



# Comparison of predictor sets for species richness and the number of rare species of butterflies and birds

James R. Thomson<sup>1</sup>, Erica Fleishman<sup>2\*</sup>, Ralph Mac Nally<sup>1</sup> and David S. Dobkin<sup>3</sup>

<sup>1</sup>Australian Centre for Biodiversity: Analysis, Policy and Management, School of Biological Sciences, Monash University 3800, Melbourne, Australia, <sup>2</sup>Center for Conservation Biology, Department of Biological Sciences, Stanford University, Stanford, CA 94305-5020, USA and <sup>3</sup>High Desert Ecological Research Institute, 15 S.W. Colorado Avenue, Suite 300, Bend, OR 97702, USA

## ABSTRACT

**Aim** Accurate inventories of biota are typically restricted to few locations within an extensive region. Accordingly, effective planning must involve some form of surrogate measures coupled with spatial modelling. We conducted a simultaneous comparison of models of both species richness and the number of rare species using three types of surrogates (indicator species, vegetation composition and structure, and topoclimate) as predictors. We evaluated each type of surrogate alone and in combination with others.

**Location** Data for our analyses were collected from 1996–2004 in three adjacent mountain ranges in the central Great Basin (Lander and Nye counties, Nevada, USA), the Shoshone Mountains, Toiyabe Range and Toquima Range.

**Methods** Data on species richness and species composition of butterflies and birds and measures of vegetation composition and structure were obtained in the field. Topoclimatic variables were derived by GIS from digital sources and satellite images. We used Poisson regression with Bayesian model averaging to predict species richness and the number of rare species. We compared the expected prediction success of all models on the basis of internal and external validation trials.

**Results** Same-taxon indicator species were the most accurate predictors of species richness and of the number of rare species of butterflies and birds. Cross-taxon indicator species and topoclimate variables were reasonably accurate predictors of species richness of butterflies and birds and of the number of rare butterfly species. Although vegetation variables were more effective for predicting species richness and number of rare species of birds than of butterflies, they were the least accurate predictors overall.

**Main conclusions** Although indicator species may provide the most accurate predictions of species richness, their practical value, like any surrogate measure, depends greatly on ecological considerations and land-use context. In general, the ability to predict numbers of rare species based on any set of candidate predictors was weaker than the ability to predict species richness, which may result from the high degree of stochasticity that often characterizes distributions of rare species. Our statistical approach for objective examination of different candidate predictors can help ensure that selection of species-richness surrogates in any system is scientifically reliable and cost-effective.

## Keywords

Bayesian model averaging, biodiversity surrogates, Great Basin, indicator species, predictive modelling, rarity, remote sensing, topoclimate, vegetation composition, vegetation structure.

\*Correspondence: Erica Fleishman, National Center for Ecological Analysis and Synthesis, 735 State Street, Suite 300, Santa Barbara, CA 93101, USA.  
E-mail: fleishman@nceas.ucsb.edu

## INTRODUCTION

Protection of locations with high native species richness generally is thought to be an efficient way of conserving overall biological diversity and sustaining key ecological functions (Scott *et al.*, 1987; Myers *et al.*, 2000). There also is inherent interest in natural and anthropogenic processes associated with species richness (Janzen *et al.*, 1976; Pickett *et al.*, 2001; Hawkins *et al.*, 2003; Yeoman & Mac Nally, 2005). Moreover, species richness has been measured in many places over long periods of time and typically is more practical to assess and to monitor than abundance or demographic variables (Link & Sauer, 1998) (for a review of the utility and limitations of species richness metrics, see Fleishman *et al.*, 2006).

Whether the underlying reason is legal mandate, influence on ecological function, or aesthetics, conservation planning often weights measures of species richness by rarity (Noss & Cooperrider, 2004; Stein *et al.*, 2000). Rarity, which can be driven by a complex array of mechanisms, has often been equated with small geographical range, low abundance or both (Brown, 1984; Gaston, 1994). Rarity is not synonymous with endemism, and species that are locally rare (uncommon) may not have narrow geographical ranges. At the landscape level, the total number of species and the number of rare species are frequently correlated, especially in nested systems (Patterson & Atmar, 1986; Wright *et al.*, 1998). However, rare species often do not occur in locations with the highest species richness (Chaplin *et al.*, 2000; Jetz & Rahbek, 2002; Orme *et al.*, 2005; Stohlgren *et al.*, 2005). The distributions of different rare species in the same taxonomic group, and especially in different taxonomic groups, may not overlap (Freitag *et al.*, 1997; Moore *et al.*, 2003). Therefore, effective methods for measuring and managing rare species may differ from those directed toward species richness (Lawler *et al.*, 2003).

Given that species-inventory data for many regions are sparse and acquisition of new data is costly, surrogate-based approaches to the estimation of biodiversity patterns have become common. Depending on taxonomic group, geographical location and scale (grain and extent in space or time), any number of environmental variables may be associated with the distribution of species richness and the number of rare species (Guisan & Thuiller, 2005). Both scientific reliability and cost-effectiveness influence the selection of variables for modelling. Inevitably, there is a trade-off between the quality of information provided by predictor variables and the cost of obtaining that information.

In this paper, we use advanced modelling approaches to compare the effectiveness of three different types of predictor variables, indicator species, vegetation and topoclimate, to predict both species richness and the number of rare species of butterflies and birds in the central Great Basin of western North America. One ecologically important and practically useful goal of this work is to determine the most reliable surrogate(s) for species richness and the number of rare species in this system. More generally, we aim to develop methods

that can be used to identify the most effective surrogates for biodiversity patterns in other systems.

Local distributional patterns of bird and butterfly assemblages might reasonably be expected to be a function of the taxonomic composition and physical structure of local vegetation (MacArthur *et al.*, 1966; Wiens & Rotenberry, 1981; Scott, 1986), but collecting detailed vegetation data requires site visits. Measures of topography or topoclimate have also been shown to be effective as explanatory and predictive variables of species richness and occurrence in some cases (e.g. Weiss *et al.*, 1988; Scott *et al.*, 2002). With GIS and appropriate software, these variables are probably the cheapest per unit area to acquire among possible predictor data because they do not require site visits. Remote sensing data for variables associated with biomass and primary productivity, such as the normalized difference vegetation index (NDVI) (Rouse *et al.*, 1973), have similar advantages (Roughgarden *et al.*, 1991; Turner *et al.*, 2001).

Workers have identified statistical relationships between species richness of a relatively large set of organisms and the occurrence patterns of a small set of 'indicator' species (Pearson, 1994; Scott, 1998; Mac Nally & Fleishman, 2004). Our recent work suggested that it is possible to identify small suites of butterflies and birds whose presence/absence patterns can be used to predict species richness of their own taxonomic group, of the other taxonomic group or the combined species richness of both groups (Fleishman *et al.*, 2005; Thomson *et al.*, 2005). Whether it is more efficient to measure the occurrence of a small number of indicator species or to conduct comprehensive species inventories depends greatly on geographical location and taxonomic group. In addition, reliable data on the occurrence of any species may require multiple site visits (Thomson *et al.*, 2005; MacKenzie *et al.*, 2006). We are not aware of any previous work examining the relationships between indicator species (as defined here) and the number of rare species, nor of efforts to compare statistically the predictive ability of indicator species against the predictive ability of other sets of variables.

## METHODS

Data for our analyses were collected in three adjacent mountain ranges that have similar biogeographical and human land-use histories, the Shoshone Mountains (1850 km<sup>2</sup>, approximate north-south boundaries 39°14' to 38°57'), the Toiyabe Range (3100 km<sup>2</sup>, 39°54' to 38°30') and the Toquima Range (1750 km<sup>2</sup>, 39°17' to 38°29') (Lander and Nye counties, Nevada, USA). Our data collection incorporated established techniques that reliably detect species presence and permit the assessment of distributional trends (Pullin, 1995; Bibby *et al.*, 2000). We provide an abbreviated description here; these methods have been described in detail in previous publications, as well as tested for sampling adequacy (e.g. Dobkin & Wilcox, 1986; Fleishman *et al.*, 2000, 2001; Mac Nally *et al.*, 2004).

We inventoried butterflies using walking transects, a standard method for temperate regions (Pollard & Yates, 1993; Pullin, 1995). Approximately every 2 weeks throughout the majority of the adult flight season (late May to August), using equal sampling effort per unit area of the site, we recorded the presence of all butterfly species in each site. In temperate regions, it is generally reasonable to interpret that a given butterfly species is absent if the area has been searched repeatedly by experienced observers during the appropriate season and weather conditions (Pullin, 1995). Sites were sufficiently large that an individual was not recorded in more than one site during a sampling round (Fleishman *et al.*, 1997).

We sampled birds during the breeding season (late May to June) using two or three 75-m variable-radius point counts in each site. Most point centres were at least 350 m apart. Point counts are an effective method of sampling birds in the Great Basin (Dobkin & Rich, 1998). Within a site, points were located in each of the dominant vegetation types (e.g. aspen, pinyon-juniper, sagebrush). During each visit, we recorded by sight or sound all birds actively using terrestrial habitat within the point. Each point was visited three times per year for 5 min per visit. Three surveys are considered sufficient to determine which species of birds are present at point count locations (Buckland *et al.*, 2001; Siegel *et al.*, 2001); species accumulation curves indicated that this was the case in our work (Betrus, 2002).

From 1996–2003, inventories for butterflies were conducted at 195 sites and inventories for birds were conducted at 84 of those sites. Site areas ranged from 1.5 to 44.4 ha, but area explained little variance in species richness (< 3%) of either taxonomic group (Mac Nally *et al.*, 2003).

We recorded 65 species of resident butterflies and 74 species of breeding birds. Site-level species richness was calculated as the total number of species recorded in the site across all years. Site-level species richness ranged from 3 to 51 for butterflies (mean 24.9) and from 5 to 34 for birds (mean 17.3). We defined species that occurred in < 20% of sites as rare. Twenty-one butterfly species and 33 bird species met this criterion. The number of rare butterfly species at a site varied from 0 to 10 (mean 1.9) and the number of rare bird species at a site varied from 0 to 8 (mean 2.3). In our system, no species of butterfly or bird is listed as threatened or endangered by the federal or state government, and there are no species-level endemics. The rare butterfly and bird species in this system are generally taxa with large geographical ranges that are locally uncommon or, in a few cases, are approaching the edge of their distribution.

To characterize vegetation composition and structure, we established two or three plots in each site that overlapped the points at which we sampled birds. We measured three radial 30-m lines, separated by 120°, from the centre of the plot. The end of each line served as the centre of a circular vegetation sampling unit with 11.3-m radius (0.04 ha). Within each circle, we recorded the identity and size of all live trees [either diameter at breast height (d.b.h.) or basal diameter, depending on plant morphology]. We also recorded the size and, where possible, the identity of standing or fallen dead trees. Vegetation at each plot

was measured once. The majority of our vegetation measures were unlikely to have changed substantially during the time period of our study (J. Chambers, pers. comm.).

We used a concave spherical densiometer to estimate the proportion of canopy cover. To estimate the cover of shrubs and ground vegetation, we used an ocular tube with measurements taken at a 45° angle downward from the line of sight (Noon, 1981). When vegetation was present, we recorded the occurrence of 12 dominant species of trees and shrubs. We collected 21 densiometer and ocular tube readings at each plot: one each at 8, 16 and 24 m along the 30-m line from the centre of the plot to the perimeter of each circle, and one while facing in each of the four cardinal directions from the centre of each circle. Cover values for each structural layer (canopy, shrubs and ground vegetation) were averaged for each plot and site. Occurrence data for dominant species were aggregated into a relative measure of frequency at each plot and site.

Sites were delineated by overlaying differentially corrected global positioning system locations on a 30-m (1 : 24,000) digital elevation model maintained on a GIS (Fleishman *et al.*, 2001). Geographical coordinates and 14 uncorrelated ( $R_{\text{Pearson}} < 0.6$ ) topographic and topoclimatic variables were measured for each site (Table 1). All of the variables could reasonably be expected, *a priori*, to have functional relationships with distribution patterns of butterflies and birds. We derived measures of NDVI from a single cloud-free, geo-corrected Landsat Thematic Mapper image (WRS 41/33) acquired in June 2000, coincident with the peak of the regional growing season and the breeding season for birds. Within our study system, NDVI in a given location and spatial patterns of NDVI are not highly variable over years to decades (K. Seto and B. Bradley, pers. comm.). For simplicity, we refer to the full set of remotely derived variables as topoclimate variables.

### Model building and prediction

We compared the effectiveness of indicator species, vegetation and topoclimate as predictors of the total number of species

**Table 1** Topoclimate variables used to model species richness and number of rare species of butterflies and birds

Geographical coordinates of the centroid of the site
Mean 'eastness' on a scale from -100 (west-facing) to 100 (east-facing)
Mean 'northness' on a scale from -100 (south-facing) to 100 (north-facing)
Mean elevation (m)
Mean slope (°)
Site length (m)
Solar insolation: standard deviation for the site at the vernal equinox (kJ)
Topographic exposure (elevation of the site compared with mean elevation within a 300 m radius): mean and standard deviation
Mean and minimum distance (m) from the centre of the site to running or standing water
NDVI: mean and standard deviation for the site

and the number of rare species of both butterflies and birds. We also tested the predictive capacity of combined environmental variables (vegetation and topoclimate) and of each environmental data set in combination with indicator species from each taxon. Thus, there were four response variables: species richness of butterflies, species richness of birds, number of rare butterfly species and number of rare bird species. There were nine predictor variables: butterfly indicator species, bird indicator species, vegetation, topoclimate, vegetation + topoclimate, butterfly indicator species + vegetation, butterfly indicator species + topoclimate, bird indicator species + vegetation and bird indicator species + topoclimate. Therefore, there were 36 combinations of responses and predictors.

We included orthogonal second-order polynomial (quadratic) transformations of each environmental (vegetation or topoclimate) variable to allow for possible non-monotonic relationships with the response variables. Given that data for indicator species are binary (presence/absence), their relationships with response variables were necessarily linear. For predictor sets that included both indicator species and environmental variables, we restricted all variables to the linear form for computational tractability.

Although we are trying to predict the number of rare species (response variable), species that are difficult to detect or that infrequently inhabit suitable sites are unlikely to be useful indicator species (predictor variable) because of the risk of false negatives, that is, erroneously concluding that a species does not occupy a site (Thomson *et al.*, 2005). In our data, false negatives mainly arise from annual variability in site occupancy by individual species. Our work suggests that species can be detected reliably when they are present, but species are not necessarily present at occupied sites (defined here as sites inhabited by a species in one or more years) every year. Given that it is important that predictor variables can be measured reliably, we excluded from the list of candidate indicator species any species with a low ( $< 0.4$ ) probability of being detected at occupied sites in any single year.

### Modelling approach

We used Poisson regression with Bayesian model averaging (Hoeting *et al.*, 1999; Wintle *et al.*, 2003) to predict, on the basis of our predictor data  $D$ , the number of species (or rare species)  $S$  present at a site  $\Pr(S|D)$ . Poisson regression is a form of generalized linear model (GLM) with a logarithmic link function and an assumed Poisson distribution for the response variable (McCullagh & Nelder, 1989). Bayesian model averaging (BMA) incorporates uncertainty in model selection into statistical inference and prediction (Wintle *et al.*, 2003). This approach has consistently been shown to produce more accurate predictions than methods that select a single best model (Raftery & Zheng, 2003; Burnham & Anderson, 2004).

The model averaging formula for estimating  $\Pr(S|D)$  is (Hoeting *et al.*, 1999)

$$\Pr(S|D) = \sum_{k=1}^K \Pr(S|M_k, D) \Pr(M_k|D)$$

where  $\Pr(S|M_k, D)$  is the posterior prediction from model  $M_k$  given the data,  $\Pr(M_k|D)$  is the posterior probability of model  $M_k$  given the data and  $K$  is the number of models considered. In other words, BMA provides an estimate of  $\Pr(S|D)$  that is a weighted average of the posterior predictions from all models considered, where the weights are the posterior model probabilities (Raftery *et al.*, 1997; Wintle *et al.*, 2003). Integrating the posterior model probabilities for all models that include a given predictor variable yields the conditional probability that the variable has a nonzero coefficient  $\Pr(\beta \neq 0)$  or, equivalently, the conditional probability that the variable is a predictor (Viallefont *et al.*, 2001). A high  $\Pr(\beta \neq 0)$  value indicates that a variable is included in the most probable model(s) and therefore contributes most to prediction; a low value indicates that the variable is included only in less probable models and therefore contributes little to prediction. In cases where it is necessary to reduce the number of predictor variables for model application (e.g. in indicator species modelling) it is appropriate to select as predictors those variables with highest  $\Pr(\beta \neq 0)$  values. We implemented BMA with the 'bic.glm' function in the 'BMA' package (Raftery *et al.*, 2005) for R (R Development Core Team, 2005). This function uses the 'leaps and bounds' algorithm (Furnival & Wilson, 1974) to identify rapidly the most probable models based on the Bayesian information criterion (BIC) approximation to Bayes factors (Raftery, 1995). bic.glm uses maximum likelihood estimation to fit individual models [and hence to estimate  $\Pr(S|M_k, D)$ ] and weights those models [i.e. estimates  $\Pr(M_k|D)$ ] according to BIC values (Raftery, 1995; Wintle *et al.*, 2003).

Indicator species are subsets of their assemblage, and only a small number of indicator species (predictor variables) can be used for prediction. Therefore, variable selection is required to choose the most useful set of indicator species (Fleishman *et al.*, 2005). We used  $\Pr(\beta \neq 0)$  values to select sets of indicator species. For each response variable, we first computed BMA with all candidate indicator species to estimate the conditional probability that each species was a predictor,  $\Pr(\beta \neq 0)$ . We selected the six species with the highest  $\Pr(\beta \neq 0)$  values as the indicator species for that response variable. We then obtained predicted values of species richness or number of rare species by applying BMA to the selected indicator species only. That is, predictions from indicator species models were weighted averages of predictions from all possible combinations of the six selected indicator species. For some response-indicator taxon combinations, the full model (all six indicator species) had posterior probability of 1 (conditional on the set of  $2^6 = 64$  models possible, given the selected indicator species), so predictions effectively were from a single model.

We used a similar two-step model-averaging procedure to select indicator species, and subsequently to predict species richness or number of rare species, for models combining

indicator species with environmental variables. For each response–predictor combination, we first computed BMA with all possible variables (e.g. all candidate bird species plus all vegetation variables) to identify the most promising indicator species in combination with a particular set of environmental variables. The six species with highest  $\Pr(\beta \neq 0)$  were selected as the indicator species. We then used BMA to predict values of species richness or number of rare species from models that combined any of the six selected indicator species with any of the relevant environmental variables (i.e. vegetation or topoclimate).

For vegetation and topoclimate variables, we assumed that the cost of obtaining data for additional variables was negligible. Therefore, all variables were used in model averaging. We made predictions using models built from any possible combination of 21 vegetation variables and 16 topoclimate variables. This assumption is valid for the types of remotely sensed topoclimate data used here, although collecting field data on many vegetation variables may be more time-consuming than collecting data on a smaller set of variables. All environmental variables were standardized prior to model fitting:  $X_{ik}' = (X_{ik} - \bar{X}_k)/s_k$ , where  $\bar{X}_k$  and  $s_k$  are the mean and standard deviation of variable  $k$  and  $X_{ik}$  is the value of variable  $k$  at site  $i$ .

### Model validation

We compared the expected prediction success of all models using both internal and external validation trials. We assessed predictive performance primarily on the basis of Spearman rank correlations between predicted and observed species richness. Rank correlation scores are more robust than parametric correlations when data do not come from a bivariate normal distribution, as is likely with species richness (counts) data. Furthermore, we believe that the ability to rank sites by relative species richness or number of rare species will often be more useful for conservation management and land-use planning, especially in our study system, than predicting the exact number of species or rare species in each site. Given that the value of different measures of predictive performance may differ among landscapes, we also evaluated the absolute accuracy of predictions by calculating root mean square error (rmse) and the mean absolute difference between observed and predicted values (represented as mean bias,  $B$ ).

#### Internal validation

We used bootstrapping to estimate the expected accuracy of predictions of species richness and number of rare species based on Bayesian model averaging. Bootstrapping validation provides almost unbiased estimates of predictive accuracy with relatively low variance (Efron & Tibshirani, 1993, 1997; Wintle *et al.*, 2005). Bootstrapping adjusts the in-sample (or naïve) validation statistic by an estimate of its optimism, which is derived from simulations of model building and model testing performed on bootstrap samples (samples drawn with replace-

ment from the model building data). We used the ‘.632 bootstrap’ method (Efron & Tibshirani, 1997) with 50 bootstrap samples to calculate adjusted validation statistics for each response–predictor combination. The 0.632 bootstrap estimate is a weighted average of the in-sample statistic and the mean ‘out-of-sample’ statistic of models fitted to the bootstrap samples. The weighting of the mean out-of-sample statistic is 0.632, which is the probability that a given case is included in a given bootstrap draw.

#### External validation

Predictions of species richness or number of rare species based on a functional relationship should be confronted with newly collected test data (Landres *et al.*, 1988; Mac Nally *et al.*, 2000). For models in which birds served as either the response variable (species richness of birds or number of rare bird species) or predictor variables (models with butterfly response variables and bird indicator species as predictors), we performed a preliminary external validation using data from 25 sites in the Toiyabe Range that have similar environmental profiles to sites used in model building and bootstrapping. The validation sites were surveyed for birds in 2004, after the model building data were compiled.

For models not involving birds as either response or predictor variables, we used the validation sites to perform a form of cross-validation, rather than independent external validation, because all relevant data were compiled concurrently and used in the original model building and bootstrapping. We rebuilt relevant models excluding the 25 validation sites from the model building data and then examined the performance of those models when applied to the 25 validation sites.

#### Alternative modelling

To ensure that our comparisons of different types of surrogate were not biased by our choice of modelling method, we also evaluated the performance of environmental variables as predictors of richness (total species richness and number of rare species) using multivariate adaptive regression splines (MARS) (Friedman, 1991; Leathwick *et al.*, 2005). MARS is an efficient method of modelling more complex functional responses and/or interactions among predictor variables. We used the ‘mars’ function in the ‘mda’ package for *R* (Hastie *et al.*, 2005) to fit MARS models for each response variable as functions of vegetation variables, topoclimate variables and combined vegetation and topoclimate variables. We fitted separately additive models and models allowing up to second- and third-order interactions. We used forward stepwise selection of basis functions and backward stepwise pruning to prevent overfitting. We evaluated MARS predictions by bootstrapping, cross-validation or external validation as appropriate. Results were similar to BMA results, although poorer overall (especially for models allowing interactions), and the relative rankings of different surrogates (vegetation, topoclimate or indicator species) were identical regardless of

whether MARS or BMA predictions were used for environmental variables. For clarity we present only BMA results.

## RESULTS

### Models for butterflies

#### Internal validation

Butterfly indicator species were the best predictors of the species richness of butterflies ( $R_S = 0.82$ ) and of the number of rare butterfly species ( $R_S = 0.70$ ). Combining environmental variables with butterfly indicator species did not improve predictions of the species richness of butterflies or of the number of rare butterfly species (Table 2a).

Bird indicator species were less successful predictors than butterfly indicator species, with rank correlations of 0.52 and 0.56 for species richness of butterflies and number of rare butterfly species, respectively. Combining topoclimate variables with bird indicator species improved slightly the predictions of the number of rare butterfly species, but not the predictions of species richness of butterflies. Combining vegetation variables with bird indicator species reduced prediction success (Table 2a).

Of the sets of predictor variables we examined, vegetation variables were the poorest predictors of butterfly species richness ( $R_S = 0.22$ ) and of the number of rare butterfly species ( $R_S = 0.21$ ). Topoclimate variables were second in accuracy to butterfly indicator species as predictors of species richness of butterflies ( $R_S = 0.64$ ), and showed similar performance to bird indicator species as predictors of the number of rare butterfly species ( $R_S = 0.52$ ). Models based on a combination of vegetation variables and topoclimate variables yielded less-reliable predictions of butterfly species richness than models based on topoclimate variables alone (Table 2a). Predictions of the number of rare butterfly species based on a combination of vegetation variables and topoclimate variables had marginally higher correlations with observed values than predictions based on topoclimate variables alone, but were less accurate absolutely (Table 2a).

#### Cross-validation

Cross-validation results for models of the species richness of butterflies and for the number of rare butterfly species were largely consistent with the internal validation results (Table 2a). Butterfly indicator species were the best predictors of species richness of butterflies ( $R_S = 0.83$ ) and of the number of rare

**Table 2** Validation results for Bayesian model averaged predictions of species richness and the number of rare species of (a) butterflies and (b) birds using indicator species, vegetation and topoclimate as predictors

Predictors	Species richness of butterflies						Number of rare species of butterflies					
	$R_{632}$	$R_v$	rmse <sub>632</sub>	rmse <sub>v</sub>	$B_{632}$	$B_v$	$R_{632}$	$R_v$	rmse <sub>632</sub>	rmse <sub>v</sub>	$B_{632}$	$B_v$
(a)												
Butterfly indicator species	0.82	0.81	5.50	5.23	0.03	-2.47	0.70	0.72	1.78	1.85	-0.04	-0.95
Bird indicator species	0.52	0.68	9.17	6.68	-0.44	3.35	0.56	0.65	2.94	1.35	-0.14	-0.27
Vegetation	0.22	-0.04	11.22	11.25	0.00	3.04	0.21	-0.15	2.86	2.17	-0.18	0.69
Topoclimate	0.64	0.54	7.60	9.11	-0.14	3.90	0.52	0.41	1.87	2.46	0.02	0.76
Vegetation + topoclimate	0.57	0.39	9.94	10.51	-0.50	2.50	0.55	0.54	3.28	2.04	-0.27	1.26
Butterfly indicator species + vegetation	0.78	0.77	5.88	6.07	-0.12	-3.26	0.68	0.75	2.11	2.47	-0.09	-0.77
Butterfly indicator species + topoclimate	0.82	0.74	5.65	6.16	-0.03	-0.38	0.70	0.69	1.86	2.59	-0.06	-1.17
Bird indicator species + vegetation	0.43	0.17	11.17	7.98	-1.00	0.43	0.45	0.20	8.19	6.21	-0.52	-1.36
Bird indicator species + topoclimate	0.51	0.52	9.04	8.34	0.06	-0.58	0.66	0.61	2.19	1.54	-0.08	-0.68
(b)												
	Species richness of birds						Number of rare species of birds					
Butterfly indicator species	0.68	0.45	4.84	7.63	-0.03	-5.28	0.32	0.57	1.85	1.39	0.00	-0.92
Bird indicator species	0.81	0.84	3.76	5.46	0.20	-4.68	0.45	0.52	1.59	1.47	-0.05	-0.93
Vegetation	0.56	0.54	11.03	6.08	-0.53	-2.80	0.20	0.03	5.32	2.85	-0.34	-0.96
Topoclimate	0.61	0.34	5.07	6.61	-0.24	-0.84	0.38	0.42	1.61	1.96	-0.03	-0.03
Vegetation + topoclimate	0.57	0.34	14.55	6.35	-0.56	-1.01	0.29	0.06	3.15	2.71	-0.13	-0.01
Butterfly indicator species + vegetation	0.68	0.45	4.54	8.48	-0.19	-3.21	0.43	0.54	2.01	1.66	-0.57	-0.98
Butterfly indicator species + topoclimate	0.51	0.42	5.49	7.54	0.23	-5.41	0.42	0.52	1.70	1.09	0.38	-0.27
Bird indicator species + vegetation	0.75	0.72	4.00	5.01	0.24	0.72	0.41	0.66	6.98	1.40	-0.33	0.66
Bird indicator species + topoclimate	0.68	0.83	4.77	5.22	0.69	-4.36	0.29	0.64	1.95	1.08	0.43	-0.50

$R$  = Spearman rank correlation between predicted and observed values, rmse = root mean square error of predicted values,  $B$  = mean bias of predicted values, calculated as the mean absolute difference between predicted and observed values.

Results of internal validations using the 0.632 bootstrap procedure (Efron & Tibshirani, 1997) are indicated by subscript '632'.

Results of external validation (for models in which birds were the response or predictor variables) or cross-validation (for models not involving birds) are indicated by subscript 'v'.

butterfly species ( $R_S = 0.74$ ) at validation sites. Bird indicator species performed better than expected given the bootstrapping results. Models based on vegetation variables alone were ineffective when applied to validation sites. Combining indicator species from either taxon with environmental variables did not improve predictions of either response variable.

#### Identity of indicator species

Twenty-two butterfly species and 20 bird species were selected as indicator species in at least one model of butterfly species richness or number of rare species of butterflies. Recall that there were six different models (i.e. response–predictor combinations) per indicator taxon and each model comprised a maximum of six indicator species. All of the indicator species contributed strongly [ $\text{Pr}(\beta \neq 0) > 0.70$  for all but three species and mostly  $> 0.90$ ] to Bayesian model-averaged predictions of species richness regardless of whether environmental variables were also included as candidate predictors. The identity of the selected indicator species depended on the response–predictor combination. Some species, such as the butterfly *Ochlodes sylvanoides*, and the violet-green swallow (*Tachycineta thalassina*), were selected as indicators for both species richness and number of rare species. Other species, including the butterfly *Icaricia lupini*, black-headed grosbeak (*Pheucticus melanocephalus*) and song sparrow (*Melospiza melodia*), were selected in both indicator species-only models and models that included environmental variables. Many species, however, were selected in only one or two models. The contributions of environmental variables to predictions of species richness when indicator species also were included as candidate predictors were often small and were much lower than when environmental variables were the only predictors.

#### Models for birds

##### Internal validation

Bird indicator species were the best predictors of the rank species richness of birds and of the rank number of rare bird species (Table 2b). Predictions of bird species richness based on bird indicator species ( $R_S = 0.81$ ) were much better than predictions of the number of rare bird species ( $R_S = 0.45$ ). The performance of bird indicator species models was not improved by the addition of either vegetation variables or topoclimate variables (Table 2b).

The utility of butterfly indicator species for predicting species richness of birds ( $R_S = 0.68$ ) was greater than that of vegetation variables ( $R_S = 0.56$ ), topoclimate variables ( $R_S = 0.61$ ), or a combination of vegetation and topoclimate variables ( $R_S = 0.57$ ). Combining butterfly indicator species with vegetation or topoclimate variables improved predictions of the number of rare bird species but not of total species richness of birds (Table 2b).

Topoclimate variables alone were better predictors of the number of rare bird species than vegetation variables, all

environmental variables or butterfly indicator species, but all sets of predictions had  $R_S < 0.50$  (Table 2b).

##### External validation

External validations confirmed that bird indicator species were the most reliable predictors of species richness of birds ( $R_S = 0.84$ ) (Table 2b). Butterfly indicator species and topoclimate variables performed worse than expected on the basis of bootstrapping results and produced poorer predictions of bird species richness than vegetation variables (Table 2b). Predictions of species richness of birds based on a combination of indicator species (either taxon) and environmental variables were no better than predictions based on indicator species alone.

A combination of bird indicator species and vegetation variables produced the best predictions of the number of rare bird species at validation sites ( $R_S = 0.66$ ). With two exceptions (vegetation and vegetation + topoclimate), predictions of the number of rare bird species were better than expected from bootstrapping results, but none of the predictions had  $R_S \geq 0.70$  (Table 2b).

##### Identity of indicator species

Thirteen butterfly species were selected as indicator species in models of bird response variables. Several of these species were consistently selected. *Satyrium sylvinum* was selected in five of six possible models and made strong contributions [ $\text{Pr}(\beta \neq 0) \geq 0.75$ ] to predicted values in each case. *Nymphalis antiopa* was selected in all six models, but made relatively small contributions [ $\text{Pr}(\beta \neq 0) < 0.20$ ] to predictions of the number of rare bird species. *Papilio multicaudatus* and *Incisalia eryphon* were selected in all three models of the species richness of birds and each made strong contributions [ $\text{Pr}(\beta \neq 0) \geq 0.90$ ] to predictions in two models.

Twenty bird species were selected as indicators of the species richness of birds or the number of rare bird species. Different sets of indicator species were selected according to the response–predictor combination, but some species [e.g. western scrub-jay (*Aphelocoma californica*) and plumbeous vireo (*Vireo plumbeus*)] were selected in multiple models. Most species selected as indicators of the species richness of birds contributed strongly to predicted values [15 of 18 cases had  $\text{Pr}(\beta \neq 0) \geq 0.70$ ], and more strongly than environmental variables in models that used a combination of indicator species and environmental predictors. The strength of contributions of bird indicator species to predictions of the number of rare bird species was more variable, but was generally greater than contributions from environmental variables in models that drew from all candidate predictors.

## DISCUSSION

Same-taxon indicator species were the most accurate surrogates for species richness and number of rare species. Cross-taxonomic indicator species provided reasonable predictions

of most response variables. They were at least as reliable as topoclimate variables, and were generally more reliable surrogates than vegetation variables. Different sets of indicator species had the strongest statistical relationships with species richness and with number of rare species. For indicator species-only models within taxonomic groups, only one butterfly species (*O. sylvanoides*) was selected as an indicator of both species richness of butterflies and number of rare species of butterflies. There was no overlap in sets of birds selected as the best indicators of species richness of birds or number of rare species of birds. In general, the ability to predict numbers of rare species based on any set of candidate predictors was weaker than the ability to predict species richness, which may result from the high degree of stochasticity that often characterizes distributions of rare species in our study system.

We acknowledge and emphasize that characterizations such as 'reliable' and 'reasonable' are relative. Whether a given threshold of explanatory or predictive accuracy (e.g. 80%) is 'good enough' inevitably depends on the context in which information is used. Land-use decisions that are relatively difficult to reverse (e.g. authorization of a new minerals mine) may demand a higher level of accuracy than situations in which land use can be modified or reversed through an iterative management process (e.g. adjusting the number of livestock permitted to graze in a given location). Decision theory provides an objective framework for choosing among alternative models in the context of how land-use or conservation actions will be determined and applied (Burgman *et al.*, 2005).

In most cases, indicator species collectively seemed to represent the variety of land-cover associations in the taxonomic group for which they conveyed information about species richness. For example, the three most influential species of butterflies in models of species richness of birds use trees as larval host plants, and those trees typically occur in different vegetation communities: dry pinyon-juniper woodland (*I. eryphon*), mixed riparian vegetation (*P. multicaudatus*) and stands of willows associated with permanent or ephemeral sources of water (*S. sylvinum*).

If indicator species reflect land-cover associations, we might expect local vegetation variables to be equally effective as predictors of species richness (Robinson & Holmes, 1984; Lee & Rotenberry, 2005). Our results suggest that this is not the case. Predictions based on vegetation variables, either alone or in combination with other types of candidate predictors, were often the least accurate in absolute terms (bootstrapped rmse values). These inaccuracies resulted from substantial variability in vegetation among sites. If model-testing sites have values that fall outside the range of values included in the model-building data, predictions for those test sites often will be inaccurate. The difficulty of making predictions for new sites with environmental characteristics that are different from environmental characteristics at sites for which species data exist highlights a potential disadvantage of using vegetation variables as predictors in regions with heterogeneous land

cover. Perhaps because many of our vegetation variables focused on structure rather than floristic composition, vegetation variables were more effective for predicting species richness and number of rare species of birds than of butterflies (Anderson & Shugart, 1974).

Topoclimate variables appeared to be reasonable predictors of species richness of both butterflies and birds and of the number of rare species of butterflies, although external validation results for birds based on topoclimate were poor. Many of the same variables were selected as predictors of species richness for the different taxonomic groups, but the strength of their contributions to the averaged predictions varied. The exceptions were elevation and mean topographic exposure within a 300-m radius. The internal and external validation results were consistent with previous work suggesting that elevation and measures of topographic heterogeneity are strong predictors of distribution patterns of butterflies, but are less closely associated with patterns for birds (Fleishman *et al.*, 2002; Mac Nally *et al.*, 2003).

Many vegetation and topoclimate variables that strongly influenced predictions when only environmental variables were included in models had weaker influences when indicator species also were included in models. For example, number of dead trees and mean distance to water contributed strongly [ $\Pr(\beta \neq 0) \geq 0.75$ ] to predictions of species richness of butterflies and number of rare species of butterflies when only environmental variables were included as predictors, but contributed little when indicator species were also included in models. Similarly, canopy cover and ground cover had much greater weights in predictions of bird species richness in models that excluded indicator species as candidate predictors. Together with the superior predictive performance of indicator species over environmental variables, these results suggest that in systems with a common pool of species but heterogeneity in species composition and land cover, the occurrence patterns of a small subset of an assemblage can provide more information about species richness patterns than environmental attributes.

One reason for the superior predictive performance of same-taxon indicator species may be that those species effectively integrate information about multiple physical and biological attributes across spatial and temporal scales relevant to the organisms of interest. In contrast, it may be difficult to measure directly environmental variables at scales relevant to biota, particularly for mobile and/or long-lived taxonomic groups.

Two considerations are fundamental to conservation practice: information content and cost efficiency (MacKenzie *et al.*, 2006). A predictive model may be extremely reliable, but its applicability will be limited unless measurement of the predictors is relatively fast and easy. Our results highlight the difficulty in identifying dependable predictors of the distribution of rare species. Successful occurrence models have been developed for individual rare species (e.g. Zabel *et al.*, 2003; Hoving *et al.*, 2005) (although they may not be more efficient

than conducting surveys directly for the target species), but it remains difficult to predict the number of rare species in a given location. Even when predictors that are 'good enough' can be identified, those predictors are likely to be different from the set selected as predictors of species richness of the same taxonomic group.

In our study system, indicator species were the most accurate tools for predicting species richness and number of rare species. We are not advocating the unqualified use of indicator species as surrogate measures. If experienced observers can detect all species in a focal group of plants or animals without expending substantially more effort than detecting a subset of species, there is little reason to implement surveys for indicator species. However, our statistical methods can be applied to any landscape and taxonomic group(s) to select indicator species potentially applicable across extensive regions on the basis of data that are collected from a small fraction of the total area. For relatively inaccessible geographical areas and cryptic taxonomic groups, observers can save time and money by searching for a subset of an assemblage rather than trying to survey all species in the assemblage (Gustafsson, 2000; Pressey *et al.*, 2000; Faith *et al.*, 2001). In these situations, validated predictive models based on indicator species may be a realistic method for deriving data on distribution patterns that can inform land-use planning (Hodkinson & Jackson, 2005).

Given that survey data for many regions are sparse and obtaining new data is costly, efforts to identify surrogate measures of species richness and the number of rare species are likely to remain imperative. The modelling approaches we used here allow investigators to compare the effectiveness of different types of surrogate predictor variables both singly and in combination. Promising models should be evaluated with external validation prior to use in management decisions. Reasons for failures, as well as successes, can be both practically and ecologically informative. Information on predictive accuracy then must be related to constraints experienced by practitioners. Practical efficiency must be an important criterion for deciding whether a surrogate measure is preferable to measuring the target directly.

## ACKNOWLEDGEMENTS

Thanks to C. Betrus, L. P. Bulluck, G. Horrocks, G. Markoff, N. McDonal, N. Losin, G. Johnson and M. Ban for collection of field data integral to this work and to J. Fay for technical support. Thanks to M. Betrus and J. Bulluck for field assistance and to W. Jetz and two anonymous referees for comments that improved the manuscript. Financial support was provided by the Australian Research Council, the Nevada Biodiversity Research and Conservation Initiative and the Joint Fire Sciences Program via the Rocky Mountain Research Station, US Department of Agriculture Forest Service and the Australian Research Council. This is contribution 96 from the Australian Centre for Biodiversity: Analysis, Policy and Management.

## REFERENCES

- Anderson, S.H. & Shugart, H.H. (1974) Habitat selection of breeding birds in an east Tennessee deciduous forest. *Ecology*, **55**, 828–837.
- Betrus, C.J. (2002) Refining the umbrella index complex: an application to bird and butterfly communities in montane canyons in the Great Basin. M.S. Thesis, Miami University, Oxford, OH.
- Bibby, C.J., Burgess, N.D., Hill, D.A. & Mustoe, S. (2000) *Bird census techniques*. Academic Press, London.
- Brown, J.H. (1984) On the relationship between abundance and distribution of species. *The American Naturalist*, **124**, 255–279.
- Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L. & Thomas, L. (2001) *Introduction to distance sampling: estimating abundance of biological populations*. Oxford University Press, New York.
- Burgman, M.A., Lindenmayer, D.B. & Elith, J. (2005) Managing landscapes for conservation under uncertainty. *Ecology*, **86**, 2007–2017.
- Burnham, K.P. & Anderson, D.R. (2004) Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods and Research*, **33**, 261–304.
- Chaplin, S.J., Gerrard, R.A., Watson, H.M., Master, L.L. & Flack, S.R. (2000) The geography of imperilment: targeting conservation toward critical biodiversity areas. *Precious heritage: the status of biodiversity in the United States* (ed. by B.A. Stein, L.S. Kutner and J.S. Adams), pp. 159–199. Oxford University Press, Oxford.
- Dobkin, D.S. & Rich, A.C. (1998) Comparison of line-transect, spot-map, and point-count surveys for birds in riparian areas of the Great Basin. *Journal of Field Ornithology*, **69**, 430–443.
- Dobkin, D.S. & Wilcox, B.A. (1986) Analysis of natural forest fragments: riparian birds in the Toiyabe Mountains, Nevada. *Wildlife 2000: modelling habitat relationships of terrestrial vertebrates* (ed. by J. Verner, M.L. Morrison and C.J. Ralph), pp. 293–299. University of Wisconsin Press, Madison, WI.
- Efron, B. & Tibshirani, R.J. (1993) *An introduction to the bootstrap*. Chapman and Hall, London.
- Efron, B. & Tibshirani, R.J. (1997) Improvements on cross-validation: the 632+ bootstrap method. *Journal of the American Statistical Association*, **92**, 548–560.
- Faith, D.P., Nix, H.A., Margules, C.R., Hutchinson, M.F., Walker, P.A., West, J., Stein, J., Kesteven, J.L., Allison, A. & Natera, G. (2001) The BioRap biodiversity assessment and planning study for Papua New Guinea. *Pacific Conservation Biology*, **6**, 279–288.
- Fleishman, E., Austin, G.T. & Murphy, D.D. (1997) Natural history and biogeography of the butterflies of the Toiyabe Range, Nevada (Lepidoptera: Hesperioidea and Papilionoidea). *Holarctic Lepidoptera*, **4**, 1–18.
- Fleishman, E., Fay, J.P. and Murphy, D.D. (2000) Upsides and downsides: contrasting topographic gradients in species richness and associated scenarios for climate change. *Journal of Biogeography*, **27**, 1209–1219.

- Fleishman, E., Mac Nally, R., Fay, J.P. & Murphy, D.D. (2001) Modeling and predicting species occurrence using broad-scale environmental variables: an example with butterflies of the Great Basin. *Conservation Biology*, **15**, 1674–1685.
- Fleishman, E., Betrus, C.J., Blair, R.B., Mac Nally, R. & Murphy, D.D. (2002) Nestedness analysis and conservation planning: the importance of place, environment, and life history across taxonomic groups. *Oecologia*, **133**, 78–89.
- Fleishman, E., Thomson, J.R., Mac Nally, R., Murphy, D.D. & Fay, J.P. (2005) Predicting species richness of multiple taxonomic groups using indicator species and genetic algorithms. *Conservation Biology*, **19**, 1125–1137.
- Fleishman, E., Noss, R.F. & Noon, B.R. (2006) Utility and limitations of species richness metrics for conservation planning. *Ecological Indicators*, **6**, 543–553.
- Freitag, S., van Jaarsveld, A.S. & Biggs, H.C. (1997) Ranking priority biodiversity areas: an iterative conservation value-based approach. *Biological Conservation*, **82**, 263–272.
- Friedman, J.H. (1991) Multivariate adaptive regression splines (with discussion). *Annals of Statistics*, **19**, 1–141.
- Furnival, G.M. & Wilson, R.W. (1974) Regression by leaps and bounds. *Technometrics*, **16**, 499.
- Gaston, K.J. (1994) *Rarity*. Chapman and Hall, London.
- Guisan, A. & Thuiller, W. (2005) Predicting species distributions: offering more than simple habitat models. *Ecology Letters*, **8**, 993–1009.
- 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.
- Hastie, T., Tibshirani, R., Leisch, F., Hornik, K. & Ripley, B.D. (2005) *mda: mixture and flexible discriminant analysis*, R package version 0.2–27. Available from <http://www.r-project.org> (accessed August 2005).
- Hawkins, B.A., Field, R., Cornell, H.V., Currie, D.J., Guégan, J.-F., Kaufman, D.M., Kerr, J.T., Mittelbach, G.G., Oberdorff, T., O'Brien, E.M., Porter, E.E. & Turner, J.R.G. (2003) Energy, water, and broad-scale geographic patterns of species richness. *Ecology*, **84**, 3105–3117.
- Hodkinson, I.D. & Jackson, J.K. (2005) Terrestrial and aquatic invertebrates as bioindicators for environmental monitoring, with particular reference to mountain ecosystems. *Environmental Management*, **25**, 649–666.
- Hoeting, J.A., Madigan, D., Raftery, A.E. & Volinsky, C.T. (1999) Bayesian model averaging: a tutorial. *Statistical Science*, **14**, 382–417.
- Hoving, C.L., Harrison, D.J., Krohn, W.B., Joseph, R.A. & O'Brien, M. (2005) Broad-scale predictors of Canada lynx occurrence in eastern North America. *Journal of Wildlife Management*, **69**, 739–751.
- Janzen, D.H., Ataroff, M., Farinas, M., Reyes, S., Rincón, N., Soler, A., Soriano, P. & Vera, M. (1976) Changes in the arthropod community along an elevational transect in the Venezuelan Andes. *Biotropica*, **8**, 193–203.
- Jetz, W. & Rahbek, C. (2002) Geographic range size and determinants of avian species richness. *Science*, **297**, 1548–1551.
- Landres, P.B., Verner, J. & Thomas, J.W. (1988) Ecological uses of vertebrate indicator species: a critique. *Conservation Biology*, **2**, 316–328.
- Lawler, J.J., White, D., Sifneos, J.C. & Master, L.L. (2003) Rare species and the use of indicator groups for conservation planning. *Conservation Biology*, **16**, 936–947.
- Leathwick, J.R., Rowe, D., Richardson, J., Elith, J. & Hastie, T. (2005) Using multivariate adaptive regression splines to predict the distributions of New Zealand's freshwater diadromous fish. *Freshwater Biology*, **50**, 2034–2052.
- Lee, P.-Y. & Rotenberry, J.T. (2005) Relationships between bird species and tree species assemblages in forested habitats of eastern North America. *Journal of Biogeography*, **32**, 1139–1150.
- Link, W.A. & Sauer, J.R. (1998) Estimating population change from count data: application to the North American Breeding Bird Survey. *Ecological Applications*, **8**, 258–268.
- Mac Nally, R. & Fleishman, E. (2004) A successful predictive model of species richness based on indicator species. *Conservation Biology*, **18**, 646–654.
- Mac Nally, R., Bennett, A.F. & Horrocks, G. (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., Fleishman, E., Fay, J.P. & Murphy, D.D. (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.
- Mac Nally, R., Fleishman, E., Bulluck, L. & Betrus, C. (2004) Comparative influence of spatial scale on beta diversity within regional assemblages of birds and butterflies. *Journal of Biogeography*, **31**, 917–929.
- MacArthur, R.W., Recher, H. & Cody, M.L. (1966) On the relation between habitat selection and species diversity. *The American Naturalist*, **100**, 319–332.
- MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L. & Hines, J.E. (2006) *Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence*. Elsevier, Burlington, MA.
- McCullagh, P. & Nelder, J.A. (1989) *Generalized linear models*, 2nd edn. Chapman and Hall, London.
- Moore, J.L., Balmford, A., Brooks, T., Burgess, N., Hansen, L.A., Rahbek, C. & Williams, P.H. (2003) Performance of sub-Saharan vertebrates as indicator groups for identifying priority areas for conservation. *Conservation Biology*, **17**, 207–218.
- Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B. & Kent, J. (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.
- Noon, B.R. (1981) Techniques for sampling avian habitats. *The use of multivariate statistics in studies of wildlife habitat* (ed. by D.E. Capen), USDA Forest Service General Technical Report RM-89, pp. 42–52. Rocky Mountain Forest and Range Experimental Station, Fort Collins, CO.

- Noss, R.N. & Cooperrider, A.Y. (2004) *Saving nature's legacy*. Island Press, Washington DC.
- Orme, C.D.L., Davies, R.G., Burgess, M., Eigenbrod, F., Pickup, N., Olson, V.A., Webster, A.J., Ding, T., Rasmussen, P.C., Ridgely, R.S., Stattersfield, A.J., Bennett, P.M., Blackburn, T.M., Gaston, K.J. & Owens, I.P.F. (2005) Global hotspots of species richness are not congruent with endemism or threat. *Nature*, **436**, 1016–1019.
- Patterson, B.D. & Atmar, W. (1986) Nested subsets and the structure of insular mammalian faunas and archipelagos. *Biological Journal of the Linnean Society*, **28**, 65–82.
- 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.
- Pickett, S.T.A., Cadenasso, M.L., Grove, J.M., Nilon, C.H., Pouyat, R.V., Zipperer, W.C. & Costanza, R. (2001) Urban ecological systems: linking terrestrial ecological, physical, and socioeconomic components of metropolitan areas. *Annual Review of Ecology and Systematics*, **32**, 127–157.
- Pollard, E. & Yates, T.J. (1993) *Monitoring butterflies for ecology and conservation*. Chapman and Hall, London.
- Pressey, R.L., Hager, T.C., Ryan, K.M., Schwarz, J., Wall, S., Ferrier, S. & Creaser, P.M. (2000) Using abiotic data for conservation assessments over extensive regions: quantitative methods applied across New South Wales, Australia. *Biological Conservation*, **96**, 55–82.
- Pullin, A.S. (ed.) (1995) *Ecology and conservation of butterflies*. Chapman and Hall, London.
- R Development Core Team (2005) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna [available from <http://www.r-project.org> (accessed August 2005)].
- Raftery, A.E. (1995) Bayesian model selection in social research (with discussion). *Sociological Methodology 1995* (ed. by P.V. Marsden), pp. 111–196. Blackwell, Cambridge, MA.
- Raftery, A.E. & Zheng, Y.Y. (2003) Discussion: performance of Bayesian model averaging. *Journal of the American Statistical Association*, **98**, 931–938.
- Raftery, A.E., Madigan, D. & Hoeting, J.A. (1997) Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, **92**, 179–191.
- Raftery, A.E., Hoeting, J.A., Volinsky, C.T. & Painter, I. (2005) *BMA: Bayesian model averaging*, R package version 3.0. <http://www.r-project.org>, <http://www.research.att.com/~volinsky/bma.html>.
- Robinson, S.K. & Holmes, R.T. (1984) Effects of plant species and foliage structure on the foraging behavior of forest birds. *Auk*, **101**, 672–684.
- Roughgarden, J., Running, S.W. & Matson, P.A. (1991) What does remote sensing do for ecology? *Ecology*, **76**, 1918–1922.
- Rouse, J.W., Haas, R.H., Schell, J.A. & Deering, D.W. (1973) Monitoring vegetation systems in the Great Plains with ERTS. *Proceedings of the third ERTS symposium*, pp. 309–317. National Aeronautics and Space Administration, Greenbelt, MD.
- Scott, J.A. (1986) *The butterflies of North America*. Stanford University Press, Stanford, CA.
- Scott, C.T. (1998) Sampling methods for estimating change in forest resources. *Ecological Applications*, **8**, 228–233.
- Scott, J.M., Csuti, B., Jacobi, J.D. & Estes, J.E. (1987) Species richness. *BioScience*, **37**, 782–788.
- Scott, J.M., Heglund, P.J., Morrison, M., Raphael, M., Haufler, J. & Wall, B. (eds.) (2002) *Predicting species occurrences: issues of scale and accuracy*. Island Press, Covello, CA.
- Siegel, R.B., Desante, D.F. & Nott, M.P. (2001) Using point counts to establish conservation priorities: how many visits are optimal? *Journal of Field Ornithology*, **72**, 228–235.
- Stein, B.A., Kutner, L.S. & Adams, J.S. (eds) (2000) *Precious heritage: the status of biodiversity in the United States*. Oxford University Press, New York.
- Stohlgren, T.J., Guenther, D.A., Evangelista, P.H. & Alley, N. (2005) Patterns of plant species richness, rarity, endemism, and uniqueness in an arid landscape. *Ecological Applications*, **15**, 715–725.
- Thomson, J.R., Fleishman, E., Mac Nally, R. & Dobkin, D.S. (2005) Influence of the temporal resolution of data on the success of indicator species models of species richness across multiple taxonomic groups. *Biological Conservation*, **124**, 503–518.
- Turner, W., Sterling, E.J. & Janetos, A.C. (2001) Contributions of remote sensing to biodiversity conservation: a NASA approach. *Conservation Biology*, **15**, 832–834.
- Viallefont, V., Raftery, A.E. & Richardson, S. (2001) Variable selection and Bayesian model averaging in case-control studies. *Statistics in Medicine*, **20**, 3215–3230.
- Weiss, S.B., Murphy, D.D. & White, R.R. (1988) Sun, slope, and butterflies: topographic determinants of habitat quality for *Euphydryas editha*. *Ecology*, **69**, 1486–1496.
- Wiens, J.A. & Rotenberry, J.T. (1981) Habitat associations and community structure of birds in shrubsteppe environments. *Ecological Monographs*, **51**, 21–41.
- Wintle, B.A., McCarthy, M.A., Volinsky, C.T. & Kavanagh, R.P. (2003) The use of Bayesian model averaging to better represent uncertainty in ecological models. *Conservation Biology*, **17**, 1579–1590.
- Wintle, B.A., Elith, J. & Potts, J.M. (2005) Fauna habitat modelling and mapping: a review and case study in the Lower Hunter Central Coast region of NSW. *Australian Journal of Ecology*, **30**, 719–738.
- Wright, D.H., Patterson, B.D., Mikkelsen, G.M., Cutler, A. & Atmar, W. (1998) A comparative analysis of nested subset patterns of species composition. *Oecologia*, **113**, 1–20.
- Yeoman, F. & Mac Nally, R. (2005) The avifaunas of fragmented, periurban, coastal woodlands in south-eastern Australia. *Landscape and Urban Planning*, **72**, 297–312.
- Zabel, C.J., Dunk, J.R., Stauffer, H.B., Roberts, L.M., Mulder, B.S. & Wright, A. (2003) Northern spotted owl habitat models for research and management application in California (USA). *Ecological Applications*, **13**, 1027–1040.

## BIOSKETCHES

**James Thomson** is a Research Fellow at the Australian Centre for Biodiversity: Analysis, Policy and Management, Monash University, Victoria, Australia. His interests include landscape, freshwater and conservation ecology, statistical analysis, experimental design and ecological modelling.

**Erica Fleishman** is a Senior Research Scientist at the Center for Conservation Biology, Stanford University. Her research focuses on integration of conservation science and land management, particularly in western North America.

**Ralph Mac Nally** is Professor of Ecology and Director of the Australian Centre for Biodiversity. His interests include landscape and conservation ecology, ecological spatial analysis and ecological computing and modelling.

**David Dobkin** is Executive Director of the High Desert Ecological Research Institute in Bend, Oregon, USA. He has conducted research on a wide variety of taxa in arid landscapes of western North America. His research focuses on the ecology of shrub-steppe landscapes, with emphasis on riparian bird communities and riparian restoration.

---

Editor: Walter Jetz