Using Indicator Species to Predict Species Richness of Multiple Taxonomic Groups

ERICA FLEISHMAN,*§ JAMES R. THOMSON,† RALPH MAC NALLY,† DENNIS D. MURPHY,‡ AND JOHN P. FAY*

*Center for Conservation Biology, Department of Biological Sciences, Stanford University, Stanford, CA 94305-5020, U.S.A.
†Australian Centre for Biodiversity: Analysis, Policy and Management, School of Biological Sciences, Monash University 3800, Victoria, Australia
‡Department of Biology/314, University of Nevada, Reno, NV 89557, U.S.A.

Abstract: *Values of species richness are used widely to establish conservation and management priorities. Because inventory data, money, and time are limited, use of surrogates such as "indicator" species to estimate species richness has become common. Identifying sets of indicator species that might reliably predict species richness, especially across taxonomic groups, remains a considerable challenge. We used genetic algorithms and a Bayesian approach to explain individual and combined species richness of two taxonomic groups as a function of occurrence patterns of indicator species drawn from either both groups or one group. Genetic algorithms iteratively screen large numbers of potential models and predictor variables in a process that emulates natural selection. The best-fitting models of bird species richness and butterfly species richness explained approximately 80% of deviances and included only indicator species from the same taxonomic group. Using species from both taxonomic groups as potential predictors did not improve model fit but slightly improved the parsimony (fewer predictors) of the model of bird species richness. The best model of combined species richness included five butterflies and one bird and explained 83% of deviance, whereas a model of combined species richness based on six butterflies as indicators explained 82% of deviance. A model of combined species richness based on birds alone explained 72% of deviance. We found that a small, common set of species could be used to predict separately the species richness of multiple taxonomic groups. We built models explaining approximately 70% of the deviance in species richness of birds and butterflies based on a common set of three bird species and three butterfly species. We also identified a set of six species of butterflies that predicted ≥66% of both bird species richness and butterfly species richness. Our approach is applicable to any assemblage or ecosystem, and may be useful both for estimating species richness and for gaining insight into mechanisms that influence diversity patterns.*

Key Words: Bayes information criterion, bird species richness, butterfly species richness, genetic algorithm, Great Basin, land cover, land use, species occurrence

Utilización de Especies Indicadoras para Predecir la Riqueza de Especies de Múltiples Grupos Taxonómicos

Resumen: *Los valores de riqueza de especies son ampliamente utilizados para definir prioridades de conservación y manejo. Debido a que los datos de inventarios, el dinero y el tiempo son limitados, se ha vuelto común el uso de sustitutos, como las especies "indicadoras," para estimar la riqueza de especies. La identificación de conjuntos de especies indicadoras que pronostiquen la riqueza de especies confiablemente, especialmente en varios grupos taxonómicos, es un reto importante. Utilizamos algoritmos genéticos y un método Bayesiano para explicar las riquezas de especies individuales y combinadas de dos grupos taxonómico como una función de patrones de ocurrencia de especies indicadoras extraídas de ambos grupos o de uno. Los algoritmos genéticos reiterativamente filtran grandes números de modelos potenciales y variables predictoras en un proceso que*

§*email efleish@stanford.edu*
Paper submitted April 8, 2004; revised manuscript accepted August 13, 2004.

Conservation Biology 1125–1137
©2005 Society for Conservation Biology
DOI: 10.1111/j.1523-1739.2005.00168.x

emula a la selección natural. Los modelos que mejor se ajustaron a la riqueza de especies de aves y de mariposas explicaron aproximadamente 80% de las anormalidades e incluyeron solo a especies indicadoras del mismo grupo taxonómico. Utilizando a especies de ambos grupos taxonómicos como predictores potenciales no mejoró el ajuste del modelo pero mejoró ligeramente la parsimonia (menos predictores) del modelo de riqueza de especies de aves. El mejor modelo de la riqueza de especies combinada incluyó a cinco especies de mariposas y una de ave y explico 83% de la anormalidad, mientras que un modelo de riqueza de especies combinadas basada en seis especies de mariposas explico 82% de la anormalidad. Un modelo de riqueza de especies combinadas basado solo en aves explicó 72% de la anormalidad. Encontramos que un conjunto pequeño, común, podría ser utilizado para pronosticar, por separado, la riqueza de especies de múltiples grupos taxonómicos. Construimos modelos que explicaron aproximadamente 70% de la anormalidad en la riqueza de especies de aves y mariposas con base en un conjunto común de tres especies de aves y tres de mariposas. También identificamos un conjunto de seis especies de mariposas que predijeron ≥ 66% de la riqueza de especies tanto de aves como de mariposas. Nuestro método es aplicable a cualquier ensamble o ecosistema, y puede ser útil tanto para estimar la riqueza de especies como para incrementar el entendimiento de los mecanismos que influyen sobre los patrones de diversidad.

Palabras Clave: algoritmo genético, cobertura de terreno, criterio Bayesiano de información, Great Basin, ocurrencia de especies, riqueza de especies de aves, riqueza de especies de mariposas, uso de suelos

Introduction

Protection of locations with a relatively large number of native species is thought to be an efficient way to maintain overall biodiversity (Scott et al. 1987; Myers et al. 2000). Data on species richness have supported a range of high-profile conservation efforts, particularly when information on endemism and changes in land use and land cover has also been available (Mittermeier et al. 1998; Hobohm 2003; Kareiva & Marvier 2003). Whether protection of locations containing large numbers of species has been effective in achieving broad conservation goals remains an open question (Christensen 2003), but extensive use of species richness as a criterion for development of conservation and management strategies almost certainly will continue.

Given that inventory data for many regions are sparse and acquisition of new data is costly, surrogate-based approaches to estimate species richness from data on land cover, land use, climate, and topography have become common (e.g., Mayer & Laudenslayer 1988; Boyce & McDonald 1999; Scott et al. 2002). Work also has been done to build functional relationships between species richness and the occurrence of "indicator" species (Pearson 1994; Scott 1998; Gustafsson 2000), which we define as a small set of species with presence or absence patterns that are correlated functionally with species richness of a larger group of organisms. If validation data sets demonstrate that the functions are effective, then—depending on the target system—it may be far more feasible to measure occurrence of indicator species than to conduct comprehensive species inventories (Gustafsson 2000).

A central challenge in predicting species richness as a function of indicator species is finding good sets of indicators. Selection of potential indicators through statistically based methods is most likely to be effective, especially when compared with selection based on ad hoc criteria such as a species' charisma or its legal protection status (Landres et al. 1998; Andelman & Fagan 2000; Mac Nally et al. 2000). An additional question is whether species from one taxonomic group (e.g., butterflies) might serve as indicators of the species richness of other taxonomic groups (e.g., herptiles or birds). Life-history characteristics that affect how species perceive and react to environments (Kotliar & Wiens 1990; Mac Nally 2005) are likely to be more different among than within major taxonomic groups. Therefore, it seems reasonable to assume that species richness of a given taxonomic group will be predicted more accurately on the basis of species drawn from that same taxonomic group than from a different taxonomic group. But conservation biologists have not quantified the magnitude of difference in predictive ability between indicators drawn from the same taxonomic group against indicators drawn from a different taxonomic group. Nor have they shown whether the fit of models of species richness for one taxonomic group can be improved by including indicator species from a second taxonomic group.

Accordingly, we addressed six related objectives concerning within-taxon and among-taxon models of species richness based on indicator species. First, we examined whether models of species richness for one taxonomic group could be improved by including indicator species from a second taxonomic group. Second, we quantified the fit of models of species richness for one taxonomic group based on indicator species drawn exclusively from a different taxonomic group. Third, we evaluated whether models of the combined species richness of two taxonomic groups could be built based on indicator species drawn either from both of those groups or exclusively from one of the groups. Fourth, we assessed whether the same set of indicator species could be used

to predict species richness of two different taxonomic groups. Fifth, we explored whether taxonomic groups differ consistently in their ability to serve as indicators of species richness. Last, we considered the ecological characteristics of selected indicator species to gain insight into why the selected species might convey information about species richness.

We used the general statistical methodology developed by Mac Nally and Fleishman (2002, 2004) to identify indicators of species richness. This approach proved highly successful in a recent case study, in which a model based on the occurrence of five butterfly species explained 88% of the deviance of species richness of 56 species of butterflies in a mountain range in western North America (Mac Nally & Fleishman 2002). Validation tests showed that predicted and observed values of species richness were highly correlated, with 93% of the observed values falling within the 95% Bayesian-credible intervals of the predictions (Mac Nally & Fleishman 2004). Here, we extend this approach by employing a genetic algorithm to select indicator species from a set of potential indicators (139 total) that is too large to be evaluated using exhaustive search methods. Genetic algorithms are heuristic, iterative search procedures that emulate the process of natural selection. They are useful for large combinatory problems and have been used successfully to select input variables in neural networks (Jefferson et al. 1997) and logistic regression models (Vinterbo & Ohno-Machado 1999). The software we used also allowed for modeling based on Poisson errors, which is most appropriate for the numerical distribution of species richness (i.e., non-negative integers).

Methods

Field Methods

Data for our analyses were collected from 1996–2003 in three adjacent mountain ranges in the central Great Basin (Lander and Nye counties, Nevada, U.S.A.) that have similar biogeographic and human land-use histories, the Shoshone Mountains (1850 km^2, approximate north–south boundaries 39°14′19″ to 38°57′32″); the Toiyabe Range (3100 km^2, 39°54′00″ to 38°30′0″); and the Toquima Range (1750 km^2, 39°17′50″ to 38°29′9″). Our data collection incorporated established techniques that detect species presence reliably and permit assessment of distributional trends (Pollard & Yates 1993; Pullin 1995; Bibby et al. 2000; Buckland et al. 2001; Siegel et al. 2001). These methods have been described in detail and tested for sampling adequacy (e.g., Fleishman et al. 1998, 2000, 2001; Mac Nally et al. 2004). Inventories for birds were conducted in 84 sites, and inventories for butterflies were conducted in 195 sites. Site areas ranged from 1.5 ha to 44.4 ha. The area of our study sites explained little variance in species richness (<3%) of either taxonomic group, perhaps because area often is not correlated with heterogeneity of topography or land cover in our study system, at least for locations smaller than mountain ranges (Mac Nally et al. 2003).

Model Building

We recorded 74 species of birds and 65 species of butterflies from our study sites (a complete list is available from E.F.). Site-level species richness ranged from 5 to 34 for birds (mean 17.3) and from 3 to 51 for butterflies (mean 24.9). We modeled the site-level species richness of birds (henceforth, bird species richness), butterflies (butterfly species richness), and both groups (combined species richness) as functions of the presence or absence of small subsets of "indicator" species. For each response variable (bird species richness, butterfly species richness, and combined species richness), we sought three sets of indicator species: birds only, butterflies only, or a combination of birds and butterflies. We limited the number of indicator species for each model to six, which is <10% of the number of species for each taxonomic group. We used Poisson regression to model species richness because Poisson error structures are appropriate for non-negative "counts" data such as number of species (McCullagh & Nelder 1989).

SELECTION OF INDICATOR SPECIES

With large sets of predictor variables, many of the screening approaches used to identify the "best" subset of variables are statistically suspect (Mac Nally 2000). Information criteria, such as Bayes information criterion (BIC; Schwarz 1978), have been recommended to identify the most efficient model(s) (Mac Nally 2000). The BIC enforces a compromise between model fit (ability to explain observed deviance in the response variable) and model complexity (number of predictor variables). The BIC is calculated as $-2 \times \log \text{likelihood} + \log(N) \times Q$, where N is the number of observations and Q is the number of model parameters. Ideally, all possible models are fitted, and the model with the lowest BIC value is retained.

Because we had such a large number of potential predictor variables (potential indicator species), it was impractical to fit all possible models, even if we considered only models with ≤ 6 predictors. We could have reduced the number of possible models by eliminating some species a priori on the basis of their distributions or other attributes (e.g., widespread, rare, or cryptic species; Mac Nally & Fleishman 2002). Only a severe pruning, however, which risks elimination of useful indicators, would have enabled us to fit all possible Poisson models. Another possibility would have been to conduct a less-severe pruning followed by an exhaustive search of ordinary multiple linear regression models, which can be

computed more rapidly, but this process would not necessarily yield the best Poisson model per se. Instead, we employed a genetic algorithm to assess candidate Poisson models.

Genetic algorithms iteratively refine a "population" of potential solutions, usually generated randomly, by a process analogous to natural selection. The value of the function to be optimized—in our case, BIC for each model—is used as a measure of individual "fitness," with fitter individuals (models) given a greater probability of "reproducing." Genetic operations such as crossover and mutation are emulated and applied to a proportion of individuals in each generation. Therefore, characteristics of individuals with high fitness are retained and recombined until very good solutions are found.

We implemented a genetic algorithm by using the *R* statistical programming software (R Development Core Team 2003; Sekhon & Mebane 1998) to search for Poisson models with low BIC values for each of the nine combinations of response and predictor variables. First, an initial population of 1000 random models (individual "genomes") is generated. Each model in the population is encoded by a binary string of Q bits, where Q is the number of candidate predictors. Each bit corresponds to a potential predictor, which may be included in (1) or excluded from (0) the model. In the genetic analogy, the strings and bits are equivalent to chromosomes and individual genetic loci, respectively. In each generation, all models in the population are fitted as Poisson regression models with generalized linear modeling functions, and BIC is calculated for each model. Each model then is assigned a fitness rank according to its BIC value (fitness increases as BIC decreases). The model in each generation with the lowest BIC is propagated into the next generation. All other models are replaced with the results of genetic operators applied to parents selected, with replacement, from the entire population. The probability that a model will be selected as a parent decreases as its fitness decreases.

To produce offspring, five genetic operators are applied to selected parents in predetermined proportions: cloning (0.125), single-point mutation (0.375), multiple-point crossover (0.250), polytope crossover (0.125), and whole mutation (0.125). A clone is an exact replica of the parent model. Single-point mutation assigns a new value to one randomly selected bit (i.e., drops or adds one indicator species). Multiple-point crossover produces two offspring models from two parent models by exchanging randomly selected segments of strings between parents. Polytope crossover creates one offspring model that is a combination of as many parent models as there are bits (species). Whole mutation assigns a new, random value to all bits (i.e., randomly generates a new genotype). Because of the applied selective pressure, species (or "genes") present in models with low BIC (individuals with high fitness) are propagated throughout the population and, through crossover and mutations, are combined with new species until combinations with very high fitness are found. After 25 generations with no improvement in fitness (i.e., no reduction in the minimum BIC), the genetic algorithm terminates and the model with the lowest BIC (the solution) is reported.

Genetic algorithms can be trapped in local optima and, as with all heuristic methods, there is no guarantee that a particular solution is the best possible solution (global optimum). The concept of an absolute best regression model is vague in any case, and the model with the lowest BIC may not necessarily be the most useful model. Accordingly, we generated solutions from five random starts for each combination of response and predictor variables and then assessed the usefulness of each unique solution (from 1–5 solutions per combination, mean = 3.5) on the basis of model fit and attributes of indicator species (see Final Model Selection section).

In addition to searching for separate sets of indicator species for each response variable (bird species richness, butterfly species richness, and combined species richness), we searched for sets of "common" indicator species. To do this, we used the genetic algorithm to seek sets of ≤ 6 indicator species (from one or both taxonomic groups) that could be used to predict both bird species richness and butterfly species richness. We searched for three sets of common indicator species: birds only, butterflies only, and any combination of birds and butterflies. We modified the genetic algorithm to fit two models, one each with butterfly species richness and bird species richness as the response variable, for each individual in the population (i.e., for each unique set of indicator species). The BIC values were calculated for both models and summed for each individual. Fitness ranks then were assigned on the basis of summed BIC values.

MODEL FITTING

We computed parameters for each model selected by the genetic algorithm runs with the WinBUGS programming framework (Spiegelhalter et al. 2003). The model used was

$$\log(\mu_i) = \alpha_0 + \sum_{k=1}^{Q} \alpha_k X_{ik},$$

$$Y_i \sim \text{Poisson}\ (\mu_i),$$

where the αs are the regression coefficients to be estimated (an intercept [α_0] and coefficients for each of the Q indicator species); μ_i is the estimated mean species richness at site i given the presence or absence (X_i) of each of the Q indicator species; and Y_i is the observed species richness at site i, which is assumed to be distributed ($\sim$) as a Poisson variable with mean μ_i.

WinBUGS uses Markov chain Monte Carlo methods to iteratively estimate model parameters by repeatedly "sampling" from distributions specified in the model (Gilks

et al. 1996). The Bayesian approach to estimating parameters formally incorporates prior knowledge about the value of a parameter and produces a posterior (or final) distribution of estimated values for each parameter. If one has virtually no prior knowledge about parameter values, it is appropriate to use a "noninformative" prior distribution for that parameter—typically a normal distribution with high variance—which means that the posterior probability distributions are dictated by the newly collected data (Lee 1989). We initially gave the αs noninformative normal priors [$\alpha \sim \text{normal}(\mu = 0, \sigma^2 = 1000)$]. Probability distributions for parameters were allowed to settle ("burn in") over 1000 iterations before posterior parameter distributions were built for another 3000 sampling iterations. Initial values were randomly generated.

We used the proportion of the posterior probability distributions of the estimated parameter values lying above or below zero (the posterior probability mass, PPM) to evaluate whether each indicator species contributed statistically to the prediction of species richness. We retained species if PPM $\geq$90%—that is, if at least 90% of the posterior probability distribution was either >0 (positive effect) or <0 (negative effect). Therefore, if PPM <90%, the corresponding species was omitted and the model was refitted. For a general overview of Bayesian statistical methods and associated terminology, see Bergerud and Reed (1998) and references therein.

FINAL MODEL SELECTION

Running the genetic algorithm and fitting models in WinBUGS generated from one to five candidate models for each response-predictor combination. Although the model with the lowest BIC is the best model found statistically (at least for the models searched), attributes of the indicator species also affect the usefulness of each model for predicting species richness. In particular, species that are difficult to detect or that have erratic patterns of occurrence among years may be unreliable indicator species even if their occupancy in at least 1 year is a good predictor of species richness. Thus, false negatives, or species recorded as absent from sites they do occupy in $\geq$1 year, decrease the accuracy of predictions of species richness. The probability of false negatives may be high for species that occupy sites infrequently even if surveys are repeated over several years and observer error is minimal (Mac Nally & Fleishman 2004). Therefore, we eliminated models containing indicator species with erratic occupancy patterns.

We used data from all sites that were surveyed for $\geq$2 years to calculate $p(- \mid \text{occupied})$, the probability of obtaining a false negative in a single year. This is the conditional probability that a species will be absent in any year from a site, given that it is known to occupy the site at some point in time. The probability of obtaining a false negative for surveys repeated over n years is $p(- \mid \text{occupied})^n$. To estimate this probability for each species, we used a Bayesian calculation to model the number of absences Y_i from occupied sites in n_i years as a Bernoulli trial with probability p_i, as follows:

$$\text{logit}(p_i) = \alpha,$$
$$Y_i \sim \text{binomial}(p_i, n_i),$$
$$p(-|\text{occupied}) = \frac{e^\alpha}{1 + e^\alpha},$$

where α is the log-odds ratio of $p(- \mid \text{occupied})$. We eliminated all models in which the maximum or mean values of $p(- \mid \text{occupied})$ for indicator species were >0.66 or >0.50, respectively. These values corresponded approximately to the tenth and fiftieth percentiles for $p(- \mid \text{occupied})$ for both birds and butterflies. We then selected the lowest BIC model for each response-predictor combination from the remaining models.

MODEL PREDICTIONS

We investigated the uncertainty of predictions of species richness made using each of the final models by applying those models to all possible combinations of indicator species presence and absence (i.e., to the 2^Q possible results of an inventory of Q indicator species at a site). We used WinBUGS to produce a posterior distribution of predicted richness values for each combination of indicators from which 95% credible intervals were calculated. Bayesian calculations produce probability distributions for model parameters, such as regression coefficients. One often uses "95% Bayesian credible intervals" (Lee 1989) to characterize the range of values for the parameter that encompasses 95% of the probability mass for that parameter. We calculated 95% credible intervals for a hypothetical, multiyear inventory in which the true occurrence of all indicator species is known (no false negatives).

Results

The genetic algorithm was an efficient method for identifying indicator species. All species selected as indicators contributed substantially (i.e., PPM > 90%) to predictions of species richness, demonstrating that the genetic algorithm identified indicator species with high explanatory capacity. Indicators in models for one taxonomic group always contributed to prediction of species richness of that group, and species selected as indicators in models for two taxonomic groups always contributed to the prediction of species richness for at least one of those groups.

We found promising models for all combinations of response variables (bird species richness, butterfly species richness, and combined species richness) and predictor variables (birds, butterflies, and both; Fig. 1) with selected models explaining 55–83% of deviances (mean = 74%; Table 1). After evaluating the false-negative probabilities of selected indicator species, we retained the lowest

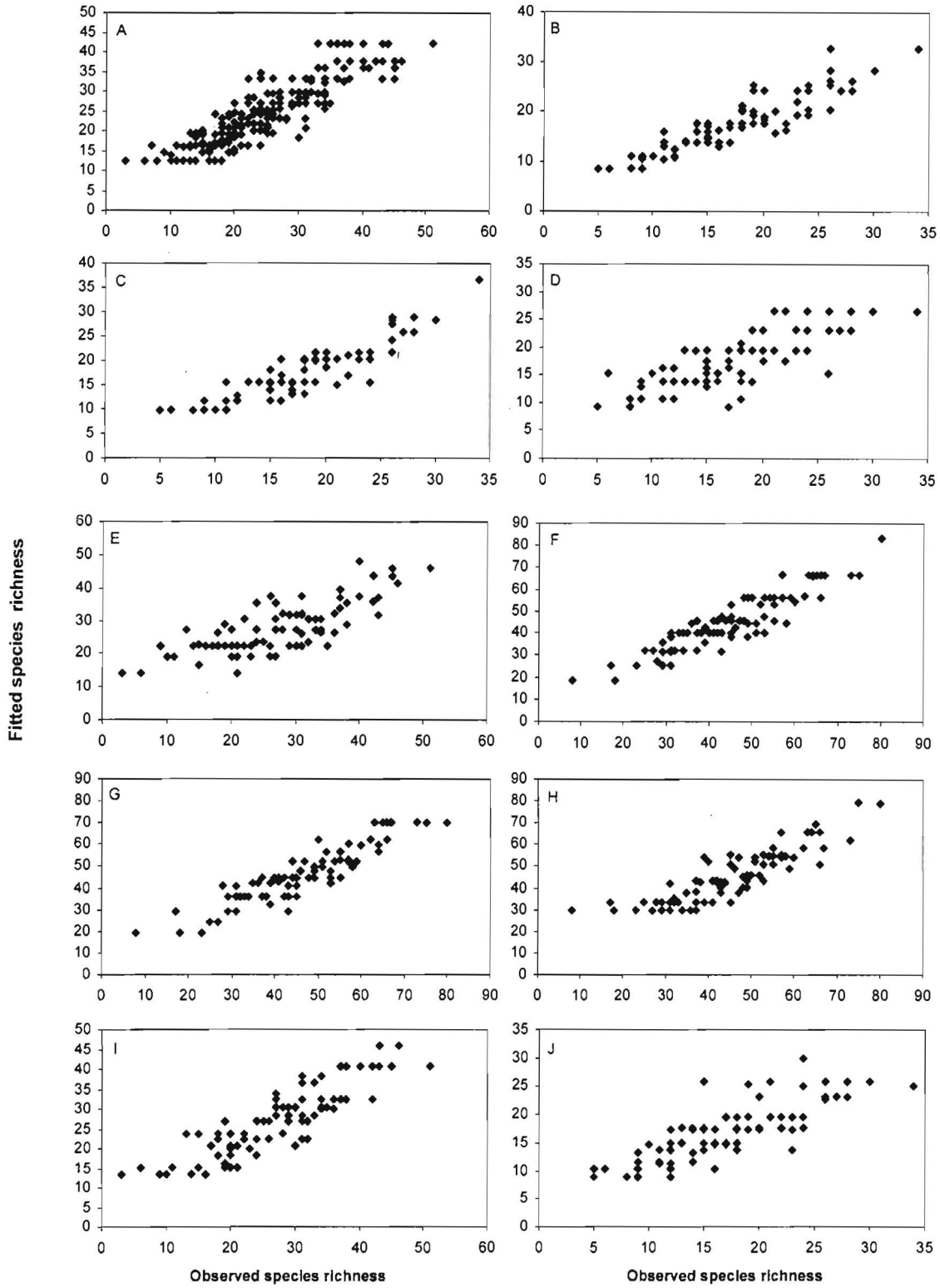

Figure 1. Fitted versus observed species richness for each combination of response variable and predictor variable: (a) butterfly species richness with butterfly indicators, (b) bird species richness with bird indicators, (c) bird species richness with combined indicators, (d) bird species richness with butterfly indicators, (e) butterfly species richness with bird indicators, (f) combined species richness with combined indicators, (g) combined species richness with butterfly indicators, (h) combined species richness with bird indicators, (i) butterfly species richness with combined "common" indicators, (j) bird species richness with combined common indicators.

Table 1. Parameter values (mean and standard deviation) for Poisson regression models of species richness of birds (bird S), butterflies (butterfly S), and both taxonomic groups (combined S) based on occurrence patterns of indicator species.

Variable (indicator species)	*Mean*	*Factor*[a]	*SD*	p(*occupied*)[b]	p(− | *occupied*)[c]
Butterfly S[d] with butterfly indicators					
(78% deviance explained, BIC[e] = 534)					
constant	2.519		0.051		
Icaricia lupini	0.266	1.30	0.037	0.69	0.44
Ochlodes sylvanoides	0.237	1.27	0.042	0.66	0.56
Papilio multicaudatus	0.203	1.23	0.035	0.51	0.47
Euchloe ausonides	0.162	1.18	0.031	0.50	0.49
Chlosyne acastus	0.113	1.12	0.031	0.49	0.27
Phyciodes mylitta	0.239	1.27	0.035	0.22	0.54
Bird S with bird indicators					
(82% deviance explained, BIC = 453)					
constant	2.150		0.072		
Black-throated Gray Warbler *(Dendroica nigrescens)*	0.222	1.25	0.067	0.68	0.23
MacGillivray's Warbler *(Oporornis tolmiei)*	0.234	1.26	0.067	0.58	0.17
Western Scrub-Jay *(Aphelocoma californica)*	0.251	1.29	0.062	0.54	0.49
Red-shafted Flicker *(Colaptes auratus)*	0.186	1.20	0.066	0.45	0.45
Western Tanager *(Piranga ludoviciana)*	0.140	1.15	0.060	0.38	0.47
Fox Sparrow *(Passerella iliaca)*	0.297	1.35	0.077	0.27	0.61
Bird S with combined indicators[f]					
(81% deviance explained, BIC = 451)					
constant	2.276		0.068		
Western Scrub-Jay *(Aphelocoma californica)*	0.289	1.33	0.059	0.54	0.49
Audubon's Warbler *(Dendroica coronata)*	0.263	1.30	0.058	0.38	0.47
Fox Sparrow *(Passerella iliaca)*	0.350	1.42	0.065	0.27	0.61
Incisalia eryphon	0.185	1.20	0.071	0.64	0.31
Satyrium sylvinum	0.240	1.27	0.070	0.13	0.44
Bird S with butterfly indicators					
(63% deviance explained, BIC = 479)					
constant	2.375		0.075		
Incisalia eryphon	0.417	1.52	0.068	0.64	0.31
Pontia beckerii	−0.167	0.85	0.064	0.55	0.33
Papilio multicaudatus	0.345	1.41	0.062	0.51	0.47
Speyeria nokomis	0.307	1.36	0.079	0.07	0.50
Butterfly S with bird indicators					
(55% deviance explained, BIC = 599)					
constant	2.946		0.046		
Rock Wren *(Salpinctes obsoletus)*	0.156	1.17	0.043	0.56	0.39
Western Tanager *(Piranga ludoviciana)*	0.203	1.22	0.043	0.38	0.47
Yellow Warbler *(Dendroica petechia)*	0.315	1.37	0.044	0.35	0.35
Black-headed Grosbeak *(Pheucticus ludovicianus)*	0.151	1.16	0.048	0.33	0.55
Chukar *(Alectoris chukar)*	0.259	1.30	0.063	0.13	0.41
Western Meadowlark *(Sturnella neglecta)*	−0.310	0.73	0.080	0.10	0.37
Combined S with combined indicators					
(83% deviance explained, BIC = 565)					
constant	2.902		0.097		
Yellow-breasted Chat *(Icteria virens)*	0.216	1.24	0.074	0.04	0.64
Hesperia comma	0.336	1.40	0.090	0.92	0.11
Incisalia eryphon	0.233	1.26	0.039	0.64	0.31
Papilio multicaudatus	0.346	1.41	0.035	0.51	0.47
Satyrium behrii	0.213	1.24	0.036	0.50	0.41
Polygonia zephyrus	0.172	1.19	0.036	0.32	0.65
Combined S with butterfly indicators					
(82% deviance explained, BIC = 568)					
constant	2.957		0.094		
Hesperia juba	0.410	1.51	0.095	0.87	0.18
Incisalia eryphon	0.217	1.24	0.041	0.64	0.31
Papilio multicaudatus	0.217	1.24	0.044	0.51	0.47
Nymphalis antiopa	0.156	1.17	0.042	0.50	0.50
Polygonia zephyrus	0.130	1.14	0.038	0.32	0.65
Phyciodes mylitta	0.165	1.18	0.044	0.22	0.54

continued

Table 1. (continued)

Variable (indicator species)	*Mean*	*Factor*[a]	*SD*	p(*occupied*)[b]	p(− | *occupied*)[c]
Combined S with bird indicators					
(72% deviance explained, BIC = 604)					
constant	3.393		0.036		
Rock Wren *(Salpinctes obsoletus)*	0.126	1.13	0.036	0.56	0.39
Western Tanager *(Piranga ludoviciana)*	0.239	1.27	0.033	0.38	0.47
Yellow Warbler *(Dendroica petechia)*	0.254	1.29	0.037	0.35	0.35
Black-headed Grosbeak *(Pheucticus ludovicianus)*	0.176	1.19	0.035	0.33	0.55
Song Sparrow *(Melospiza melodia)*	0.175	1.19	0.040	0.29	0.33
Red-tailed Hawk *(Buteo jamaicensis)*	0.247	1.28	0.065	0.06	0.61
Butterfly S with "common" indicators[g]					
(73% deviance explained, BIC = 540)					
constant	2.605		0.070		
Gray Flycatcher *(Empidonax wrightii)*	0.121	1.13	0.049	0.43	0.46
Warbling Vireo *(Vireo gilvus)*	0.182	1.20	0.051	0.39	0.09
Ochlodes sylvanoides	0.385	1.47	0.068	0.66	0.46
Satyrium behrii	0.302	1.35	0.047	0.50	0.41
Lycaena rubidus	0.231	1.26	0.047	0.19	0.57
Bird S with "common" indicators					
(68% deviance explained, BIC = 479)					
constant	2.173		0.085		
Mountain Chickadee *(Poecile gambeli)*	0.279	1.32	0.055	0.58	0.31
Gray Flycatcher *(Empidonax wrightii)*	0.247	1.28	0.064	0.43	0.46
Warbling Vireo *(Vireo gilvus)*	0.533	1.70	0.066	0.39	0.09
Ochlodes sylvanoides	0.161	1.17	0.075	0.66	0.46
Lycaena rubidus	0.108	1.11	0.061	0.19	0.57

[a] *A multiplicative factor (the exponentiated mean, e^{mean}) by which the presence of the indicator species affects the predicted species richness.*
[b] *Probability that a site is ever occupied by that species (i.e., the proportion of sites occupied).*
[c] *Conditional probability that a species will be absent in any year from a site given that it is known to occupy the site at some point in time.*
[d] *Species richness.*
[e] *Bayes information criterion.*
[f] *Combined indicators include both birds and butterflies.*
[g] *A single set of species that can be used to predict species richness of both birds and butterflies.*

BIC models for all but two response-predictor combinations. The lowest BIC models predicting butterfly species richness and combined species richness with butterfly indicators both included one species with a false negative probability that we deemed unacceptably high, *Colias philodice* [$p(-$ | occupied$) = 0.77$]. As a result, we rejected the latter models in favor of the models with second-lowest BIC values. In both cases, the retained model included many of the same indicator species as the rejected model (three for combined species richness and four for butterfly species richness) and had similar fit (within 2% explained deviance).

Model predictions with 95% credible intervals are shown in Fig. 2. Credible intervals were relatively narrow but widened with increasing predicted species richness because the response variables had Poisson distributions (i.e., mean equal to variance).

Species Richness of One Taxonomic Group

The best-fitting models of bird species richness and butterfly species richness included only indicator species from the same taxonomic group as the response variable. The best models of bird species richness with birds as indicators and butterfly species richness with butterflies as indicators each included six indicator species and explained 82% and 78% of deviances, respectively (Table 1). Sample sizes for models of butterfly species richness (195 sites) were much larger than those for models of bird species richness (84 sites).

Using species from both taxonomic groups as potential predictors did not improve explanatory capacity for butterfly species richness. Genetic algorithm solutions did not include bird species as predictors of butterfly species richness if butterflies were also available as predictors. Only the 84 sites surveyed for birds as well as butterflies could be included in searches for models using both taxonomic groups as predictors. As a result, the predictors and fit of the best model of butterfly species richness with combined indicators, which included only butterflies as predictors, were not identical to those of the best model of butterfly species richness with butterflies as indicators (data from 195 sites). The two models did, however, have three indicator species in common (*Icaricia lupini, Ochlodes sylvanoides,* and *Chlosyne acastus*). We present only the latter model because it includes all available data and is likely to be the most useful (Table 1).

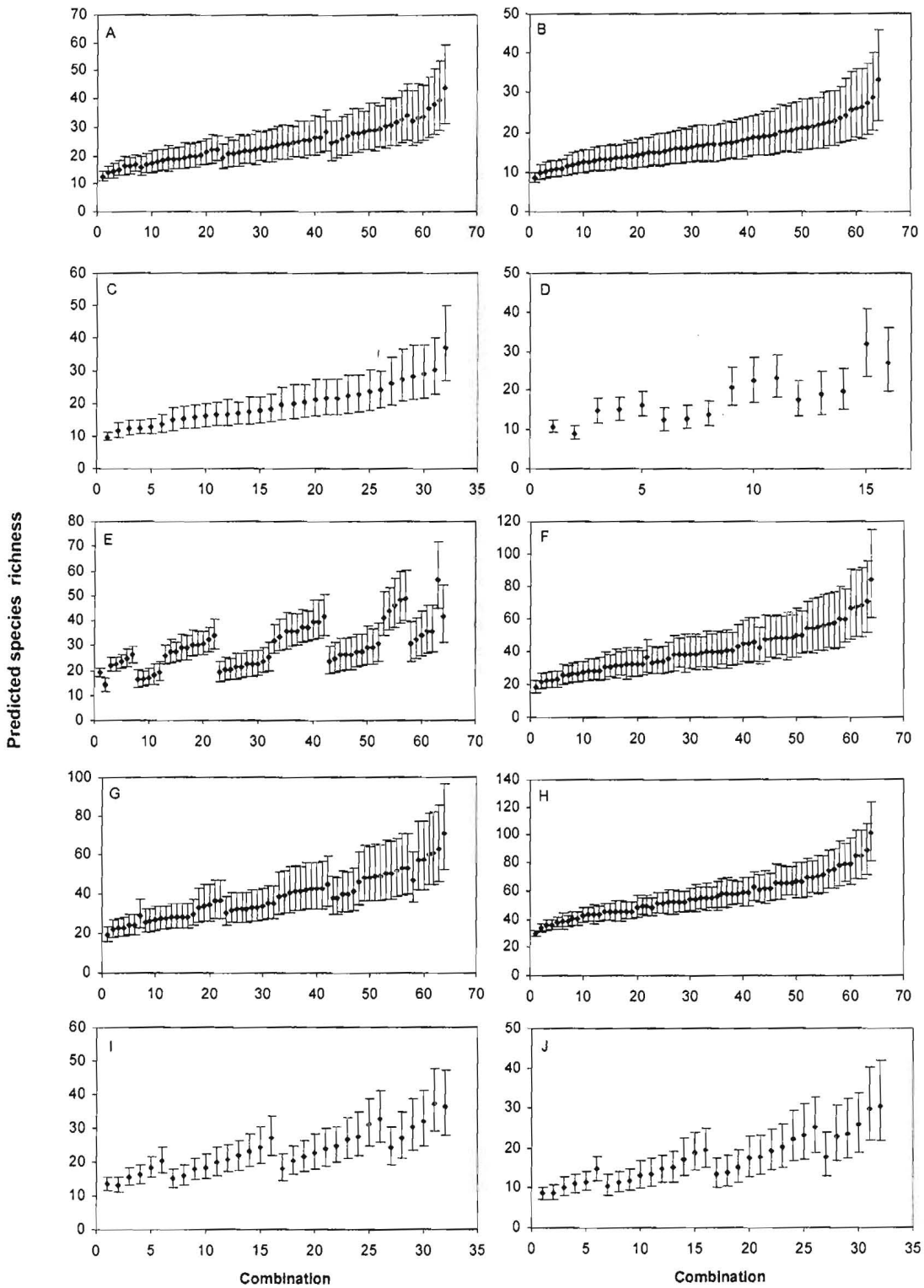

Figure 2. Predicted species richness values and 95% credible intervals versus each possible combination of indicator species presence and absence: (a) butterfly species richness with butterfly indicators, (b) bird species richness with bird indicators, (c) bird species richness with combined indicators, (d) bird species richness with butterfly indicators, (e) butterfly species richness with bird indicators, (f) combined species richness with combined indicators, (g) combined species richness with butterfly indicators, (h) combined species richness with bird indicators, (i) butterfly species richness with combined "common" indicators, (j) bird species richness with combined common indicators.

Using both taxonomic groups as predictors of bird species richness resulted in a slight improvement in model parsimony. The fit of a model including three bird species (two of which were included in the model of bird species richness with birds as indicators) and two butterfly species as predictors (81% deviance explained) was similar to a six-species model of bird species richness with birds as indicators (82% deviance explained; Table 1).

Models of species richness of one taxonomic group based on indicator species drawn exclusively from a different taxonomic group had poorer fits than within-taxon models or combined models; nonetheless, we identified potentially useful indicators. A model with just 4 butterfly indicator species explained 63% of the deviance in bird species richness (Table 1). This model was the only solution produced by five runs of the genetic algorithm, suggesting that these 4 butterfly species were substantially better predictors of bird species richness than were other butterfly species. The latter inference was supported by the results of an additional genetic algorithm run that was constrained to optimize, with ≤ 6 indicator species, model fit (percent deviance explained) rather than BIC. The best solution added 2 species to the existing model but had only marginally better fit (66% deviance explained).

The best model of butterfly species richness with birds as indicators explained 55% of deviance and included 6 indicator species (Table 1). One of these species, Western Meadowlark (*Sturnella neglecta*), had a negative coefficient, indicating a negative association with butterfly species richness. Prediction of butterfly species richness with bird indicator species could be improved by including ≥ 6 indicator species in the model. For example, models with 7 bird indicator species explained >60% of deviance in butterfly species richness.

Species Richness of Two Taxonomic Groups

Butterfly species were better indicators of combined species richness than were bird species (Table 1). The best model of combined species richness included five butterfly species and one bird species and explained only 1% more deviance (83%) than a model with six butterfly species (82%). Nonetheless, bird species alone could be used to predict combined species richness, with a six-species model explaining 72% of deviance.

Our search for a common set of species that could be used to predict both bird and butterfly species richness yielded promising results. We built models explaining 68% and 73% of species richness of birds and butterflies, respectively, with a common set of three bird species and three butterfly species. Four of these species (two birds, Gray Flycatcher [*Empidonax wrightii*] and Warbling Vireo [*Vireo gilvus*], and two butterflies, *Ochlodes sylvanoides* and *Lycaena rubidus*) were included in both models, whereas the remaining bird (Mountain Chickadee [*Poecile gambeli*]) and butterfly (*Satyrium behrii*) were retained only in models predicting richness of their own taxonomic group.

We also identified a set of six species of butterflies that might reasonably predict both bird species richness and butterfly species richness. This set included the four indicator species in the best model of bird species richness and two additional species, *Ochlodes sylvanoides* and *Lycaena heteronea*. These six species explained 70% of the deviance in butterfly species richness. The additional two species did not improve the fit of the model of bird species richness. Thus, a set of six butterfly indicator species yielded good models of both bird species richness (four indicator species, 66% of deviance explained) and butterfly species richness (six indicator species, 70% of deviance explained).

The best set of six common bird indicator species explained 66% of deviance in bird species richness and 48% of deviance in butterfly species richness. It probably would be necessary to increase the total number of bird indicator species to adequately predict both bird species richness and butterfly species richness. Three of the six bird species identified as indicators of butterfly species richness, Western Tanager (*Piranga ludoviciana*), Yellow Warbler (*Dendroica petechia*), and Black-headed Grosbeak (*Pheucticus ludovicianus*), also had capacity to predict bird species richness (a model including only the latter three species explained 54% of deviance of bird species richness). Western Tanager, Yellow Warbler, and Black-headed Grosbeak could be combined with three additional species, Western Scrub-Jay (*Aphelocoma californica*), Red-shafted Flicker (*Colaptes auratus*), and Song Sparrow (*Melospiza melodia*), to produce a model that explained 77% of deviance of bird species richness. Thus, a set of nine bird indicator species yielded a good model of bird species richness (six indicator species, 77% of deviance explained) and a reasonable model of butterfly species richness (six indicator species, 55% of deviance explained).

Discussion

Our work suggests that although indicator species probably are most effective within taxonomic groups, it is possible to identify suites of species whose occurrence patterns explain high percentages of deviance—sometimes more than 80%—in species richness of multiple taxonomic groups. Occurrence of species in one taxonomic group could be used to explain species richness of a different taxonomic group, and species richness of two taxonomic groups could be explained using species drawn from either one or both of those groups. Moreover, our work demonstrated that a common set of species can provide information on the species richness of two

taxonomic groups, implying that it may be possible to use relatively few indicator species to predict separately the species richness of multiple taxonomic groups. Our modeling framework is applicable to any ecosystem and may be a useful tool for estimating cross-taxonomic species richness at scales pertinent to conservation and land management. Whether a given threshold of explanatory or predictive accuracy (e.g., 80%) is "good enough" is inevitably context specific. Each real-world situation must be considered to determine the optimum trade-off between the ability and need to make predictions and the accuracy of those predictions.

The genetic algorithm, an iterative screening approach that emulates the process of natural selection, appeared to be an excellent method for rapidly identifying useful combinations of predictor species from astronomically large numbers of possible combinations, reducing the need to use inappropriate model error structures or to fit all possible models. Exhaustive model searches will be impossible for many systems, especially when multiple taxonomic groups are considered. Genetic algorithms can be used to identify indicator species from any system. Indeed, genetic algorithms may prove useful in any modeling task that requires identification of a suitable subset of a large number of potential predictor variables.

A model that explains considerable deviance in the data used for its creation should not be used to guide conservation and land management unless it also makes effective predictions. Thus, prediction of species richness based on any functional relationship, including occurrence of indicator species, should be regarded as a testable hypothesis to be confronted with newly collected test data (Landres et al. 1998; Mac Nally et al. 2000; Fleishman et al. 2003; Mac Nally & Fleishman 2004). Based on much experience with real data sets, we believe cross-validation is inadequate for evaluating the utility of predictive models. Accordingly, we are in the process of collecting validation data to facilitate tests of the models presented here. In terms of both biogeography and land use, mountain ranges are an appropriate scale for developing and testing the transferability (sensu Leftwich et al. 1997) of indicator species models in this geographic region. If the models successfully predict species richness, we will combine the model building and validation data sets to yield an improved or "updated" model that more effectively represents species richness patterns over a large area (Hilborn & Mangel 1997; Mac Nally & Fleishman 2004). Previous work on prediction of butterfly species richness with butterfly indicators gives us much cause for optimism about the success of the current models (Mac Nally & Fleishman 2004).

Species Richness of One Taxonomic Group

Models of species richness of single taxonomic groups based on same-taxon indicator species explained approximately 80% deviance and did not appear to be highly sensitive to number of locations used to build the model. Regardless of response-predictor combination, effective indicator species collectively represented the variety of land-cover associations in the taxonomic group(s) for which they conveyed information about species richness. For example, the model of bird species richness with birds as predictors included two species, each associated with relatively dry vegetation (Western Scrub-Jay, Black-throated Gray Warbler [*Dendroica nigrescens*]), riparian vegetation (MacGillivray's Warbler [*Oporornis tolmiei*], Fox Sparrow [*Passerella iliaca*]), and mixed vegetation (Red-shafted Flicker, Western Tanager). Similarly, the model of species richness of butterflies with butterflies as predictors included species whose larvae feed on a full range of vegetation growth forms (Fleishman et al. 1997): grasses (fed on by *Ochlodes sylvanoides*), forbs (*Icaricia lupini, Chlosyne acastus, Phyciodes mylitta*), shrubs (*Euchloe ausonides*), and trees (*Papilio multicaudatus*). Moreover, each of the latter six butterfly species uses at least two of the three most prominent local topographic features (dry canyons, riparian canyons, and uplands), and the six species represent all five families of butterflies present in the Great Basin.

The fit of models of species richness for one taxonomic group was not improved by including indicator species from a second taxonomic group. Models of butterfly species richness, in particular, were not improved by including birds. A potential explanation may be that in the Great Basin, most if not all land-cover types used by birds are also used by butterflies. Butterflies, however, are typically diverse and abundant in meadows, a land-cover type that provides relatively few resources for birds, especially when subject to livestock grazing and recreational use (Warkentin & Reed 1999). Differences in occurrence patterns of birds and butterflies with respect to land cover may explain why genetic algorithm solutions did not include birds as predictors of butterfly species richness if butterflies were also available as predictors and why butterflies were better indicators of combined species richness than birds. Consistent with this hypothesis, the model of bird species richness with butterflies as indicators included one species primarily associated with low-elevation grasses and shrubs, *Pontia beckerii*. Sites at which *P. beckerii* was present usually had lower bird species richness than sites at which it was absent.

There was a slight improvement in model parsimony (i.e., fewer indicator species could be used to explain a similar percentage of deviance) when indicator species from two taxonomic groups were used to predict bird species richness. In practice, the advantages of surveying or monitoring indicator species from a single taxonomic group likely would outweigh the benefit of including one fewer indicator species in a model of species richness.

Practical Applications of Predictions for Multiple Taxonomic Groups

Our results suggest that information from different taxonomic groups can be pooled to predict species richness of a comprehensive biota. Among the most promising outcomes of our work is the discovery that it may be possible to use a common set of species to predict species richness for more than one taxonomic group. For example, 68% of bird species richness and 73% of butterfly species richness could be explained using a set of three species of birds and three species of butterflies, four of which were included in both models. If sufficient data existed, our methods certainly could be applied to a larger number of taxonomic groups in an effort to identify reliable indicators of species richness for a comprehensive set of plants and animals in a given ecosystem or geographic region. Alternatively, it may be more effective to use our methods to select, in turn, indicators within each taxonomic group of conservation concern.

Given that rare species often do not occur in locations with greatest species richness, it may not be possible to draw inferences about the presence of rare species on the basis of indicators of species richness with any method (Lawler et al. 2003). None of the birds or butterflies in our study system is listed as threatened or endangered, and there are no species-level endemics. Also, definition and estimation of rarity in the context of large-scale land management presents substantial challenges. In our study system, occurrence rates of individual species often vary among mountain ranges (Fleishman et al. 2003). Moreover, low occurrence rates may not equate to extinction risk: the relatively uncommon resources used by some species of birds and butterflies with low rates of occurrence, such as plants with patchy distributions, are not directly threatened by current land uses.

Efforts to identify surrogate-based approaches to estimate species richness and other measures of ecological status are popular because inventory data for many regions are insufficient to directly inform conservation and land-use planning decisions. Accordingly, a method for selecting indicator species that requires relatively complete species data might appear to be anything but a shortcut. Our methods, however, select indicator species potentially applicable across an extensive landscape on the basis of data that, although representative of the larger system, are collected from a small fraction of the total land area. Across planning landscapes of tens to hundreds of thousands of square kilometers, we believe the method is a realistic approach to deriving estimates of species richness that can inform strategies for allocation of multiple land uses, including conservation. The Great Basin, for example, covers more than 400,000 km^2 and includes approximately 350 major mountain ranges. If indicator species identified on the basis of data collected from sections of three mountain ranges can be used to predict species richness at any location even within the 30,000 km^2 biogeographic subregion in which our work was centered (Austin & Murphy 1987), we believe that the method can serve as an effective and practical tool for prioritizing management activities.

Comparative values of species richness are used as a primary criterion for establishing conservation and management priorities. The lack of data, money, and time has prompted scientists and practitioners to seek shortcuts for estimating species richness. Our work is the first to demonstrate empirically not only that species richness of one taxonomic group can be explained using indicators drawn from the same taxonomic group, but also that individual and combined species richness of two taxonomic groups can be explained using indicator species drawn either from both of those groups or exclusively from one of the groups. Our approach to selecting indicator species may help provide reliable data on species richness for meeting biological objectives in diverse assemblages and ecosystems. In addition, by developing and testing hypotheses to explain why particular species convey information about a larger biota, we gain insight into patterns and drivers of species richness and the nature of ecological assemblages.

Acknowledgments

We thank C. Betrus, L. P. Bulluck, and N. McDonal for collection of field data integral to this work; M. Betrus and J. Bulluck for field assistance; and R. Blair, D. Dobkin, E. Main, and three anonymous reviewers for insightful comments and suggestions. Financial support for this research was provided by the Nevada Biodiversity Research and Conservation Initiative; the Joint Fire Sciences Program via the Rocky Mountain Research Station, U.S. Department of Agriculture Forest Service; the Australian Research Council; and Miami University. This is contribution 59 from the Australian Centre for Biodiversity: Analysis, Policy and Management.

Literature Cited

Andelman, S. J., and W. F. Fagan. 2000. Umbrellas and flagships: efficient conservation surrogates, or expensive mistakes? Proceedings of the National Academy of Sciences 97:5954–5959.

Austin, G. T., and D. D. Murphy. 1987. Zoogeography of Great Basin butterflies: patterns of distribution and differentiation. Great Basin Naturalist 47:186–201.

Bergerud, W. A., and W. J. Reed. 1998. Bayesian statistical methods. Pages 89–104 in V. Sit and B. Taylor, editors. Statistical methods for adaptive management studies. Land management handbook 42. British Columbia Ministry of Forests, Forestry Division Services Branch, Victoria, Canada. Also available from http://www.for.gov.bc.ca/hfd/pubs/docs/lmh/lmh42.htm. (accessed March 2004).

Bibby, C. J., N. D. Burgess, D. A. Hill, and S. Mustoe. 2000. Bird census techniques. Academic Press, London.

Boyce, M. S., and L. L. McDonald. 1999. Relating populations to habitats using resource selection functions. Trends in Ecology & Evolution **14**:268-272.

Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas. 2001. Introduction to distance sampling: estimating abundance of biological populations. Oxford University Press, New York.

Christensen, J. 2003. Auditing conservation in an age of accountability. Conservation in Practice **4**:12-19.

Fleishman, E., G. T. Austin, and D. D. Murphy. 1997. Natural history and biogeography of the butterflies of the Toiyabe Range, Nevada (Lepidoptera: Papilionoidea). Holarctic Lepidoptera **4**:1-18.

Fleishman, E., G. T. Austin, and A. D. Weiss. 1998. An empirical test of Rapoport's rule: elevational gradients in montane butterfly communities. Ecology **79**:2482-2493.

Fleishman, E., J. P. Fay, and D. D. Murphy. 2000. Upsides and downsides: contrasting topographic gradients in species richness and associated scenarios for climate change. Journal of Biogeography **27**:1209-1219.

Fleishman, E., R. Mac Nally, and J. P. Fay. 2003. Validation tests of predictive models of butterfly occurrence based on environmental variables. Conservation Biology **17**:806-817.

Fleishman, E., R. Mac Nally, J. P. Fay, and D. D. Murphy. 2001. Modeling and predicting species occurrence using broad-scale environmental variables: an example with butterflies of the Great Basin. Conservation Biology **15**:1674-1685.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter. 1996. Introducing Markov chain Monte Carlo. Pages 1-16 inW. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors. Markov chain Monte Carlo in practice. Chapman and Hall, New York.

Gustafsson, L. 2000. Red-listed species and indicators: vascular plants in woodland key habitats and surrounding production forests in Sweden. Biological Conservation **92**:35-43.

Hilborn, R., and M. Mangel. 1997. The ecological detective: confronting models with data. Princeton University Press, Princeton, New Jersey.

Hobohm, C. 2003. Characterization and ranking of biodiversity hotspots: centres of species richness and endemism. Biodiversity and Conservation **12**:279-287.

Jefferson, M. F., N. Pendleton, S. B. Lucas, and M. A. Horan. 1997. Comparison of a genetic algorithm neural network with logistic regression for predicting outcome after surgery for patients with nonsmall cell lung carcinoma. Cancer **79**:1338-1342.

Kareiva, P., and M. Marvier. 2003. Conserving biodiversity coldspots. American Scientist **91**: 344-351.

Kotliar, N. B., and J. A. Wiens. 1990. Multiple scales of patchiness and patch structure: a hierarchical framework for the study of heterogeneity. Oikos **59**:253-260.

Landres, P. B., J. Verner, and J. W. Thomas. 1988. Ecological uses of vertebrate indicator species: a critique. Conservation Biology **2**:316-328.

Lawler, J. J., D. White, J. C. Sifneos, and L. L. Master. 2003. Rare species and the use of indicator groups for conservation planning. Conservation Biology **16**:936-947.

Lee, P. M. 1989. Bayesian statistics. Edward Arnold, London.

Leftwich, K. N., P. L. Angermeier, and C. A. Dolloff. 1997. Factors influencing behavior and transferability of habitat models for a benthic stream fish. Transactions of the American Fisheries Society **126**:725-734.

Mac Nally, R. 2000. Regression and model-building in conservation biology, biogeography, and ecology: the distinction between—and reconciliation of—"predictive" and "explanatory" models. Biodiversity and Conservation **9**:655-671.

Mac Nally, R. 2005. Scale and an organism-centric focus for studying interspecific interactions in landscapes. In press in J. A. Wiens and M. R. Moss, editors. Issues in landscape ecology. Cambridge University Press, New York.

Mac Nally, R., and E. Fleishman. 2002. Using 'indicator' species to model species richness: model development and predictions. Ecological Applications **12**:79-92.

Mac Nally, R., and E. Fleishman. 2004. A successful predictive model of species richness based on indicator species. Conservation Biology **18**:646-654.

Mac Nally, R., A. F. Bennett, and G. Horrocks. 2000. Forecasting the impacts of habitat fragmentation. Evaluation of species-specific predictions of the impact of habitat fragmentation on birds in the box-ironbark forests of central Victoria, Australia. Biological Conservation **95**:7-29.

Mac Nally, R., E. Fleishman, L. Bulluck, and C. Betrus. 2004. Comparative influence of spatial scale on beta diversity within regional assemblages of birds and butterflies. Journal of Biogeography **31**:917-929.

Mac Nally, R., E. Fleishman, J. P. Fay, and D. D. Murphy. 2003. Modelling butterfly species richness using mesoscale environmental variables: model construction and validation for mountain ranges in the great basin of western North America. Biological Conservation **110**:21-31.

Mayer, K. E., and W. F. Laudenslayer. 1988. A guide to the wildlife habitats of California. California Department of Forestry and Fire Protection, Sacramento.

McCullagh, P., and J. A. Nelder. 1989. Generalized linear models. 2nd edition. Chapman and Hall, London.

Mittermeier, R. A., N. Myers, J. Thomsen, G. A. B. Da Fonseca, and S. Olivieri. 1998. Biodiversity hotspots and major tropical wilderness areas: approaches to setting conservation priorities. Conservation Biology **12**:516-520.

Myers, N., R. A. Mittermeier, C. G. Mittermeier, G. A. B. da Fonseca, and J. Kent. 2000. Biodiversity hotspots for conservation priorities. Nature **403**:853-858.

Pearson, D. L. 1994. Selecting indicator taxa for the quantitative assessment of biodiversity. Philosophical Transactions of the Royal Society of London—Series B: Biological Sciences **345**:75-79.

Pollard, E., and T. J. Yates. 1993. Monitoring butterflies for ecology and conservation. Chapman and Hall, London.

Pullin, A. S., editor. 1995. Ecology and conservation of butterflies. Chapman and Hall, London.

R Development Core Team. 2003. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from http://www.R-project.org (accessed March 2004).

Schwarz, G. 1978. Estimating the dimension of a model. Annals of Statistics **6**:461-464.

Scott, C. T. 1998. Sampling methods for estimating change in forest resources. Ecological Applications **8**:228-233.

Scott, J. M., B. Csuti, J. D. Jacobi, and J. E. Estes. 1987. Species richness. BioScience **37**:782-788.

Scott, J. M., P. J. Heglund, M. Morrison, M. Raphael, J. Haufler, and B. Wall, editors. 2002. Predicting species occurrences: issues of scale and accuracy. Island Press, Covello, California.

Sekhon, J. S., and W. R. Mebane. 1998. Genetic optimization using derivatives. Theory and application to non-linear models. Political Analysis **7**:187-210.

Siegel, R. B., D. F. Desante, and M. P. Nott. 2001. Using point counts to establish conservation priorities: how many visits are optimal? Journal of Field Ornithology **72**:228-235.

Spiegelhalter, D. J., A. Thomas, and N. Best. 2003. WinBUGS. Version 1.4. Bayesian inference using Gibbs sampling. Manual. MRC Biostatistics Unit, Institute for Public Health, Cambridge, United Kingdom.

Vinterbo, S., and L. Ohno-Machado. 1999. A genetic algorithm to select variables in logistic regression: example in the domain of myocardial infarction. Journal of the American Medical Informatics Association 1999 Symposium **Supplement**:984-988.

Warkentin, I. G., and J. M. Reed. 1999. Effects of habitat type and degradation on avian species richness in Great Basin riparian habitats. The Great Basin Naturalist **59**:205-212.