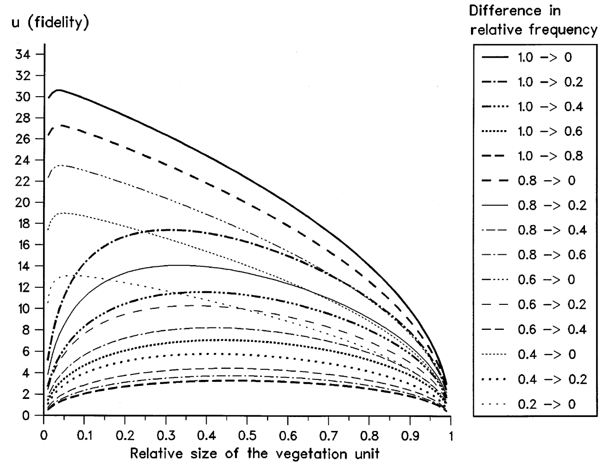

Fig. 6. *u* as a function of the relative size of the vegetation unit *P* (with *N* = 1000). All combinations of relative frequency show a maximum.

u as a function of the relative size of the vegetation unit P

The hitherto described properties of *u* correspond more or less with the traditional understanding of phytosociology, whereas the dependence on the relative size of the vegetation unit *P* is surprising. Fidelity of the combinations presented in Fig. 6 varies, depending on how many relevés of the data set belong to the vegetation unit under consideration (with *N* = constant). All graphs show decreasing fidelity when *P* approaches 1.00. In this case, the vegetation unit comprises almost the entire data set. The same applies to *P* approaching 0, where the vegetation unit has only very few relevés. In this range of *P*, a species that is equally faithful in a smaller vegetation unit needs to have a larger difference in relative frequency than a species in a larger vegetation unit. This can be illustrated by an example from Fig. 6: a species with a difference in relative frequency of 0.6 (the 1.0-> 0.4 graph) attains a fidelity of approximately *u* = 10 when the relative size of the vegetation unit is 20% (*P* = 0.2). If the relative size is reduced to 5% (*P* = 0.05), a fidelity of *u* = 10 corresponds to a difference in relative frequency of 0.8 (the 1.0->0.2 graph).

The fact that all graphs show a maximum implies a phenomenon that has gone unnoticed in vegetation science. In a given data set, vegetation units comprising a certain proportion of the data set are more likely to yield differential species with maximum fidelity than other proportions. In Table 1, maximum values of *u* are presented for data sets of various size and for various combinations of relative frequency. The smaller the difference between relative frequency inside and outside the vegetation unit, the more P_{max} is shifted towards *P* = 0.5: in other words, the maximum value of *u* is achieved if the relative size *P* of the vegetation unit comprises about half of the data set. With only small differences in relative frequency, P_{max} remains virtually constant if the data set is extended (*N* approaching 100 000). With large differences in relative frequency (1.0->0, 0.8->0 etc.), P_{max} is closer to 0, the larger the difference in relative frequency. In these cases, P_{max} decreases noticeably if the size of the data set *N* is enlarged. At the same time, the absolute value of *u* favours combinations with only few occurrences outside the vegetation unit. In Table 1 maximum *u* is achieved for *P* = 0.003, *N* = 100 000 and a difference in relative frequency of 1.0 -> 0. This corresponds to a vegetation unit of 300 relevés in which the species is

Table 1. Maximum values of *u* for data sets of various size *N* and various combinations of relative frequency. P_{max} is the relative size of the vegetation unit where maximum *u* is achieved.

	N = 100		*N* = 1000		*N* = 10 000		*N* = 100 000	
	P_{max}	*u*	P_{max}	*u*	P_{max}	*u*	P_{max}	*u*
1.0 -> 0	0.097	8.96	0.031	30.61	0.010	99.00	0.003	315.23
1.0 -> 0.2	0.320	5.37	0.310	17.43	0.309	55.26	0.309	174.80
1.0 -> 0.4	0.395	3.55	0.388	11.58	0.388	36.74	0.387	116.22
1.0 -> 0.6	0.443	2.14	0.437	7.09	0.436	22.53	0.437	71.28
1.0 -> 0.8	0.477	0.95	0.473	3.31	0.472	10.55	0.472	33.38
0.8 -> 0	0.108	7.90	0.035	27.27	0.011	88.44	0.004	281.83
0.8 -> 0.2	0.345	4.31	0.335	14.09	0.333	44.70	0.333	141.42
0.8 -> 0.4	0.423	2.49	0.415	8.24	0.414	26.18	0.414	82.84
0.8 -> 0.6	0.470	1.08	0.465	3.75	0.464	11.97	0.464	37.89
0.6 -> 0	0.124	6.69	0.040	23.48	0.013	76.45	0.004	243.95
0.6 -> 0.2	0.380	3.10	0.367	10.30	0.366	32.72	0.366	103.52
0.6 -> 0.4	0.459	1.28	0.451	4.45	0.449	14.20	0.450	44.94
0.4 -> 0	0.151	5.26	0.049	18.98	0.016	62.24	0.005	199.00
0.4 -> 0.2	0.430	1.66	0.416	5.80	0.414	18.51	0.414	58.57
0.2 -> 0	0.208	3.38	0.069	13.11	0.022	43.71	0.007	140.42

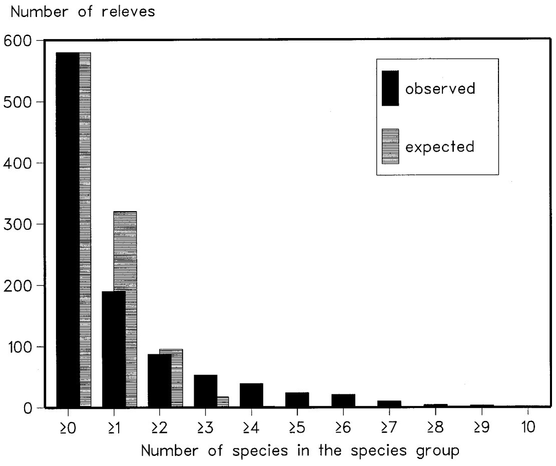

Fig. 7. Observed and expected distribution function exemplified by the species group of step 9 in Table 2. Note the point of intersection between a minimum number of 2 and 3.

present in all relevés and to a complete absence of the species outside these relevés. In terms of phytosociology, such a species would be called a perfect character species. The properties of *u* (1) to achieve a maximum with small *P* and with as few occurrences as possible outside and (2) to show a amplification of this trend with increasing size of the data set coincide with the demands for a character species.

Considering all its properties, *u* is a very appropriate measure of fidelity, and thus, allows detection of the most faithful species for a vegetation unit of any given size in any given data base.

As will be shown below, these faithful species can be used as differential species to define vegetation units. It is appropriate to use only species as differential species that exceed a specific minimum fidelity. This minimum fidelity is called 'threshold *u*' by Bruelheide & Jandt (1995: 322). In selecting the threshold *u*, the size and composition of the group of differential species are determined indirectly. Examples for species groups defined in a such a way are presented by Bruelheide (1995) Bruelheide & Jandt (1995) and Jandt (1999).

Extracting species groups from a data base

Using differential species to define vegetation units

The previous section described a method of finding differential species for a given vegetation unit; this section tackles the question of how to exactly define this vegetation unit using differential species.

Under the condition that a group of differential species has to be used for definition, various possibilities must be considered. In the most simple case, a vegetation unit is defined by a distinct combination of differential species. A relevé thus belongs to a vegetation unit if all differential species (2, 3, 4, etc.) are present simultaneously. In this case, increasingly few relevés meet the demand for having all differential species, the greater the number of differential species. This can be concluded from combinatorial calculations that specify the probability for distinct species combinations (Bruelheide 1995: 37). As Pignatti (1980: 182) stated, this probability becomes "lower and lower with each new species added to the combination". Consequently, the possible number of relevés that meet this requirement becomes lower and lower.

These considerations contradict the requirement of using as many differential species as possible to define a vegetation unit (Braun-Blanquet 1921: 322). Therefore, another course is followed: only a distinct minimum number out of all differential species under consideration has to be present in a relevé to allocate it to the vegetation unit in question. Since this minimum number can be realized by different species combinations, many possibilities for occurrence of such relevés are computed in data bases. For example, if three out of four differential species (a, b, c, d) are required for a relevé to belong to the vegetation unit, there are five possible combinations (abc, abd, acd, bcd and abcd). In demanding two out of four differential species, there are even 11 possible combinations. Since each combination corresponds to a distinct probability, the probability for any given minimum number in any given species group can be calculated recurrently (Bruelheide 1995: 38). With known frequency of differential species in the whole data set, the number of relevés that are expected to have this minimum number can be calculated. Using an example from Bruelheide (1995: 138, *Helianthemum ovatum*-group), Fig. 7 shows the expected minimum numbers of relevés. Based on the frequency of the 13 species in the data set (N = 580 relevés) and on the supposition of independent distribution of these species amongst all relevés, 300 relevés are expected to have one or more of these 13 species. With an increasing minimum number of species the number of expected relevés decreases drastically.

In Fig. 7, the expected number of relevés is compared with the actually observed number of relevés. Both graphs can be characterized as cumulative distribution functions. Expected and observed distribution intersect between minimum number 2 and 3. This point of intersection can be used to derive the minimum number of species from the condition for the vegetation unit to contain more relevés with this minimum number than would be expected under random distribution. Formulated mathematically, the minimum number has to be only slightly greater than the point of intersection between observed and expected distribution function. The example in Fig. 7 shows that a number of at least three out of 13 species of the species group is required. All the relevés of the data set are allocated to the vegetation unit defined by the species group that have at least three of the group's species.

The COCKTAIL algorithm for extracting species groups

The previous chapters have shown that (1) differential species can be detected for a given vegetation unit by calculating the value of *u* and (2) a vegetation unit can be defined by a given group of differential species. Now, the problem of circularity addressed in the introduction can be avoided by using an algorithm of optimization. This algorithm is implemented in the program COCKTAIL (Bruelheide 1995). There are two possibilities to start the algorithm:

la. To start with a *preselected vegetation unit* is useful when a group of relevés in a data base is known to describe a specific vegetation unit, e.g. the typical relevés in this unit including the type relevé. Then the algorithm begins calculating all species' fidelities to that type and takes the species with the highest value of u as the starting species group.

1b. To start with a *preselected species group* is useful when certain species combinations are to be tested for their ability to produce species groups. Such combinations may be obtained either from previous numeric analyses, e.g. from a species similarity matrix, or by knowledge from literature, e.g. lists of character or differential species. Any species combination may be tested although not all will produce species groups as defined in the previous chapter. The algorithm can proceed from a single species or from a group of *k* species.

2. The number of species of the species group is counted in each relevé. Expected and observed cumulative distribution functions for relevés having 0, 1, 2, ...*k* species are calculated. The distributions' intersection defines the required minimum number *m* of species a relevé has to contain in order to belong to the vegetation unit. The vegetation unit is defined by all relevés having *m* or more species belonging to the species group. The algorithm aborts if there is no intersection between observed and expected cumulative distribution, which is the case when species with fewer co-occurrences than expected were chosen as starting species group.

3. The occurrences of each species inside this vegetation unit are counted and the value of *u* is calculated.

4. For all species actually included in the species group the procedure tests whether the *u* of all of them exceeds the (initially) fixed threshold *u*. If yes, the algorithm proceeds to step 5; if not, there are two possibilities:

4a. One of the initially selected species does not exceed the threshold *u*. In this case, the species group is rejected and the algorithm aborts.

4b. The last species added to the species group has caused one of the species' *u* value to decrease below the threshold *u*. In this case this species is removed and marked for exclusion in the next cycle. The algorithm does not try to add this species to the species group until the species group has been changed by adding another species.

5. All species not belonging to the species group are sorted according to their *u* value. If there are species among them which exceed the threshold *u*, the algorithm proceeds to step 6. If no species, other than those that are already members of the species group and those that are to be disregarded, exceeds the threshold *u*, the algorithm stops. The species group is optimized when all species that are more faithful than the preset fidelity threshold have been included in the species group. The species group cannot be optimized when there are still species marked for exclusion which have high fidelity but which cannot be included in the species group without decreasing the other species' fidelity below the threshold *u*.

6. The species group is enlarged by including the species with the highest value of *u*. The algorithm proceeds stepwise and only includes one species at a time. It continues with step 2.

Note that step 4b is necessary to guarantee that the species group's composition is not changed to such a degree by newly included species that the initial species are no longer the ones with the highest fidelity. Without this restriction, the species group would change its initial composition and the algorithm would yield only one solution to form an optimal species group, i.e. a group where fidelity is maximized. With the restriction 4b, the formation of more than one species group is possible; such groups would differ in species composition and in the maximal fidelity attained by a species in the species group. Although this restriction allows for the formation of a number of species groups, not every species group can be optimized. This is the case if species which do not co-occur more than expected are chosen for a starting group (see step 2).

Table 2 gives an example for the algorithm, starting with two species,*Viola canina* and *Galium pumilum*, and a threshold *u* of 9.00. Note that each newly included species increases the size of the vegetation unit reaching its maximum size in step 8 with 56 relevés. In step 9 the required minimum number of species a relevé has to contain shifts from 2 to 3, thus decreasing the size of the vegetation unit from 87 to 53 relevés. This increase in the minimum number was the result of including *Knautia arvensis*, a species with a high absolute frequency. Note that not all species which exceed the threshold *u* are included in the species group. For example, *Genista*

Table 2. Example of optimizing a species group in the dataset of Bruelheide (1995) with $N = 580$. The starting species group comprised *Viola canina* and *Galium pumilum*, which showed highest fidelity in a species-by-species fidelity matrix. On the left: distribution of relevés among the classes 0, 1, 2, … k species of the species group. The horizontal line marks the intersection between observed (obs.) and expected (exp.) cumulative distribution function. The minimum number of species required for a relevé to be assigned to the vegetation unit is marked in grey. On the right: Number of relevés in the vegetation unit (VU) and in the rest of the data set and the the number of occurrences of each species in these two categories for all species transgressing the threshold (set at u = 9.00). The species actually belonging to the species group are marked in grey.

1. Start with two species

Nr of relevés	**obs.**	**exp.**	**Start**	**rest**	**VU**	***u***
with ≥ 0	580	580	Nr of relevés	562	18	
with ≥ 1	57	72	*Viola canina*	10	18	18.12
with ≥ 2 species	18	2	*Galium pumilum*	29	18	13.49
			Helianthemum ovatum	17	10	9.61

2. After including *Helianthemum ovatum*

Nr of relevés	**obs.**	**exp.**	1.	**rest**	**VU**	***u***
with ≥ 0	580	580	Nr of relevés	556	24	
with ≥ 1	68	96	*Viola canina*	8	20	17.40
with ≥ 2	24	5	*Galium pumilum*	25	22	14.32
with ≥ 3 species	10	0	*Helianthemum ovatum*	11	16	13.90
			Thymus pulegioides	9	12	11.65
			Festuca ovina	31	17	10.52
			Genista tinctoria	10	10	9.74
			Scabiosa columbaria	2	6	9.18

3. After including *Thymus pulegioides*

Nr of relevés	**obs.**	**exp.**	2.	**rest**	**VU**	***u***
with ≥ 0	580	580	Nr of relevés	549	31	
with ≥ 1	70	114	*Thymus pulegioides*	2	19	16.86
with ≥ 2	31	9	*Viola canina*	7	21	15.97
with ≥ 3	17	0	*Helianthemum ovatum*	7	20	15.45
with ≥ 4 species	5	0	*Galium pumilum*	23	24	13.61
			Festuca ovina	25	23	12.79
			Avena pratensis	3	9	10.09
			Alchemilla glaucescens	11	12	9.52
			Genista tinctoria	9	11	9.38
			Hieracium pilosella	26	16	9.09

4. After including *Festuca ovina*

Nr of relevés	**obs.**	**exp.**	3.	**rest**	**VU**	***u***
with ≥ 0	580	580	Nr of relevés	542	38	
with ≥ 1	88	153	*Helianthemum ovatum*	2	25	17.68
with ≥ 2	38	18	*Thymus pulegioides*	1	20	15.98
with ≥ 3	27	1	*Festuca ovina*	18	30	15.37
with ≥ 4	14	0	*Viola canina*	6	22	15.02
with ≥ 5 species	4	0	*Galium pumilum*	23	24	12.04
			Avena pratensis	1	11	11.33
			Alchemilla glaucescens	10	13	9.26
			Genista tinctoria	8	12	9.21
			Hieracium pilosella	24	18	9.20

5. After including *Avena pratensis*

Nr of relevés	**obs.**	**exp.**	4.	**rest**	**VU**	***u***
with ≥ 0	580	580	Nr of relevés	541	39	
with ≥ 1	88	161	*Helianthemum ovatum*	2	25	17.43
with ≥ 2	39	20	*Thymus pulegioides*	1	20	15.76
with ≥ 3	30	1	*Festuca ovina*	18	30	15.14
with ≥ 4	18	0	*Viola canina*	6	22	14.80
with ≥ 5	6	0	*Galium pumilum*	22	25	12.43
with ≥ 6 species	2	0	*Avena pratensis*	0	12	12.33
			Alchemilla glaucescens	10	13	9.12
			Genista tinctoria	8	12	9.07
			Hieracium pilosella	24	18	9.04

6. After including *Alchemilla glaucescens*

Nr of relevés	**obs.**	**exp.**	5.	**rest**	**VU**	***u***
with ≥ 0	580	580	Nr of relevés	538	42	
with ≥ 1	95	178	*Helianthemum ovatum*	2	25	16.74
with ≥ 2	42	26	*Viola canina*	4	24	15.66
with ≥ 3	31	2	*Thymus pulegioides*	1	20	15.14
with ≥ 4	22	0	*Festuca ovina*	17	31	15.05
with ≥ 5	12	0	*Galium pumilum*	22	25	11.87
with ≥ 6	3	0	*Avena pratensis*	0	12	11.84
with ≥ 7 species	1	0	*Alchemilla glaucescens*	7	16	11.13
			Pimpinella saxifraga	38	24	9.32
			Hieracium pilosella	23	19	9.20

7. After including *Pimpinella saxifraga*

Nr of relevés	**obs.**	**exp.**	6.	**rest**	**VU**	***u***
with ≥ 0	580	580	Nr of relevés	527	53	
with ≥ 1	122	221	*Viola canina*	2	26	15.05
with ≥ 2	53	42	*Helianthemum ovatum*	2	25	14.72
with ≥ 3	38	4	*Festuca ovina*	15	33	14.08
with ≥ 4	24	0	*Thymus pulegioides*	1	20	13.31
with ≥ 5	20	0	*Pimpinella saxifraga*	27	35	12.71
with ≥ 6	7	0	*Galium pumilum*	18	29	12.25
with ≥ 7	3	0	*Alchemilla glaucescens*	4	19	11.87
with ≥ 8 species	1	0	*Avena pratensis*	0	12	10.42
			Erophila verna	5	14	9.37
			Danthonia decumbens	17	20	9.20

8. After including *Erophila verna*

Nr of relevés	**obs.**	**exp.**	7.	**rest**	**VU**	***u***
with ≥ 0	580	580	Nr of relevés	524	56	
with ≥ 1	124	232	*Viola canina*	2	26	14.59
with ≥ 2	56	47	*Helianthemum ovatum*	2	25	14.27
with ≥ 3	41	5	*Festuca ovina*	15	33	13.62
with ≥ 4	27	0	*Pimpinella saxifraga*	25	37	13.12
with ≥ 5	21	0	*Thymus pulegioides*	1	20	12.91
with ≥ 6	11	0	*Alchemilla glaucescens*	3	20	12.20
with ≥ 7	4	0	*Galium pumilum*	18	29	11.83
with ≥ 8	3	0	*Erophila verna*	2	17	11.39
with ≥ 9 species	0	0	*Avena pratensis*	0	12	10.11
			Knautia arvensis	97	47	9.20

9. After including *Knautia arvensis*

Nr of relevés	**obs.**	**exp.**	8.	**rest**	**VU**	***u***
with ≥ 0	580	580	Nr of relevés	527	53	
with ≥ 1	190	320	*Helianthemum ovatum*	2	25	14.72
with ≥ 2	87	95	*Viola canina*	3	25	14.39
with ≥ 3	53	17	*Festuca ovina*	16	32	13.58
with ≥ 4	39	2	*Thymus pulegioides*	1	20	13.31
with ≥ 5	24	0	*Alchemilla glaucescens*	3	20	12.59
with ≥ 6	21	0	*Pimpinella saxifraga*	28	34	12.27
with ≥ 7	10	0	*Erophila verna*	2	17	11.75
with ≥ 8	4	0	*Galium pumilum*	19	28	11.75
with ≥ 9	3	0	*Avena pratensis*	0	12	10.42
with ≥ 10 species	0	0	*Knautia arvensis*	97	47	9.64

tinctoria exceeded the threshold in steps 2 - 5 but fell below u = 9.00 after *Alchemilla glaucescens* was included in step 6. After step 9 the species group is optimized: all species with a value of u above 9.00 are included; no other species exceeds this threshold.

The final result of the optimization process is a species group containing all species with high fidelity: no species with comparably high fidelity remains outside this group. When starting with preselected vegetation units, these units are optimized in such a way that they are defined by differential species groups *a posteriori*. In doing so, the final compositon of relevés in the vegetation unit may be somewhat different than at the beginning. It should be pointed out that not all presumed syntaxa can actually be defined by groups of differential species. Many syntaxa do not have any differential species or are defined by dominance of a species rather than by floristic composition.

Combinations of species groups

In a given data set, this optimization method yields various groups of differential species that are formed independently of each other. In further steps, these species groups can be evaluated in various ways, e.g. by correlation with ecological measurements. Another possibility is to combine them to further subdivide the data set.

For instance, species groups can be used to reproduce existing classifications (syntaxa of the Braun-Blanquet system). In this connection, note that a relevé can be allocated to several species groups but only to one single syntaxon on a given level of the hierarchy of the Braun-Blanquet system. For example, in grassland relevés in the Harz mountains, 24 species groups were encountered that, together, would allow $2^{24} = 16777216$ possible combinations (Bruelheide 1995: 152). These combinations had to be assigned to six syntaxa on the alliance level. The definition of a syntaxon needs not to be limited only to require the presence of distinct species groups but also to require its absence. For example, it is common practice to demand the absence of certain *Molinio-Arrhenatheretea* species in order to define the *Violion caninae* communities or to not allow *Calthion palustris* species to occur in *Caricion fuscae* communities.

When assigning combinations, the combination of species groups that corresponds to a syntaxon must be clearly stated. This can be best performed by using formal logic as was suggested by Bruelheide (1997). For future classifications covering large areas and different vegetation types, the development of expert systems would be desirable. Given a number of species groups in a large data set and the known assignment of a set of reference relevés to specific syntaxa, certain algorithms of machine learning can be adapted to produce general assignment rules using a minimum number of conditions (Ziarko & Shan 1996).

Discussion

At present, there is a clear need for the evaluation of methods in the context of designing vegetation classification on a European scale. The European Vegetation Survey, whose objective is to unify existing national classifications (Dierschke 1992) has to deal with more than 1 million relevés (Rodwell 1995). Although the TURBOVEG computer program (Hennekens 1996; Schaminée & Hennekens 1995) offers a highly efficient data base management package, there is unclarity about the methodology of evaluation (Mucina et al. 1993).

The method described in the current paper is suitable for detecting species groups in large data bases by using the species' distribution in all relevés. Moreover, laying down the basic criteria and establishing an optimization procedure ensure a high degree of operationalization which renders results reproducable.

To date, experience in defining syntaxa by means of species groups has been gathered in various plant communities. Species groups have been computed for montane grassland syntaxa (*Polygono-Trisetion flavescentis*, *Arrhenatherion elatioris*, *Violion caninae*, *Calthion palustris* and *Caricion nigrae*; Bruelheide 1995), for limestone grasslands on a regional scale (*Gentiano-Koelerietum*; Bruelheide & Jandt 1995) and on a national scale (*Festuco-Brometea*; Jandt 1999, in press), for floodplain meadows (*Cnidion*; Burkart 1998), for deciduous forests (*Fagion*; Pflume 1999) and for swamp forests (*Alnion glutinosae* and *Alno-Ulmion*; Mast 1999). Further applications are under development. In all of them, the species group method has proved its merit in redefining vegetation types of the Braun-Blanquet system which, in most cases, had not been clearly defined before.

The method was tested systematically against agglomerative cluster techniques and traditional character-species-based classification in randomly compiled test data sets (Bruelheide & Jandt 1997). The species group method proved to produce stable results. Although the degree of correct assignment was somewhat lower than with other methods, the species group method's main advantage was to produce robust allocation criteria that allowed to clearly discriminate the communities in the data set. The feature of the species group method to extract allocation criteria is similar to the abilities of TWINSPAN (Hill 1979). The two methods were compared in their ability to classify wet meadows (*Calthion*, incl. *Filipendulenion*) in a large data set (9196 relevés) from Germany and the Czech Republic (Bruelheide & Chytrý 2000). TWINSPAN revealed almost no correspondence between the classifications derived from the two national data sets; whereas COCKTAIL produced a fairly good correspondence.

With the transition from smaller to larger scales (from local through regional and national up to the European level, and from small to large data bases), the demands for an appropriate classification method will change. One aspect is robustness when the data set is enlarged. Many classification programs have proved to be very sensitive even to slight changes in the data set, sometimes even to a few outliers that had been included inadvertently. According to the experience in the Working Group for the European Vegetation Survey such outliers are no exception but the rule in large data sets. To remove outliers in data sets of several thousand relevés will be neither possible nor desired. Similarly, other features of real data render classifications in large data sets more difficult. For example, TWINSPAN has proved to be very sensitive to unbalanced data sets (Bruelheide & Chytrý 2000).

Another aspect for providing applicable vegetation classifications is simplicity. The European Vegetation Survey relies on the participation of many scientists who have to make their results available to others. Therefore, the classification criteria should be clear and transferable from one data set to another.

The species group method represents a very flexible approach. It allows the use of the immense amount of classification data accumulated by phytosociology in the 20th century to redefine existing units. It has proved to be able to produce robust results in data sets of various size, a feature which is mainly due to the fidelity measure taken as its basis. The statements about the characteristics of *u*, especially its dependence on the relative size of vegetation units, indicates that regional species groups do not necessarily loose their validity in large data bases. This was proved when classifications were transferred from regional to national scale or vice versa (Jandt 1999). Therefore, it could be possible to find species groups of similar species composition on regional as well as on European scale. This would overcome one of the reservations often raised against fidelity, i.e. the problem of its limited geographical validity (Braun-Blanquet 1921: 319; Du Rietz & Gams 1924: 279; Westhoff & van der Maarel 1973: 658).

Acknowledgements. I thank Mrs. Ute Jandt and Mr. Thomas Flintrop for constructive discussions on the *u*-value over the last six years. I am very grateful to Laco Mucina and four anonymous referees for their very helpful comments and justified criticism on the previous version, to Mike Dale and Milan Chytrý who encouraged me with many useful ideas to rewrite the paper and to D.F. Whybrew who polished my English.

References

Barkman, J.J. 1989. Fidelity and character-species, a critical evaluation. *Vegetatio* 85: 105-116.

Becking, R.W. 1957. The Zürich-Montpellier school of phytosociology. *Bot. Rev.* 23: 411-488.

Bortz, J., Lienert, G.A. & Boehnke, K. 1990. *Verteilungsfreie Methoden in der Biostatistik.* Springer, Berlin.

Braun-Blanquet, J. 1918. Eine pflanzensoziologische Exkursion durchs Unterengadin und in den schweizerischen Nationalpark. *Beitr. Geobot. Landesaufn. Schweiz* 4: 1-80.

Braun-Blanquet, J. 1921. Prinzipien einer Systematik der Pflanzengesellschaften auf floristischer Grundlage. *Jahrb. St. Gall. Naturwiss. Ges.* 57: 305-351.

Braun-Blanquet, J. 1925. Zur Wertung der Gesellschaftstreue in der Pflanzensoziologie. *Vierteljahrsschr. Naturforsch. Ges. Zürich* 70: 122-149.

Braun-Blanquet, J. 1964. *Pflanzensoziologie. Grundzüge der Vegetationskunde.* 3rd. ed. Springer, Berlin.

Brisse, H., De Ruffray, P., Grandjouan, G. & Hoff, M. 1995: The phytosociological data base ‘SOPHY’. *Ann. Bot.* (*Roma*) 53: 177-223.

Brockmann-Jerosch, H. 1907. *Die Pflanzengesellschaften der Schweizeralpen. I. Teil. Die Flora des Puschlav (Bezirk Bernina, Kanton Graubünden) und ihre Pflanzengesellschaften.* Engelmann, Leipzig.

Bruelheide, H. 1995. Die Grünlandgesellschaften des Harzes und ihre Standortsbedingungen. Mit einem Beitrag zum Gliederungsprinzip auf der Basis von statistisch ermittelten Artengruppen. *Diss. Bot.* 244: 1-338.

Bruelheide, H. 1997. Using formal logic to classify vegetation. *Folia Geobot. Phytotax.* 32: 41-46.

Bruelheide, H. & Chytrý, M. 2000. Towards unification of national vegetation classifications: a comparison of two methods for analysis of large data sets. *J. Veg. Sci.* 11: 295-306.

Bruelheide, H. & Jandt, U. 1995. Survey of limestone grassland by statistically formed groups of differential species. *Coll. Phytosociol.* 23: 319-338.

Bruelheide, H. & Jandt, U. 1997. Demarcation of communities in large data bases. *Phytocoenologia* 27: 141-159.

Burkart, M. 1998. Die Grünlandvegetation der unteren Havelaue in synökologischer und syntaxonomischer Sicht. *Arch. Naturw. Diss.* 7: 1-157.

Dierschke, H. 1992. European Vegetation Survey – ein neuer Anlauf für eine Übersicht der Pflanzengesellschaften Europas. *Tuexenia* 12: 381-383.

Dierschke, H. 1994. *Pflanzensoziologie. Grundlagen und Methoden.* Ulmer, Stuttgart.

Du Rietz, G.E. 1936. Classification and nomenclature of vegetation units 1930-1935. *Sven. Bot. Tidskr.* 30: 580-589.

Du Rietz, G.E. & Gams, H. 1924. Zur Bewertung der Bestandestreue bei der Behandlung der Pflanzengesellschaften. *Vierteljahrsschr. Nat. forsch. Ges. Zuer.* 69: 269-280.

Feoli, E. & Orlóci, L. 1979. Analysis of concentration and detection of underlying factors in structured tables. *Vegetatio* 40: 49-54.

Goodall, D.W. 1953. Objective methods for the classification of vegetation. II. Fidelity and indicator value. *Aust. J. Bot.* 1: 434-456.

Goodall, D.W. 1973. Numerical classification. In: Whittaker, R.H. (ed.) *Ordination and classification of communities,* pp. 575-615. Junk, The Hague.

Hennekens, S.M. 1996. *TURBO(VEG). Software package for input, processing, and presentation of phytosociological data.* IBN-DLO, Wageningen & University of Lancaster.

Hill, M.O. 1979. *TWINSPAN – A FORTRAN program for arranging multivariate data in an ordered two-way table by classification of the individuals and attributes.* Cornell University, Ithaca, NY.

Jancey, R.C. 1979. Species ordering on a variance criterion. *Vegetatio* 39: 59-63.

Jandt, U. 1999. Kalkmagerrasen am Südharzrand und im Kyffhäuser. Gliederung im überregionalen Kontext, Verbreitung, Flora und Standortsverhältnisse. *Diss. Bot.* 322: 1-246.

Jandt, U. In press. *Application of the species group method to the data base of calcareous grasslands in Germany*. Proc. I.A.V.S. Symposium Uppsala.

Kaiser, T., Baier, V., Grünewald, I. & Haas, S. 1998. Erfassungsdefizite bei Vegetationsaufnahmen mesophiler Laubwälder in Abhängigkeit vom Aufnahmezeitpunkt. *Tuexenia* 18: 51-61.

Lausi, D. & Feoli, E. 1979. Hierarchical classification of European salt marsh vegetation based on numerical methods. *Vegetatio* 39: 171-184.

Lepš, J. & Hadincová, V. 1992. How reliable are our vegetation analyses. *J. Veg. Sci.* 3: 119-124.

Mast, R. 1999. *Vegetationsökologische Untersuchung der Feuchtwald-Gesellschaften im niedersächsischen Bergland - mit einem Beitrag zur Gliederung der Au-, Bruch- und Moorwälder in Mitteleuropa.* PhD Thesis Univ. Göttingen.

Molenaar, W. 1970. Approximations to the Poisson, binomial and hypergeometric distribution functions. *Math. Centre Tracts* 31: 1-159.

Mucina, L., Rodwell, J.S., Schaminée, J.H.J. & Dierschke, H. 1993. European Vegetation Survey: Current state of some national programmes. *J. Veg. Sci.* 4: 429-438.

Pflume, S. 1999. *Laubwaldgesellschaften im Harz - Gliederung, Ökologie und Verbreitung.* PhD Thesis Univ. Göttingen.

Pignatti, S. 1980. Reflections on the phytosociological approach and the epistemological basis of vegetation science. *Vegetatio* 42: 181-185.

Poore, M.E.D. 1955. The use of phytosociological methods in ecological investigations. I. The Braun-Blanquet system. *J. Ecol.* 43: 226-244.

Rodwell, J.S. 1995. The European Vegetation Survey questionnaire: an overview of phytosociological data, vegetation survey programmes and data bases in Europe. *Ann. Bot. (Roma)* 53: 87-98.

Rodwell, J.S., Pignatti, S., Mucina, L. & Schaminée, J.H.J. 1995. European Vegetation Survey: Update on progress. *J. Veg. Sci.* 6: 759-762.

Schaminée, J.H.J. & Hennekens, S.M. 1995. Update on the installation of Turboveg in Europe. *Ann. Bot. (Roma)* 53: 159-161.

Szafer, W. & Pawlowski, B. 1927. Die Pflanzenassoziationen des Tatra-Gebirges. A. Bemerkungen über die angewandte Arbeitstechnik. In: Szafer, W., Kulczynski, B., Pawlowski, B., Stecki, K. & Sokolowski, A.W. (eds.) *Die Pflanzenassoziationen des Tatra-Gebirges.* III., IV. und V. Teil. pp. 1-12. Bull. Int. Acad. Polon. Sci. Lettres B 3, Suppl. 2, Cracovie.

Tüxen, R. 1972. Kritische Bemerkungen zur Interpretation pflanzensoziologischer Tabellen. In: van der Maarel, E. & Tüxen, R. (eds.) *Grundfragen und Methoden in der Pflanzensoziologie*, pp. 168-182. Junk, The Hague.

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

Westhoff, V. & van der Maarel, E. 1973. The Braun-Blanquet approach. In: Whittaker, R.H. (ed.) *Ordination and classification of communities,* pp. 617-737. Junk, The Hague.

Whittaker, R.H. 1973. Dominance-types. In: Whittaker, R.H. (ed.) *Ordination and classification of communities,* pp. 387-402. Junk, The Hague.

Williams, W.T. & Lambert, J.M. 1959. Multivariate methods in plant ecology. I. Association-analysis in plant communities. *J. Ecol.* 47: 83-101.

Williams, W.T. & Lambert, J.M. 1960. Multivariate methods in plant ecology. II. The use of an electronic digital computer for association-analysis. *J. Ecol.* 48: 689-710.

Ziarko, W. & Shan, N. 1996. A method for computing all maximally general rules in attribute-value systems. *Computational Intelligence* 12: 223-234.

Received 10 March 1999;
Revision received 16 September1999;
Accepted 13 October 1999;
Coordinating Editor: L. Mucina.