

Capacity of large-scale, long-term biodiversity monitoring programmes to detect trends in species prevalence

Scott E. Nielsen · Diane L. Haughland · Erin Bayne · Jim Schieck

Received: 15 April 2008 / Accepted: 16 March 2009 / Published online: 1 April 2009
© Springer Science+Business Media B.V. 2009

Abstract There is a critical need for monitoring programmes to assess change or trends in species status to inform conservation. A key aspect in developing such programmes is evaluating their statistical power—the ability to detect a real change. Here we examine the capacity of a broad-scale biodiversity monitoring programme in Alberta, Canada to measure changes in species prevalence. Using observed variation in detectability and prevalence for 252 species monitored at 85 sites, we simulated 3% annual declines and evaluated sample size (6 different sizes) and length of monitoring (5 different durations) necessary to detect change with a 90% certainty (power) at an α of 0.1. Our results suggest that after four monitoring cycles (e.g., 20 years for a 5-year cycle) a power of 90% can be expected for 99% of species when monitoring 1,625 sites, 65% of species for 300 sites, 27% of species for 75 sites, and 8% of species for 25 sites. We found that 66% detectability and 50% prevalence were needed to ensure that 3% annual change is detected at 50 sites over a 20-year period. Our results demonstrate that broad-scale monitoring programmes cannot effectively detect trends in all species at all spatial scales. The time period and spatial scale necessary to detect a real change at a specified level needs to be provided to stakeholders to ensure the short-term success of biodiversity monitoring programmes and to ensure that the most robust indicators of biodiversity are selected.

Keywords Alberta · Biodiversity monitoring · Detectability · Prevalence · Statistical power

S. E. Nielsen (✉)

Department of Renewable Resources, University of Alberta, 751 General Services Building,
Edmonton, AB T6G 2H1, Canada
e-mail: scott@ualberta.ca

D. L. Haughland · E. Bayne · J. Schieck

Department of Biological Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada

J. Schieck

Alberta Biodiversity Monitoring Institute, Alberta Research Council, P.O. Bag 4000, Vegreville,
AB T9C 1T4, Canada

Introduction

Many countries are striving to reduce the rate of biodiversity loss by 2010 (Secretariat of the Convention on Biological Diversity 2005). To achieve this target, signatories require monitoring programmes capable of measuring trends in biodiversity. However, biodiversity cannot be measured in its entirety, forcing the use of surrogates. Biodiversity monitoring sometimes relies on coarse filter surrogates, such as habitat quantity or quality, measured via remote sensing or field mapping (Duro et al. 2007; Lengyel et al. 2008a). However, habitats may persist even though key elements of biodiversity are lost (Huggard et al. 2006) making habitat an inappropriate measure of status for some species. For instance, in Africa numerous mammals were displaced through hunting activities not correlated with human density and land use, while exotic species contributed to losses of mammals in Australia (Ceballos and Ehrlich 2002). To ensure coarse filter surrogates of biodiversity do not overlook important on-the-ground changes it is necessary to monitor biota.

Approaches to monitoring biota can be categorized as either targeted or surveillance monitoring (Nichols and Williams 2006). Targeted monitoring tends to be hypothesis-driven, stressor-specific, and restricted to a few well-studied, charismatic, or rare species. Surveillance monitoring is taxonomically-broad, species-rich, and most frequently hypothesis-free (Yoccoz et al. 2001; Nichols and Williams 2006). Although targeted monitoring appears to be an attractive short-cut for indexing biodiversity condition, indicator species are rarely reliable surrogates for larger groups of species or other taxa (e.g., Prendergast et al. 1993; Simberloff 1998; Lindenmayer et al. 2002; Favreau et al. 2006). Accurate measures of biodiversity necessitate that more than a few indicator species be considered. Designing a monitoring programme to measure and detect trends for multiple species and taxa is, however, challenging. Benefits gained by monitoring large numbers of species need to be weighed against the loss of precision and accuracy associated with rapid-assessment survey protocols.

Prospective power analyses has become standard practise when determining how to allocate monitoring effort, but most published reports describe how to optimize effort for single species (Taylor and Gerrodette 1993) or a single taxon (Roy et al. 2007; Van Strien et al. 1997; Archaux and Bergès 2008; Manley et al. 2004, 2005). The design attributes that affect statistical power for a single species or taxon are the same for taxonomically-broad monitoring programs, including: number of sites monitored, duration of monitoring, process and sample variability, statistical method, choice of α and β levels, and effect size (Fox 2001; Field et al. 2004; Legg and Nagy 2006). The difference lies in the constraints faced by taxonomically-broad programmes. Monitoring effort cannot be optimized for all species simultaneously. When multi-stakeholder monitoring is designed to be large-scale and long-term, the often-prescribed solution of monitoring a few sites intensively and many sites superficially is not desirable (e.g., reduced-effort schemes, Roy et al. 2007). It is better to focus on minimizing and understanding sampling error and employing design attributes that allow the programme to make robust inferences to the population of sites, such as a random or systematic sampling design.

At present, few programmes have been implemented over long enough periods or at large enough scales to gauge their capability to identify biodiversity loss. We report here on the prospective capacity of a large-scale taxonomically-broad monitoring programme to identify trends in the prevalence of species (proportion of monitoring sites a species was detected) using a combination of real-world field data from Alberta, Canada and numerical

simulations. Species prevalence was chosen as our measure of abundance because it is simple to communicate and measure, although it may have lower statistical power than other metrics such as relative abundance or density (Purvis and Hector 2000). Our objectives were to: (1) determine the number of monitoring sites and duration of monitoring necessary to detect a 3% annual change in species prevalence; (2) evaluate the degree to which detectability and prevalence influence statistical power; (3) develop a simple predictive model to estimate how statistical power was affected by species prevalence, detectability, number of monitoring sites, number of repeated visits, and an α of 0.1; and (4) determine whether certain taxonomic groups are more effective at detecting trends than others.

Methods

Study area and monitoring design

We simulated statistical power for a large-scale, long-term biodiversity monitoring initiative recently initiated in Alberta, Canada by the Alberta Biodiversity Monitoring Institute (ABMI; Stadt et al. 2006). ABMI's monitoring design consists of 1,656 sites evenly spaced across Alberta using a 20-km grid. Site locations are permanent with exact GPS coordinates recorded, so they can be surveyed repeatedly over time. Sites have been randomly divided into five panels with each panel to be surveyed once every 5 years. During 2003–2005, prototype data for ABMI were collected at 85 of the 1,656 permanent monitoring sites, primarily in central Alberta's boreal forest (Fig. 1). Ten of the 85 sites were re-surveyed each year for 3 years to allow assessments of the reliability of monitoring protocols, including species detectability.

Survey methods

At each of the 85 monitoring sites, the presence–absence (detected/non-detected) and relative abundance of songbirds, vascular plants, and bryophytes were recorded (ABMI 2007). Songbirds were surveyed using single-visit point counts to each monitoring site during the breeding season (June). An omni-directional microphone was used to digitally record singing birds for 10 min at each of nine stations. The stations were spaced 300 m apart in a 3×3 grid. All audio recordings were interpreted by a single expert in a standardized laboratory setting. Vascular plants and bryophytes were surveyed within a 1-ha square plot that was centred on a permanent ABMI site marker. The 1-ha plot was flagged into four 0.25-ha sub-plots, and vascular plants were surveyed during July for each sub-plot using area-restricted, 20-min searches. Species not identified in the field were collected for expert identification in the laboratory. During the same visit in July, presence–absence of bryophytes was determined using time-limited searches of microhabitats. Technicians searched the 1-ha plot and created a list of all pre-defined microhabitats found (types of lowland and upland substrates, trees and stumps, downed woody material, and rocks). A technician then searched microhabitats for bryophytes over a two hour period. Samples of each moss and liverwort that appeared distinctive were collected for identification by expert bryologists in the laboratory. The cumulative list of species observed over all sub-samples (bird points, sub-plots, or microsites) was used as our measure of presence–absence for the monitoring site.

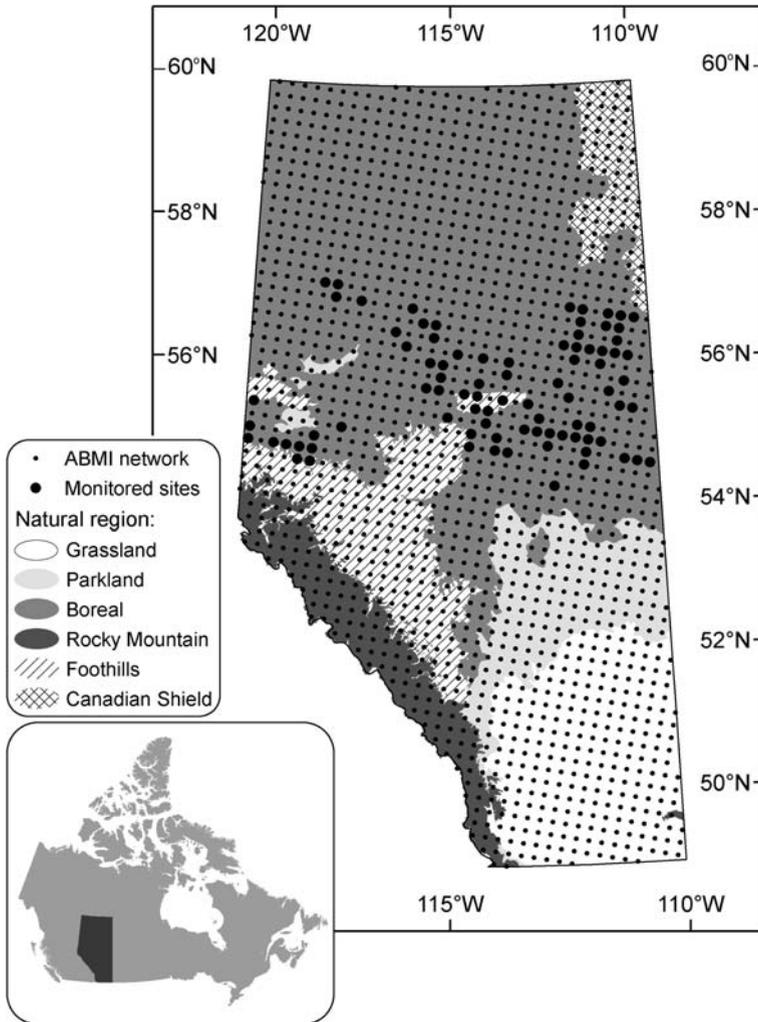


Fig. 1 Location of sites from the Alberta Biodiversity Monitoring Institute (ABMI) used to evaluate programme monitoring power. The ABMI network is a 20 km × 20 km systematic grid of permanent monitoring sites. Location of Alberta in Canada shown in the *lower left* of the figure

Simulations and analysis of trends

We simulated 3% annual declines in the prevalence of each species for those with an initial prevalence where detected of 0.1 or greater. Although 3% is arbitrary, a 3% annual decline for 10 years would result in a 27% overall decline, which was viewed as a meaningful change by managers in Alberta. The prevalence of each species at the 85 sites we sampled in ABMI was used to populate our simulation at year t . In our simulation, species prevalence was tracked each year at each site. We then “sampled” from this known population as per the systematic sampling design used by the ABMI. Declines in prevalence were

Table 1 Number of monitoring sites and panel sizes (number of monitoring sites within a panel) considered in assessments of statistical power for detecting 3% annual changes in species prevalence

Number of monitoring sites	Panel size	Area (km ²) ^a	Regional scale represented
25	5	10,000	Forest management agreement
50	10	20,000	
75	15	30,000	Natural sub-region
100	20	40,000	
300	60	120,000	Natural region
1,625	325	650,000	Province of Alberta

^a Scales represented in Alberta based on a 20-km by 20-km systematic grid are provided

modelled as a deterministic reduction in the number of sites where a species occurred by reducing the number of occupied sites by 3% each year. Sampling variation at sites where species were present was introduced to the simulations by detection probabilities (i.e. probability that a species is detected given that it is present). Detectability was estimated from the mean occurrence of a species at each of the ten sites that were re-surveyed for each of the three years. Species detectability for each site ranged from 0.33 (recorded only once during the three years) to 1 (recorded all three years). A uniform random number was generated at each step in the simulation to determine whether a species was detected at a site given that it was present. If the random number was less than the species detectability value, the simulation treated that site as an absence (zero) in subsequent analysis (i.e. false negative). As ABMI sites were permanent and protocols constant, the detectability value we used in simulations represents a combination of observer/method-associated variability. It also includes natural inter-annual variation that could have included short-term seasonal changes or year-to-year variation in occurrence or detection probability at a site.

A total of 500 simulations at 3% annual declines in species prevalence were generated for each of 252 species and 30 scenarios representing six sample sizes (25, 50, 75, 100, 300, and 1,625 sites; Table 1) and five time-horizons for monitoring (10, 20, 30, 40, and 50 years). All sample sizes and monitoring periods were divisible by 5, as Alberta protocols call for grouping sites into five panels that are each re-monitored in a single year (Table 1). Given the number of species and scenarios examined, 7,560 unique estimates of statistical power were determined based on 3,780,000 data points (500 replicates per scenario).

For each species and scenario combination we assessed the power of detecting a trend in the species prevalence when simulated using population-averaged panel-data via Generalized Estimating Equation (GEE) in STATA 9 (StataCorp 2005). GEEs allow for direct identification of data correlations associated with longitudinal measurements and observations that are clustered around a common group or panel (Liang and Zeger 1986). Each simulated monitoring site was randomly grouped into one of five panels and each panel ‘surveyed’ during a different year. This mimicked the ABMI monitoring design consisting of five rotating panels where surveys were completed once every 5 years. Year of simulation was set to identify a longitudinal data series. A binomial family and logit link GEE with an autoregressive order 1 within-group correlation structure for panels was used to account for temporal autocorrelation of grouped monitoring sites and the variable *year* tested for a trend. To efficiently estimate models and manage the results of simulations, we used the STASBY command in STATA where GEE models were repeated for each of the

500 simulations per scenario and species combination and the statistics from each model collected (saved). We recorded the number of times a trend parameter for year exceeded the critical F -value assuming an α of 0.1. Because the consequences of either declaring a spurious change significant (type I error) or overlooking a significant change (type II error) can be equally undesirable (Fairweather 1991; Mapstone 1995; but see Field et al. 2004), we balanced α (type I error) and β (type II error) by setting β at 0.1. Therefore, when 90% of simulations in a species-monitoring scenario (number of monitoring sites and length of monitoring) exceeded the critical α of 0.1, we recorded the scenario as correctly detecting a trend.

Influence of number of monitoring sites and length of monitoring on statistical power

Power of detecting change for monitoring scenarios were plotted as taxa-specific means (± 1 SD) based on sample size and length of monitoring. We also recorded the percent of species within a taxonomic group that correctly identified a trend ($\beta = 0.1$, $\alpha = 0.1$). To guide evaluation of monitoring effectiveness, we report the sample size and monitoring period necessary to correctly identify a trend for 50, 75, and 100% of monitored species.

Influence of prevalence and detectability on statistical power

Statistical power was compared among taxonomic groups (songbirds, vascular plants, and bryophytes) using a Kruskal–Wallis rank sum test for the sample size and monitoring period of 50 sites and 20 years (scales chosen at what appeared to be the most pronounced differences among taxa). Power was estimated for each scenario and species combination as the proportion of the 500 simulations exceeding the critical F -value at an α of 0.1. Using species prevalence, detectability, panel sample size (number of sites monitored each year or 20% of the total sample size), and number of monitoring cycles (years/5), we estimated a generalized linear model with a logit link and robust variance estimators (StataCorp 2005) to predict the statistical power of the monitoring programme. Variables were standardized to a mean of zero and standard deviation of one and the model re-parameterized to estimate standardized coefficients (β Z-StanVar) to evaluate the relative importance of individual predictors. Percent change in odds ratio for a standard deviation increase in covariate X (%StdX) was also used to evaluate variable contribution. Prior to model estimation, data from simulations were partitioned into a model training and model testing data set using a 75–25% allocation respectively. Training and testing sets were systematically selected based on the ranked order relative to species prevalence and detectability. Every fourth species (in ranked order) was removed from the training dataset and retained for later evaluation. Predictive accuracy of model estimates were determined by regressing observed power against predicted power for the testing (removed) dataset and the regression fit tested for a slope of 1 and intercept of 0 using an F -test (Haefner 2005).

To explore how statistical power would increase in the short-term if we accepted a greater probability of error by including rare, cryptic species, we re-ran simulations for two scenarios. The first represented power for common, easily-detected species, while the second represented power for rare, difficult to detect species. Each scenario was simulated for a 10-year period, an α and β of 0.2, and for the same six sample sizes used previously and outlined in Table 2.

Table 2 Prevalence and detectability of species used in simulations, as measured at selected ABMI monitoring sites in the boreal natural region of Alberta

Taxonomic group	Number of species	Prevalence		Detectability	
		Mean	SD	Mean	SD
Birds	62	0.41	0.25	0.59	0.20
Bryophytes	65	0.42	0.23	0.63	0.16
Vascular plants	125	0.35	0.24	0.69	0.22

Results

Monitored species

A total of 663 species (119 bird species; 166 bryophytes and 378 vascular plants) were detected at 85 monitoring sites. Bird species detected include 85% of the 87 Passeriformes known to inhabit Alberta's boreal forest (J. Schieck, unpublished data), as well as one introduced species (domestic chicken). Bryophytes detected represent 37% of 219 boreal species (R. Belland, unpublished data), while vascular plants detected represent 39% of 854 species (Moss 1994; J. Gould, unpublished data) previously recorded in Alberta's boreal. In addition to species native to the boreal, we recorded 26 vascular plant species introduced to North America and 20 plant species native to Alberta, but not previously recorded in the boreal ecoregion. We detected a total of 61 species considered to be sensitive (13 bird species), potentially at risk (1 bird species), or of conservation concern (41 bryophyte and 6 vascular plant species) (Alberta Natural Heritage Information Centre 2006a, b). Of 663 species detected, 252 were present at 10% or more of the sites. Prevalence and detectability data from these 252 species were used to examine statistical power.

Influence of number of monitoring sites and length of monitoring on statistical power

Using average statistical power by species within taxonomic groups for monitoring scenarios (number of sites and length of monitoring period), species within all three taxonomic groups achieved an average power of 90% when 1,625 sites were monitored over 20 years and for 300 sites monitored over 30 years (Fig. 2). In fact, when considering the percent of species within a taxonomic group, all but one species (a bryophyte) achieved a 90% power after 20 years of monitoring on 1,625 (provincial scale) sites with the majority of species reaching 90% power after only 10 years of monitoring (Table 3). Assuming 300 monitoring sites (the approximate scale of natural regions in Alberta), 50 years of monitoring resulted in at least 90% power for all sampled species. More than 75% of the species reached 90% power within 30 years and the majority of species achieved this target after 20 years (Fig. 2). We found that 50 years of monitoring were required to achieve an average of 90% power if 75 or fewer sites were monitored for songbirds and vascular plants and if 50 or fewer sites were monitored for bryophytes (Fig. 2). The majority of species reached 90% power after 40 years of monitoring at 75 sites and 50 years of monitoring at 50 sites (Table 3). When monitoring was restricted to 25 sites (the approximate scale of forest management areas in Alberta), about one-third of all species tested reached 90% power after 50 years and about one-tenth reached 90% power after 20 years of monitoring (Table 3).

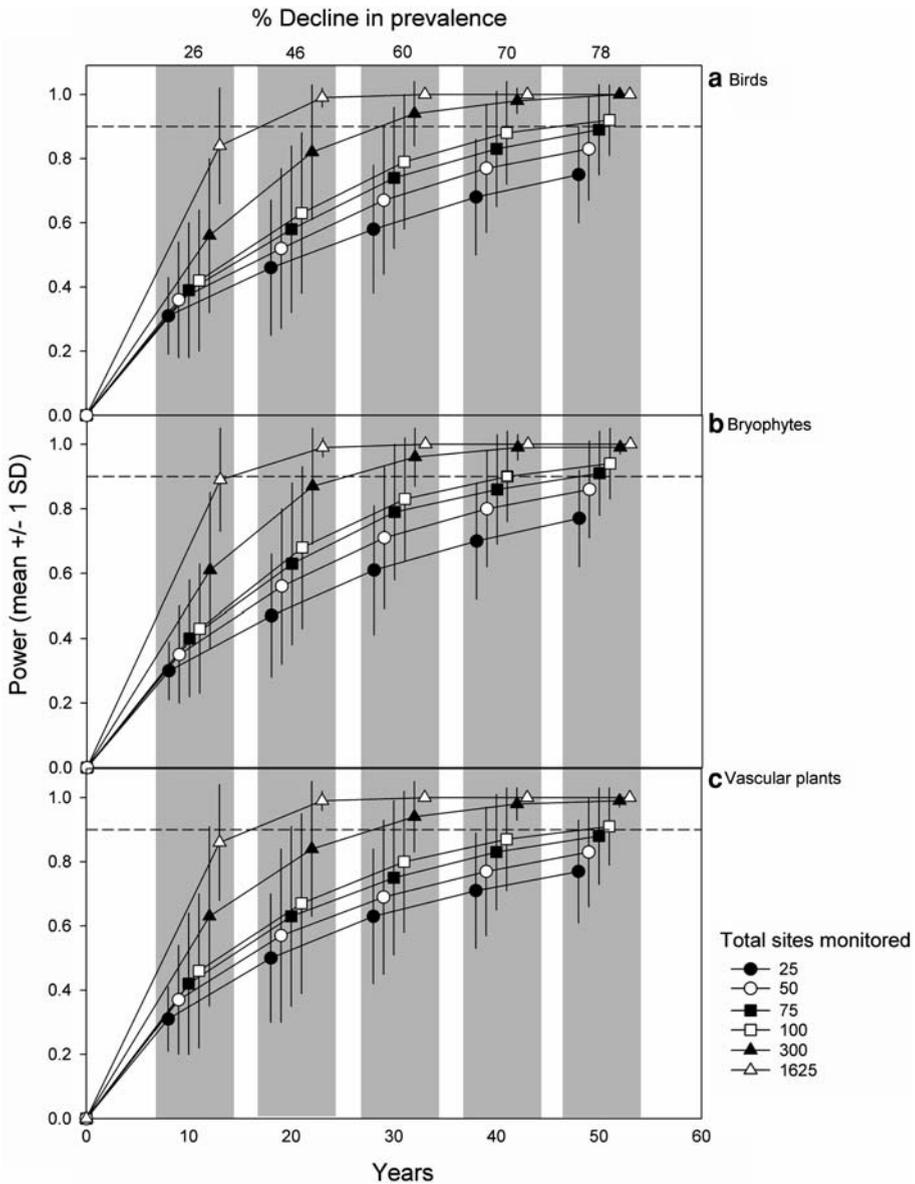


Fig. 2 Power curves by taxonomic group (**a** birds, **b** bryophytes, **c** vascular plants) for six sample sizes according to years of monitoring. Percent decline in prevalence for each time period based on a 3% annual decline is illustrated and the results for a specific length of monitoring are staggered for ease of visualization

Influence of prevalence and detectability on statistical power

Assuming 50 monitoring sites and a 20-year period of monitoring, no difference in statistical power was evident among taxonomic groups ($\chi^2 = 1.11, df = 2, P = 0.575$). Prevalence and detectability both influenced statistical power, although the influence of prevalence was

Table 3 Percent of species where 3% annual declines were detected at a power of 90% ($\beta = 0.1$) and an $\alpha = 0.1$ for combinations of sample size (number of monitoring sites) and period of monitoring (years and number of re-visits) by taxonomic group

Taxa	Years	Number of re-visits	Number of monitoring sites (Number per panel)					
			25 (5)	50 (10)	75 (15)	100 (20)	300 (60)	1625 (325)
Birds (62 spp.)	10	2	0	3	6	13	18	62
	20	4	11	18	21	27	64	100
	30	6	16	29	39	55	80	100
	40	8	21	42	60	66	97	100
	50	10	31	58	69	77	100	100
Bryophytes (65 spp.)	10	2	0	2	2	6	23	70
	20	4	6	18	23	40	71	98
	30	6	18	37	48	60	91	100
	40	8	23	48	68	71	97	100
	50	10	37	66	74	85	100	100
Vascular plants (125 spp.)	10	2	0	2	5	12	34	66
	20	4	8	28	32	38	62	100
	30	6	26	36	46	54	82	100
	40	8	30	45	56	63	95	100
	50	10	32	53	65	76	100	100
			10000	20000	30000	40000	120000	650000
			Spatial scale (km ²) at 20-km spacing					

Light and dark grey shading represents scenarios where at least 50 and 75% of species respectively reached desired power. A box outlines those scenarios that reached an optimal success of 100% of species

dependent on the level of detectability. At 60% detectability or less, prevalence had little effect on statistical power (Fig. 3). However, at high levels of detectability (i.e. >60%), power was positively related to species prevalence, especially when prevalence ranged between 0.4 and 0.6 (Fig. 3). All taxa had species with high prevalence and detectability making no particular group more statistically powerful than another.

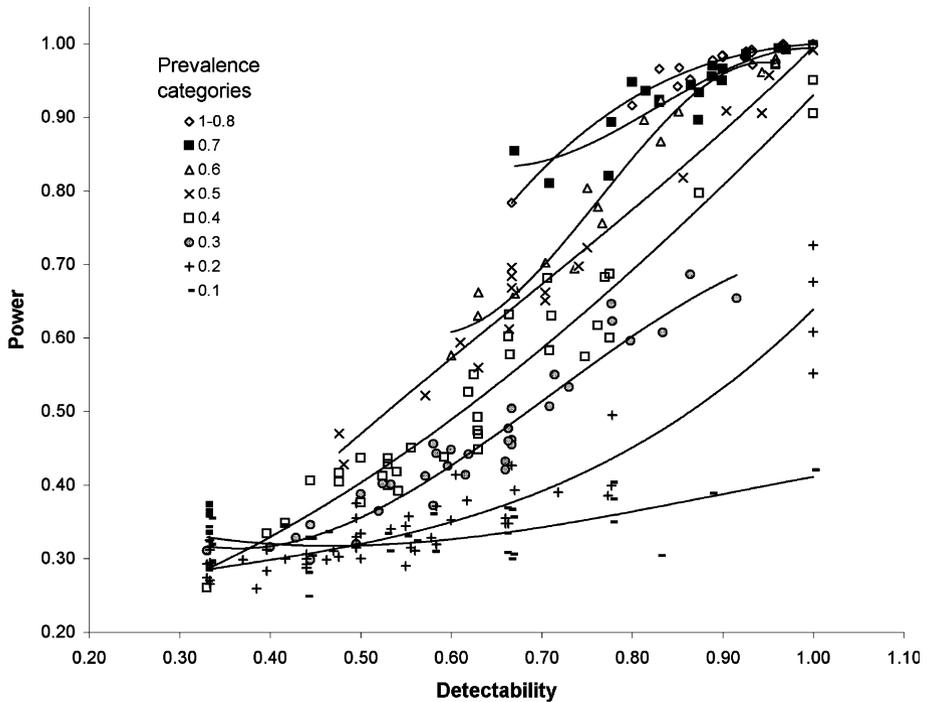


Fig. 3 Relationship between statistical power and species prevalence and detectability for a 3% annual decline at 50 monitoring sites and for 20 years of monitoring

Predicting statistical power for monitoring changes in species prevalence

The model that best described statistical power for monitoring changes in species prevalence included number of monitoring sites per panel, length of monitoring period (number of monitoring cycles), prevalence of the species, and detectability. The model accounted for 87% of the deviance in the original data (r^2 calculated following Menard (2000), Table 4). As would be expected, statistical power increased in a non-linear manner as number of monitoring sites, length of monitoring, prevalence, and detectability increased. Interaction terms among detectability and prevalence and number of monitoring sites per panel and the number of monitoring cycles were also important determinants of statistical power (Table 4). Based on Z-standardized β coefficients and the percent change in odds ratio for a standard deviation increase in each variable (Table 4), the factors most influencing on statistical power were programme design variables with the most important being an interactive effect between the number of monitoring sites and length of monitoring (number of monitoring cycles) followed by the individual factors of number of monitoring sites and length of monitoring. For example, even if detectability and prevalence were at unity, if monitoring was restricted to five sites per panel and two monitoring cycles (10 years), power reached a maximum of 0.88. However, even with minimal detectability and prevalence (0.33 and 0.1, respectively), the desired power (0.9) was exceeded when monitoring 325 sites per panel for three monitoring cycles (15 years). The most important species variables were the interaction between prevalence and detectability (Table 4).

Table 4 Estimated coefficients (β_i), standard errors (SE), P values (Wald z statistic), percent change in odds ratio per unit increase in covariate X (%), coefficients for Z-standardised variables (β Z-Std Var), percent change in odds ratio for a standard deviation increase in covariate X (%Std X), and standard deviation of X (SD of X) for estimating the power of a monitoring programme ($\alpha = 0.05$, $df = 5,665$, 190 clusters, $r^2 = 0.87$)

Variable	Unstandardized variables				Standardized variables		
	β	SE	P	(%)	β Z-Std Var	%Std X	SD of X
Monitoring design variables							
Number of monitoring sites per panel	0.025	0.002	<0.001	37.7	2.85	1,624	113.2
Number of monitoring cycles	0.624	0.035	<0.001	68.7	1.77	485	2.8
Number of monitoring sites per panel squared	-0.770 ^a	0.048 ^a	<0.001	36.8	-2.97	-94.9	38,660
Number of monitoring cycles squared	-0.027	0.002	<0.001	35.8	-0.95	-61.3	34.6
Species variables							
Prevalence	3.036	0.18	<0.001	766	0.7	101.4	0.231
Detectability	0.273	0.214	0.202	48.3	0.06	5.7	0.204
Prevalence squared	-3.432	0.304	<0.001	1.2	-0.74	-52.3	0.216
Detectability squared	1.035	0.189	<0.001	104	0.28	32.5	0.272
Interaction terms							
Detectability and prevalence	4.314	0.339	<0.001	2,751	1	172.4	0.232
Number of monitoring sites per panel and number of monitoring cycles	0.508 ^b	0.066 ^b	<0.001	37	3.95	5,103	777.5
Constant	-4.534	0.161	<0.001	0.4	2.89		

^a Coefficients and SE 10,000 times original value

^b Coefficients and SE 100 times original value

There was high correlation ($r^2 = 0.91$) between predicted and observed power for species withheld during model building (25% of total), indicating excellent model prediction. However, the slope and intercept between predicted and observed power differed from 1 and 0 respectively ($F = 52.88 > F_{2,1853} = 2.99$), suggesting slight biases in predictions over some ranges of data. This may reflect an artifact of non-converged GEE models for scenarios with small sample size and short monitoring periods on species with low prevalence and detectability. The consequence was an over-estimate of statistical power for these scenario combinations.

Influence of α and β

For rare, cryptic (low detectability) species, power to detect change over a 10-year monitoring period increased by approximately 10%, when α was doubled to 0.2. This increase in power was consistent regardless of number of monitoring sites assessed (Fig. 4). Power reached a maximum of 0.6 ($\beta \sim 0.4$) when the number of monitoring sites was at 1,625. Minimum β was therefore approximately twice as large as α for rare species at the scale equal to the province of Alberta. For common, easily detected species, a larger α also resulted in increased power, although the difference between the two levels of α diminished as power approached 100% (Fig. 4). Using an error probability of 0.1, 60 sites per panel

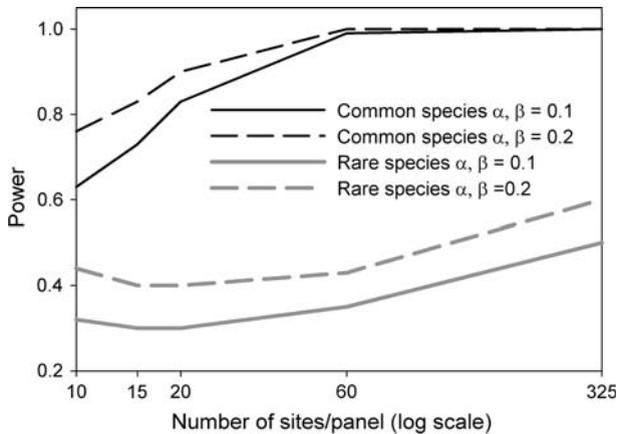


Fig. 4 Power to detect change over time for $\alpha = \beta$ at 0.1 and 0.2. These analyses were conducted for a common, easily-detected species (prevalence = 0.74, detectability = 0.89) and a rare, difficult to detect species (prevalence = 0.13, detectability = 0.33)

were required to reach a power of 90% for common species, while 10–15 sites per panel were required to reach a power of 80% when using an α of 0.2.

Discussion

Monitoring programmes having the goal of detecting trends or changes in species populations have traditionally focused on measures of local abundance (count-based monitoring), with area-occupied and presence–absence monitoring only more recently gaining popularity (Marsh and Trenham 2008). Despite a reduction in local information, presence–absence is viewed by many as a good indicator of change in population size and species range (Gaston 1994), especially in heterogeneous environments (Karr and Chu 1999). In contrast to count-based (abundance) monitoring of a few select species or single taxonomic group, much less is known about the effectiveness of monitoring trends in presence–absence for taxonomically-broad monitoring programmes that have numerous species within numerous taxonomic groups. In Alberta, Canada, a large-scale (1,656 sites with 20-km spacing), long-term (100-year) systematic biodiversity monitoring programme managed by ABMI (Stadt et al. 2006) is attempting to detect 3% annual declines in species prevalence (presence–absence across multiple sites) over a 20-year period and ideally at spatial scales as small as forest management areas.

Based on our simulations and sampling design, we found that 3% annual declines were detected for all but 1 of 252 species over a 20-year period of monitoring at a provincial scale (1,656 sites) and that declines could be detected for the majority of species (65%) at the level of a natural region (300 sites). However, at a forest management scale (25 sites) short-term trends were detected for only 8% of the species examined (species with a prevalence of 10% or more). Because our measure of detectability was across years, not within a season, change may have occurred during the period of monitoring making our estimates conservative. Detectability was also based on only ten permanent monitoring sites surveyed each year over a three year period resulting in detectability rates of species for any one site being either 33%, 66%, or 100% (low, moderate, or high). In contrast to

detectability, prevalence was based on 85 monitoring sites where surveys were completed in at least one of the three years. Since our intent was to evaluate how changes in monitoring design variables (number of monitoring sites and duration of monitoring) and species variables (detectability and prevalence) affected statistical power for a long-term monitoring programme using species parameterized by best-available data, we do not see these limitations as problematic. Indeed, prevalence and detectability could have been randomly assigned to simulated species using some pre-defined distribution. This, however, would have compromised our goals of determining whether differences existed among taxonomic groups and whether the ABMI programme had sufficient power to detect short-term trends at local scales.

Our results are similarly to those of Manley et al. (2004, 2005) who detected ($\alpha = 0.2$, power = 0.8) a 20% one-time change in species occurrence for 66% of Lake Tahoe vertebrates at $\sim 2,760$ sites using the United States Forest Service Multiple Species Inventory and Monitoring protocol. In the short-term, statistical power for Manley et al. (2004) was slightly higher than what we found due to the large number of sites surveyed (higher grid density) and higher α . Choice of α is arbitrary. An a priori discussion of what should be deemed a statistically significant trend needs to be agreed upon prior to implementation of a biodiversity monitoring programme. Grid density, on the other hand, can be modified to improve short-term detection of trends. ABMI has a low systematic grid density (20-km spacing) and although it would be impractical to increase density for the entire Province due to monitoring costs and logistics, certain regions of the Province that are more threatened could be delineated and density of sites increased to improve short-term detection of biodiversity trends.

Indeed, one of the strengths of a systematic sampling design is that users can draw inferences about trends at different spatial scales. However, as the spatial scale decreases, so does the power to detect trends over short periods of monitoring. Prevalence of the species is also important. Detecting trends in rare species, especially in localized regions, is more difficult than for common species, resulting in fewer short-term indicators. We suggest that detectability needs to be at least 66% and species prevalence above 50% to detect short-term (~ 20 year) changes at relatively few sites (~ 25 sites). Using this rule-of-thumb, 11% of the 252 species evaluated (all common species) would be considered candidates for monitoring short-term trends in a localized region (i.e., a forest management area of Alberta). No single taxon (songbirds, bryophytes, and vascular plants) had a statistical advantage for detecting change. Assuming similar rates and encompassing all taxa included in ABMI (vascular plants, mosses, fungi, lichens, phytoplankton, birds, mammals, fish, springtails, mites, zooplankton, and benthic invertebrates), we estimate that ~ 200 species could be used as short-term indicators for regional assessments. For some taxa where sub-samples are collected at a site, statistical power would likely improve if sub-samples were used as units of replication (with a random effect for site).

One possible solution for monitoring short-term trends in rare species is to extrapolate trends from common species to rare species who share ecological similarities or vulnerabilities (Edwards et al. 2004; MacKenzie et al. 2005a). However, if the emphasis of the monitoring programme is ecosystem-wide, long-term biodiversity monitoring, then taxonomically-broad surveillance techniques that emphasize common species should be supported. Monitoring common species does not presently fit the priorities of most conservation groups or existing monitoring programmes (for instance the Red List Index by Butchart et al. 2005). However, recent work suggests that even small proportional declines in common species can result in significant modifications to ecosystem structure, function, and services (Gaston and Fuller 2008). We suggest that species-based monitoring

programmes need to include the complete range of species, including common species, and that long-term, ecosystem-wide monitoring cannot be used as a replacement for existing monitoring initiatives focused on rare species (e.g., Canada's Species At Risk Act) or specific management concerns.

Regardless of the prevalence of species being monitored, modifications can be made to the monitoring and analytical designs of the programme to enhance statistical power. This may include one or more of the following changes: (1) reduce observer error through staff training and modify survey protocols to increase detection probabilities; (2) account for detection probabilities of less than one through statistical analyses; (3) augment the traditional rotating panel survey design with monitoring sites that are visited repeatedly to account for natural annual variation; (4) stratify monitoring sites to better account for variance; (5) use community-level information to evaluate trends rather than analyses of individual species; and (6) integrate raw information or estimates from other monitoring programmes. Below we discuss possible advantages of each modification and practical challenges for implementation in a monitoring programme like that of the ABMI.

If a species has low detectability, it may be possible to modify survey methods to improve detectability and consequently increase power. Methods to reduce observer error will also boost detectability and increase statistical power (Strand 1996; Thompson and Mapstone 1997; Lotz and Allen 2007). For example, better training of field staff may be a cost-effective approach to minimizing observer error and increasing detectability. Gains from increased training however, may be limited for some taxa. For instance, even highly trained experts failed to detect 33% of lichen species that were known to occur at a site (McCune et al. 1997). Reducing observer error through modifications of survey methods or increased staff training should clearly be a priority for monitoring programmes, but should not be considered the only solution for monitoring species with low detectability.

In addition to increased training and modification of survey methods to reduce observer error, statistical approaches can be used to account for low species detectability when sites are resurveyed within a single monitoring period (MacKenzie et al. 2005b). However, revisiting sites on a different date would substantially increase monitoring costs (double the cost for one revisit and triple the cost for two revisits) making these methods improbable for large-scale monitoring programmes that occur in remote locations such as the ABMI programme. Reduced-effort schemes that select only a sub-set of sites for intense, repeated sampling (Roy et al. 2007) may provide one compromise assuming that intensely-sampled sites were representative of reduced-effort sites. Repeat visits, however, assume a "closed" status where occupancy does not change between repeated surveys (MacKenzie et al. 2005b). For sedentary species (i.e. vascular plants), increasing the sample effort (i.e. time spent surveying) to enhance detectability will often be more cost-effective than returning to a site on a different date because of travel costs (Drapeau et al. 1999). In some cases, re-sampling existing data may act as a surrogate for repeated visits, although this changes the definition of the spatial and/or temporal scale of sampling. For instance, bird songs are permanently recorded as 10-min surveys in the ABMI programme allowing re-sampling of the surveys into sub-samples for estimation of detection probabilities and corrections of occupancy at a site. For other taxa, use of multiple, independent observers during a single survey visit may prove to be a cost-effective solution to estimating detection probabilities.

Another approach to boosting statistical power of monitoring programmes that are based on rotating panel or serially alternating monitoring designs is to incorporate an augmented design. In augmented designs, a select set of sites (typically 20–50% of total) are repeatedly visited during a monitoring cycle (either in consecutive years for multiple panels or one panel selected for annual surveys) to measure and account for natural annual

variation. By accounting for annual variation in augmented designs, trends can be documented more effectively for a set number of monitoring sites than if the same effort was spent visiting new sites (Urquhart and Kincaid 1999). Since in our example the number of monitoring sites was fixed and the detectability and prevalence of species was critical to short-term detection of trends, we expect that an augmented rotating panel design will result in only small gains in power, yet substantially increase monitoring costs. Our expectation has not been tested.

When there is *a priori* knowledge about a species preferred habitats, stratified sampling that emphasizes important habitats can be employed to increase statistical power at any given level of effort. Rare species are often associated with particular habitats making stratified surveys especially effective for these species. However, stratified designs are problematic for long-term biodiversity monitoring programmes, since stratification must remain constant in the presence of natural and anthropogenic-induced ecosystem change (climate change, natural disturbances, etc.). In the presence of these changes, the extent and location of strata will change resulting in the loss of initially optimized statistical power. Although a systematic sampling grid may not initially be as optimal for detecting trends in some species as compared to a stratified design, the systematic monitoring design will not be dependent on initial conditions and will maintain its power over time. Systematic designs are therefore often favoured over stratified designs for long-term monitoring initiatives.

An alternative to analysis of trends for individual species is the integration of analyses among species. Relationships between biota and environmental conditions often are stronger for community-level analyses than for single species analyses (e.g., McCune et al. 1997) and the inclusion of additional species strengthens relationships for single species analyses (Plattner et al. 2004; Clarke and Murphy 2006). Metrics of species intactness (Buckland et al. 2005; Nielsen et al. 2007) can also be considered. For example, Lamb et al. (2009) found species intactness to be more effective for detecting trends than common measures of community diversity (i.e. Shannon or Simpson diversity) or multivariate analyses (i.e. Mantel test) of community change. As such, community analyses that build upon statistical models of intactness are expected to better elucidate changes for rare species than individual analyses of species using prevalence.

Statistical power of monitoring programmes can also be enhanced by integrating raw information or parameter estimates from other monitoring programmes (Henry et al. 2008; Lengyel et al. 2008b). Analytical techniques, such as meta-analysis or weighted analysis, can be used to combine information from other monitoring programmes allowing for more immediate assessments of biodiversity change (Henry et al. 2008). Critical to this integration is clear definitions of survey methods. In the long-term, creation of a standardized, international monitoring network or a common set of biodiversity monitoring protocols is needed to facilitate direct comparisons and reporting of biodiversity trends (Henry et al. 2008; Lengyel et al. 2008b).

In summary, we advocate that biodiversity monitoring programmes include a continuum of species from rare to common. The inclusion of rare species will build support for the monitoring programme because loss of rare species is easily understood by the general public (Biggs 2000). However, large-scale biodiversity monitoring programmes can not be expected to be a replacement for existing monitoring programmes that focus on rare species, since detectability and observer error from more general survey methods are likely to compromise detection of trends. Furthermore, emphasis on rare species may be misplaced ecologically since even small reductions in abundance of common species will have

profound effects on ecosystems (Gaston and Fuller 2008), making such species important components of monitoring programmes. Inclusion of common species should also increase statistical power of the programme and facilitate community-based metrics of biodiversity. We think that taxonomically-broad monitoring programmes will be of great value to managers and increase in value with time. For instance, it will be possible to evaluate the effects of climate change on species distribution (geographic range) using information from long-term, large-scale monitoring programmes that measure a broad spectrum of species. Taxonomically-broad monitoring is also more likely to facilitate a proactive monitoring and management approach to conservation by recognizing trends early on and thereby maintaining common species while being common, as opposed to a reactive monitoring and management paradigm that focuses on rescuing rare species that have already declined markedly.

References

- Alberta Biodiversity Monitoring Institute (2007) Terrestrial field data collection protocols, Version 2.1. Alberta Biodiversity Monitoring Program, Alberta, Canada. Available via DIALOG <http://www.abmi.ca>. Cited 24 April 2007
- Alberta Natural Heritage Information Centre (2006a) Bird tracking list. Available via DIALOG <http://tprc.alberta.ca/parks/heritageinfocentre/default.aspx>. Cited 1 December 2007
- Alberta Natural Heritage Information Centre (2006b) Plant tracking and watch lists. Available via DIALOG <http://tprc.alberta.ca/parks/heritageinfocentre/default.aspx>. Cited 1 December 2007
- Archaux F, Bergès L (2008) Optimising vegetation monitoring. A case study in a French lowland forest. *Environ Monit Assess* 141:19–25. doi:10.1007/s10661-007-9874-0
- Biggs BG (2000) What is significant—Wollemi pine or the southern rushes? *Ann Mo Bot Gard* 87(1):72–80. doi:10.2307/2666210
- Buckland ST, Magurran AE, Green RE, Fewster RM (2005) Monitoring change in biodiversity through composite indices. *Philos Trans R Soc B-Biol Sci* 360(1454):243–254
- Butchart SHM, Stattersfield AJ, Baillie J, Bennun LA, Stuart SN, Akcakaya HR, Hilton-Taylor C, Mace GM (2005) Using Red List Indices to measure progress towards the 2010 target and beyond. *Philos Trans R Soc B-Biol Sci* 360(1454):255–268
- Ceballos G, Ehrlich PR (2002) Mammal population losses and the extinction crisis. *Science* 296(5569):904–907. doi:10.1126/science.1069349
- Clarke RT, Murphy JF (2006) Effects of locally rare taxa on the precision and sensitivity of RIVPACS bioassessment of freshwaters. *Freshw Biol* 51(10):1924–1940. doi:10.1111/j.1365-2427.2006.01611.x
- Drapeau P, Leduc A, McNeil R (1999) Refining the use of point counts at the scale of individual points in studies of bird-habitat. *J Avian Biol* 30(4):367–382. doi:10.2307/3677009
- Duro D, Coops NC, Wulder MA, Han T (2007) Development of a large area biodiversity monitoring system driven by remote sensing. *Prog Phys Geogr* 31(3):235–260. doi:10.1177/0309133307079054
- Edwards TC, Cutler DR, Geiser L, Alegria J, McKenzie D (2004) Assessing rarity of species with low detectability: lichens in Pacific Northwest forests. *Ecol Appl* 14(2):414–424. doi:10.1890/02-5236
- Fairweather PG (1991) Statistical power and design requirements for environmental monitoring. *Aust J Mar Freshwater Res* 42(5):555–567. doi:10.1071/MF9910555
- Favreau JM, Drew CA, Hess GR, Rubino MJ, Koch FH, Eschelbach KA (2006) Recommendations for assessing the effectiveness of surrogate species approaches. *Biodivers Conserv* 15(12):3949–3969. doi:10.1007/s10531-005-2631-1
- Field SA, Tyre AJ, Jonzen N, Rhodes JR, Possingham HP (2004) Minimizing the cost of environmental management decisions by optimizing statistical thresholds. *Ecol Lett* 7(8):669–675. doi:10.1111/j.1461-0248.2004.00625.x
- Fox DR (2001) Environmental power analysis—a new perspective. *Environmetrics* 12(5):437–449. doi:10.1002/env.470
- Gaston KJ (1994) *Rarity*. Chapman & Hall, London
- Gaston KJ, Fuller RA (2008) Commonness, population depletion and conservation biology. *Trends Ecol Evol* 23(1):14–19. doi:10.1016/j.tree.2007.11.001
- Haefner JW (2005) *Modeling biological systems*, 2nd edn. Springer, New York

- Henry PY, Lengyel S, Nowicki P, Julliard R, Clobert J, Čelik T, Gruber B, Schmeller DS, Babij V, Henle K (2008) Integrating ongoing biodiversity monitoring: potential benefits and methods. *Biodivers Conserv* 17(14):3357–3382. doi:[10.1007/s10531-008-9417-1](https://doi.org/10.1007/s10531-008-9417-1)
- Huggard DJ, Dunsworth GB, Herbers JR, Klenner W, Kremsater LL, Serrouya R (2006) Monitoring ecological representation in currently non-harvestable areas: four British Columbia case studies. *For Chron* 82(3):383–394
- Karr JR, Chu EW (1999) *Restoring life in running waters: better biological monitoring*. Island Press, Washington, DC
- Lamb EG, Bayne E, Holloway G, Schieck J, Boutin S, Herbers J, Haughland DL (2009) Indices for monitoring biodiversity change: are some more effective than others? *Ecol Indicators* 9(3):432–444. doi:[10.1016/j.ecolind.2008.06.001](https://doi.org/10.1016/j.ecolind.2008.06.001)
- Legg CJ, Nagy L (2006) Why most conservation monitoring is, but need not be, a waste of time. *J Environ Manag* 78(2):194–199. doi:[10.1016/j.jenvman.2005.04.016](https://doi.org/10.1016/j.jenvman.2005.04.016)
- Lengyel S, Deri E, Varga Z, Horvath R, Tothmeresz B, Henry PY, Kobler A, Kutnar L, Babij V, Seliskar A, Christia C, Papastergiadou E, Gruber B, Henle K (2008a) Habitat monitoring in Europe: a description of current practices. *Biodivers Conserv* 17(14):3327–3339. doi:[10.1007/s10531-008-9395-3](https://doi.org/10.1007/s10531-008-9395-3)
- Lengyel S, Kobler A, Kutnar L, Framstad E, Henry PY, Babij V, Gruber B, Schmeller D, Henle K (2008b) A review and framework for the integration of biodiversity monitoring at the habitat level. *Biodivers Conserv* 17(14):3341–3356. doi:[10.1007/s10531-008-9359-7](https://doi.org/10.1007/s10531-008-9359-7)
- Liang KY, Zeger SL (1986) Longitudinal data analysis using generalized linear models. *Biometrika* 73(1):13–22. doi:[10.1093/biomet/73.1.13](https://doi.org/10.1093/biomet/73.1.13)
- Lindenmayer DB, Manning A, Smith PL, Possingham HP, Fischer J, Oliver I, McCarthy MA (2002) The focal species approach and landscape restoration: a critique. *Conserv Biol* 16(2):338–345. doi:[10.1046/j.1523-1739.2002.00450.x](https://doi.org/10.1046/j.1523-1739.2002.00450.x)
- Lotz A, Allen CR (2007) Observer bias in anuran call surveys. *J Wildl Manag* 71(2):675–679. doi:[10.2193/2005-759](https://doi.org/10.2193/2005-759)
- MacKenzie DI, Nichols JD, Sutton N, Kawanishi K, Bailey LL (2005a) Improving inferences in population studies of rare species that are detected imperfectly. *Ecology* 86(5):1101–1113. doi:[10.1890/04-1060](https://doi.org/10.1890/04-1060)
- MacKenzie DI, Nichols JD, Pollock KH, Royle JA, Bailey LL, Hines JE (2005b) *Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence*. Academic Press, Amsterdam
- Manley PN, Zielinski WJ, Schlesinger MD, Mori S (2004) Evaluation of a multiple-species approach to monitoring species at the ecoregional scale. *Ecol Appl* 14(1):296–310. doi:[10.1890/02-5249](https://doi.org/10.1890/02-5249)
- Manley PN, Schlesinger MD, Roth JK, Van Horne B (2005) A field-based evaluation of a presence-absence protocol for monitoring ecoregional-scale biodiversity. *J Wildl Manag* 69(3):950–966. doi:[10.2193/0022-541X\(2005\)069\[0950:AFEOAP\]2.0.CO;2](https://doi.org/10.2193/0022-541X(2005)069[0950:AFEOAP]2.0.CO;2)
- Mapstone BD (1995) Scalable decision rules for environmental-impact studies—effect size, Type-I, and Type-II errors. *Ecol Appl* 5(2):401–410. doi:[10.2307/1942031](https://doi.org/10.2307/1942031)
- Marsh DM, Trenham PC (2008) Current trends in plant and animal population monitoring. *Conserv Biol* 22(3):647–655. doi:[10.1111/j.1523-1739.2008.00927.x](https://doi.org/10.1111/j.1523-1739.2008.00927.x)
- McCune B, Dey JP, Peck JE, Cassell D, Heiman K, Will-Wolf S, Neitlich PN (1997) Repeatability of community data: species richness versus gradient scores in large-scale lichen studies. *Bryologist* 100(1):40–46
- Menard S (2000) Coefficients of determination for multiple logistic regression. *Am Stat* 54(1):17–24. doi:[10.2307/2685605](https://doi.org/10.2307/2685605)
- Moss EH (1994) *Flora of Alberta*. In: Packer JG (ed), 2nd edn. University of Toronto Press, Toronto
- Nichols JD, Williams BK (2006) Monitoring for conservation. *Trends Ecol Evol* 21(12):668–673. doi:[10.1016/j.tree.2006.08.007](https://doi.org/10.1016/j.tree.2006.08.007)
- Nielsen SE, Bayne EM, Schieck J, Herbers J, Boutin S (2007) A new method to estimate species and biodiversity intactness using empirically derived reference conditions. *Biol Conserv* 137(3):403–414. doi:[10.1016/j.biocon.2007.02.024](https://doi.org/10.1016/j.biocon.2007.02.024)
- Plattner M, Birrer S, Weber D (2004) Data quality in monitoring plant species richness in Switzerland. *Community Ecol* 5(1):135–143. doi:[10.1556/ComEc.5.2004.1.13](https://doi.org/10.1556/ComEc.5.2004.1.13)
- Prendergast JR, Quinn RM, Lawton JH, Eversham BC, Gibbons DW (1993) Rare species, the coincidence of diversity hotspots and conservation strategies. *Nature* 365(6444):335–337. doi:[10.1038/365335a0](https://doi.org/10.1038/365335a0)
- Purvis A, Hector A (2000) Getting the measure of biodiversity. *Nature* 405(6783):212–219. doi:[10.1038/35012221](https://doi.org/10.1038/35012221)
- Roy DB, Rothery P, Brereton T (2007) Reduced-effort schemes for monitoring butterfly populations. *J Appl Ecol* 44(5):993–1000. doi:[10.1111/j.1365-2664.2007.01340.x](https://doi.org/10.1111/j.1365-2664.2007.01340.x)
- Secretariat of the Convention on Biological Diversity (2005) *Handbook of the convention on biological diversity including its cartagena protocol on biosafety*, 3rd edn. UNEP, Montreal

- Simberloff DA (1998) Flagships, umbrellas, and keystones: is single-species management passé in the landscape era? *Biol Conserv* 83(3):247–257. doi:[10.1016/S0006-3207\(97\)00081-5](https://doi.org/10.1016/S0006-3207(97)00081-5)
- Stadt JJ, Schieck J, Stelfox HA (2006) Alberta biodiversity monitoring program—monitoring effectiveness of sustainable forest management planning. *Environ Monit Assess* 121(1):33–46. doi:[10.1007/s10661-005-9075-7](https://doi.org/10.1007/s10661-005-9075-7)
- StataCorp (2005) Stata statistical software: release 9. StataCorp, College Station
- Strand GH (1996) Detection of observer bias in ongoing forest health monitoring programmes. *Can J Res* 26(9):1692–1696. doi:[10.1139/x26-191](https://doi.org/10.1139/x26-191)
- Taylor BL, Gerrodette T (1993) The uses of statistical power in conservation biology: the Vaquita and Northern Spotted Owl. *Conserv Biol* 7(3):489–500. doi:[10.1046/j.1523-1739.1993.07030489.x](https://doi.org/10.1046/j.1523-1739.1993.07030489.x)
- Thompson AA, Mapstone BD (1997) Observer effects and training in underwater visual surveys of reef fishes. *Mar Ecol Prog Ser* 154:53–63. doi:[10.3354/meps154053](https://doi.org/10.3354/meps154053)
- Urquhart NS, Kincaid TM (1999) Designs for detecting trend from repeated surveys of ecological resources. *J Agric Biol Environ Stat* 4(4):404–414
- Van Strien AJ, Vande Pavert R, Moss D, Yates TJ, Van Swaay CAM, Vos P (1997) The statistical power of two butterfly monitoring schemes to detect trends. *J Appl Ecol* 34(3):817–828. doi:[10.2307/2404926](https://doi.org/10.2307/2404926)
- Yoccoz NG, Nichols JD, Boulinier T (2001) Monitoring of biological diversity in space and time. *Trends Ecol Evol* 16(8):446–453. doi:[10.1016/S0169-5347\(01\)02205-4](https://doi.org/10.1016/S0169-5347(01)02205-4)