Environmental Conservation 33 (2): 118–127 © 2006 Foundation for Environmental Conservation doi:10.1017/S0376892906002967

The influence of sampling intensity on vegetation classification and the implications for environmental management

A. COOPER[1]*, T. MCCANN[1] AND R. G. H. BUNCE[2]
[1]*School of Biological and Environmental Sciences, University of Ulster, Coleraine BT52 1SA, Northern Ireland, and* [2]*ALTERRA, Wageningen, the Netherlands*
Date submitted: 11 March 2005 Date accepted: 17 March 2006 First published online: 7 July 2006

SUMMARY

As part of a programme of landscape-scale habitat surveillance in the United Kingdom (UK), the effect of grassland sampling intensity on the outcome of numerical classification was assessed. Sample quadrats from two regions of the UK were available for post priori analysis; a random sample from Great Britain (GB), with grasslands sampled in proportion to area, and an independent stratified random sample from Northern Ireland (NI), with similar numbers of quadrats from agricultural and semi-natural grassland habitat strata. Classification of a combined area-proportional (balanced) random sample from GB and NI showed the species composition of UK grasslands to be determined largely by climate, landscape structure and land-use intensity. The classification was influenced primarily by the greater number of eutrophic agricultural grassland quadrats and semi-natural grassland quadrats of the larger GB study area. The semi-natural grasslands of NI, represented by a small number of quadrats, had little influence. Classification of a stratified NI sample combined with an area-proportional GB sample was influenced most by the NI semi-natural grassland quadrats. The structure of the classifications depended on sampling intensity. Vegetation classification should be derived from a balanced sample so that it is representative and its application does not lead to decisions being directed at classes of vegetation (or estimates derived from them) that are weighted by sampling intensity. Area-proportional sample design linked explicitly to landscape structure satisfies the requirement for a balanced classification. The issue of data-balance is relevant in conservation management and environmental assessment, where stratification is a commonly accepted procedure to reduce sampling effort, or is carried out to sample rare or ecologically interesting vegetation. It applies to landscape-scale vegetation classifications used for environmental assessments and to classifications that compare plant communities between regions (as in phytosociological studies). The issue is also important when combining environmental databases from international sources for classification purposes.

*Correspondence: Dr Alan Cooper Tel: +44 02870 324692 Fax: +44 02870 324911 e-mail: a.cooper@ulster.ac.uk

Keywords: countryside survey, data-balance, DCA, environment policy, grassland, sample design, TWINSPAN

INTRODUCTION

Multivariate classification and ordination are key techniques for describing plant communities and assessing relationships between vegetation, the environment and management practices. Comparative experiments on the distorting effects of different algorithms on vegetation data sets, which are usually either artificially constructed (Gauch *et al.* 1977) or are taken from vegetation with restricted species assemblages (Austin & Orloci 1966; Podani 1989) have guided their development. It is widely understood that sample clustering algorithms and decisions on the weighting attached to plant species cover and rarity, influence analysis (Gauch 1982). There are no field data-driven landscape-scale investigations, however, into the effects of sampling on the outcome of multivariate classification or the effects of combining extant data sets from different studies or different types of landscape. The issue is of general ecological relevance and is important where the resulting classification is used to guide decisions on environmental management.

Random sampling designs eliminate systematic error (Greig-Smith 1983), but the sampling intensities needed for reliable landscape-scale ecological studies are usually high. The impracticality of random sampling is the main reason why examples are few at the landscape-scale of field survey. Sample stratification based on prior knowledge of the distribution patterns of vegetation increases sampling efficiency by allocating sample quadrats to each stratum (Smartt & Grainger 1974). Noy-Meir (1971) stratified a vegetation survey by 10 land types to sample 240 000 km^2 of semi-arid vegetation in southern Australia. Orloci and Stanek (1979) stratified a vegetation sample of 900 km^2 of Alaskan Highway by previously mapped ecoregions and terrain types. Smith and Bunce (1978) used a multivariate land classification of Great Britain (GB) to define sampling strata for vegetation survey. Cooper *et al.* (1997) carried out a statistically structured assessment of regional variation in the distribution, species composition and management of blanket bog in Northern Ireland (NI) based on multivariate land class and habitat strata.

The concept of sample redundancy (Gauch 1982) recognizes that a vegetation sampling intensity greater than necessary to meet objectives is inefficient, but the effects of over-sampling, as in the generalized case of sampling ecologically interesting habitats more intensively, have not been assessed. The effect of this on interpreting or applying the results of numerical classification, has received little attention (Kenkel *et al.* 1989). This issue became relevant following ecological survey to coordinate and integrate environmental management policy across GB and NI (i.e. the United Kingdom; Cooper & McCann 2000; Haines-Young *et al.* 2000). The survey was carried out to inform land-use decisions relating to the biodiversity and conservation management of agricultural landscapes.

Independent grassland sample data sets from NI (Cooper & McCann 1994) and GB (Barr *et al.* 1993), recorded using comparable methods and similar sampling procedures, were available for post priori analysis. While the surveys were specific to the grassland vegetation of the UK, the general principle of assessing the effects of sampling intensity on vegetation classifications emerged. The main aim of our paper is to investigate the influence of sampling intensity on the outcome of multivariate vegetation classification, using regionally recorded field data sets, typically available to decision-makers in conservation management and environmental assessment. Specific objectives are to assess the implications for sampling vegetation at the landscape-scale, comparing vegetation between regions and combining metadata sets from different studies.

METHODS

The United Kingdom (UK) comprises Northern Ireland and Great Britain (England, Scotland and Wales). The coasts of NI and GB are, at their closest, separated by 15 km of the Irish Sea. At a regional scale there are differences between the two countries relating to the more oceanic climate of Ireland and its smaller flora, a result of its earlier separation from Europe following the last period of glaciation. Ecological survey of the grassland vegetation of these two regions was carried out independently by the Centre for Ecology and Hydrology (for GB grassland) and the University of Ulster (for NI grassland).

In the GB field survey (Barr *et al.* 1993), plant species composition was recorded in 256 1-km^2 sample grid squares. The sampling programme was stratified by a multivariate land classification of Great Britain (Bunce *et al.* 1996) derived from the analysis of 1-km^2 grid squares based on environmental attributes such as altitude and climate. Within each sample square, up to 22 quadrats (200-m^2 and 4-m^2 square quadrats and 10-m^2 linear quadrats) were placed at random in the open countryside and species composition was recorded. From the data set of about 12 500 quadrats, a single classification was constructed (Bunce *et al.*1999) using the polythetic divisive algorithm TWINSPAN (Hill 1994). The stopping criterion was a minimum group size of 30. This produced 100 vegetation classes. From the grassland vegetation classes, 407 of the 200-m^2 square quadrats (Table 1) were drawn at random with a frequency approximately proportional to the estimated area of each class. The data set approximates to a land area-proportional random sample of GB grasslands.

In the NI field survey (Murray *et al.* 1992), land cover in 628 0.25-km^2 sample grid squares, stratified by multivariate land classification (Cooper 1986) was mapped and estimates of the area of each land cover type were made. Grassland was mapped as three types of agricultural grassland (ryegrass, mixed grassland and other agricultural grassland) and five types of semi-natural grassland (species-rich dry grassland, species-rich wet grassland, hill pasture, calcareous grassland and fen meadow). Murray *et al.* (1992) and Cooper *et al.* (1994) give descriptions these grassland types. The species composition of each grassland stratum was sampled randomly to a similar intensity, with between 36 and 77 random 200-m^2 quadrats (Table 1) giving a sample size of 421 (Cooper *et al.* 1994). The data set is thus non-proportional in relation to the land area.

To achieve comparability, species nomenclature was standardized between the NI and GB data sets, with recording procedure, taxonomic species composites and aggregates taken into account. Botanical nomenclature follows Stace (1991). Analysis of the two data sets was by species presence/absence in the quadrats. Species with less than five occurrences were deleted from each data set to reduce the influence of rare species. Because the data sets were characterized by a continuous gradient structure, the classifications were stopped with simple rules that were readily interpretable (Dale 1988). These were the third level in the classification hierarchy or <30 quadrats in an end-group. At this level, assessment of the heterogeneity of each end-group by detrended correspondence analysis (DCA) run under CANOCO 4.5 (ter Braak & Smilauer 1998) showed no major gradient structures or group discontinuities.

A random sub-set of the NI quadrats, sampled in proportion to the estimated area of each grassland habitat (Table 1) was created to carry out a combined area-proportional NI/GB classification. Because NI is smaller than GB, the area-proportional combined data set gives a small number of NI quadrats. The NI sample was therefore adjusted to give an arbitrary 20% representation. Of the 421 NI quadrats, the area-proportional sub-sample comprised 115 quadrats. In terms of sample design, this was directly comparable with the area-proportional GB data set. This enabled us to test the hypothesis that NI and GB grasslands were similar in terms of species composition and occurred in the same relative amounts in the farmed landscape. Two combined NI/GB quadrat classifications were carried out to assess the effects of sampling effort: (1) an area-proportional balanced sample data set to compare NI and GB, and (2) a stratified NI sample combined with an area-proportional GB sample. The known

https://doi.org/10.1017/S0376892906002967 Published online by Cambridge University Press

Table 1 Number of GB and NI samples and sub-samples and cross-tabulations of grassland sampling strata with a TWINSPAN classification of a combined area-proportional sample of NI and GB strata. GB grassland strata = G1–G10 (strata are named by species with the highest frequencies) and NI grassland habitat types = N1–N8 (Cooper & McCann 2000).

Survey class code	*Survey class name*	*No. of samples*	*Number of sub-samples*	*TWINSPAN group*					
				1	*2*	*3*	*4*	*5*	*6*
GB									
G1	*Lolium/Trifolium/Cynosurus*	115		–	–	5	109	1	–
G2	*Agrostis/Cerastium/Lolium*	47		2	–	5	39	1	–
G3	*Lolium/Trifolium/Poa*	42		28	13	–	1	–	–
G4	*Lolium/Trifolium/Cerastium*	34		17	8	–	9	–	–
G5	*Lolium/Agrostis/Holcus*	28		–	4	2	22	–	–
G6	*Trifolium/Holcus/Cynosurus*	26		–	–	6	6	13	1
G7	*Holcus/Ranunculus*	22		–	–	14	2	6	–
G8	*Plantago/Lotus/Achillea*	17		–	–	–	5	12	–
G9	*Plantago/Festuca/Lotus*	11		–	–	–	–	1	10
G10	Other grassland classes	65		–	2	4	29	15	15
GB total		407							
NI									
N1	Ryegrass	57	57	1	35	11	10	–	–
N2	Mixed grassland	41	6	–	–	2	2	2	–
N3	Other agricultural grassland	58	45	–	10	13	17	5	–
N4	Species-rich dry grassland	66	1	–	–	–	–	1	–
N5	Species-rich wet grassland	77	3	–	–	1	1	1	–
N6	Hill pasture	42	2	–	–	–	–	2	–
N7	Fen meadow	44	0	–	–	–	–	–	–
N8	Calcareous grassland	36	1	–	–	–	–	1	–
NI total		421	115	–	–	–	–	–	–

ecology of species (Stace 1991; Webb *et al.* 1996) and the mean weighted quadrat Ellenberg indicator values for pH, nitrogen and water (Ellenberg 1988; Hill *et al.* 1999) were used as diagnostics to aid interpretation.

RESULTS

Area-proportional NI and GB data set classification

The first division in the TWINSPAN classification (Fig. 1) separated 87 quadrats (groups 5 and 6) characterized by indicator species of British semi-natural grasslands, a high mean number of species and low mean weighted Ellenberg pH and nitrogen indicator values (Table 2). Group 6 was composed entirely of GB quadrats with a high frequency of *Thymus praecox* and *Erica cinerea* (Table 2), species characteristic of well-drained upland grassland. Group 5 represents GB and NI wet grassland, with a high frequency of *Juncus effusus* and *Poa trivialis* (species indicative of wet grassland) and a high Ellenberg water indicator value. It was the only group to contain the statistically expected number of NI and GB quadrats (Table 3).

At the second level in the TWINSPAN hierarchy, the 435 quadrats of groups 1–4 had a high frequency of the agriculturally preferred (i.e. sown) species *Lolium perenne*. Groups 1 and 2 were distinguished from groups 3 and 4 by a low mean number of species (Table 2) and high Ellenberg nitrogen indicator values. Group 1 contained predominantly GB dry eutrophic grassland quadrats, characterized by the TWINSPAN indicator species *Dactylis glomerata*, historically sown for its drought-resistance in southern Britain. Group 2 represented wet eutrophic NI grassland. It had a significantly higher than expected number of NI quadrats (Table 3). TWINSPAN indicators distinguishing it from group 1 included *Poa trivialis*, *Agrostis stolonifera* and *Alopecurus geniculatus*, species of wet soils. Group 2 also had a higher Ellenberg water indicator value than group 1.

Groups 3 and 4 were characterized by species indicative of less eutrophic grasslands. Group 3, with a significantly higher than expected number of NI quadrats, had a high frequency of wet grassland species (*Juncus effusus* and *Cardamine pratensis*) and a high Ellenberg water indicator value. Group 4, the largest grassland group, comprised less eutrophic samples mainly from GB, characterized by indicator species of well-drained soils (such as *Dactylis glomerata* and *Achillea millefolium*).

The main grassland groups derived from the classification were thus: eutrophic dry grassland (group 1) largely from GB; eutrophic wet grassland, largely from NI (group 2); less eutrophic dry grassland (group 4), largely from GB; less eutrophic wet grassland, with a relatively large number of quadrats from NI (group 3); and semi-natural species-rich grasslands (groups 5 and 6), with one group (group 6) characteristic of the GB marginal uplands.

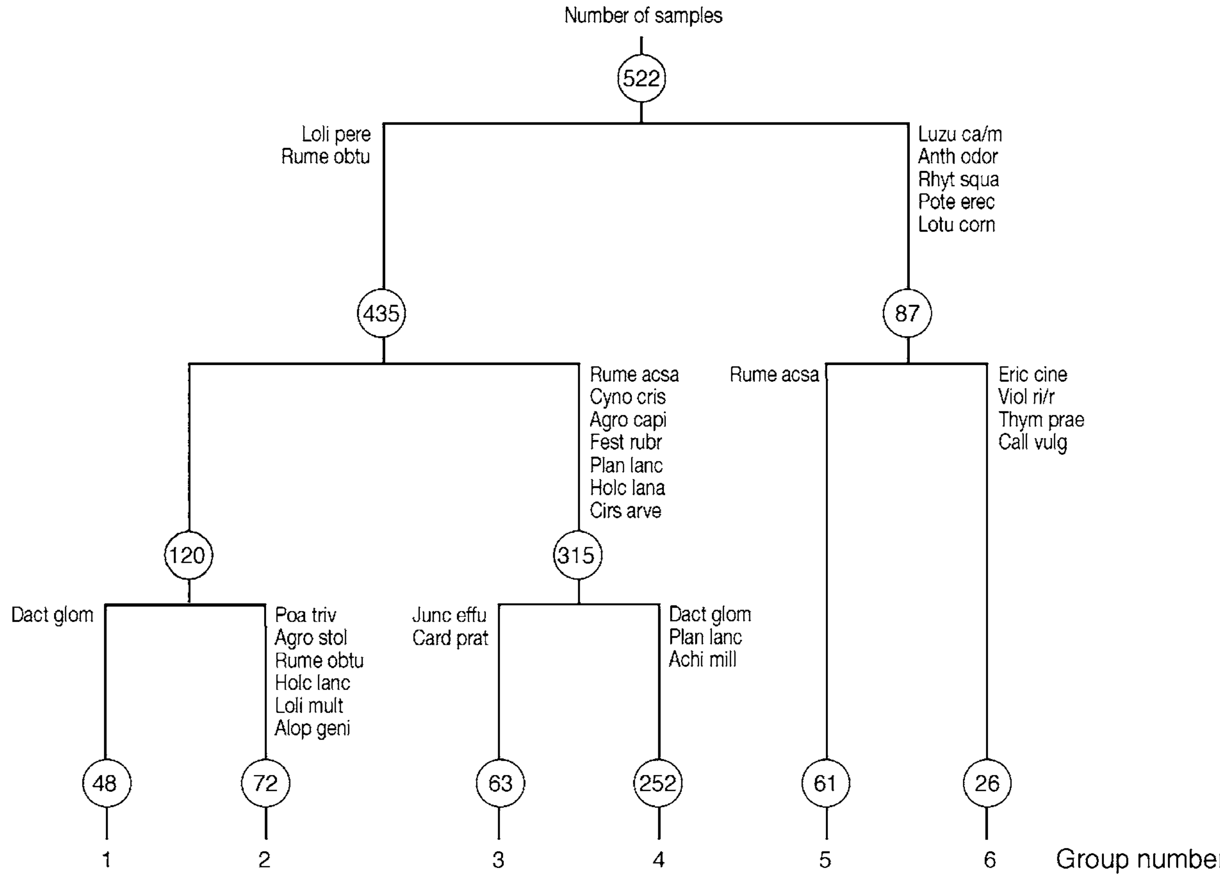

Figure 1 Classification of GB and NI quadrats sampled in proportion to land area. TWINSPAN indicator species are given for each division in the classification. The number of quadrats in a group is circled. Species names are: Achi mill = *Achillea millefolium;* Agro capi = *Agrostis capillaris;* Agro stol = *Agrostis stolonifera;* Alop geni = *Alopecurus geniculatus;* Anth odor = *Anthoxanthum odoratum;* Call vulg = *Calluna vulgaris;* Card prat = *Cardamine pratensis;* Cirs arve = *Cirsium arvense;* Cyno cris = *Cynosurus cristatus;* Dact glom = *Dactylis glomerata;* Eric cine = *Erica cinerea;* Fest rubr = *Festuca rubra;* Holc lana = *Holcus lanatus;* Junc effu = *Juncus effusus;* Loli mult = *Lolium multiflorum;* Loli pere = *Lolium perenne;* Lotu corn = *Lotus corniculatus;* Luzu ca/m = *Luzula campestris/multiflorum;* Plan lanc = *Plantago lanceolata;* Poa triv = *Poa trivialis;* Pote erec = *Potentilla erecta;* Rhyt squa = *Rhytidiadelphus squarrosus;* Rume acsa = *Rumex acetosa;* Rume obtu = *Rumex obtusifolius;* Thym prae = *Thymus praecox*; and Viol ri/r = *Viola riviniana/reichenbachiana*.

Stratified NI and area-proportional GB samples

The first TWINSPAN division (Fig. 2) separated off groups 1–3, with a high frequency of the agriculturally preferred *Lolium perenne* (Table 4). Groups 1 and 2 had a low mean number of species (Table 4) and high Ellenberg water indicator values. Group 1 was composed mainly of GB samples, with group 2 mainly NI samples characterized by a high frequency of *Alopecurus geniculatus* and *Agrostis stolonifera* (species indicative of wet soils) and a high Ellenberg water indicator value. Group 3 was relatively species-rich grassland, probably related to a less intensive management regime.

Groups 4–7 were largely NI samples (Table 5) characterized by semi-natural grassland species, a high mean number of species and low Ellenberg nitrogen indicator values. Group 4 had a high frequency of *Cynosurus cristatus*. It contained a high proportion of the NI species-rich dry grassland samples. Group 5 had a high frequency of the wet grassland species *Ranunculus flammula* and *Carex nigra* and had a high Ellenberg water indicator value. It comprised mainly NI species-rich wet grassland samples. Group 6 contained most of the NI calcareous grassland samples. Group 7 had a high frequency of *Carex panicea*, *Succisa pratensis* and *Pedicularis sylvatica*, species of wet soils. It contained a high proportion of the NI hill pasture and fen meadow samples and had low Ellenberg pH and nitrogen indicator values.

The main groups derived from the classification were thus: eutrophic grassland largely from GB (group 1); eutrophic grassland with wet grassland indicators, largely from NI (group 2); less eutrophic grassland (group 3); and four groups (groups 4–7) representing the semi-natural grasslands of NI.

DISCUSSION

The UK classification of a combined NI and GB area-proportional random quadrat sample gives a balanced description of grassland variation, determined largely by the

https://doi.org/10.1017/S0376892906002967 Published online by Cambridge University Press

Table 2 Group attributes of a TWINSPAN classification of a combined area-proportional sample of GB and NI quadrats. The percentage frequency of TWINSPAN indicator species (I) and species with a frequency >80% in any one group are shown. Species are arranged by DCA order.

Species code	*Name*	*Indicator species*	*TWINSPAN group*					
			1	*2*	*3*	*4*	*5*	*6*
Loli mult	*Lolium multiflorum*	I	12	50	11	6	–	–
Alop geni	*Alopecurus geniculatus*	I	–	36	32	8	3	–
Rume obtu	*Rumex obtusifolius*	I	35	88	46	36	7	–
Loli pere	*Lolium perenne*	I	98	99	79	87	49	4
Poa triv	*Poa trivialis*	I	17	83	63	35	46	4
Agro stol	*Agrostis stolonifera*	I	8	58	70	51	43	23
Ranu repe	*Ranunculus repens*		35	65	92	77	77	35
Trif repe	*Trifolium repens*		90	74	95	86	92	62
Dact glom	*Dactylis glomerata*	I	50	12	11	63	33	19
Cera font	*Cerastium fontanum*		38	44	84	81	89	50
Cirs arve	*Cirsium arvense*	I	4	7	22	48	31	12
Holc lana	*Holcus lanatus*	I	12	54	95	86	92	73
Cyno cris	*Cynosurus cristatus*	I	2	10	56	56	69	23
Card prat	*Cardamine pratensis*	I	–	6	57	6	30	4
Agro capi	*Agrostis capillaris*	I	15	19	67	70	90	65
Rume acsa	*Rumex acetosa*	I	–	11	62	56	90	27
Junc effu	*Juncus effusus*	I	–	–	76	12	46	12
Achi mill	*Achillea millefolium*	I	4	4	6	42	43	58
Fest rubr	*Festuca rubra*	I	2	–	38	48	69	58
Plan lanc	*Plantago lanceolata*	I	2	1	11	52	66	92
Anth odor	*Anthoxanthum odoratum*	I	–	4	59	32	95	85
Rhyt squa	*Rhytidiadelphus squarrosus*	I	–	–	35	21	82	62
Luzu ca/m	*Luzula campestris/multiflorum*	I	–	–	19	11	85	54
Lotu corn	*Lotus corniculatus*	I	–	–	5	21	46	92
Pote erec	*Potentilla erecta*	I	–	–	8	7	54	73
Viol ri/r	*Viola riviniana/reichenbachiana*	I	–	–	2	3	21	96
Thym prae	*Thymus praecox*	I	–	–	–	1	8	77
Call vulg	*Calluna vulgaris*	I	–	–	–	1	5	58
Eric cine	*Erica cinerea*	I	–	–	–	–	5	73
	Mean number of species		9	11	23	22	31	36
	Number of quadrats	NI	1	45	27	30	12	0
		GB	47	27	36	222	49	26
	Mean weighted Ellenberg water value		5.13	5.39	5.88	5.31	5.62	5.49
	Mean weighted Ellenberg pH value		6.20	6.20	5.67	5.96	5.36	4.99
	Mean weighted Ellenberg nitrogen value		5.97	6.18	5.03	5.16	4.36	3.35

Table 3 Percentage frequency of quadrats in the vegetation groups from a TWINSPAN classification of a combined area-proportional NI and GB sample. χ^2 values > 6.64 are significant at $p < 0.01$. The null hypothesis was that each group contains the same proportion of NI and GB quadrats (1:5) as their representation in the sampling programme.

	TWINSPAN group					
	1	*2*	*3*	*4*	*5*	*6*
NI quadrats	2.1	62.5	42.9	11.9	19.7	0.0
GB quadrats	97.9	37.5	57.1	88.1	80.3	100.0
χ^2 value	11.12	68.65	15.91	15.04	0.20	7.35

key environmental variables of climate, landscape structure and agricultural land-use intensity. The statistically different ratio of NI and GB quadrats in five of the six vegetation classes, compared with the ratio of quadrats expected from the area-proportional sample, is a formal demonstration of significant differences in species composition between the two regions. The area-proportional classification is dominated by the greater number of quadrats from the larger GB study area, where dry eutrophic grassland predominates and semi-natural grasslands are mainly in upland landscapes. Eutrophic agricultural grassland quadrats from NI are also represented in the classification, characterized by indicator species of wet soils. This is a function of the more oceanic climate of Ireland. The difference in community composition between eutrophic grasslands in GB and NI is apparent despite the small size of NI. This is because sown eutrophic grassland in NI is the main grassland type in the area-proportional sample. The small influence of NI semi-natural grassland quadrats on the classification compared with GB semi-natural grassland quadrats is due largely to their relatively small area

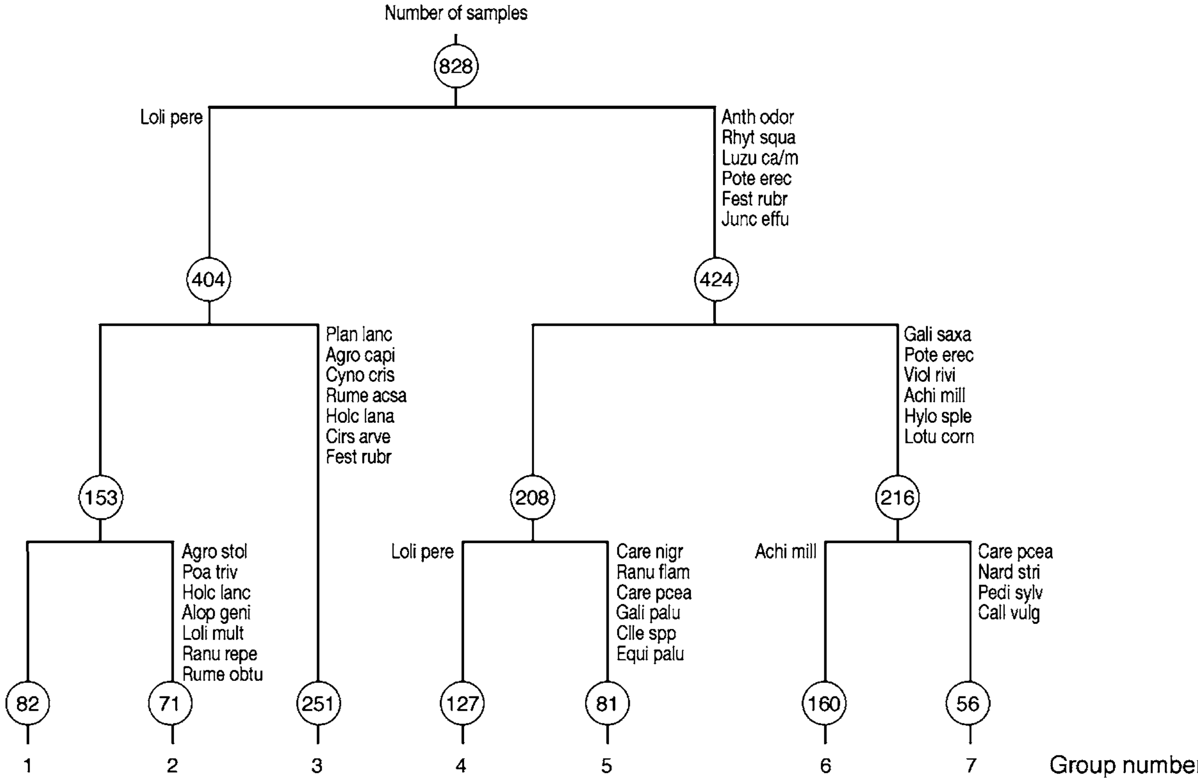

Figure 2 Classification of NI stratified sample and a GB area-proportional sample. The number of quadrats in a group is circled. Species names are: Achi mill = *Achillea millefolium;* Agro capi = *Agrostis capillaris;* Agro stol = *Agrostis stolonifera*; Alop geni = *Alopecurus geniculatus*; Anth odor = *Anthoxanthum odoratum*; Call vulg = *Calluna vulgaris*; Care nigr = *Carex nigra*; Care pcea = *Carex panacea*; Cirs arve = *Cirsium arvense*; Clle sp. = *Calliergon* sp.; Cyno cris = *Cynosurus cristatus*; Equi palu = *Equisetum palustre*; Fest rubr = *Festuca rubra*; Gali palu = *Galium palustre*; Gali saxa = *Galium saxatile*; Holc lana = *Holcus lanatus*; Hylo sple = *Hylocomium splendens*; Junc effu = *Juncus effusus*; Loli mult = *Lolium multiflorum*; Loli pere = *Lolium perenne*; Lotu corn = *Lotus corniculatus*; Luzu ca/m = *Luzula campestris/multiflorum*; Nard stri = *Nardus stricta*; Pedi sylv = *Pedicularis sylvatica*; Plan lanc = *Plantago lanceolata*; Poa triv = *Poa trivialis*; Pote erec = *Potentilla erecta*; Ranu flam = *Ranunculus flammula*; Ranu repe = *Ranunculus repens*; Rhyt squa = *Rhytidiadelphus squarrosus*; Rume acsa = *Rumex acetosa*; Rume obtu = *Rumex obtusifolius*; and Viol ri/r = *Viola riviniana/reichenbachiana.*

and therefore low sample representation compared with sown grasslands.

In the combined classification of the area-proportional GB quadrat sample and the uniformly sampled NI strata, quadrats from the NI semi-natural grassland habitats have the greatest influence on vegetation classification. The classification is influenced by the weighting imposed by the relatively high sampling intensity of the NI semi-natural grassland strata. As in the combined NI and GB area-proportional classification, the influence of climate is apparent (although only in the eutrophic grasslands) but the influence of landscape structure is not. Our main conclusion is that sampling intensity influences pattern recognition, the principal aim of classification (Kenkel *et al.* 1989) and that this is important in studies where representativeness is a priority. We note, however, that while weighted sampling is not appropriate if landscape-scale representation is the aim, it is acceptable for characterizing the species composition of strata.

Sampling intensity has also been shown to influence the outcome of the numerical taxonomy of organisms. Clifford and Williams (1973) showed that by varying the number of identical taxonomic units in a numerical classification of grasses, added weight was given to the more intensively sampled groups and this influenced the classification and changed it. We show that the same applies to vegetation samples, in other words, sampling effort influences pattern recognition.

A model of plant community composition (Austin & Smith 1989) provides the conceptual basis of balanced sampling in community ecology. It is based on landscape-scale variation in environment gradient structure and distribution pattern and shows that similar communities (associated groups of species) can be delimited across study areas, provided that they have the same repeating patterns of landscape structure and environment gradients in common. In terms of this generally accepted model, balanced sampling linked to landscape structure allows representative classes of vegetation to be defined so that valid comparisons can be made. We conclude that vegetation classification should be derived from a balanced sample so that it is representative. Area-proportional sample

Table 4 Group attributes of a TWINSPAN classification of a combined area-proportional sample of GB quadrats and a stratified sample of NI quadrats. The percentage frequency of TWINSPAN indicator species (I) and species with a frequency >80% in any one group are shown. Species are arranged by DCA order.

Species code	*Name*	*Indicator species*	*TWINSPAN group*						
			1	*2*	*3*	*4*	*5*	*6*	*7*
Loli mult	*Lolium multiflorum*	I	12	55	8	2	1	2	–
Rume obtu	*Rumex obtusifolius*	I	49	87	39	27	16	9	–
Loli pere	*Lolium perenne*	I	99	94	88	52	7	42	4
Alop geni	*Alopecurus geniculatus*	I	2	44	12	16	20	1	2
Cirs arve	*Cirsium arvense*	I	7	10	48	19	4	29	2
Poa triv	*Poa trivialis*	I	29	89	41	65	65	31	16
Agro stol	*Agrostis stolonifera*	I	16	77	54	45	74	25	32
Ranu repe	*Ranunculus repens*	I	40	82	80	91	72	64	46
Cera font	*Cerastium fontanum*		45	56	82	87	49	81	39
Trif repe	*Trifolium repens*		89	75	89	89	81	86	64
Holc lana	*Holcus lanatus*	I	26	70	88	98	96	89	86
Achi mill	*Achillea millefolium*	I	7	1	38	13	–	68	11
Agro capi	*Agrostis capillaris*	I	20	27	72	81	46	92	79
Rume acsa	*Rumex acetosa*	I	6	20	57	87	67	74	50
Cyno cris	*Cynosurus cristatus*	I	9	13	59	80	48	72	48
Ranu acri	*Ranunculus acris*		6	11	34	70	83	40	34
Plan lanc	*Plantago lanceolata*	I	1	3	51	62	27	79	66
Fest rubr	*Festuca rubra*	I	2	1	42	85	65	87	70
Anth odor	*Anthoxanthum odoratum*	I	–	15	32	91	88	85	100
Junc effu	*Juncus effusus*	I	2	7	15	70	86	36	48
Lotu corn	*Lotus corniculatus*	I	–	–	20	23	1	69	29
Rhyt squa	*Rhytidiadelphus squarrosus*	I	2	1	16	59	62	87	89
Gali palu	*Galium palustre*	I	–	–	–	9	58	3	4
Clle sp.	*Calliergon sp.*	I	–	–	4	33	74	21	25
Junc ac/a	*Juncus articulatus/acutiflorus*		–	1	6	46	86	14	52
Luzu ca/m	*Luzula campestris/multiflorum*	I	–	1	6	49	52	78	86
Equi palu	*Equisetum palustre*	I	–	–	–	5	42	1	9
Gali saxa	*Galium saxatile*	I	–	–	3	3	5	51	57
Ranu flam	*Ranunculus flammula*	I	–	1	2	8	72	2	30
Care nigr	*Carex nigra*	I	–	–	1	15	83	11	46
Pote erec	*Potentilla erecta*	I	–	–	4	15	47	71	98
Viol ri/r	*Viola riviniana/reichenbachiana*	I	–	–	3	2	2	44	36
Care pcea	*Carex panacea*	I	–	–	–	9	57	11	84
Nard stri	*Nardus stricta*	I	–	–	1	6	21	10	71
Hylo sple	*Hylocomium splendens*	I	–	–	1	2	10	41	57
Succ prat	*Succisa pratensis*		–	–	–	6	42	38	80
Call vulg	*Calluna vulgaris*	I	–	–	–	–	9	16	71
Pedi sylv	*Pedicularis sylvatica*	I	–	–	–	1	10	4	57
	Mean number of species		9	13	22	29	34	38	38
	Number of quadrats	NI	5	54	57	85	79	97	44
		GB	77	17	194	42	2	63	12
	Mean weighted Ellenberg pH value		5.19	5.49	5.34	5.78	6.76	5.40	6.23
	Mean weighted Ellenberg nitrogen value		6.16	6.15	5.96	5.65	5.28	5.45	4.55
	Mean weighted Ellenberg water value		5.97	6.01	5.17	4.76	4.17	4.18	3.35

design linked explicitly to landscape structure satisfies the requirement for a balanced classification.

The importance of using structured methods of defining landscape variation and balanced sampling in comparative studies of vegetation, was first developed in Cumbria, England (Smith & Bunce 1978), where a multivariate land classification of kilometre grid squares, derived from variables such as climate, elevation and soil type was developed for stratifying area-proportional ecological sampling. The context for this was regional land-use planning and conservation. The approach was developed further to construct the Institute of Terrestrial Ecology (ITE) land classification of GB (Bunce *et al.* 1996) and used to monitor and assess the ecological consequences of land cover change as part of a

Table 5 Number of GB and NI samples and sub-samples and cross-tabulations of grassland sampling strata with a TWINSPAN classification of a combined NI stratified sample and a GB area-proportional sample. GB grassland strata = G1–G10 (strata are named by species with the highest frequencies) and NI grassland habitat types = N1–N8 (Cooper & McCann 2000).

Survey class code	*Survey class name*	*TWINSPAN group*							
		1	*2*	*3*	*4*	*5*	*6*	*7*	*Total*
GB									
G1	*Lolium/Trifolium/Cynosurus*	1	–	101	10	–	3	–	115
G2	*Agrostis/Cerastium/Lolium*	4	2	37	1	–	3	–	47
G3	*Lolium/Trifolium/Poa*	38	3	1	–	–	–	–	42
G4	*Lolium/Trifolium/Cerastium*	25	6	3	–	–	–	–	34
G5	*Lolium/Agrostis/Holcus*	8	4	16	–	–	–	–	28
G6	*Trifolium/Holcus/Cynosurus*	–	–	3	8	–	15	–	26
G7	*Holcus/Ranunculus*	–	1	4	16	–	1	–	22
G8	*Plantago/Lotus/Achillea*	–	–	3	1	–	13	–	17
G9	*Plantago/Festuca/Lotus*	–	–	–	–	–	8	3	11
G10	Other grassland classes	1	1	26	6	2	20	9	65
NI									
N1	Ryegrass	5	38	12	2	–	–	–	57
N2	Mixed grassland	–	1	19	13	2	6	–	41
N3	Other agricultural grassland	–	15	21	16	5	1	–	58
N4	Species-rich dry grassland	–	–	2	29	2	27	6	66
N5	Species-rich wet grassland	–	–	1	19	51	1	5	77
N6	Hill pasture	–	–	1	5	2	19	15	42
N7	Fen meadow	–	–	1	–	–	43	–	44
N8	Calcareous grassland	–	–	–	1	17	–	18	36
	Total	82	71	251	127	81	160	56	–

national strategic assessment of ecological resources. Balanced sampling was an integral element of these studies.

A further example of balanced sampling at the landscape-scale is that of Austin and Heyligers (1989), who used an explicit rule-based method of stratified sampling to incorporate major gradients of environmental variation (gradsects) in a conservation evaluation of the forest vegetation of 20 000 km^2 of New South Wales (Australia). Gradsects were selected to represent the main regional environmental variables of geology, altitude and rainfall. The range of each variable was divided into classes and the resulting three-dimensional cells were sampled randomly. The gradsect approach ensured that a balanced range of environments was sampled in approximate proportion to their area.

If the principle of data balance is not applied to vegetation classification, comparison of the plant communities of regional landscapes can be misleading. Separate phytosociological studies of shrub heath across Europe, for example, give a descriptive summary of the range of communities (Oudhof & Barendregt 1987), but estimates of the proportions of the heath vegetation classes produced by a joint classification of the data sets would be invalid because the sampling design is not balanced. The main reason for lack of data balance in cases such as these is that the areas selected for study are usually of inherent phytosociological interest and therefore tend to be over-sampled. In France, for example, the heaths of Brittany are represented by large numbers of relevees, whereas the disturbed pine forest heaths of the Landes region have few.

The issue of representative versus weighted sampling (data balance) is particularly relevant in conservation management and environmental assessment, where stratification is a commonly accepted procedure to reduce sampling effort, or is carried out to sample (i.e. characterize) rare or ecologically interesting vegetation. The principle of data balance, however, is not routinely considered or applied by community ecologists or environmental managers. For example, Chytry (2001) has shown, by analysing the Czech National Phytosociological Database consisting of 40 000 relevees by different authors from different districts, that data on species-poor vegetation are more scarce than for species-rich vegetation and that the database may be severely biased towards relevees with a higher biodiversity. The National Vegetation Classification (NVC) of GB (Rodwell 1991), which has been applied extensively for comparative vegetation mapping by environmental managers in GB and also in NI (a region outside the sampling domain of the NVC), was not derived from a balanced data set.

Data balance can be approximated using vegetation data that have already been collected, provided that information on sampling decisions is available, that preparatory analysis is carried out to ensure comparability and that an even distribution of samples through the strata can be achieved. If details of sampling intensity and study area boundaries (strata) are not stated explicitly in phytosociological studies, a balanced sample can not be extracted post priori for quantitative assessment. Preparatory analysis relating to sampling intensity associated with a phytosociological study of European shrub heath relevees, is exemplified by Oudhof and Barendregt (1987). By processing relevees into homogeneous agglomerative clusters, employing sorting procedures that discarded outliers, and by displaying relationships on ordination diagrams, Oudhof and Barendregt (1987) reduced the influence of the unbalanced sampling design resulting from over-sampling.

A further example of how approximations to address data balance can be made where the distribution of a habitat or landscape has not been mapped or is not available, is that of Duckworth *et al.* (1998). In a comparison of vegetation in different parts of Europe, Duckworth *et al.* (1998) aggregated relevées from existing studies by cells of 0.5° of latitude and longitude, to address the problem of an uneven distribution of sampling in their study area. The Global Observation Research Initiative in Alpine Environments project (Bayfield *et al.* 2005) is a worked example of how data from multiple sources, working from standardized sampling protocols, can be screened to get reliable results.

The sampling design issue emerging from this discussion is that a vegetation classification should be linked, by area-proportional sampling, to landscape structure to maintain data balance, so that its application is not influenced by sampling intensity. Policy decisions relating to biodiversity assessment and conservation have a requirement for balanced sampling and statistical robustness (Cooper *et al.* 1997; Haines-Young *et al.* 2000). If the principle of data balance is not followed, conclusions could be unrepresentative, that is they could be directed at classes of vegetation or estimates (such as area or species diversity) derived from them, which are a function of sampling intensity. If vegetation sampling is linked explicitly to landscape structure (Austin & Smith 1989), valid comparisons of vegetation classes within and across landscapes can be made. Biodiversity policy development in the UK has been guided by field survey based on this approach (Cooper *et al.* 1997; Haines-Young *et al.* 2000). An environmental classification of Europe has similarly been developed (Metzger *et al.* 2005) based on climate, geomorphology and geographical position (ocean influence and day length). It allows assessment and monitoring through stratified random sampling and provides a basis for environmental reporting and assessing changes in habitats and biodiversity.

Careful consideration of data balance is also relevant if extant data from different studies are combined for classification purposes. It has particular relevance to the coordination of sampling procedures in vegetation assessment and management across larger regions such as Europe (Neuhausl 1990). Data balance is especially important in view of current imperatives to use extant electronic data sets to inform land use, conservation and biodiversity policy and the ease with which meta-datasets can be combined and used for ecological resource assessment or conservation evaluation (Metzger *et al.* 2005). We emphasize that sampling balance needs to be applied to surveys employing such strata so that valid comparisons across European landscapes and internationally can be made.

ACKNOWLEDGEMENTS

The study was supported by the Environment and Heritage Service of the Department of the Environment for Northern Ireland and the UK Department of the Environment. It formed part of the Northern Ireland Countryside Survey.

References

Austin, M.P. & Orloci, L. (1966) Geometric models in ecology II. An evaluation of some ordination techniques. *Journal of Ecology* **54**: 217–227.

Austin, M.P. & Heyligers, P.C. (1989) Vegetation survey design for conservation - GRADSECT sampling of forest in NE New South Wales. *Biological Conservation* **50**: 13–32.

Austin, M.P. & Smith, T.M. (1989) A new model for the continuum concept. *Vegetatio* **83**: 35–47.

Barr, C.J., Bunce, R.G.H. Clarke, R.T., Fuller, R.M., Furse, M.T., Gillespie, M.K., Groom, G.B., Hallam, C.J., Hornung, M., Howard, D.C. & Ness, M.J. (1993) *Countryside Survey 1990: Main Report*. London, UK: Department of Environment.

Bayfield, N., Brooker, R. & Turner, L. (2005) Some lessons to be learned from the ECN, GLORIA and SCANNET networks for international environmental monitoring. In: *Mountains of Northern Europe, Conservation, Management, People and Nature*, ed. D.B.A. Thompson, M.F. Price & C.A. Galbraith, pp. 213–222. Edinburgh, UK: TSO Scotland.

Bunce, R.G.H., Barr, C.J., Clarke, R.T. Howard, D.C. & Lane, A.M.J. (1996) Land classification for strategic ecological survey. *Journal of Environmental Management* **47**: 37–60.

Bunce, R.G.H., Barr, C.J. Gillespie, M.K., Howard, D.C., Smart, S.M., Scott, W.A., Watkins, J.W. & van de Poll, H.M. (1999) *The Vegetation of the British Countryside: the Countryside Vegetation System*. London, UK: HMSO.

Chytry, M. (2001) Phytosociological data give biased estimates of species richness. *Journal of Vegetation Science* **12**: 439–444.

Clifford, H.T. & Williams, W.T. (1973) Classificatory dendrograms and their interpretation. *Australian Journal of Botany* **21**: 151–162.

Cooper, A. (1986) The Northern Ireland land classification. Report, University of Ulster, Coleraine, UK.

Cooper, A. & McCann, T.P. (1994) Botanical composition of grassland land cover types in Northern Ireland. Report, University of Ulster, Coleraine, UK.

Cooper, A. & McCann, T.P. (2000) NICS2000: summary report on Broad Habitats. Report, University of Ulster, Coleraine, UK.

Cooper, A., McCann, T.P. & Power, J. (1997) Regional variation in the cover, species composition and management of blanket bog. *Landscape and Urban Planning* **37**: 19–28.

Dale, M.B. (1988) Knowing when to stop: cluster concept - concept cluster. *Coenoses* **3**: 11–32.

Duckworth, J.C., Bunce, R.G.H. & Malloch, A.J.C. (1998) Vegetation gradients in Atlantic Europe: the use of existing phytosociological data in preliminary investigations on the potential effects of climatic change on British vegetation. *Global Ecology and Biogeography* **9**: 187–199.

Ellenberg, H. (1988) *Vegetation Ecology of Central Europe*. Cambridge, UK: Cambridge University Press.

Gauch, H.G. (1982) *Multivariate Analysis in Community Ecology*. Cambridge, UK: Cambridge University Press.

Gauch, H.G., Whittaker, R.H. & Wentworth, R.T. (1977) A comparative study of reciprocal averaging and other ordination techniques. *Journal of Ecology* **65**: 157–174.

Greig-Smith, P. (1983) *Quantitative Plant Ecology*, Third edition. Oxford, UK: Blackwell.

Haines-Young, R.H., Barr, C.J., Black, H.I.J., Briggs, D.J., Bunce, R.G.H., Clarke, R.T., Cooper, A., Dawson, F.H., Firbank, L.G., Fuller, R.M., Furse, M.T., Gillespie, M.K., Hill, R., Hornung, M., Howard, D.C., McCann, T., Morecroft, M.D., Petit, S., Sier, A.R.J., Smart, S.M., Smith, G.M., Stott, A.P., Stuart, R.C. & Watkins, J.W. (2000) *Accounting for Nature: Assessing Habitats in the UK Countryside*. London, UK: Department of the Environment, Transport and the Regions.

Hill, M.O. (1994) *DECORANA and TWINSPAN, for Ordination and Classification of Multivariate Species Data: a New Edition, Together with Supporting Programs, in FORTRAN 77*. Huntingdon, UK: Institute of Terrestrial Ecology.

Hill, M.O., Mountford, J.O., Roy, D.B. & Bunce, R.G.H. (1999) *Ellenberg's Indicator Values for British Plants*. Huntingdon, UK: Institute of Terrestrial Ecology.

Kenkel, N.C., Juhasz-Nagey, P. & Podani, J. (1989) On sampling procedures in community and population ecology. *Vegetatio* **83**: 195–207.

Metzger, M.J., Bunce, R.G.H., Jongman, R.H.G., Mucher, C.A. & Watkins, J.W. (2005) A climatic stratification of the environment of Europe. *Global Ecology and Biogeography* **14**: 549–563.

Murray, R., McCann, T.P. & Cooper, A. (1992) A land classification and landscape ecological survey of Northern Ireland. Report, University of Ulster, Coleraine, Northern Ireland.

Noy-Meir, I. (1971) Multivariate analysis of the semi-natural vegetation in southeastern Australia: nodal ordination by components analysis. *Proceedings of the Ecological Society of Australia* **6**: 159–93.

Neuhausl, R. (1990) Unified classification of European natural forests: the approach of the vegetation map of Europe. *Vegetatio* **89**: 173–182.

Orloci, L. & Stanek, W. (1979) Vegetation survey of the Alaska Highway, Yukon territory: types and gradients. *Vegetatio* **41**: 1–56.

Oudhof, J.A.E. & Barendregt, A. (1987) Processing of vegetation relevees to obtain a classification of south-west European heathland. In: *Data Analysis in Community and Landscape Ecology*, ed. R.H.G. Jongman, C.J.F. ter Braak & O.F.R. van Tongeran, pp. 258–265. Cambridge, UK: Cambridge University Press.

Podani, J. (1989) Comparison of ordinations and classifications of vegetation data. *Vegetatio* **83**: 111–128.

Rodwell, J.S. (1991) *British Plant Communities: Woodland and Scrub*. Cambridge, UK: Cambridge University Press.

Smartt, P.F.M. & Grainger, J.A.E. (1974) Sampling for vegetation survey: some aspects of the behaviour of unrestricted, restricted and stratified techniques. *Journal of Biogeography* **1**: 193–206.

Smith, R.S. & Bunce, R.G.H. (1978) An ecological survey of Cumbria. Report, Cumbria County Council and Lake District Special Planning Board, Kendal, UK.

Stace, C.A. (1991) *New Flora of the British Isles*. Cambridge, UK: Cambridge University Press.

ter Braak, C.J.F. & Smilauer, P. (1998) *CANOCO Reference Manual and User's Guide to Canoco for Windows: Software for Canonical Community Ordination (Version 4.5)*. Ithaca, New York, USA: Microcomputer Power.

Webb, D.A., Parnell, J. & Doogue, D. (1996) *An Irish Flora*. Dundalk, Ireland: Dundalgan Press.