Appendix 1

Methods: Variable selection, Independence, and Year Effects

We describe here several aspects of our methods in greater detail than we can do in the paper itself. We first add further details regarding the explanatory effect of tree cover in comparison to other variables. We also discuss the problems of independence in the data set and of year effect in pooling data.

Selecting an explanatory landscape variable

We used percentage tree cover because it was best overall, as discussed in Cunningham and Johnson (2011). Percentage tree cover and edge density were equivalent in their influence on species. Because of the relatively high values for the best-explained species, percentage tree cover had the highest overall average R^2_L values. This finding was consistent at 5 different scales (Table A1.1). Because edge density is scale-dependent, in that it is influenced by scale and grain of analysis (Wu et al. 2002), we used percentage tree cover as our landscape descriptor for subsequent analysis.

Table A1.1. Average R^2 value for each variable for all species at 5 landscape scales, using quadratic models. The strongest measures at each scale are bolded.

	Scale (m)				
Variable	200	400	800	1200	1600
Pct tree cover	**0.14**	**0.12**	**0.10**	**0.08**	**0.08**
Edge density	0.13	0.11	0.09	**0.08**	0.07
Cohesion	0.13	0.10	0.07	0.06	0.05
Largest patch index	0.11	0.08	0.05	0.05	0.05
Core area	0.04	0.02	0.01	0.01	0.01
Mean patch size	0.06	0.06	0.05	0.05	0.05

Amount of tree cover was correlated with other measures of fragmentation. Percentage tree cover was strongly and positively correlated with edge density (Pearson's r = 0.82, using tree cover calculated within 200 m) and largest patch index (r = 0.81). Percentage tree cover was moderately correlated with cohesion (r = 0.58), percentage core area (r = 0.50), and maximum patch size on a segment (r = 0.68). Correlations were also strong between measures of tree cover calculated at different scales: percentage tree cover within 200 m was strongly correlated with that within 400 m (r = 0.93) and within 1200 m (r = 0.72).

Independence

A potential concern in our analysis is that we did not account for possible dependence between variables on adjacent segments of transects. A transect segment will be intrinsically more similar to an adjacent segment than to one some distance away. Moreover, adjacent segments have nearly identical surrounding landscapes, so they are non-independent in that manner, as well. The implication of this for analytical purposes is primarily that Independence allows one to compute the probability of a series of events as the product of the probabilities of the individual events. Reliably computing probability of significant results is especially important in computing probabilities under specified hypotheses.

Where hypothesis testing is not the aim, independence is not always a requisite. Consider an example in which one wishes to estimate the average height of male students in high school classes. Suppose one of the classes includes a set of identical twins. Clearly their heights are not independent. One could eliminate that non-independence by (randomly) choosing one of the two students and excluding his height from the calculation. However, if heights of identical twins differ from non-twins, then elimination of one of the twins results in a biased estimate of average height. Hence the need to ascertain whether or not independence of observations is necessary or even desirable in some applications.

A commonly used approach to deal with non-independence is to use only a fraction of the data set, say every fifth segment in our case, so that segments can more realistically be considered independent. Suppose we did that, using only segments 1, 6, 11, 16, etc. We could estimate the curves and other outputs we show based on this fraction of independent data. Then we could repeat the process, next using segments 2, 7, 12, 17, etc. Ultimately we would obtain five different curves, each of which is based on a set of (more-or-less) independent observations.

We used this approach and compared incidence plots (LOESS curves) from five subsamples of our data to the entire data set. Subsamples were extracted by taking every fifth transect segment, as noted above. Thus subsamples 1 includes segments 1, 6, 11, 16, etc.; subsamples 2 includes segments 2, 7, 12, 17, and so on. The results are plotted below for the five subsamples (in color), each of one-fifth of the data, as well as the LOESS curve based on the entire data set (in black: Fig. A1.2). Which of the five curves should be used? Each has equal credibility. Alternatively, we could somehow average the curves, to obtain a single curve that reflects all of the observations. But this is fundamentally the same as using all of the data initially, which is what we were trying to avoid.

We repeated this process for 16 species with at least 20 observations in each subset (Fig. A1.3). Three conclusions are evident from these plots: 1) the curve based on the entire data set is in fact representative of the overall pattern manifested by the five individual curves; 2) the curve based on the entire data set is, as would be expected, smoother than curves based on partial data sets; and, most importantly, 3) a single curve based on partial data (note curve 5 in the American robin example) may not be representative of the patterns shown by the majority of the curves. To the latter point, curve 5 suggests that the occurrence of American robins peaks at about 10 percent tree cover and is indifferent as tree cover ranges from 25 percent to 70 percent. All other curves show an increasing likelihood of occurrence with increasing tree cover. Using only a fraction of the data would definitely be wasteful of information, resulting in unjustified jaggedness, and could well be misleading, depending on how representative the selection fraction of the data are. It would also be unnecessary, because independence of the observations is not a requirement for such summaries of data.

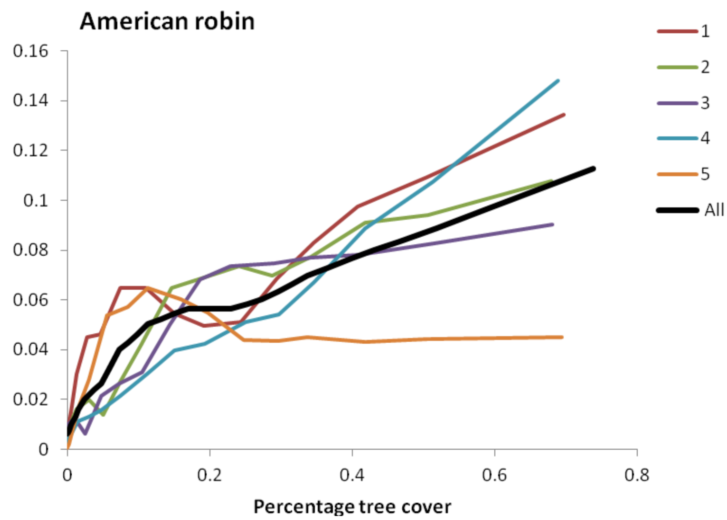

Figure A1.2 Incidence plots calculated using subsamples of the data and all data for American robin.

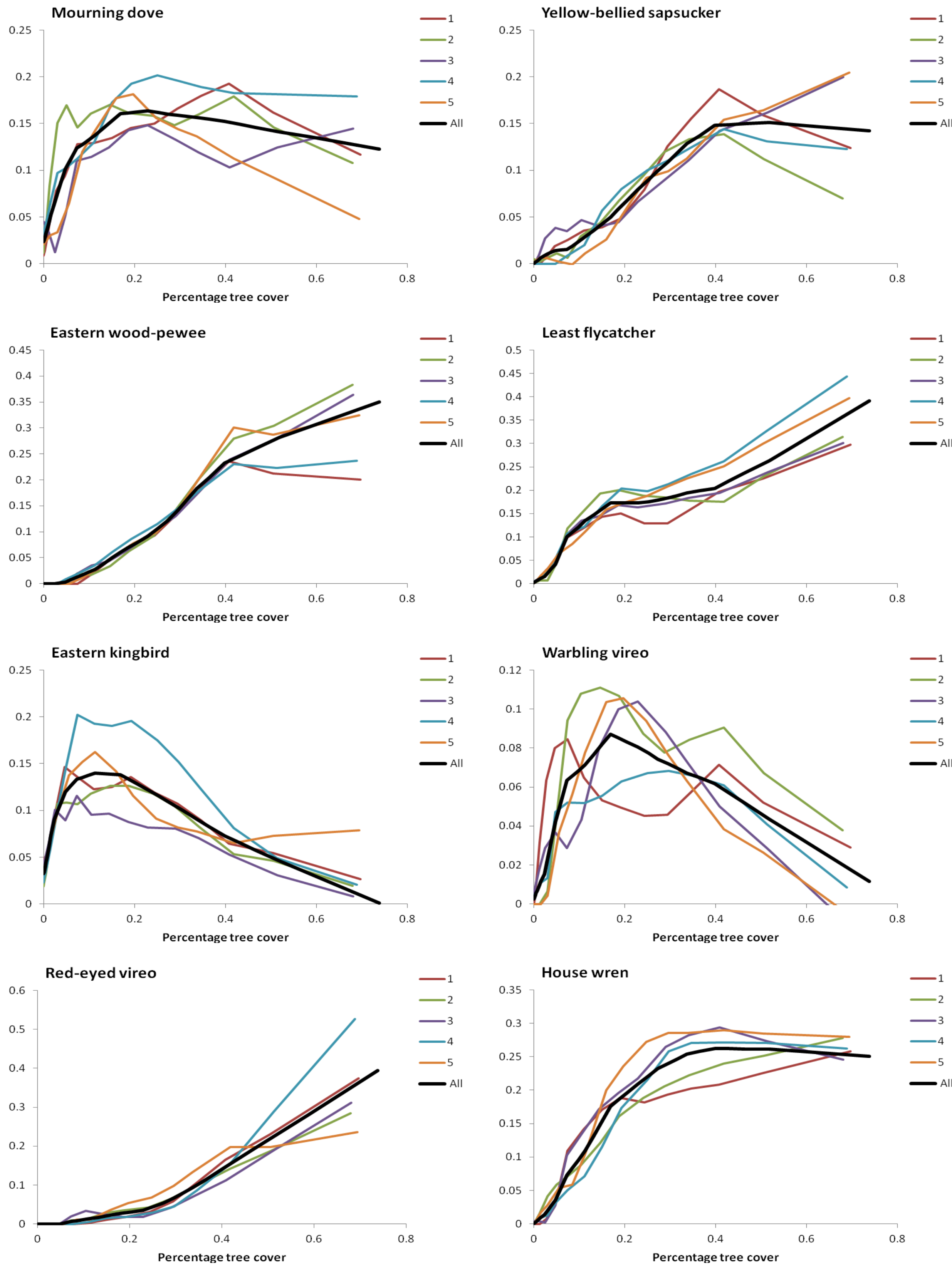

Figure A1.3a Incidence plots calculated using subsamples of the data and all data for 8 of 16 species.

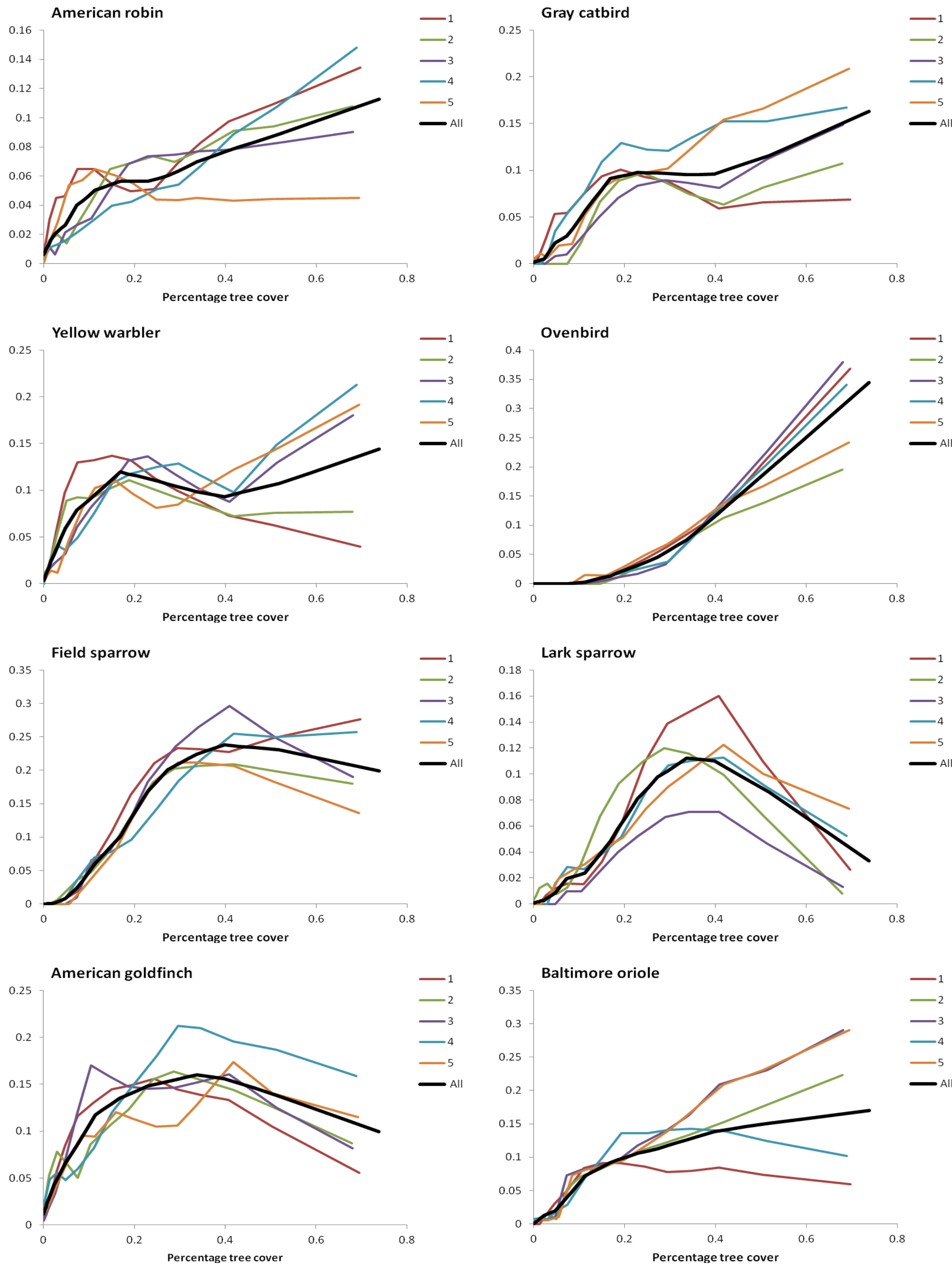

Figure A1.3b Incidence plots calculated using subsamples of the data and all data for 8 of 16 species.

Year effects

Because the abundance and occupancy rates of birds can vary dramatically among years, it is important to consider that variation when estimating preferences for habitat types. As an extreme example, suppose that in one year tree-favoring scarlet tanagers were absent, and if that was the only year that many heavily wooded transects were surveyed, one would obtain mostly zero occupancy values in most of the heavily wooded segments and conclude that the species avoids trees. Numerically, suppose scarlet tanagers were observed on 5 of 500 segments (overall occurrence rate = 0.01) in one year. Next suppose the species was much more common and widely distributed the following year, occurring on 20 of 400 segments (overall occurrence rate = 0.04). We likely would have less-favorable segments occupied than in the previous year, so any preference or selection for certain segments would be less evident.

We can reduce that potentially biasing effect by dividing the presence or absence value (1 or 0) for each segment by the overall occurrence rate in that year. Hence an occurrence value in the first year would be divided by 0.01, producing values of 100 or 0 for each segment. In the second year occurrence rates would be divided by 0.04, yielding values of 25 or 0. This adjustment scales upward occurrence values in the first year, when the species was less ubiquitous.

Mathematically, suppose the frequency of occurrence of a particular species in year *t* on segment *j* is f_{jt} (= 0 or 1). Then the overall occurrence rate of that species in year *t* is the number of segments on which the species was recorded, divided by the number of segments surveyed in year *t*: $f_{\cdot t} = \sum_i f_{jt} / N_t$. Then scaled occupancy values $f'_{jt} = f_{jt} / f_{\cdot t}$ will account for annual variation in occupancy when used to develop incidence plots.

We compared incidence plots developed from both standard and scaled occupancy values (f_{jt} and f'_{jt}, respectively) for species with relatively even distributions among years (Fig. A1.4) and for species with uneven distributions among years (Fig. A1.5). Generally the profiles were very similar, regardless of whether standard or scaled occupancy values were used. Thus, for simplicity we present results based on the more familiar 1/0 occupancy values.

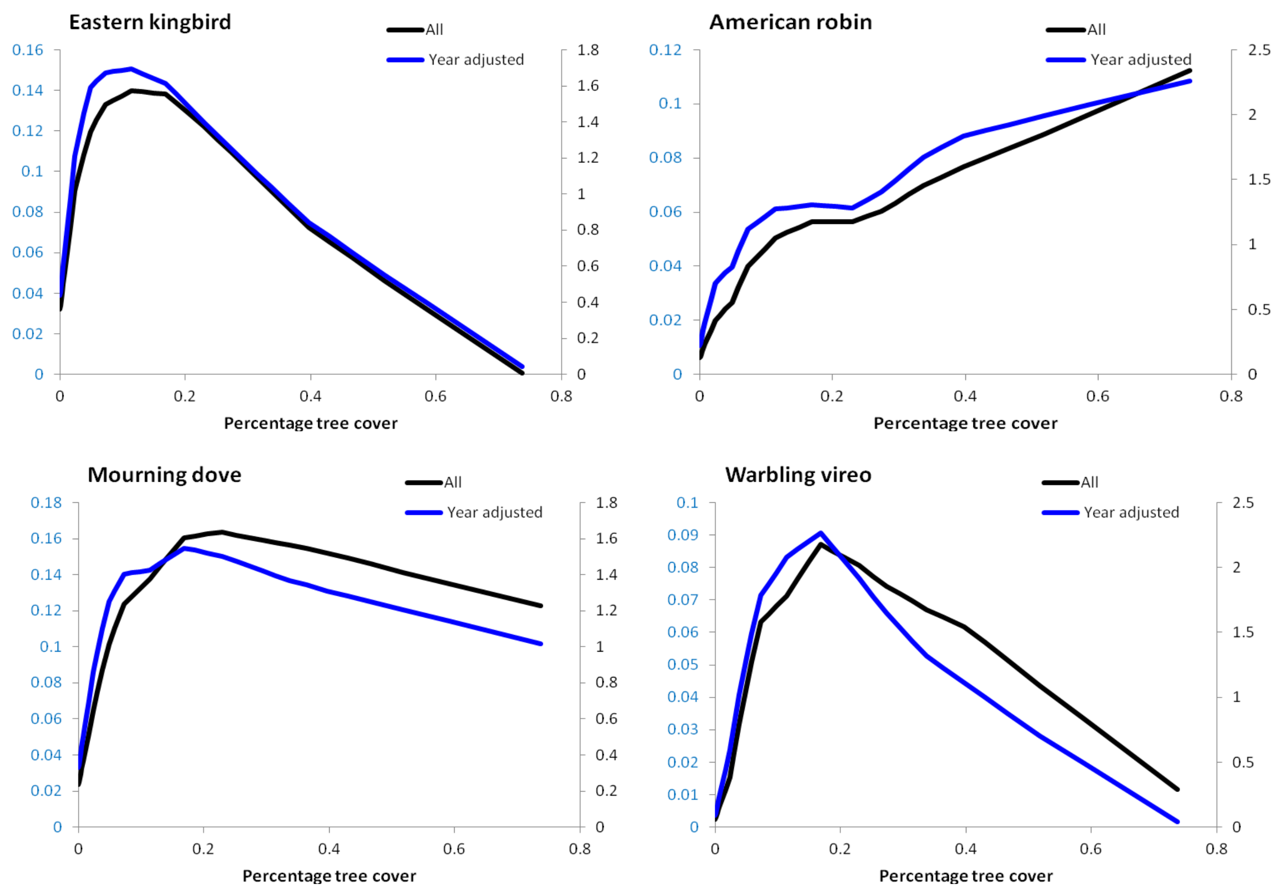

Figure A1.4 Standard (1/0) and year-adjusted results for four species *evenly* distributed among years

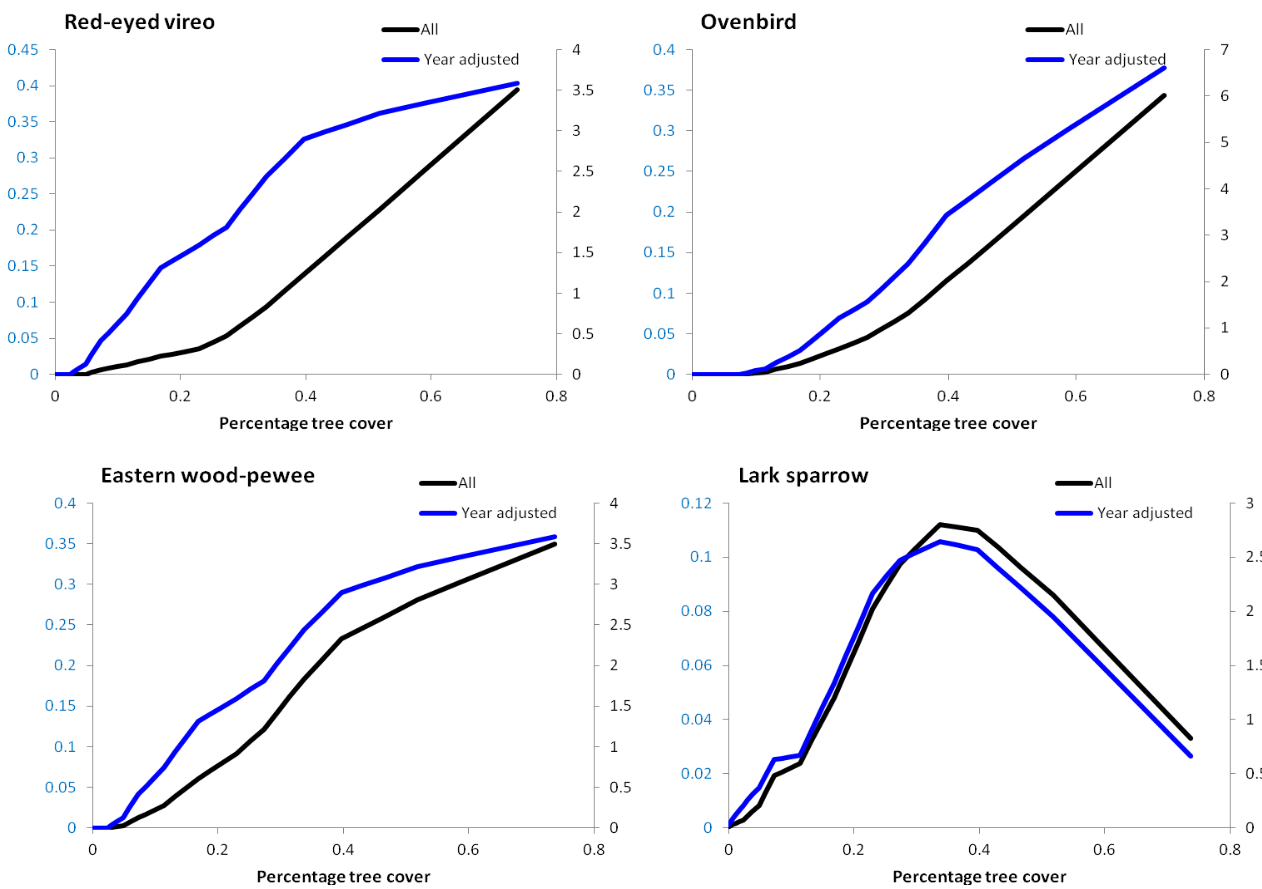

Figure A1.5 Standard (1/0) and year-adjusted results for four species *unevenly* distributed among years.

References

Cunningham M.A. and D. H. Johnson. 2011. Seeking parsimony in landscape metrics. Journal of Wildlife Management 75: 692-701.

Wu J., W. Shen, W. Sun W, and P. T. Tueller. 2002. Empirical patterns of the effects of changing scale on landscape metrics. Landscape Ecology 17: 761–782.