European Arachnology 2002 (F. Samu & Cs. Szinetár eds.), pp. 143-150.
© Plant Protection Institute & Berzsenyi College, Budapest, 2004. ISBN 963 214 791 X
(Proceedings of the 20th European Colloquium of Arachnology, Szombathely 22-26 July 2002)

Does stochasticity in emergent pioneer communities reduce the value of indicator species analysis?

JAMES R. BELL[1],EMMA M. SHAW[2] & C. PHILIP WHEATER[2]
[1] *Horticulture Research International, Wellesbourne, Warwick, CV35 9EF. England.*
(james.bell@hri.ac.uk)
[2] *Department of Environmental & Geographical Sciences, The Manchester Metropolitan University, Manchester, M1 5GD, England.*

Abstract

There is a need to understand more about how community dynamics affect the value of spiders as indicator species. In this paper, we discuss how community dynamics interact with indicator values, using the program 'IndVal', a non-hierarchical analysis tool. IndVal expresses a percentage indicator value for each species in each habitat group, based upon combining their relative abundance and frequency. Spiders were sampled for a six year period using pitfall traps which collected individuals from two contrasting habitats lying at perturbation extremes. Natural daleside (ND) was compared with two landform replication sites R5 and R8, which were created through blasting post-quarry faces. Whilst the fauna were different between ND, R5 and R8, species richness and dominance were comparable. Significant differences were found between the rates of species turnover. Whilst no significant difference lay between R5 and R8, both had significantly higher rates of turnover than ND. Conversely, the number of indicator species was highest in ND compared to both R5 and R8. Rates of species turnover and IndVal values are indirectly linked: when turnover rates are higher, IndVal percentages fall. This is an artefact of the fidelity of a species to a particular habitat which is skewed strongly in favour of species from mature habitats. By their very nature, the types of species that are common to emergent pioneer habitats have stochastic colonisation patterns. Pioneer species often indicate important components of the habitat (e.g. bare ground) and are generally of high indicator value, despite their lower fidelity values. Therefore, there is a need to adopt a less stringent interpretation of IndVal values when emergent pioneer habitats are included in any analysis.

Key words: IndVal, species turnover, DECORANA, spiders

INTRODUCTION

Historically, plants have been used as biological indicators of habitat conditions. Some plant indicators have made their way into folklore suggesting a long association with man, but more recent is the bio-indicative use of insects and spiders (Speight 1986; Maelfait et al. 1989; Maelfait 1996). Spider species have been suggested to have value in indicating habitat stress (Maelfait & Hendrickx 1998), habitat quality (Vollrath 1988) and habitat management (McIver et al. 1990; Wheater et al. 2000; Samu & Szinetár 2002), for example. Whilst some spiders are eurytopic, others are stenotopic and strongly associated with a set of environmental conditions. In such circumstances, a stenotopic distribution which is related to one or more habitat components is a useful tool for both conservation and research arachnology. However, interpretation of the merit given to a potential indicator species has sometimes been based on anecdotal, rather than on strict scientific grounds and has only served to slow progress. To overcome human bias, statistical software has been developed to elucidate upon whether particular species within a community may have some biological value in describing the environmental conditions present. There are two different forms which include hierarchical (TWINSPAN) and non-hierarchical (IndVal) approaches. The non-hierarchical approach is favoured because it is seen to be more statistically robust, less arbitrary and not at all divisive (see Dufrệne & Legendre 1997 and references therein).

In an earlier paper on indicator species from quarry habitats, we analysed a similar data set using TWINSPAN (Wheater et al. 2000: the data set used were sampled over fewer years but more sites). Separately, we also investigated the role of species turnover in quarry habitats (Bell et al. 1998). However, we did not formerly bring these two concepts together, believing them to be mutually exclusive. In this paper, the implications of using IndVal where there is both a strong habitat gradient present in the data set and disproportionate rate of species turnover between habitats will be discussed.

MATERIALS AND METHODS

Study sites and collection techniques

All the quarry sites under investigation are situated within a 5 x 2 km area, 3 km east of the town of Buxton, Derbyshire, lying just outside the boundary of the Peak District National Park in the English uplands. Two types of sites were selected for this study: natural daleside (ND) acted as the control and was compared with two landform replication sites (R5 and R8) which were separated by 300 m and located within Tunstead, a limestone quarry. ND supports a rich semi-natural *Festuca ovina – Avenula pratensis* grassland with prominent rock buttresses, headwalls and both bare and vegetated screes. Conversely, R5 and R8 have been created using 'restoration blasting', a method which seeks to replicate the local landscape after quarrying. R5 and R8 were blasted on the 25th November 1988 and 21st April 1989 respectively. Both were hydroseeded to produce a *Festuca ovina – Avenula pratensis* grassland.

Pitfall traps were used on areas of vegetated scree cones and slopes at Tunstead quarry and surrounding area for six years from 1992 to 1995 and 1997 to 1998. The pitfall traps were of standardised dimensions: plastic cups of 65 mm diameter and 90 mm depth. The killing and preserving liquid was an aqueous solution of 5% formalin with a small amount of detergent added (~1%). Two groups of five pitfall traps per site were placed in a line and flush with the soil surface at ND, R5 and R8 (i.e. N=10 pitfall traps x 3 sites). During the six years, pitfall trapping took place for the first two weeks of every month, leaving a gap of two weeks before the next collection. Trapping ran from early April until early October.

Statistical analysis

IndVal, a new statistical non-hierarchical analysis tool, was used to generate an indicator species list. For each species in each site an indicator value was calculated, based upon combining their relative abundance and frequency (see Dufrệne & Legendre 1997). IndVal uses a simple suite of equations relating to species *i* and sites *j* included within dataset (i. e. the rows and columns of a spreadsheet). The degree of habitat specificity (A_{ij}) is defined first by using:

$A_{ij}= \mathrm{Nindividuals}_{ij}/\mathrm{Nindividuals}_{i}$

This translates as the mean number of species across sites of a defined group divided by the sum of the mean number of individuals of a species across all groups. Having established that species *i* may have a degree of habitat specificity, IndVal then defines how faithful species *i* is to any one group, known as the measure of fidelity (B_{ij}), given by:

$B_{ij}=\mathrm{Nsites}_{ij}/\mathrm{Nsites}_{j}$

The measure of habitat fidelity is defined by the number of sites in a group where species *i* is present divided by the total number of sites in that group. IndVal is then able to compute IndVal$_{ij}$, the indicator value, which ranges from 0 (no indication) to 100 (perfect indication). This value is at its maximum when a species is confined to a single habitat type, is always abundant there and never absent. IndVal$_{ij}$ is effectively a percentage of the loyalty of a species to a group, given by the equation:

IndVal$_{ij}$= habitat specificity (A$_{ij}$) x species fidelity (B$_{ij}$) x100

Once IndVal$_{ij}$ has calculated the index for each species across the suite of habitats, the largest value will be indicative of a species that can casually be associated with a particular habitat type. This value is known as *IVmax* – literally, the maximum IndVal$_{ij}$ observed across the habitat gradient. That a habitat has been singled out as potentially indicative may be informative but is only inferential. To know whether this value (i.e. *IVmax*) is of some statistical importance ($<P=0.05$), Monte Carlo randomisation tests are used to reallocate these data to see if the observed *IVmax* could be an artefact of chance (i.e. a random effect). We refer to *IVmax* implicitly throughout the text as the 'indicator value' - meaning that a species shows a statistical degree of habitat specificity.

To describe the community dynamics, diversity indices and species turnover between habitats (i.e. ND, R5 and R8) were calculated. Diversity (Simpson, Margalef) and similarity indices (Jaccard index) were used to show that the composition of the spider communities were equitable. Whilst diversity indices are simple to calculate, species turnover may be more complex. Species turnover was calculated using DEtrended CORrespondence ANAlysis (DECORANA known as DCA) using a method described by Bell et al. (2002). In short, each of the six alternate fortnightly collections were summed for each year generating a matrix of 36 samples by 76 species of total count data. Once DCA produced a biplot, Euclidean distances were measured from the axes scores for consecutive between year shifts to calculate species turnover (e.g. NDB1>NDB2; NDB2>NDB3; NDB3>NDB4; NDB4>NDB5; NDB5>NDB6) for each of the sites (i.e. ND; R5; R8). These distances indicated the species turnover in *n* dimensional space over time. Once calculated, Euclidean distances were analysed using a two-way ANOVA with fixed (habitat, year) effects to determine if there was a significant difference in rates between the control (ND) and the landform replication sites (R5; R8). The factorial structure of the ANOVA was simply 3 sites x 5 consecutive between year shifts. F_{max} and Kolmogorov-Smirnoff tests were used to check for homoscedasticity and normality of data: in all cases, the ANOVA assumptions were met. However, whilst all other criteria were satisfied for a normal ANOVA design, we analysed Euclidean distances which were derived from a DCA ordination, raising questions about the independence of the data. This occurred because the distance between the DCA points were measured on the same xy coordinate grid (i.e the same problem as if one was to measure distance between points on a map), and therefore were dependent on each other. To make sure that the normal ANOVA was not subject to a Type I or II error as a result of this violation, we used a Monte Carlo checking procedure which generates its own 'F' distribution. Monte Carlo randomisation tests (30,000 randomisations) were used to establish whether the null hypothesis (i.e. that the result could be generated from a random distribution) should be rejected (Manly 1991). The level of significance in a Monte Carlo test was expressed as the percentage of values which are equal to, or higher than can be found in a randomised distribution. If the percentage of values that exceeded the observed mean square was less than 5%, then the null hypothesis would be rejected. In simple terms, if the outcomes were the same between the normal ANOVA and the Monte Carlo ANOVA, it would suggest that a Type I or II error had probably not occurred.

RESULTS

The number of statistically significant indicator species for each habitat revealed that much higher numbers were associated with ND (N =16) than with either R5 (N =6) or R8 (N =1) (Table 1). Furthermore, the percentage indicator values had a much higher ceiling in ND (94%) than either R5 (69%) or R8 (59%) with seven indicator species in excess of a 74% fidelity to ND (Table 1).

Whilst the dominance and richness indices varied, there was no real magnitude of difference, suggesting that these measures were broadly equitable between habitats (Table 2). Divergence was indicated by the Jaccard index which suggested that the composition of the fauna between the two landform replication sites was more similar than any comparison with the natural daleside (Table 2), in agreement with a previous study using DECORANA ordinations (Bell et al. 1998).

Two-way ANOVAs were used to compare the species turnover for the totals distance matrix derived from the DCA axes scores. Species turnover was significantly different between habitats ($F_{2,15}$=6.31, P=0.0102). Post hoc tests revealed that the ND habitat had significantly lower means than either the R5 ($F_{1,15}$=8.71, P=0.0098) or the R8 ($F_{1,15}$=10.15, P=0.0061) habitats. However, a comparison of means between R5 and R8 revealed that they were not significantly different in size ($F_{1,15}$=0.05, *P*=NS). Monte Carlo tests underpinned these ANOVA results: there was strong evidence that observed test results could not be generated randomly (i.e. only 0.42% of randomisations exceeded the observed mean square and therefore the null hypothesis should be rejected).

Table 1. Indicator species with their respective overall indicator value (*IVmax*) are expressed as a percentage. The indicator value elucidates upon the statistically significant association for one of the three habitats under study (ND, R5 and R8). A breakdown of the indicator values (IndVal_{ij}) across the sites is also given. * Statistically significant (*P*<0.05), ** Statistically significant (*P*<0.01)

	Habitat indicator	*IVmax*	Distribution of indicator values (%) across the three habitat types		
			ND	R5	R8
Alopecosa pulverulenta	ND	95**	94	0	2
Agroeca proxima	ND	93**	93	0	1
Ozyptila atomaria	ND	92**	92	0	0
Gonatium rubens	ND	83**	83	0	0
Drassodes cupreus	ND	80**	79	3	5
Trochosa terricola	ND	76**	76	5	10
Pardosa pullata	ND	74**	74	2	15
Euophrys frontalis	ND	67**	67	0	0
Zelotes apricorum	ND	67**	67	0	0
Meioneta saxatilis	ND	55**	54	1	0
Ceratinella brevis	ND	49**	49	1	1
Pocadicnemis pumila	ND	42*	42	3	0
Clubiona diversa	ND	42**	42	0	0
Hahnia montana	ND	37*	37	2	3
Heliophanus flavipes	ND	33*	33	0	0
Tegenaria silvestris	ND	33*	33	0	0
Coelotes atropos	R5	70**	1	69	19
Typhocrestus digitatus	R5	65**	0	65	3
Pardosa palustris	R5	63**	26	63	0
Tegenaria agrestis	R5	50**	0	49	6
Erigone atra	R5	41*	0	41	10
Haplodrassus signifer	R5	33*	0	33	0
Pardosa amentata	R8	59**	0	7	59

Table 2. Comparative community structure of the spider communities between habitats.

	Number of species	Number of individuals	Simpson	Margalef	Jaccard
			Dominance	*Richness*	*Similarity*
ND	46	1349	6.551	6.244	
R5	49	846	8.717	7.121	
R8	41	546	9.328	6.347	
ND v R5					0.357
ND v R8					0.338
R5 v R8					0.579

Consequently, the normal ANOVA was not likely to be subject to a type I or II error suggesting that the approach was valid.

DISCUSSION

IndVal analysis is a simple, but extremely effective tool used to calculate the level of indicator species - habitat association by combining the relative frequency and relative abundance (Dufrệne & Legendre 1997). From a statistical view point, IndVal overcomes many of the problems experienced with TWINSPAN, the alternative indicator analysis program. IndVal deals effectively with irregular habitat gradients, is more sensitive at identifying indicator species and is not restrained by any hierarchical structure which may impose itself on the identification of good indicator species. However, without wishing to detract from these improvements, the identification of pioneer indicator species from a dataset with a strong habitat gradient needs further attention. In our example, ecologists might draw the conclusion that there are fewer indicator species in the pioneer habitats R5 and R8 when compared to ND (Table 1). At face value, this is true, but IndVal selects against longer lists in pioneer habitats because of the statistical approach employed. In simple terms:

IndVal= habitat specificity x species fidelity x100 (see Dufrệne & Legendre 1997).

Habitat specificity for pioneer species is generally strong, in that species have evolved to cope with harsh environmental conditions. Ignoring vagrants, most species could be described as 'stenotopic' of pioneer habitats (i.e. restricted to this type of habitat for ecological reasons). Whilst all this maybe true, pioneer species are predisposed to less fidelity than indicators of mature habitats because of their nature to disperse, arguably to spread risk (den Boer 1968), whereas more mature habitats attract species which are less vagile over large areas. In the example given here, R5 and R8 were dominated by linyphiid spiders which are well known super-colonisers of new or highly disturbed habitats (e.g. Meijer 1977; De Keer & Maelfait 1988). But ballooning dispersal is random (Meijer 1977) accounting for the variation between years and the high turnover of individuals at R5 and R8 compared to ND. Therefore, species turnover and species fidelity are intuitively linked and indirectly proportional to one another. As turnover increases as a consequence of habitat disturbance, the faithfulness of most, if not all, species decrease. Organisms which live in harsh environments are more likely to be subject to higher mortalities because life is more stressful.

Let us assume, however, that the ability to locate and colonise any habitat is the same for all spiders in all habitats. Despite this equality, other factors still count against spiders once they have arrived. Poor overwintering potential also ensures lower fidelity. Many spiders overwinter either as immatures (e.g. *Pardosa* species) or as adults (e.g. *Tegenaria* species). Spiders respond to the lowering of air temperatures during winter by moving into sheltered vegetation, such as litter, for protection from temperature extremes (Edgar & Loenen 1974). However, overwintering sites were limited on R5 and R8 because they were removed by quarrying and then blasting. Since there was a lack of large amounts of dead vegetation under which to find cover, it is not known exactly where the colonists overwintered. The nearest overwintering habitat was on the quarry top, a steep climb of about 10 m. In contrast, there was excellent overwintering potential on the natural daleside which displayed a scree and soil mix up to 300 mm deep in some places. This depth allowed easy downward migration in winter and encouraged a high overwintering potential, increasing population stability between years and therefore higher fidelity values.

Prey quality and quantity should also be briefly considered. Although spiders have the ability to starve themselves, the lack of prey increases ballooning motivation (Weyman & Jepson 1994) which only serves to increase turnover and decrease the level of fidelity. In short, community ecology by its very nature is complex and it would be remiss to cite only a handful of factors that contribute to the differences between the turnover rates of mature and pioneer habitats - there are many. Why one individual stays and another leaves is in part, genetic, based on the experiences of the previous generation, partly cued into the environment when a spider senses that it physically can leave, and also partly inexplicable but probably related to the theory of bet-hedging (Hopper 1999).

The philosophical question arises then, 'how do we define an indicator species of a pioneer habitat?'. The answer must surely be, less stringently than for mature habitats. In the example given here, there were two cases where IndVal selected against potential good pioneer indicator species at R5, based on statistical grounds. *Erigone promiscua* is an indicator of pioneer habitats, particularly on frontal dunes and other barren landscapes, but despite being numerous (N =52) this spider only occurred at R5 in year 3. Arguably, this is a local extinction which should not be ignored. *Arctosa perita* is another associate of dunes, and had just colonised R5 in year 5, building up only a small population of 12 individuals by that time. Since both are known to be strong indicators of pioneer habitats, there is a case in favour of including them based not on statistical, but on ecological grounds.

But the situation is also confused, and less clear in parts making interpretation and generalisations difficult. For example, if the numbers of reoccurring species (i.e. those that arrive in one year, disappear in another and then recolonise) is analysed, the numbers found doing so were comparable between habitats (ND=14, R5=16, R8=15). Whilst there is a reasonable argument to suggest that within pioneer communities the process of recolonisation is habitat driven, it could be the case that the level of absenteeism in the mature habitat is related to trapping error. For example, the mature grasslands of the ND are much more structured and complex than the simple patchy tussocks of R5 and R8, which probably lead to a lower trap encounter rate by spiders in the ND. If this is true, should we in turn recognise this trapping disparity and include among the indicators those spiders of the ND whose trappability is lower? This could be easily done, but one must also recognise that those species with a lower trap encounter rate are likely to be less suitable as ecological indicators than those already selected. After all, we have a long list to choose from at ND, which is not the case for R8, for example (Table 1).

In short, we have unearthed a multivariate problem that may have a suite of probable explanations, all of which are currently untested. There is a need to understand these disparities between indicators of different habitats. Equally, we should also consider what it is that we want our indicator species to indicate. Then, having done so, we will be able to select more precisely those species that fit the paradigm. In the meantime, how do we go about dealing with the problems faced by indicator species analysis? At a practical level, there is still uncertainty on how adjustments can be made to the data to support the inclusion of more pioneer species. Regular data transformations (e.g. power transformations, logs, etc) will not solve the problem, as these only serve to dampen heterogeneity, bringing extreme values closer together. If it is the express aim to retain each year as a separate entity for later turnover analyses, then there is no simple statistical route on which to advise. However, a technique known as ' Beal's smoothing' may be useful in a few limited circumstances. Beal's smoothing is used where heterogeneous community data sets present problems and typically include a large number of zeros (McCune 1994). Beal's smoothing seeks to enhance the strongest patterns in the data by computing the "favourability" of each sample for each species, regardless of whether the species was present in the sample. However, if the whole data set were transformed using this technique, the result would be that meaningless Monte Carlo randomisations were generated when *IVmax* was iterated (McCune pers. comm.). Clearly, there is a needed for a professional biostatistical input at this level. However, the procedure might be to transform only the most problematic part of the data set to fill in the empty cells where there is a clearly identified problem. If there is no need to calculate turnover between years and therefore retain data sets as separate entities then a simple solution is to add the data sets together and remove the between year differences. A good example of how this has been achieved for spider communties has been demonstrated by Szita et al. (2004).

The general conclusion to be drawn is that whilst IndVal and other techniques have much to offer, the philosophy which should underpin the statistical approach is currently weak. IndVal can be used in most situations with confidence, except where there is a steep habitat gradient (e.g. pioneer to mature habitats). Here, IndVal should be used with caution, with some emphasis on alternative 'manual' ecological selection of indicator species from pioneer habitats.

ACKNOWLEDGEMENTS

We would like to thank all those at MMU who were helpful in more ways than can be described here, particularly Rod. We also thank Bruce McCune for discussions on Monte Carlo methods in ecology. Lastly, we thank two peer reviewers and Ferenc Samu for enlightening academic exchanges.

REFERENCES

Bell, J.R., Cullen, W.R. & Wheater, C.P. 1998. Structure of spider communities in limestone quarry environments. In: *Proceedings of the 17th European Colloquium of Arachnology*. (P.A. Selden ed.), pp. 253-259. British Arachnological Society, Burnham Beeches, UK.

Bell, J.R., Haughton, A.J., Boatman, N.D. & Wilcox, A. 2002. Do incremental increases of the herbicide glyphosate have indirect consequences for spider communities? *Journal of Arachnology* 30, 288-297.

den Boer, P.J. 1968. Spreading of risk and stabilization of animal numbers. *Acta Biotheoretica* 18, 165-194.

De Keer, R. & Maelfait, J.-P. 1988. Observations on the life cycle of *Erigone atra* (Araneae, Erigoninae) in heavily grazed pasture. *Pedobiologia* 32, 201-212.

Dufrêne, M. & Legendre, P. 1997. Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecological Monographs* 67, 345-366.

Edgar, W. & Loenen, M. 1974. Aspects of the overwintering habitat of the wolf spider *Pardosa lugubris. Journal of Zoology* 172, 383-388.

Hopper, K.R. 1999. Risk-spreading and bet-hedging in insect population biology. *Annual Review of Entomology* 44, 535-560.

McCune, B. 1994. Improving community analysis with the Beal's smoothing function. *Ecoscience* 1, 82-86.

McIver, J.D., Moldenke, A.R. & Parsons, G.L. 1990. Litter spiders as bio-indicators of recovery after clearcutting in a western coniferous forest. *Northwest Environmental Journal* 6, 410-412.

Maelfait, J.-P. 1996. Spiders as bioindicators. In: *Bioindicator Systems for Soil Pollution* (N.M. van Straalen & D.M. Krivolutsky eds.), pp. 165-178. Kluwer Academic Publishers, Dordrecht.

Maelfait, J.P., Desender, K. & Baert, L. 1989. Some examples of the practical use of spiders and carabid beetles as ecological indicators. In: *Comptes rendus du symposium 'Invertebres de Belgique'* (Anon ed.), pp. 437-442. Brussels, Belgium.

Maelfait, J-P. & Hendrickx, F. 1998. Spiders as bioindicators of anthropogenic stress in natural and semi-natural habitats in Flanders (Belgium): some recent developments. In: *Proceedings of the 17th European Colloquium of Arachnology*. (P.A. Selden ed.), pp. 293-300. British Arachnological Society, Burnham Beeches, UK.

Manly, B.F.J. 1991. *Randomization and Monte Carlo methods in Biology*. Chapman & Hall, London.

Meijer, J. 1977. The immigration of spiders (Araneida) into a new polder. *Ecological Entomology* 2, 81-90.

Samu, F. & Szinetár, Cs. 2002. On the nature of agrobiont spiders. *Journal of Arachnology* 30, 389-402.

Speight, M.C.D. 1986. Criteria for the selection of insects to be used as bio-indicators in nature conservation research. In: *Proceedings of the 3rd European Congress of Entomology, Amsterdam*. (H.H.W. Velthuis ed.), pp. 485-488. Amsterdam, The Netherlands.

Szita, É., Samu, F., Fetykó, K. & Szirányi, A. 2004. Testing the origin of agrobiont spiders: spiders in agricultural and natural grassland habitats of the Körös-Maros National Park, Hungary In: *Proceedings of the 20th European Colloquium of Arachnology*. (F. Samu & Cs. Szinetár eds.), pp. 319-326. Plant Protection Institute & Berzsnyi College, Budapest.

Vollrath, F. 1988. Spider growth as an indicator of habitat quality. *Bulletin of the British Arachnological Society* 7, 217-219.

Weyman, G.S. & Jepson, P.C. 1994. The effect of food supply on the colonisation of barley by aerially dispersing spiders (Araneae). *Oecologia* 100, 386-390.

Wheater, C.P., Cullen, W.R. & Bell, J.R. 2000. Spider communities as tools in monitoring reclaimed limestone quarry landforms. *Landscape Ecology* 15, 401-406.