

Appendix 1

Numerical simulation study

Simulation methods

We carried out a small Monte Carlo simulation study in order to illustrate the differences between correlation and indicator value measures when using them to select the correct site-group combination. For simplicity we compared the presence-absence case only, and generated a four-cluster structure with clusters being equal in size. This latter choice makes group-equalized indices to give the same results as unequalized indices, so the comparison can focus on the difference between the point-biserial correlation (r_{pb}) and indicator value ($IndVal_{ind}$) measures. Three distinct patterns of presence-absence-species were simulated: small patterns, where the species is related to one group (sp1); intermediate patterns, where the species is related to the combination of two groups (sp2); large pattern, where the species is related to the combination of three groups (sp3). Species-site group relationships were simulated in different degrees by setting the relative frequency of the species inside the related pattern and the relative frequency of the species outside it ranging from 0.00 up to 0.99 (Table A1). Only positive associations were studied (i.e. the inside relative frequency was always equal or greater than the outside relative frequency). We generated 1000 random binary (presence-absence) species patterns by using the Bernoulli law. We then assessed how often the maximum association value, obtained among the values corresponding

to all possible combinations of site groups, was identifying the true species related pattern.

Simulation results

When the species of interest is absent or almost absent outside the target pattern and with low to moderate frequency inside, we detected more easily small patterns than large patterns (Table A1). This occurred because with large patterns and low to moderate frequencies inside allow that sub-patterns of the target pattern arise by chance. Analogously, if the species of interest was moderate or highly frequent outside the target cluster combination, we identified more easily large cluster combinations than small ones. When increasing the number of sites, these differences tend to disappear and the chance of selecting the correct cluster combinations is equal for all three species patterns.

When the species was absent outside the target cluster both indicator value and correlation indices performed similarly, although $IndVal_{ind}$ was slightly more powerful in some cases. The differences between the indicator value and correlation measures were more evident as the relative frequency of the species outside the target cluster combination increased. In this case $IndVal_{ind}$ lost its ability to detect the 'true' related pattern because it sometimes considered that a larger site combination (i.e. larger pattern) was more appropriate to represent the species pattern. For very high species frequencies the indicator value measure concludes that the appropriate pattern is the whole set of sites. On the contrary, the point-biserial correlation index was able to detect any difference in frequency as long as the number of sites was enough.

Table A1. Numerical simulation results. Pattern – Associated species pattern (see text); in – relative frequency of the species inside the associated pattern; out – relative frequency of the species outside the associated pattern, $E(IndVal_{ind})$ – expected indicator value according to the simulated situation, $E(r_{pb})$ – expected point-biserial correlation coefficient according to the simulated situation. Values in the remaining columns are percentages of successful identification of the related pattern for the total number of sites indicated (n).

Pattern	Simulated situation		$\sqrt{E(IndVal_{ind})}$	Indicator value			$E(r_{pb})$	Correlation value		
	in	out		n = 20	n = 100	n = 1000		n = 20	n = 100	n = 1000
<i>sp1</i>	0.01	0.00	0.100	6	22	91	0.087	5	22	92
<i>sp2</i>	0.01	0.00	0.100	1	4	60	0.071	0	7	59
<i>sp3</i>	0.01	0.00	0.100	0	1	42	0.050	0	1	23
<i>sp1</i>	0.10	0.00	0.316	40	93	100	0.277	41	93	100
<i>sp2</i>	0.10	0.00	0.316	17	65	100	0.229	14	64	99
<i>sp3</i>	0.10	0.00	0.316	6	42	100	0.164	5	27	96
<i>sp1</i>	0.50	0.00	0.707	97	100	100	0.655	97	100	100
<i>sp2</i>	0.50	0.00	0.707	75	100	100	0.577	75	99	100
<i>sp3</i>	0.50	0.00	0.707	60	98	100	0.447	46	96	100
<i>sp1</i>	0.90	0.00	0.949	100	100	100	0.933	100	100	100
<i>sp2</i>	0.90	0.00	0.949	99	100	100	0.905	98	100	100
<i>sp3</i>	0.90	0.00	0.949	98	100	100	0.832	97	100	100
<i>sp1</i>	0.99	0.00	0.995	100	100	100	0.993	100	100	100
<i>sp2</i>	0.99	0.00	0.995	100	100	100	0.990	100	100	100
<i>sp3</i>	0.99	0.00	0.995	100	100	100	0.980	100	100	100
<i>sp1</i>	0.01	0.01	0.050	3	10	3	0.000	4	11	10
<i>sp2</i>	0.01	0.01	0.071	0	3	4	0.000	0	3	7
<i>sp3</i>	0.01	0.01	0.087	0	1	11	0.000	0	4	4
<i>sp1</i>	0.10	0.01	0.277	34	65	100	0.220	34	68	100
<i>sp2</i>	0.10	0.01	0.302	14	51	100	0.197	14	50	99
<i>sp3</i>	0.10	0.01	0.311	8	38	99	0.146	5	24	94
<i>sp1</i>	0.50	0.01	0.687	91	100	100	0.626	90	100	100
<i>sp2</i>	0.50	0.01	0.700	75	99	100	0.562	74	98	100
<i>sp3</i>	0.50	0.01	0.705	59	99	100	0.438	42	94	100
<i>sp1</i>	0.90	0.01	0.933	100	100	100	0.912	100	100	100
<i>sp2</i>	0.90	0.01	0.943	99	100	100	0.894	99	100	100
<i>sp3</i>	0.90	0.01	0.947	98	100	100	0.824	97	100	100
<i>sp1</i>	0.99	0.01	0.980	100	100	100	0.974	100	100	100
<i>sp2</i>	0.99	0.01	0.990	100	100	100	0.980	100	100	100
<i>sp3</i>	0.99	0.01	0.993	100	100	100	0.974	100	100	100
<i>sp1</i>	0.10	0.10	0.158	9	3	0	0.000	13	9	8
<i>sp2</i>	0.10	0.10	0.224	6	5	0	0.000	6	7	9
<i>sp3</i>	0.10	0.10	0.274	4	9	0	0.000	3	4	6
<i>sp1</i>	0.50	0.10	0.559	49	79	100	0.433	57	95	100
<i>sp2</i>	0.50	0.10	0.645	50	90	100	0.436	55	94	100
<i>sp3</i>	0.50	0.10	0.685	51	97	100	0.354	28	85	100
<i>sp1</i>	0.90	0.10	0.822	89	100	100	0.756	93	100	100
<i>sp2</i>	0.90	0.10	0.900	94	100	100	0.800	97	100	100
<i>sp3</i>	0.90	0.10	0.932	95	100	100	0.756	92	100	100
<i>sp1</i>	0.99	0.10	0.872	95	100	100	0.824	97	100	100
<i>sp2</i>	0.99	0.10	0.948	99	100	100	0.894	99	100	100
<i>sp3</i>	0.99	0.10	0.979	99	100	100	0.912	100	100	100
<i>sp1</i>	0.50	0.50	0.354	1	0	0	0.000	10	9	9
<i>sp2</i>	0.50	0.50	0.500	2	0	0	0.000	7	8	7

<i>sp3</i>	0.50	0.50	0.612	12	1	0	0.000	5	7	8
<i>sp1</i>	0.90	0.50	0.581	2	0	0	0.354	41	85	100
<i>sp2</i>	0.90	0.50	0.761	15	2	0	0.436	55	95	100
<i>sp3</i>	0.90	0.50	0.871	44	19	0	0.433	52	95	100
<i>sp1</i>	0.99	0.50	0.627	2	0	0	0.438	50	95	100
<i>sp2</i>	0.99	0.50	0.811	23	7	0	0.562	72	98	100
<i>sp3</i>	0.99	0.50	0.921	53	37	10	0.626	91	100	100
<i>sp1</i>	0.90	0.90	0.474	0	0	0	0.000	4	7	8
<i>sp2</i>	0.90	0.90	0.671	0	0	0	0.000	6	6	6
<i>sp3</i>	0.90	0.90	0.822	0	0	0	0.000	10	9	8
<i>sp1</i>	0.99	0.90	0.515	0	0	0	0.146	7	27	94
<i>sp2</i>	0.99	0.90	0.720	0	0	0	0.197	16	53	98
<i>sp3</i>	0.99	0.90	0.872	2	0	0	0.220	37	65	100
<i>sp1</i>	0.99	0.99	0.497	0	0	0	0.000	0	1	8
<i>sp2</i>	0.99	0.99	0.704	0	0	0	0.000	0	3	8
<i>sp3</i>	0.99	0.99	0.862	0	0	0	0.000	3	8	8