Ann. Bot. Fennici 37: 219–226 ISSN 0003-3847
Helsinki 27 September 2000

Ecological indicator values of British species: an application of Gaussian logistic regression

David B. Roy, Mark O. Hill, Peter Rothery & Robert G.H. Bunce

Roy, D. B., Hill, M. O., & Rothery, P., Centre for Ecology and Hydrology, Monks Wood, Abbots Ripton, Huntingdon, PE28 2LS, UK
Bunce, R. G. H., Centre for Ecology and Hydrology, Merlewood Research Station, Grange-over-Sands, Cumbria, LA11 6JU, UK

Received 25 January 2000, accepted 7 April 2000

Roy, D. B., Hill, M. O., Rothery, P. & Bunce, R. G. H. 2000: Ecological indicator values of British species: an application of Gaussian logistic regression. — *Ann. Bot. Fennici* 37: 219–226.

In a large ecological survey of Britain,13 841 quadrats were sampled in 508 1-km squares. The quadrats included 1 132 species of vascular plants, of which 643 occurred in 10 or more quadrats. Applying the method of Gaussian logistic regression to data from this survey, ecological optima and tolerances of species were estimated for Ellenberg's seven ecological indicator variables. Tolerances showed very little relation to the original scales. Most optima were within the range of the original scales but a few species lacked optima for some variables. Optima showed a strong positive relation to original values, but the resulting scale was compressed. We propose a locally-weighted trend line to convert each optimum value to an estimate of the original value. Reprediction using methods based on large-scale quadrat samples offers a very good means of extending Ellenberg's values to a new geographic area such as Britain.

Keywords: ecological survey, Ellenberg value, environmental calibration, optimum, response function, tolerance

INTRODUCTION

Ellenberg's indicator values have been widely used to summarize the habitat preferences of vascular plants in Central Europe (Wittig & Durwen 1982, Ellenberg 1988, Roo-Zielinska & Solon 1988, Thimonier *et al*. 1994). These indicator values, sometimes known by their German name *Zeigerwerte*, allow each species to be placed on a scale according to its response to certain climatic and edaphic factors in the field. Ellenberg values define the realized niche of plants, not their fundamental niche (Thompson *et al*. 1993). They have been widely used in Germany (Ellenberg *et al*. 1991) and the Netherlands (Van der Maarel *et al*. 1985, Melman *et al*. 1988).

In Britain, indicator values have been less widely used, but can undoubtedly be useful in some contexts (Sparks *et al*. 1996, Hill & Carey 1997). A major attraction of them is that they can be used to monitor change in the countryside, particular when data from large repeated surveys are available (Thimonier *et al*. 1994). Doubts about the value of Ellenberg values outside Central Europe may have prevented them from being used more. They must inevitably become less reliable as one moves away from the region for which they were developed (Van der Maarel 1993). Not only will new species be represented (rather a small number of such in Britain), but species' preferences will change.

Thompson *et al*. (1993) have suggested that the ecological optima of species are dependent on the presence or absence of potential competitors, which change with geographic location. The ecological amplitude of species may also be narrower at the edge of their range; for example, *Hedera helix*, widespread and often growing as a liane in Britain, becomes restricted to the ground in parts of Europe with colder winters (Iversen 1944).

Ellenberg *et al*. (1991) recommended testing or calibrating indicator values in other regions. For countries where this has been attempted, there is good agreement with original values (Van der Maarel 1993, Diekmann 1995, Ertsen *et al*. 1998). For the purposes of monitoring changes in the British countryside, a standardized set of British values would be useful. Ter Braak and Gremmen (1987) have suggested that a technique based on Gaussian logistic regression could be used to extend the original values to a new area. We have tested their method on British data.

DATA AND METHODS

The Countryside Survey 1990

Countryside Survey 1990 (hereafter referred to as CS90) was a comprehensive survey of the British countryside, conducted in 1990 (Barr *et al*. 1993). A stratified random sample of 508 1-km squares was drawn from 32 relatively homogeneous strata called 'Land Classes'. In each 1-km square, records were made of land cover, landscape features, habitats and vegetation. The vegetation data are used here. Three plot types were used to record vegetation (Table 1):

1. main plots placed at random throughout the 1-km squares,
2. linear plots placed along hedgerows, streams and verges,
3. habitat plots targeted to provide additional information on areas of semi-natural vegetation.

Nomenclature of taxa follows Stace (1991).

Indicator values

Indicator values were taken from the standard published source (Ellenberg *et al*. 1991), which provides scores for the large majority of species found in Britain. Ecological indicator values are available for seven scales, here given their German initials: L = light, T = temperature, K = continentality, F = moisture, R = reaction (pH), N = nitrogen, S = salinity. Using data from the Park

Table 1. Types and numbers of vegetation plots

Plot type	Max. per 1-km square	Total
X = Main plots (200 m^2)	5	3 805
Y = Habitat plots (4 m^2)	5	2 531
H = Hedge plots (1 m × 1 m)	2	847
B = Boundary plots (10 m × 1 m)	5	1 805
R = Verge plots (10 m × 1 m) — random	2	1 145
V = Additional verge plots (10 m × 1 m)	3	1 165
S = Streamside plots (10 m × 1 m) — random	2	1 258
W = Additional streamside plots (10 m × 1 m)	3	1 285
Total		13 841

Grass Experiment, Hill and Carey (1997) have already shown that the indicator values for N are as much an indication of overall productivity as of nitrogen.

Gaussian logistic regression

Ter Braak and Looman (1986) proposed Gaussian logistic regression as a means of estimating ecological indicator values and amplitudes of species, given a series of samples with measured values of an ecological variable and data on species' presence or absence. Simulations have shown this method to be generally more reliable than simple weighted average of presence-absence data alone (Ter Braak & Looman 1986). The response of a species describes the probability, *p(x)*, that the species occurs as a function of an environmental variable *x*. The Gaussian-logit curve models the presence-absence response of a species:

$$\log_e [p(x)/1 - p(x)] = b_0 + b_1x + b_2x^2 = a - 0.5(x - u)^2/t^2 \quad (1)$$

where u is the species optimum or indicator value (the value of x with highest probability of occurrence) and t is its tolerance (a measure of ecological amplitude). The parameters b_0, b_1 and b_2 can be estimated by logistic regression to obtain the following:

$$\text{optimum } u = -b_1/2b_2 \quad (2)$$

$$\text{tolerance } t = 1/\sqrt{(-2b_2)} \quad (3)$$

$$\text{maximum probability } p_{\max} = p(u) = 1/[1 + \exp(-b_0 - b_1u - b_2u^2)] \quad (4)$$

The Gaussian-logit response curve is bell-shaped and symmetrical, and therefore its optimum is identical to its mean. Gaussian logistic regression was performed by the statistical package Genstat (Genstat 5 Committee 1993).

Reprediction of indicator values from an existing set

A procedure using the Gaussian-logit model has been used to determine the amplitude of plant species responses and to test the internal consistency of Ellenberg's indicator values for moisture in the Netherlands (Ter Braak & Gremmen 1987). For any ecological indicator variable, the procedure consists of two steps.

1. For each sample quadrat, calculate the mean indicator score for those species which have an initial ecological indicator value.
2. For each species, use Gaussian logistic regression to calculate an optimum and tolerance, based on the quadrat means defined at stage 1.

The essential feature of this method is that it treats the mean values of species indicator values as if they were the value of a measured variable. The method was applied to CS90 data as a means of extending Ellenberg's ecological indicator values to Britain.

If the optimum lies outside or near the edge of the sampled range, the optimum is poorly estimated. When testing the Gaussian-logit response of species to Ellenberg indicator scales, the sampled range was restricted to the original range of the indicator values. In cases where the optimum lay outside the range of the indicator variable, the response curve was necessarily truncated and a sigmoid curve (linear logit curve) is fitted.

RESULTS

Distribution of quadrat indicator values

The distribution of quadrat indicator values (stage 1 of the reprediction procedure described above) differed widely between the indicator variables (Fig. 1). The variables *L*, *T*, *K* and *F* showed a simple unimodal pattern. Light values, *L*, had a negative skew, corresponding to the fact that low-light conditions are relatively rare in the British countryside, which is poorly wooded.

Both *R* and, especially, *N* showed a bimodal pattern of variation. This reflects the two types of countryside present in Britain, namely the intensively-used land of the lowland zone and the extensive, mostly acid countryside of the uplands.

High *S* values are rarely found in the British countryside except on the coast. It is perhaps surprising that there was an apparently unimodal distribution of *S* values. It should be noted that the

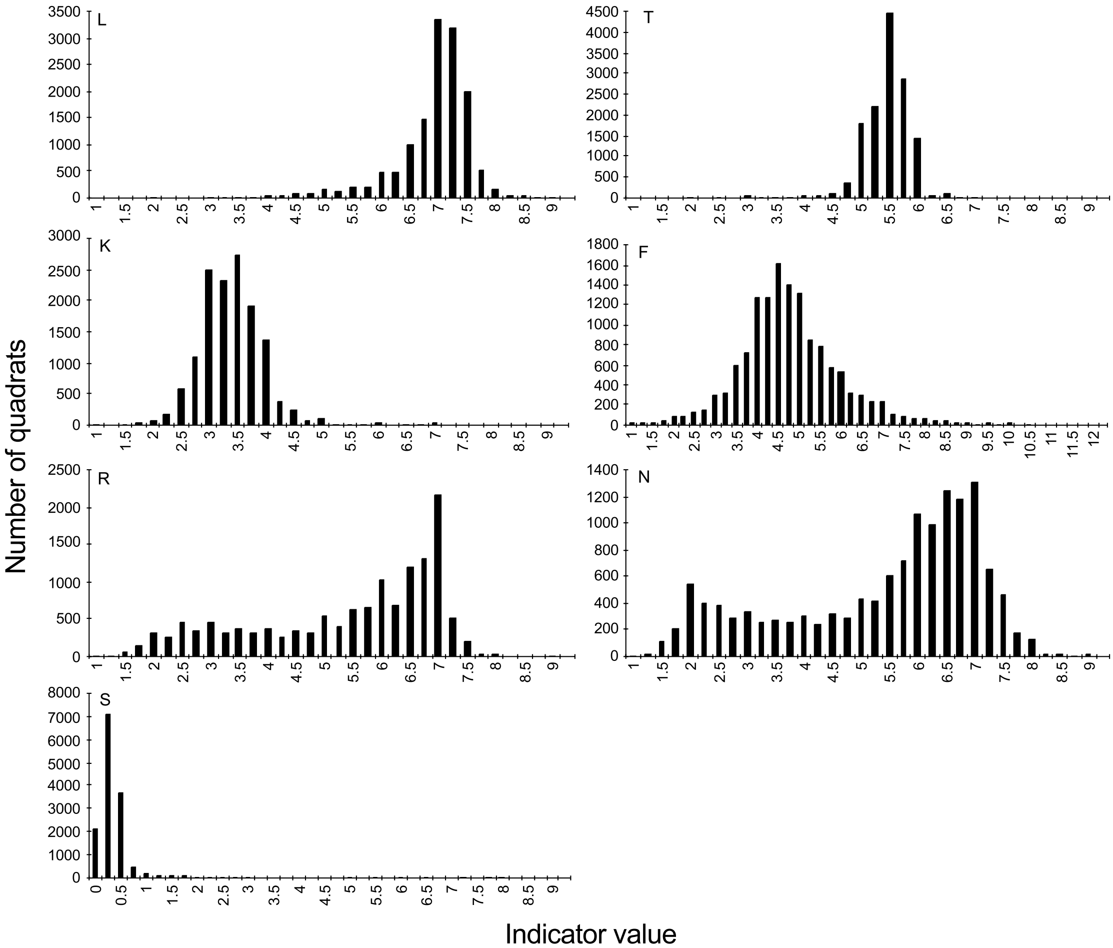

Fig. 1. Mean ecological indicator values for 13 841 quadrats sampled in Britain during the Countryside 1990 survey.

mode falls at 0.2, which does not indicate salty conditions, but merely that at least one species of plant present has some salt tolerance, at least in coastal ecotypes.

Categories of response

In total, 1 132 vascular plant species were recorded in the vegetation data, of which 643 occurred in 10 or more quadrats (Table 2). The most common species, *Holcus lanatus*, occurred in 5 855 quadrats, 43% of the total.

Species responses can be listed in six categories (Table 2). The significance of an optimum can be judged by whether the quadratic coefficient b_2 is significantly less than 0. The majority of species optima were significant at the 5% level (Good optima) or non-significant but with negative b_2 and estimated optimum falling within the range of Ellenberg values (Weak optima). For many species, such as *Cirsium vulgare* (Fig. 2) a significant optimum was estimated for each Ellenberg scale.

For a few species, the fitted optimum was beyond the maximum or minimum of the original scale. These are indicated as Truncated good optima (b_2 significantly less than 0) or Truncated weak optima (b_2 negative but not significantly so). The truncation consisted of reassigning their optimum to the maximum or minimum of the original scale.

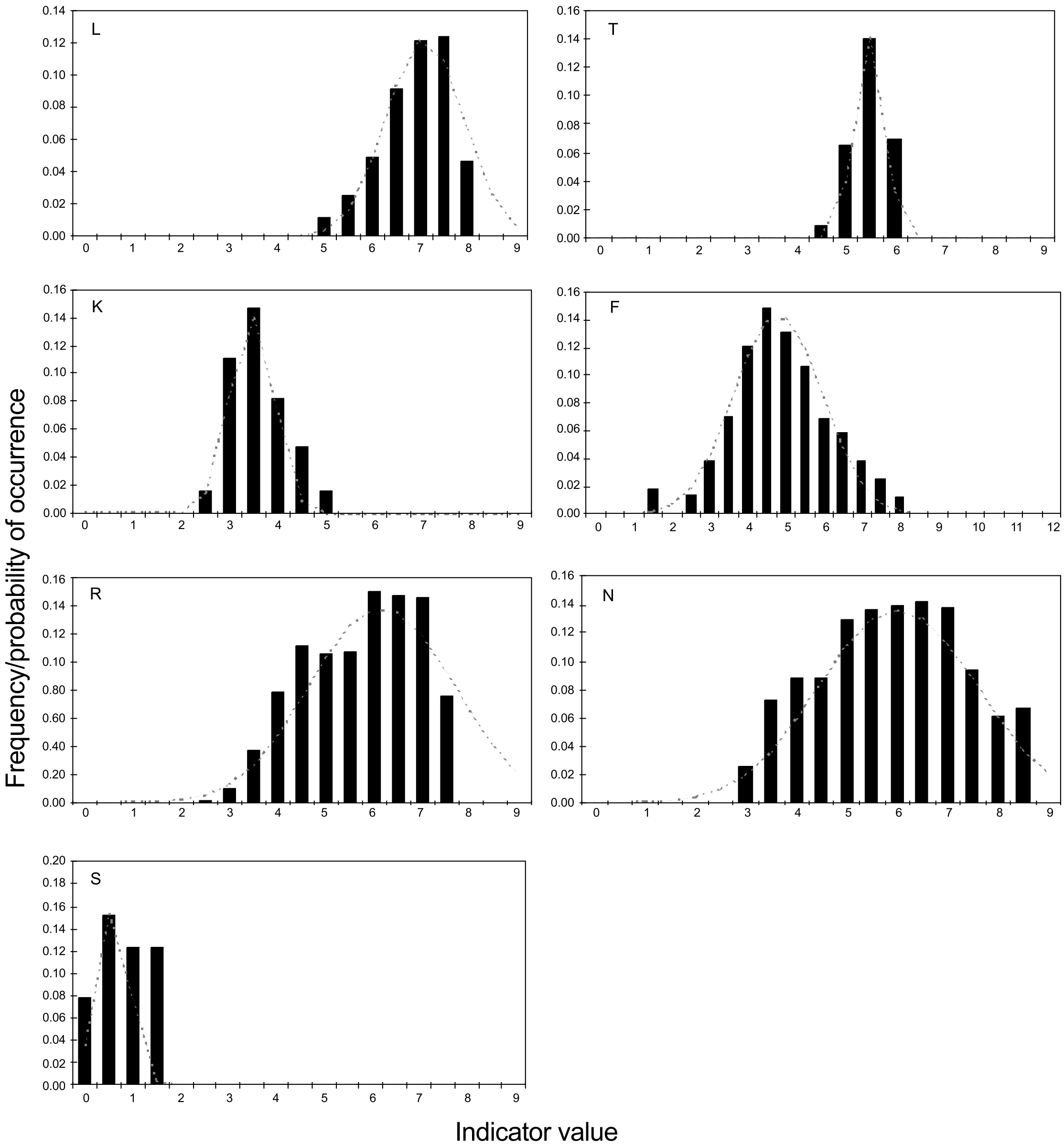

Fig. 2. Observed (solid bars) and fitted (dashed line) response frequencies for *Cirsium vulgare* for the ecological indicator variables *L*, *T*, *K*, *F*, *R*, *N*, *S*.

A small number of species did not have optima; i.e. the logit quadratic coefficient was positive. Where these species had significant linear logistic regression coefficients (b_1 significantly positive or negative), they were assigned to the category Linear logit (Table 2). An example of such a response is presented (Fig. 3). Species showing a linear logit response were given a predicted value which was the maximum or minimum value for the scale in question.

Finally, those species where the logit quadratic coefficient was positive and the linear logit curve was not significant were assigned to the category Trough (Table 2). A striking example is provided by heather *Calluna vulgaris*, which has a wide tolerance for moisture (Fig. 4). Even this

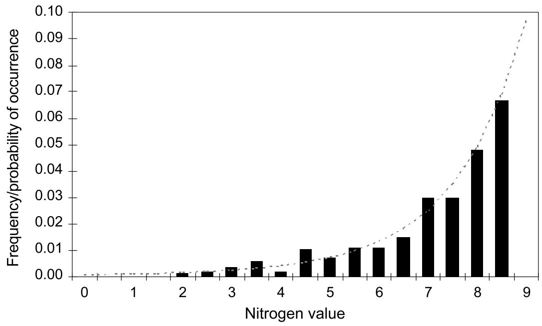

Fig. 3. Observed (solid bars) and fitted (dashed line) response frequencies for *N* (Nitrogen) for *Phragmites australis*.

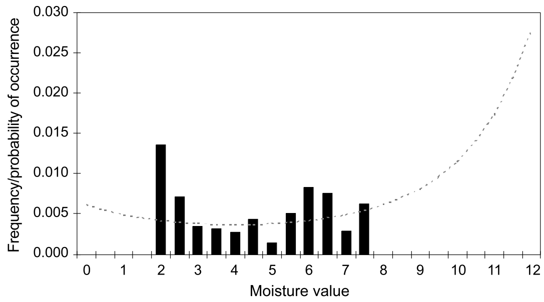

Fig. 4. Observed (solid bars) and fitted (dashed line) response frequencies for *F* (Moisture) for *Calluna vulgaris*.

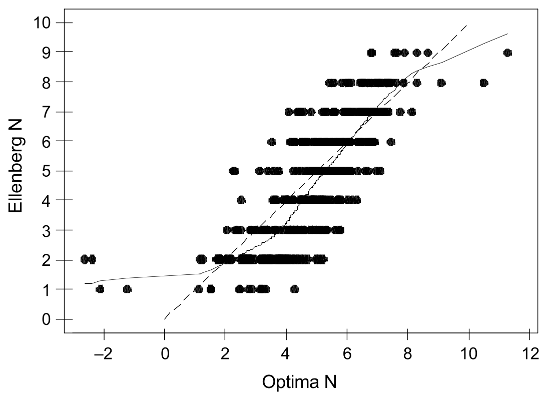

Fig. 5. Optima for *N* (nitrogen) in relation to original Ellenberg values. The curved line (solid) is a locally-weighted mean of original values, calculated for each possible optimum value. The straight line (dashed) shows the relationship that should apply if the optima were exactly equal to the original Ellenberg values.

wide tolerance has its limits. Heather is common on peat bogs, but only on hummocks; it does not occur in situations where there is prolonged submergence ($F > 8$). This is a small but not completely negligible class of samples (Fig. 1).

Optima and tolerances in relation to original values

There was some general agreement between the Gaussian logistic regression point estimates and the original values. Of 488 species for which both an original Ellenberg and Gaussian logistic regression estimate of N is available, 11 species had the same value, 294 had a difference of one or less, 482 of two or less. However, the relation between the optima and original values was not linear (Fig. 5), nor did Gaussian regression prevent the shrinkage which is often regarded as a fault of the weighted-averaging method of calibration (Hill & Gauch 1980, Ter Braak & Gremmen 1987).

When only the species with good or weak optima are considered, the relation between optima and original values appears to be even weaker (Table 3). For the variable *K*, the mean optimum rose

Table 2. Numbers of species in classes of logistic regression fit

Class of fit	*L*	*T*	*K*	*F*	*R*	*N*	*S*
Good optima	436	497	474	444	455	513	351
Weak optima	176	139	146	152	134	102	185
Truncated good optima	1	0	0	3	3	8	0
Truncated weak optima	14	0	3	15	39	12	16
Linear logit	11	3	13	20	11	6	73
Trough	5	4	7	9	1	2	18
Total	643	643	643	643	643	643	643

from 2.9 to 3.7 when the value of the original variable rose from 2 to 5. Although this is perhaps an extreme case, reflecting the unsuitability of Central European *K* values in Britain (Preston & Hill 1997), it shows that the optima are not necessarily a good reprediction.

Tolerance values varied little along the gradient (Fig. 6). There was only a very slight tendency for supposedly wide-tolerance species, which were accorded the rating *x* by Ellenberg, to have broader observed tolerances in Britain. Species signified by * in Fig. 6, which were not included at all in Ellenberg's enumerations showed tolerances almost identical to those which were rated *x*.

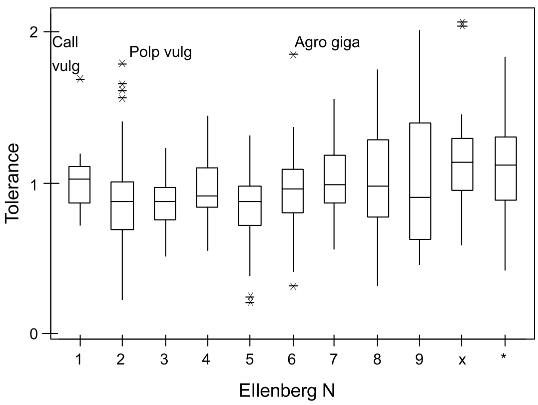

Fig. 6. Tolerance for *N* (nitrogen) in relation to original Ellenberg values; the value *x* denotes a species designated as wide-ranging for *N*, the value * denotes a species not scored for *N*.

DISCUSSION AND CONCLUSIONS

Although field naturalists commonly use the associates of a given species to infer its ecological attributes, this method has several limitations if quadrat data are all that are available. Loss of information on both context and plant architecture makes light values particularly difficult to infer from quadrat lists. In principle, plant size and life form could be used to find out whether a species was small and growing in the shade of larger ones, but this would necessitate a methodology that was quite specific to light. Even then, the Ellenberg light values for trees are defined to be the values in which young trees can develop. They cannot be inferred by examining mature stands.

For all variables, the optima were compressed to a smaller range than that of the original variables. This compression means that the optima cannot be used directly to reconstruct values where these are not known. A good way to do this is by means of the locally-weighted trend line shown in Fig. 5. Each optimum value can be converted to an estimate of the original value on the trend line. For example *Ulex europaeus* has its optimum for *N* estimated at 4.3. This would then be

Table 3. Optima and numbers of species for those species whose estimated optimum falls within the range of original values (good and weak optima); species lacking an original value are indicated by the symbol *, those originally signified as wide-ranging by *x*.

	Ellenberg indicator value															
	0	1	2	3	4	5	6	7	8	9	10	11	12	*	*x*	All
L		4.9	4.6	4.9	5.2	5.6	6.4	6.7	7.5	7.9				6.3	6.1	6.6
		(1)	(7)	(15)	(33)	(36)	(76)	(184)	(135)	(22)				(92)	(10)	(612)
T			3.1	3.9	4.7	5.2	5.7	6.1	6.3					5.3	5.0	5.3
			(5)	(9)	(23)	(139)	(181)	(23)	(3)					(100)	(153)	(636)
K		2.5	2.9	3.2	3.5	3.7	5.2	4.2						3.6	3.7	3.4
		(15)	(92)	(214)	(37)	(61)	(4)	(14)						(97)	(86)	(620)
F				3.8	4.1	4.5	5.3	6.2	6.7	7.5	8.4	9.8	8.5	5.0	3.5	5.3
				(30)	(67)	(128)	(61)	(47)	(56)	(62)	(15)	(4)	(2)	(87)	(35)	(594)
R		2.0	3.1	3.7	4.2	4.9	5.7	6.3	6.4	5.9				5.4	5.6	5.4
		(6)	(27)	(40)	(40)	(32)	(46)	(123)	(44)	(4)				(96)	(130)	(589)
N		2.5	3.1	4.0	4.7	5.3	5.7	6.3	7.1	7.9				5.3	5.1	5.1
		(15)	(75)	(57)	(54)	(74)	(74)	(67)	(43)	(11)				(101)	(44)	(615)
S	0.4	0.7	1.7	1.2	2.0	2.8	3.1	4.5	6.8	6.4				0.7		0.7
	(383)	(40)	(4)	(3)	(5)	(3)	(4)	(3)	(7)	(2)				(82)		(536)

corrected to $N = 3.6$. Likewise, the extreme values can be converted. The optimum for *Trichophorum cespitosum* is –1.3 but would be converted to $N = 1.4$.

Progress in extrapolating ecological indicator values to new geographical areas will be greatest when the variables can be defined by external criteria. Some of the original variables may prove to be poorly defined. It is clear, also, that a more detailed study of the patterns of species occurrence in environmental space is required. A univariate approach such as that used here takes no account of the correlation between different variables. The method also takes no account of variations in species richness between differing parts of a given gradient. In spite of these difficulties, the estimated values obtained from the corrected optima were generally good, with root-mean-square differences from the original values of 1.0, 0.3, 0.9, 0.8, 2.0 and 1.4 for *L*, *T*, *K*, *F*, *R* and *N* respectively. In the end repredictions must be checked for reliability, because some of the optima were only marginally significant and because some species, especially vernal geophytes, may have been missed in many samples. There is no doubt, however, that reprediction using methods based on large-scale quadrat samples is the best means of extending Ellenberg's values to a new area such as Britain.

ACKNOWLEDGEMENTS: This work was funded by the U.K. Natural Environment Research Council through the thematic programme URGENT, award number GST/02/1979 and by the U.K. Department of the Environment, Transport and the Regions ECOFACT project, contract CR0175.

REFERENCES

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*. — HMSO, London. 174 pp.

Diekmann, M. 1995: Use and improvement of Ellenberg's indicator values in deciduous forests of the Boreo-nemoral zone in Sweden. — *Ecography* 18: 178–189.

Ellenberg, H. 1988: *Vegetation ecology of Central Europe* (4 ed). — Cambridge Univ. Press, Cambridge. 731 pp.

Ellenberg, H., Weber, H. E., Düll, R., Wirth, V., Werner, W. & Paulissen, D. 1991: Zeigerwerte von Pflanzen in Mitteleuropa. — *Scripta Geobotanica* 18: 1–248.

Ertsen, A. C. D., Alkemade, J. R. M. & Wassen, M. J. 1998: Calibrating Ellenberg indicator values for moisture, acidity, nutrient availability and salinity in the Netherlands. — *Pl. Ecol.* 135: 113–124.

Genstat 5 Committee 1993: *Genstat 5 Release 3 Reference Manual*. — Clarendon Press, Oxford. 796 pp.

Hill, M. O. & Gauch, H. G. 1980: Detrended correspondence analysis: an improved ordination technique. — *Vegetatio* 42: 47–58.

Hill, M. O. & Carey, P. D. 1997: Ellenberg indicator values predict yield in the Rothamsted Park Grass Experiment. — *J. Veg. Sci.* 8: 579–586.

Iversen, J. 1944: *Viscum*, *Hedera* and *Ilex* as climatic indicators. — *Geologiska Föreningens Förhandlingar* 66: 463–483.

Melman, T. C. P., Clausman, P. H. M. A. & Udo de Haes, H. A. 1988: The testing of three indicator systems for trophic state in grasslands. — *Vegetatio* 75: 143–152.

Preston, C. D. & Hill, M. O. 1997: The geographical relationships of British and Irish vascular plants. — *Bot. J. Linn. Soc.* 124: 1–120.

Roo-Zielinska, E. & Solon, J. 1988: Phytosociological typology and bioindicator values of plant communities, as exemplified by meadows in the Nida valley, southern Poland. — *Docum. Phytosociol.* 11: 543–554.

Sparks, T. H., Greatorex-Davies, J. N., Mountford, J. O., Hall, M. L. & Marrs, R. H. 1996: The effects of shade on the plant communities of rides in plantation woodland and implications for butterfly conservation. — *Forest Ecol. Managem.* 80: 197–207.

Stace, C. A. 1991: *New flora of the British Isles*. — Cambridge Univ. Press, Cambridge. 1226 pp.

Ter Braak, C. J. F. & Gremmen, N. J. M. 1987: Ecological amplitudes of plant species and the internal consistency of Ellenberg's indicator values for moisture. — *Vegetatio* 69: 79–87.

Ter Braak, C. J. F. & Looman, C. W. N. 1986: Weighted averaging, logistic regression and the Gaussian response model. — *Vegetatio* 65: 3–11.

Thimonier, A., Dupouey, J. L., Bost, F. & Becker, M. 1994: Simultaneous eutrophication and acidification of a forest ecosystem in North–East France. — *New Phytologist* 126: 533–539.

Thompson, K., Hodgson, J. G., Grime, J. P., Rorison, I. H., Band, S. R. & Spencer, R. E. 1993: Ellenberg numbers revisited. — *Phytocoenologia* 23: 277–289.

Van der Maarel, E. 1993: Relations between sociological-ecological species groups and Ellenberg indicator values. — *Phytocoenologia* 23: 343–362.

Van der Maarel, E., Boot, R., Van Dorp, D. & Rijntjes, J. 1985: Vegetation succession on the dunes near Oostvoorne, The Netherlands; a comparison of the vegetation in 1959 and 1980. — *Vegetatio* 58: 137–187.

Wittig, R. & Durwen, K.-J. 1982: Ecological indicator-value spectra of spontaneous urban floras. — In: Bornkamm, R., Lee, J. A. & Seaward, M. R. D. (eds.), *Urban ecology*: 23–31. Blackwell, Oxford.