

OPTIMIZING ALLOCATION OF MONITORING EFFORT UNDER ECONOMIC AND OBSERVATIONAL CONSTRAINTS

SCOTT A. FIELD,¹ The Ecology Centre, University of Queensland, St Lucia, 4072, Queensland, Australia
ANDREW J. TYRE,² The Ecology Centre, University of Queensland, St Lucia, 4072, Queensland, Australia
HUGH P. POSSINGHAM, The Ecology Centre, University of Queensland, St Lucia, 4072, Queensland, Australia

Abstract: Efforts to design monitoring regimes capable of detecting population trends can be thwarted by observational and economic constraints inherent to most biological surveys. Ensuring that limited resources are allocated efficiently requires evaluation of statistical power for alternative survey designs. We simulated the process of data collection on a landscape, where we initiated declines over 3 sample periods in species of varying prevalence and detectability. Changing occupancy levels were estimated using a technique that accounted for effects of false-negative errors on survey data. Declines were identified within a frequentist statistical framework, but the significance level was set at an optimal level rather than adhering to an arbitrary conventional threshold. By varying the number of sites sampled and repeat visits made, we show how managers can design an optimal monitoring regime that maximizes statistical power within fixed budget constraints. Results show that 2 to 3 visits/site are generally sufficient unless occupancy is very high or detectability is low. In both cases, the number of required visits increase. In an example of woodland bird monitoring in the Mt. Lofty Ranges, South Australia, we show that, although the budget required to monitor a relatively rare species of low detectability may be higher than that for a common, easily detectable species, survey design requirements for common species may be more stringent. We discuss implications for multi-species monitoring programs and application of our methods to more complex monitoring problems.

JOURNAL OF WILDLIFE MANAGEMENT 69(2):473–482; 2005

Key words: decision theory, detectability, false-negative errors, monitoring, Mt. Lofty Ranges, optimization, statistical power, survey design, woodland birds.

The ability to accurately monitor the status of populations of conservation concern is of increasing interest to managers seeking to manage biodiversity at a landscape scale (Dixon et al. 1998, Pollock et al. 2002). Reliably demonstrating trends in survey data is, however, notoriously difficult; efforts to do so have spawned a wide variety of design and analysis methods (Thomas 1996, Dixon et al. 1998, Williams et al. 2002). Even if survey sites are appropriately stratified in space (Thompson 1992), imperfect detection remains a problem (Thompson and Seber 1994, MacKenzie et al. 2002, Royle and Nichols 2003). Most surveys are unlikely to record all individuals present within a sample unit, leading to false-negative errors. These can have substantial effects on bias and precision of population parameter estimates (Tyre et al. 2003) and, thus, reduce statistical power to detect trends. Although numerous options exist for estimating detectability (e.g., distance sampling, mark–recapture techniques, and repeated site visits; Lancia et al. 1994, Yoccoz et al. 2001, MacKenzie et al. 2002, Williams et al. 2002, Royle and Nichols 2003, Tyre et al. 2003),

they can be expensive to implement on large scales. Consequently, when conservation managers attempt to implement landscape-scale monitoring under tight financial constraints, statistical power is often the first casualty.

Set in this context, the need to optimize efficiency of survey designs is paramount. Optimal survey design seeks maximization of statistical power or minimization of financial costs within constraints determined by management objectives, budgets, and idiosyncrasies of the system under study. Basic requirements include the ability to quantify detectability and determine costs of alternative survey configurations. These must be combined in an analysis of trade-offs among those configurations that increase statistical power and those that reduce financial costs.

As few species are likely to have a detectability = 1.0 (Mackenzie et al. 2002), biological surveys are prone to observer errors that reduce accuracy of parameter estimates and thereby reduce statistical power. Therefore, any attempt to optimize survey efficiency must consider such errors. Mackenzie et al. (2002) and Tyre et al. (2003) showed how occupancy and detectability could be simultaneously estimated from presence–absence survey data using a Zero-Inflated Binomial model (ZIB, Hall 2000). Tyre et al. (2003) showed that this method was capable of substantially reducing

¹ E-mail: scott.field@adelaide.edu.au

² Present address: School of Natural Resource Sciences, University of Nebraska-Lincoln, Lincoln, NE 68583-0819, USA

bias and improving precision in parameter estimation from presence-absence data when false negatives occur. An obvious consequence is an increase in statistical power for a given survey effort—an issue of key importance in the scenario we consider here.

Statistical power is also strongly influenced by significance level or Type I error rate α . In threatened species monitoring, α is the probability of declaring a population decline when it has not happened. Ecological studies commonly fix $\alpha = 0.05$ and accept whatever statistical power results. This implies that Type I errors (triggering recovery actions when no occupancy decline has occurred) are of greater biological or management importance than Type II errors, β (failing to detect an occupancy decline), which is not necessarily true. Management objectives, rather than arbitrary statistical conventions, should determine conclusions drawn from data and actions thus triggered. Numerous authors have suggested that costs (e.g., in economic, political, social, or environmental terms) of each kind of error should be used to determine values of α and β (e.g., Toft and Shea 1983, Peterman 1990, Yoccoz 1991, Peterman and M'Gonigle 1992, Steidl et al. 1997, Burgman 2000, Di Stefano 2001). Several approaches for implementing this principle have been proposed (reviewed in Field et al. 2004). For example, Mapstone (1995) suggested a method in which an initial α is chosen, the corresponding Type II error rate, β , is calculated, and α iteratively changed until some target $\alpha:\beta$ ratio is achieved.

We present an analysis of monitoring optimization, followed by a case study based on woodland bird monitoring in the Mt. Lofty Ranges, South Australia. Our aim was to calculate statistical power for various configurations of survey effort across a landscape and identify survey designs that optimized statistical power with respect to financial costs given monitoring objectives. We also evaluated sensitivity of results to survey design parameters: number of sites visited and repeat visits made to each site. Finally, we examined the relevance of our results to analysis of more complex problems such as optimizing survey design across a multi-species assemblage.

METHODS

We considered a scenario in which a conservation manager sought to detect a decline, d , of a specified a priori magnitude (i.e., the effect size) in occupancy, p , defined as the proportion of sites

in a landscape that are occupied by a species of conservation concern. In practice, d might be set at the level of decline that would be considered serious enough to trigger recovery actions, for example, the decline recommended by the International Union for Conservation of Nature and Natural Resources (IUCN) as warranting an upgrade of conservation status (IUCN 2001). We presumed that investigators would establish survey sites across the landscape and collect data by making >1 visits to each site within a single sample period (e.g., the breeding season) and also assumed site occupancy remained static. We presumed a fixed budget, B , over 3 sample periods. Management could seek to achieve objectives by varying allocation of the budget between establishing survey sites and making repeat visits to those sites within each period.

We used a simulation model to explore how different survey designs influenced statistical power. This model had 3 key components: (1) a procedure for reducing effects of false-negative survey errors on parameter estimation, (2) a function specifying relative costs of sites and visits, and (3) a module that simulated collection and analysis of datasets from simple virtual landscapes. We describe these components below and then go on to describe procedures for determining optimum designs and evaluating sensitivity of results to design parameters.

Accounting for False Negative Survey Errors

We characterized detectability of a species, q , defining it as the probability of successfully recording a species on a particular survey visit, given that it inhabits the site (i.e., it is the probability of avoiding a false-negative survey error). In the simple case considered here, we assumed that q could differ among species but remained constant for each species across the landscape for the duration of the study (see Mackenzie et al. (2002) and Tyre et al. (2003) for a method of including habitat-specific variation in q). We did not distinguish among the various reasons that an observer might fail to observe a resident species (e.g., cryptic behavior, temporary absence from the survey site in other parts of the home range). In practice, estimates of q will vary according to the manner and intensity of searching and, therefore, are specific to the survey method used.

In each survey period, we assumed that n sites were visited m times each. The number of observations of a species at a given site was s ($s \leq m$). If

the species was observed at least once, then the likelihood of this observation was

$$L(s > 0) = p \binom{m}{s} q^s (1-q)^{m-s}, \quad (1)$$

which was the probability of s successes in m trials, a binomial distribution, multiplied by the probability that the site was occupied. If the species was not observed at a site, the likelihood was:

$$L(s = 0) = (1-p) + p(1-q)^m, \quad (2)$$

which was the probability that it was not present plus the probability that it was there but was not observed in m visits. We summed negative logarithms of these likelihoods over all sites and minimized this value to find maximum likelihood estimates for the 2 unknown parameters: p and q . We implemented this procedure in C++ and used the Nelder-Mead simplex algorithm (Press 2002) to find the maximum likelihood estimates.

Calculating Survey Costs

We assumed that the overall cost of conducting annual surveys had 2 components: cost of establishing new sites, c_e , and cost of making a repeat visit to an existing site, c_v . Although true random selection of sites would be preferred, we recognized that, given a choice among sites of equal suitability, a manager might initially tend to establish sites with lowest access cost. Therefore, we assumed that cost of establishing a new site was an increasing function of the number already selected:

$$c_n = c_0 \exp^{\gamma n}, \quad (3)$$

where

c_0 = cost of establishing first site,
 c_n = cost of establishing n^{th} site, and

γ = constant determining how rapidly cost of adding a new site increases.

Defining $c_{n_{\text{max}}}$ as the cost of the last possible (most expensive) site that could be chosen in the landscape and n_{max} as the number of possible sites, we solved equation 3 for γ :

$$\gamma = \ln[(c_{n_{\text{max}}}/c_0)]/n_{\text{max}}. \quad (4)$$

We used $\gamma = 0.00135$ in our simulations based on approximate values for these parameters in bird surveys undertaken by the authors in the Mt. Lofty Ranges, South Australia: $n_{\text{max}} = 300$, $c_{n_{\text{max}}} = \$75$, $c_0 = \$50$, $c_v = \$44$ (costs in Australian dollars). Varying $c_{n_{\text{max}}}$ between \$50 and \$100 (and thus γ

between 0 and 0.00231) had no substantive effect on conclusions obtained from the simulations described below.

Although the cost of adding new sites may increase, the cost of making a repeat visit to a particular site will remain approximately constant no matter how many visits are made. Thus, overall survey cost, which we assumed must equal the budget, was given by

$$B = \sum_{i=0}^{n-1} c_0 \exp^{\gamma i} + nmrc_v, \quad (5)$$

which is the summed cost of establishing n sites plus the cost of making m repeat visits to those n sites in each of r years. We assumed that all sites were chosen prior to the first season of surveying and none were added later.

Statistical Power Simulations

We calculated statistical power by sampling simulated datasets. We simulated decline in species' occupancy rates over 3 sample periods, $t = 0, 1,$ and 2 , where p_t was modeled using the logit link function (i.e., analogous to a logistic regression model).

$$\ln[p_t / (1 - p_t)] = a + bt \quad (6)$$

At $t = 0$, the occupancy state for each of n sites was generated as a Bernoulli random variable with the probability of a success (occupancy) $p = p_0$. The simulation then sampled the landscape, recording observation of a given species on each visit with probability q , if present. After m visits to each of n sites, the resulting dataset consisted of a vector of n random variables. This process was repeated for $t = 1$ and $t = 2$, where the probability of occupancy was p_1 and p_2 , respectively. The value used for a was obtained from Equation (6) by specifying p_0 at $t = 0$ (i.e., $a = \ln(p_0 / (1 - p_0))$) and b was chosen to achieve a specific percent decline between the first and last periods, $d = 1 - (p_2 / p_0)$.

We estimated parameters for 2 models; first where $Pr(\text{success}) = \hat{p}$ in equations 1 and 2 was assumed to be constant across the 3 survey periods, and a second model where a separate \hat{p}_t was fit for each period. We compared these models by calculating the difference in log-likelihoods of the 2 models and compared this likelihood ratio statistic to a chi-squared distribution with 1 degree of freedom. Where the model with time-varying \hat{p}_t had a significantly higher likelihood (at the chosen significance level, see below), we identified a negative trend.

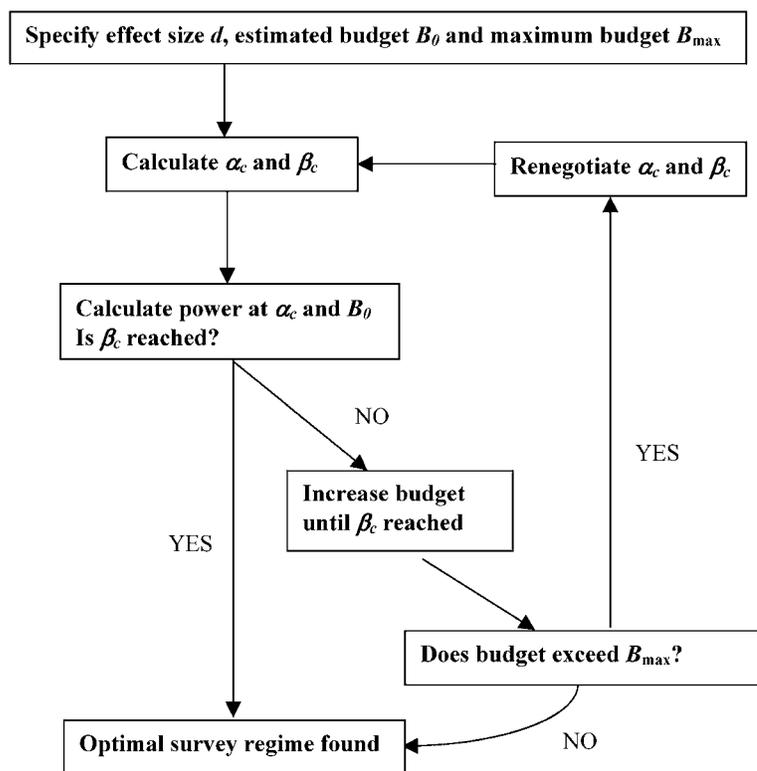


Fig. 1. Decision process for identifying the optimal survey regime to detect a population change of specific magnitude over a specific timeframe, given flexible statistical thresholds and a flexible budget.

We assumed no temporal or spatial autocorrelation. Temporal autocorrelation could arise in this scenario if a site occupant had not died or dispersed or if a new occupant had not arrived in an empty site. Increasing the period between surveys reduces this problem; therefore, we assumed the 3 surveys were spread out over sufficient time (e.g., 0, 5, 10 yrs) to avoid substantial temporal autocorrelation. Spatial autocorrelation arises through correlations in habitat quality and because of dispersal of individuals and is present in every ecological study. It may lead to increased Type I error rates if a statistical correction is not applied (Cerioli 1997). Although an interesting extension of our work, in the interests of simplicity, we omitted autocorrelation from consideration.

We varied model parameters B , n , d , p_0 , and q , to evaluate a plausible range of scenarios that a manager might face (see Sensitivity Analysis and Case Study below). We determined m using equation 5 and the specified values for c_o , c_v , n , and B . Whenever m was not an integer, we allocated additional surveys to the remaining fraction of

sites (e.g., $B = \$50,000$ and $n = 140$, $m = 2.25$) so all sites received at least 2 visits and 25% of randomly chosen sites received a third visit.

For each combination of parameters, we ran 5,000 simulations and recorded the number in which a trend was detected as a measure of statistical power (i.e., the probability of detecting an occupancy decline) given a decline had occurred. We fit smooth B-splines with 9 degrees of freedom through simulation results using the function “smooth.spline” in the “modreg” library of the statistical package *R* version 1.6.2 (Venables, W. N., D. M. Smith and the *R* Development Core Team, Vienna, Austria, 1999–2001). These functions reduced small stochastic variations from simulations and accurately represented the curves in the vicinity of the optima at relatively small m . At large m ($m > 5$) there are fewer examples; therefore, less emphasis should be placed on results in this range.

regularly represented the curves in the vicinity of the optima at relatively small m . At large m ($m > 5$) there are fewer examples; therefore, less emphasis should be placed on results in this range.

Finding the Minimum Budget

We assumed either that $\alpha = \beta$ or that unequal values were directly specified by stakeholders. Following Mapstone (1995), exact (or “critical”) values of α and β used in simulations are denoted by α_c and β_c . Once α_c and β_c were set, we found the optimal survey regime by iteratively changing B until the desired power was achieved at minimum cost (Fig. 1). For the specified d , β was calculated at an initial estimated budget B_0 that was increased until β_c was reached and the optimal survey regime attained. However, for some maximum budget allocated to the project, B_{max} , if $B > B_{max}$, then the process of setting α_c and β_c would have to be revisited (e.g., by renegotiation among stakeholders) and either α_c or β_c relaxed until the minimum required budget fit within the specified financial constraints. Optimal budgets were

calculated to the nearest \$1,000 and α_c and β_c to an accuracy of 10^{-3} .

Sensitivity Analysis and Case Study

We performed sensitivity analysis on parameters that could, in practice, be manipulated or measured by a manager and assessed how the optimal survey regime responded. Our baseline scenario and reference point was: $B = \$75,000$, $p = 0.5$, $q = 0.5$, $d = 0.25$. We considered a series of practical questions that a manager might pose and performed sensitivity analyses over relevant parameter combinations.

*Question 1: At what value should I set α ?—*For a manager with a fixed budget but unable to specify α_c and β_c a priori, we used the default $\alpha:\beta$ ratio of $k = 1$ and employed Mapstone's (1995) method to derive α_c by iteration for the baseline scenario.

*Question 2: How much should I spend on monitoring?—*For the case considered above, where $k = 1$, we determined α_c and β_c for budgets between \$50,000 and \$150,000. For each B , we plotted β_c and efficiency, defined as power achieved per dollar spent (β_c/B), to assess how returns diminished with increased resources.

To explore a case where decisions about acceptable Type I and Type II error rates could be made in advance, we assumed stakeholders had specified that avoiding Type II errors was twice as important as avoiding Type I errors, and the maximum acceptable Type II error rate was 10% (i.e., $\alpha_c = 0.20$ and $\beta_c = 0.10$). We used the method described in Fig. 1 to derive the minimum budget, B^* , necessary to detect an occupancy decline at baseline parameter settings.

*Question 3: How should I allocate the budget between sites and visits?—*We explored the trade-off between n and m by plotting profiles of β as a function of m for species with different values of p and q in situations where the objective was to detect d of varying magnitude. Using the value of α_c for $k = 1$ determined in Question 2 for the baseline budget of \$75,000, we examined how optimal combinations of n and m (designated n^* and m^*) changed as we varied levels of p , q , and d , 1 at a time (p and q to 0.25 and 0.75; d to 0.1 and 0.4), while the other 2 were held at their baseline levels.

*Question 4: How should I design a survey for individual species?—*We applied these methods to optimizing survey design for woodland bird species: the superb fairy wren (*Mahurus cyaneus*) and the yellow-tailed black cockatoo (*Calyptorhynchus funereus*). Data to estimate p and q were collected in the Mt. Lofty Ranges, South Aus-

tralia, during September–December 2000 using the 20-min, 2-ha timed active area search method (Loyn 1986). Data came from 38 sites described by Field et al. (2002) plus 68 additional sites that were added to ensure that major forest and woodland habitats were represented in approximate proportion to their prevalence in the region (Field et al. unpublished data). Details of survey methods are as described by Field et al. (2002), and each site was visited 3 times during the survey. The selected species provided contrasting examples: *M. cyaneus* was ubiquitous ($\hat{p} = 0.96$) and highly observable ($\hat{q} = 0.74$); whereas, *C. funereus* was less widespread ($\hat{p} = 0.5$) and less detectable ($\hat{q} = 0.35$). Although *C. funereus* is a large and conspicuous species, its low q resulted from its greater mobility and large home-range size relative to the size of the survey unit.

For both species we set $\alpha_c = 0.20$ and $\beta_c = 0.10$ (i.e., $k = 2$) and aimed to detect an occupancy decline that would represent a change of current regional conservation status (vulnerable to endangered for *C. funereus* and least concern to vulnerable for *M. cyaneus*). Using IUCN criterion A2 (IUCN 2001), with p as an index of population size, this entailed a decline of 50% for *C. funereus* and 30% for *M. cyaneus*. Using the procedure outlined in Fig. 1, we calculated B^* , n^* , and m^* at which the required α_c and β_c could be achieved for both species.

*Question 5: Can I optimize survey design for >1 species?—*Given that data on multiple species can be collected simultaneously, a manager might be interested in finding a compromise design that meets objectives across an entire assemblage. Differences in p , q , and d among species mean that the optimal design for 1 species is likely to fall short for other species, requiring an increase in B . We examined this question by calculating β_c for *C. funereus* and *M. cyaneus* using the other species' optimal design and the increase in B required to achieve β_c in each case.

RESULTS

*Question 1: At what value should I set α ?—*Using the baseline scenario, $\alpha = 0.05$ yielded $\beta = 0.48$, well above 0.05 as stipulated by $k = 1$. Through iteration, we found $\alpha = \beta = 0.218$ (Fig. 2). It is noteworthy that between $\alpha = 0.05$ and $\alpha = 0.20$, β decreased at a faster rate than α increased (i.e., the gain in statistical power was proportionally greater than the increase in Type I errors; Fig. 2).

*Question 2: How much should I spend on monitoring?—*For $\alpha_c = \beta_c$, larger budgets increased power

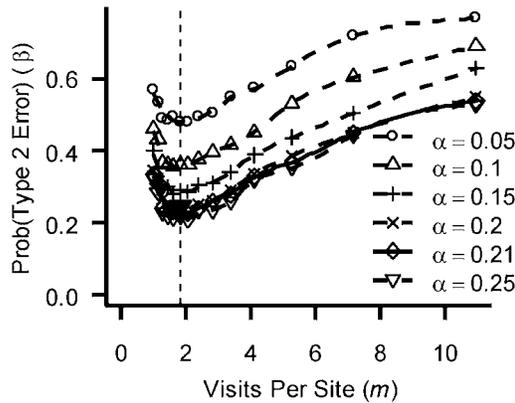


Fig. 2. Profile of Type II error rate, β , for various values of Type I error rate, α , using baseline parameters (\$75K budget, $p = 0.5$, $q = 0.5$, $d = 0.25$) in 5,000 simulation runs. Solid curve is the profile at optimum α -level ($\alpha = \beta = 0.218$); dashed curves, from top to bottom, represent increasing values of α as indicated. A vertical dashed line indicates the optimum, which is the same for each curve. Curves are smooth B-splines with 9 degrees of freedom fit through simulation results.

asymptotically but yielded sharply diminishing returns (Fig. 3). When $\beta_c = 0.1$ was specified in advance, by iteration we found $B^* = \$123,000$, $n^* = 302$ sites, and $m^* = 2.04$ visits (Fig. 4). Lower levels of each of p , q , and d resulted in lower power that could only be redressed by increasing the budget.

Question 3: How should I allocate the budget between sites and visits?—Optimal values of m (m^*) were influenced markedly by p and q but less so by d ; whereas, a decrease in p had little impact, high p increased m^* (Fig. 5a). Variations in q had the opposite effect: high q had little effect; whereas, low q substantially increased m^* . However, at low q the profile was rather flat; therefore, gains from

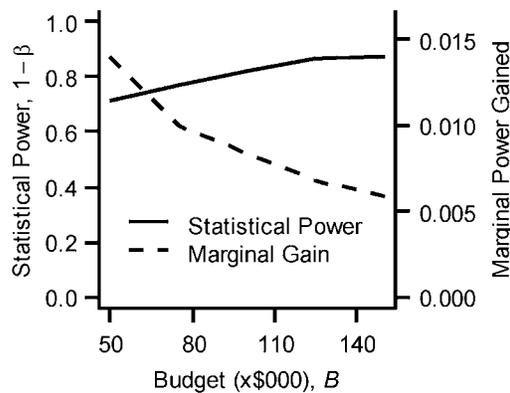


Fig. 3. Simulation results for maximum statistical power and marginal power gained (maximum power per unit cost) as a function of monitoring budget using baseline parameters ($p = 0.5$, $q = 0.5$, $d = 0.25$).

choosing the optimum were less (Fig. 5b). Varying d had little effect on m^* ; in this case, the shape of the profile was more affected (Fig. 5c). At high and low d , profiles were very flat; only for moderate d was substantial increase in power gained by choosing m^* .

Question 4: How should I design a survey for individual species?—Optimal budget and combinations of sites and visits for *M. cyaneus* and *C. funereus* differed substantially (Fig. 6). For *M. cyaneus*, the target was achieved with $m^* = 3.4$ visits to $n^* = 22$ sites at a cost of $B^* = \$11,000$; whereas, for *C. funereus*, $m^* = 2.2$ visits to $n^* = 105$ sites at a cost of $B^* = \$36,000$ were required (Fig. 6).

Question 5: Can I optimize survey design for >1 species?—The optimal budget and design for *M. cyaneus* fell well short of achieving the objective for *C. funereus* ($\beta = 0.42$). Maintaining the *M. cyaneus* optimum $m^* = 3.4$ visits required $n^* = 73$ sites at a cost of $B^* = \$37,000$ to reach the target of $\beta = 0.1$ for *C. funereus*. In contrast, using the *C. funereus* budget and design to monitor *M. cyaneus* easily surpassed the objective ($\beta = 0.002$), which made this design the most cost-effective for meeting both objectives.

DISCUSSION

Our analysis yielded interesting results regarding the trade-offs between allocating a fixed amount of monitoring effort to more sites (n) as opposed to more repeat visits (m) to those sites. For most parameter values studied, Type II error rate, β , exhibited a sharp profile around m^* (Figs. 2, 4–6), demonstrating that this trade-off can be

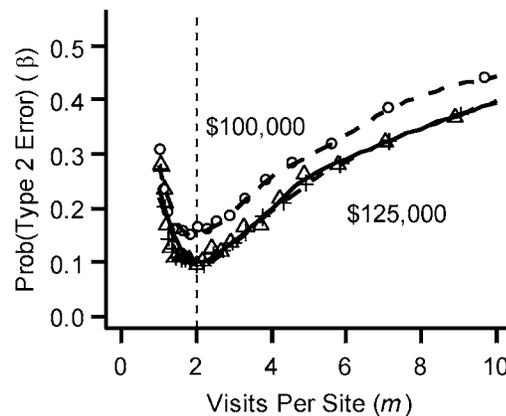


Fig. 4. Profile of Type II error rate, β , for various budgets, using baseline parameters ($p = 0.5$, $q = 0.5$, $d = 0.25$) and $\alpha_c = 0.20$ in simulations. Curves, from top to bottom, represent increasing budgets. A vertical dashed line indicates the optimum, which is the same for each curve. Curves are smooth B-splines with 9 degrees of freedom fit through results.

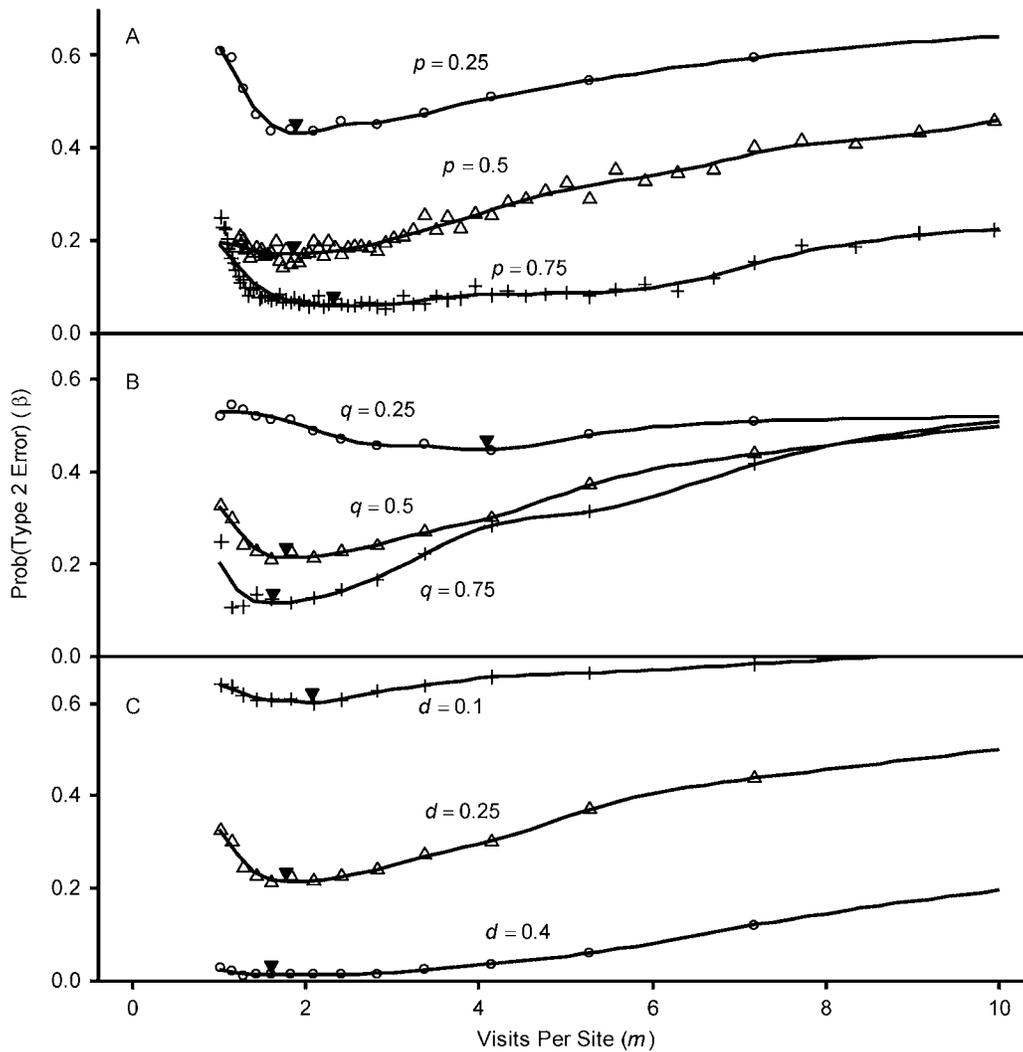


Fig. 5. Profile of Type II error rate, β , for 3 values of: (a) initial occupancy p ; (b) detectability q ; and (c) magnitude of occupancy decline d , with other parameters at baseline ($p = 0.5$, $q = 0.5$, $d = 0.25$) and $\alpha = \alpha_c = 0.218$. Curves, from top to bottom, represent increasing values of p , q or d , as indicated. Solid arrowheads indicate the optima for individual curves.

critical to monitoring efficacy. Except for very large occupancy declines, the slope of the curve was steeper to the left of the optimum, indicating that too few visits would incur a greater penalty in lost statistical power than sampling too few sites. Another robust result was that 2 visits/site usually sufficed. Of parameter combinations studied, those with sharp profiles generally had an optimum close to 2 visits/site, while the broad optimal zone of those with flatter profiles usually included 2 visits/site (Figs. 5–6). Exceptions to the 2 visits/site rule occurred when p was high (Fig. 5a) or q was low (Fig. 5b).

The effect of high occupancy can be seen in the profiles at $p = 0.75$ in Fig. 5a and $p = 0.96$ for *M. cyaneus* in Fig. 6. In both cases m^* is substantially elevated and approaches four visits per site. In other words, when the landscape is close to fully occupied ($p \approx 1.0$), it is more useful to sample intensively at relatively few sites than to sample widely across the landscape. This effect occurs because unoccupied sites contain no information about false negatives. When a manager samples a series of unoccupied sites and, inevitably, records a string of negative results, she or he can not be sure that failure to record the species is due to its

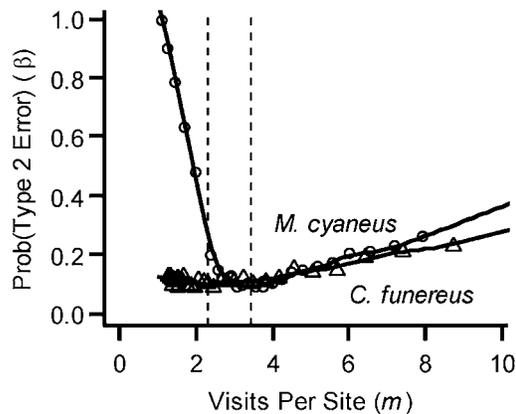


Fig. 6. Profile of Type II error rate, β , at the optimal budget for the superb fairy wren (*M. cyaneus*) and the yellow-tailed black cockatoo (*C. funereus*). Vertical dashed lines indicate optima for individual curves.

genuine absence (low p) rather than a run of bad luck with false-negative errors caused by poor detectability (low q). For this reason, landscapes with low p make it harder to reduce variance around the estimate of p , which is essential to achieving increased power. The best strategy, then, is to sample more widely across the landscape when p is low, to maximize the number of occupied sites in the sample. In contrast, for high p , a large proportion of informative (occupied) sites should usually be visited. Sampling widely will, therefore, be less of an imperative and sampling intensively to reduce uncertainty around any zeroes in the sample is relatively more valuable.

The same logic explains the need for more repeat visits when q is low and the risk of obtaining false zeroes is correspondingly high. Sampling intensively improves knowledge of p by increasing the chance that, if the species is present, it will eventually be observed. If the species is absent, sampling intensively increases confidence in the negative result obtained. For example, if $q = 0.25$ and the species has not been recorded at a site after 4 visits, the probability that this is a true negative result will have risen from 0.25 after 1 visit to $1.0 - 0.75^4 = 0.68$ after 4 visits. Increasing knowledge about the true status of the site leads to a better estimate of p and a corresponding increase in power. This is borne out in Fig. 5a: where $q = 0.25$, m^* has risen dramatically to 4.04. In contrast, for $q = 0.75$, $m^* = 1.81$ and power decays rapidly with increasing m . Large numbers of repeat visits are of little value here because the probability is high ($1.0 - 0.25^2 =$

0.94) that the species will be seen by the second visit.

Varying d in either direction from its baseline level flattened the power profile with little change to m^* (Fig. 5c). Large declines were so easy to detect that 1–3 visits were nearly equally satisfactory, and the penalty for straying further from the optimum was relatively modest. On the other hand, small declines were so hard to detect that little was gained by using the optimal design.

The case study also revealed some interaction among p , q , and d in their influence on survey design. Although higher q for *M. cyaneus* (0.74) relative to *C. funereus* (0.35) should have tended to reduce m^* (Fig. 5b), extremely high p (0.94) for *M. cyaneus*, increased m^* (Fig. 5a) and overwhelmed the effect of high q (Fig. 6).

Predictably, the required budget for the relatively rare, less detectable species (*C. funereus*) was much higher than that required for the ubiquitous, highly detectable species (*M. cyaneus*). Nevertheless, somewhat counter intuitively, survey design requirements for the latter were much more stringent (i.e., sensitive to suboptimal design parameters). This result arose largely from the severe penalty, as discussed above, for insufficient repeat visits for *M. cyaneus*. The larger d required for *C. funereus* also played a role by flattening its profile and making power obtained less sensitive to deviations from the optimal design.

The best compromise design for monitoring the 2 species simultaneously was to use the budget and design for *C. funereus*, as this comfortably met the objective for *M. cyaneus*. In effect, the much larger budget for *C. funereus* lowered the *M. cyaneus* curve to $\beta = 0.1$. This illustrates that, in general, it may be the harder (i.e., rarer, less detectable) species that determines the design to be used, as it will usually require a higher budget to achieve the same level of power (Fig. 5a,b shows how much less power is obtained for rare, less detectable species when the budget is fixed). However, if 2 species with different optima required similar budgets, the shape of the respective power profiles would decide the issue. A sharper profile means a greater penalty in lost power (and thus extra budget required to restore it) as the design moves away from that species' optimum. Such species would, therefore, receive higher priority than those with flatter profiles. A multi-species design should also account for differences in utility of detecting declines resulting from economic value or level of threat faced. Fully analyzing this problem for >2 species is

clearly a much more complex undertaking to be addressed in future work.

Future Work

We limited our analysis to a rather narrow scenario to provide a simple demonstration of method and process for monitoring optimization. Therefore, we offer the following caveats, as well as suggestions for extending this work.

Alternative Sources of Monitoring Data.—In monitoring studies, investigators usually are ultimately interested in population size trends. Although presence-absence data may provide an adequate indicator, its sensitivity in diagnosing population change depends on the relationship between abundance and occupancy (Holt et al. 2002). It is likely, for example, that changes in presence-absence may not be detected until a catastrophic decline in abundance has already occurred. For this reason, trend detection based on abundance data will often be more statistically powerful than presence-absence data (but see van Horne 1983), although potentially more expensive. Therefore, exploration of optimal survey design using an abundance data equivalent of the ZIB model (e.g., Dobbie and Welsh 2001a,b) within realistic economic constraints would be useful.

Reproductive and mortality data also improve knowledge of population trends with the added advantage of yielding ecological information useful in devising a management response. However, these data are even more costly to collect than abundance data and are often beyond the means of management agencies to acquire in meaningful quantities from a representative network of sites across an entire landscape. Nevertheless, relative efficiency of these different data sources and optimal allocation of scarce budgetary resources among them, remains an interesting issue for further study.

Nonlinear Declines.—In reality, declines will not occur in a simple linear fashion as we have assumed here, but they will be subject to multiple stochastic influences that combine to generate irregular dynamics characteristic of natural populations. Trend detection under these conditions is considerably more complicated. Consequently, testing the generality of our present conclusions using a model including population dynamics subject to environmental stochasticity would be a useful extension.

Variable Budgets and Sampling Intervals.—Our examples optimized sampling within a fixed budget over a fixed (unspecified) period. However, it

is easy to envisage situations in which the budget, period, or both are variable. For example, Haight et al. (2000) optimized a set of decisions for a translocation problem with uncertain future budgets and Urquhart (1998) studied how frequency of sampling in relation to magnitude of change affected power to detect trends. Adding an open-ended future to the problem would dramatically increase complexity and require temporal discounting of costs. How, or whether, to do this for endangered species is an open topic.

MANAGEMENT IMPLICATIONS

Optimal monitoring design for a given species inevitably depends on idiosyncrasies associated with its ecology and habitat and the agency undertaking the surveys. Nevertheless, we provide a framework that managers can use to significantly improve chances of detecting important population declines once they have preliminary information on the species in question. Even in the absence of preliminary data, our analyses yield several general recommendations to improve monitoring efficacy.

Set Optimal Significance Levels

The conventional $\alpha = 0.05$ has no basis in statistics or ecology, and its slavish use can lead to serious errors and wasted resources in threatened species management (Mapstone 1995, Di Stefano 2003, Field et al. 2004). Instead, if frequentist statistics are used, an $\alpha:\beta$ ratio should be specified that reflects relative costs of Type I and Type II errors. If occupancy and detectability can be estimated, methods described here to generate the statistical power surface enable design of an optimal survey to detect a change in conservation status for a given species. Their routine use could substantially improve monitoring study quality and lessen risk of injudicious allocation of scarce conservation resources.

Estimate p and q and Conduct Repeat Visits

Optimal survey design requires estimating occupancy and detectability in advance. Encouragingly, our results suggest that a rule of thumb of 2 to 3 visits to each site would perform adequately for most species. Still, substantial increases in power were gained by choosing the optimal survey design in some circumstances, notably, if attempting to detect a modest occupancy decline when detectability is very low or occupancy is very high. In general, making fewer than the optimal number of visits resulted in a harsher penalty

than making more than the optimal number due to decreased ability to estimate detectability. Therefore, if species occupancy and detectability rates are uncertain, more visits are preferable to more sites to reduce power lost by straying from the optimum.

ACKNOWLEDGMENTS

We thank D. MacKenzie, J. Rhodes and 2 anonymous reviewers for comments that substantially improved the paper. We also thank the numerous landholders and surveyors who contributed to fieldwork. This work was supported by funding from the Australian Research Council, South Australian Department of Environment and Heritage, and Western Mining Corporation.

LITERATURE CITED

- BURGMAN, M. A. 2000. Population viability analysis for bird conservation: Prediction, heuristics, monitoring and psychology. *Emu* 100:347–353.
- CERIOLI, A. 1997. Modified tests of independence in 2×2 tables with spatial data. *Biometrics* 53:619–628.
- DI STEFANO, J. 2001. Power analysis and sustainable forest management. *Forest Ecology and Management* 154:141–153.
- . 2003. How much power is enough? Against the development of an arbitrary convention for statistical power calculations. *Functional Ecology* 17:707–709.
- DIXON, P. M., A. R. OLSEN, AND B. M. KAHN. 1998. Measuring trends in ecological resources. *Ecological Applications* 8:225–227.
- DOBBIE, M. J., AND A. H. WELSH. 2001a. Modelling correlated zero-inflated count data. *Australian and New Zealand Journal of Statistics* 43:431–444.
- . 2001b. Models for zero-inflated count data using the Neyman type A distribution. *Statistical Modelling* 1:65–80.
- EBERHARDT, L. L. 2003. What should we do about hypothesis testing? *Journal of Wildlife Management* 67:241–247.
- FIELD, S. A., A. J. TYRE, N. JONZÉN, J. R. RHODES, AND H. P. POSSINGHAM. 2004. Minimizing the cost of environmental management decisions by optimizing statistical thresholds. *Ecology Letters* 7:669–675.
- , H. P. POSSINGHAM, AND A. J. TYRE. 2002. Estimating bird species richness: how should repeat surveys be organized in time? *Austral Ecology* 27:624–629.
- HAIGHT, R. G., K. RALLS, AND A. M. STARFIELD. 2000. Designing species translocation strategies when population growth and future funding are uncertain. *Conservation Biology* 14:1298–1307.
- HALL, D. 2000. Zero-inflated Poisson and binomial regression with random effects: A case study. *Biometrics* 56:1030–1039.
- HOLT, A., K. J. GASTON, AND F. HE. 2002. Occupancy-abundance relationships and spatial distribution: a review. *Basic and Applied Ecology* 3:1–13.
- INTERNATIONAL UNION FOR THE CONSERVATION OF NATURE AND NATURAL RESOURCES. 2001. IUCN red list categories and criteria: version 3.1. International Union for the Conservation of Nature and Natural Resources, Gland, Switzerland and Cambridge, United Kingdom.
- LOYN, R. 1986. The 20 minute search—a simple method for counting forest birds. *Corella* 10:58–60.
- MACKENZIE, D. I., J. D. NICHOLS, G. B. LACHMAN, S. DROEGE, J. A. ROYLE, AND C. A. LANGTIMM. 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology* 83:2248–2255.
- MAPSTONE, B. 1995. Scalable decision rules for environmental impact studies: effect size, type 1, and type 2 errors. *Ecological Applications* 5:401–410.
- PETERMAN, R. 1990. Statistical power analysis can improve fisheries research and management. *Canadian Journal of Fisheries and Aquatic Sciences* 47:2–15.
- , AND M. M'GONIGLE. 1992. Statistical power analysis and the precautionary principle. *Marine Pollution Bulletin* 24:231–234.
- POLLOCK, K. H., J. D. NICHOLS, T. R. SIMONS, G. L. FARNSWORTH, L. L. BAILEY, AND J. R. SAUER. 2002. Large scale wildlife monitoring studies: statistical methods for design and analysis. *Environmetrics* 13:105–119.
- PRESS, W. H. 2002. *Numerical recipes in C++: the art of scientific computing*. Cambridge University Press, Cambridge, England, United Kingdom.
- ROYLE, J. A., AND J. D. NICHOLS. 2003. Estimating abundance from repeated presence-absence data or point counts. *Ecology* 84:777–790.
- STEIDL, R., J. HAYES, AND E. SCHAUBER. 1997. Statistical power analysis in wildlife research. *Journal of Wildlife Management* 61:270–279.
- THOMAS, L. 1996. Monitoring long-term population change: why are there so many analysis methods? *Ecology* 77:49–58.
- Thompson, S. K. 1992. *Sampling*. Wiley, New York.
- , AND G. A. F. SEBER. 1994. Detectability in conventional and adaptive sampling. *Biometrics* 50:712–724.
- TOFT, C., AND P. SHEA. 1983. Detecting community-wide patterns: estimating power strengthens statistical inference. *American Naturalist* 122:618–625.
- TYRE A. J., B. TENHUMBERG, S. A. FIELD, D. NIEJALKE, K. PARRIS, H. P. POSSINGHAM. 2003. Improving precision and reducing bias in biological surveys: Estimating false-negative error rates. *Ecological Applications* 13(6):1790–1801.
- URQUHART, N. S., S. G. PAULSEN, AND D. P. LARSEN. 1998. Monitoring for policy-relevant regional trends over time. *Ecological Applications* 8:246–257.
- VAN HORNE, B. 1983. Density as a misleading indicator of habitat quality. *Journal of Wildlife Management* 47:893–901.
- WILLIAMS, B. K., J. D. NICHOLS, AND M. J. CONROY. 2002. *Analysis and management of animal populations: modeling, estimation, and decision making*. Academic Press, San Diego, California, USA.
- Yoccoz, N. 1991. Use, overuse and misuse of significance tests in evolutionary biology and ecology. *Bulletin of the Ecological Society of America* 72:106–111.
- , J. D. NICHOLS, AND T. BOULINIER. 2001. Monitoring of biological diversity in space and time. *Trends in Ecology & Evolution* 16:446–453.

Associate Editor: Lubow.