Robust and realistic power analysis for cost-effective monitoring of occupancy trends

Kari E. Norman[1,2*], Jamie Sanderlin[3], Benjamin R. Goldstein[4], Connor M. Wood[5], Rebecca Kirby[6], Shaula Hedwall[7], Gavin M. Jones[1,2,8]

[1]*USDA Forest Service, Rocky Mountain Research Station, Albuquerque, NM*
[2]*Center for Fire Resilient Ecosystems and Society, University of New Mexico, Albuquerque, NM*
[3]*USDA Forest Service, Rocky Mountain Research Station, Flagstaff, AZ*
[4]*International Union for Conservation of Nature, Washington, DC*
[5]*K. Lisa Yang Center for Conservation Bioacoustics, Cornell Lab of Ornithology, Cornell University, Ithaca, NY*
[6]*USDA Forest Service, Southwestern Region, Albuquerque, NM*
[7]*US Fish and Wildlife Service, Arizona Ecological Services Field Office, Flagstaff, AZ*
[8]*School of Earth, Environment, and Sustainability, University of Iowa, Iowa City, IA*

*Corresponding authors: kari.norman@usda.gov

Keywords: acoustic monitoring, cost optimization, monitoring design, occurrence trend, power analysis

Abstract

Structured monitoring programs play a key role in identifying whether a population is declining, stable, or increasing, and power analyses allow program planners to determine the number of sampling units required to reliably estimate a specified trend (e.g., 1% annual decline) with a specified error rate (e.g., α=0.20, or 80% power). Increasingly, monitoring programs track trends in site occupancy (i.e., the proportion of sampling units occupied by the target species) using dynamic occupancy models, but practical and adequate guidance for conducting power analysis using this framework is extremely limited. Here, we explored the sensitivity of sample size recommendations from Bayesian occupancy-based power analysis to variation in trend generation approach (using equilibrium, recursive, and constant survival methods), power check criteria (ranging from inclusive to strict), and a range of biological scenarios (i.e., differences in initial occupancy, survival, and detectability) and explored how these technical choices would translate to cost-optimized monitoring designs. We used a proposed real-world acoustic monitoring program aimed at tracking the recovery of the federally threatened Mexican spotted owl (*Strix occidentalis lucida*) as our motivating example, which requires that a monitoring program can detect a 25% occupancy decline over a 10-year period with 90% power. We demonstrated that power to detect occupancy trends is highly sensitive to trend generation approach and power check criteria. In the proposed Mexican spotted owl monitoring program, differences in trend-generation approach alone resulted in a difference in the number of sampling units required to achieve sufficient power by up to 72% (n=695 for the constant survival approach; n=1200 for the equilibrium approach) for one biological scenario, which translates to nearly $500,000 in initial acoustic equipment costs. Our study offers practical advice to practitioners interested in developing occupancy-based power analyses that allow them to robustly track realistic population changes in a cost-effective manner.

Keywords: acoustic monitoring, cost optimization, monitoring design, occurrence trend, power analysis

Introduction

Monitoring is a foundational data gathering mechanism for conservation biology. Using repeated sampling across space and time in a structured or probabilistic sampling framework, monitoring programs document where species are, how their populations are changing, and the drivers of that change. Data produced from such programs are critical for identifying conservation concerns and have an outsized impact on policy development (Hughes et al. 2017). Recognition of the importance of moving from haphazard and bespoke sampling to systematic sampling for biodiversity conservation is now driving multi-lateral international agreements on monitoring goals including Article 7 of the Kunming-Montreal Global Biodiversity Framework, which emphasizes the importance of systematic monitoring for species of urgent conservation concern (Kunming-Montreal; Gonzalez et al., 2023). It has also prompted the development of accessible tools for robust monitoring design (Griffith et al. 2024; Norman & Poisot 2025).

Power analysis is a common approach for assessing the ability of a potential monitoring scheme to detect a target phenomenon of a given effect size (Sanderlin et al. 2019; Weiser et al. 2021; Leung & Gonzalez 2024). Power analysis can be used to inform minimum required sample size for a monitoring program design by evaluating the rate at which models applied to simulated data recover the true simulated values of the parameter of interest. For such power analyses to be useful however, the parameters used to simulate the ecological process must be sufficiently realistic to inform implementation, which requires extensive knowledge of target species' natural history (often informed by pilot data), the policy context, and logistical constraints (Williams et al. 2002). Realistic power analysis is especially important for monitoring of at-risk species, where potential misestimation of population trends could have dire implications for species persistence by misdirecting conservation investments or management actions. Appropriately designing monitoring programs in such high stakes settings therefore requires careful translation of species recovery requirements to statistical terms, simulation of potential population trends with high-ecological realism, and design of sampling schemes that operate within logistical and budgetary constraints (i.e. Sanderlin et al., 2014).

Following the development of occupancy models that correct for imperfect detectability in the early 2000's (MacKenzie et al. 2002, 2003), an increasingly common population monitoring goal is to detect trends in site occupancy (White 2019; Moussy et al. 2022). For estimating trends in occupancy a popular choice is the dynamic occupancy model (MacKenzie et al. 2003; Royle & Kéry 2007; Goldstein et al. 2024a), which models changes to the latent (true) occupancy state at a site in a certain season as the product of the stochastic process of colonization and persistence, conditional on the site's occupancy state in the preceding season. While simulating data from the dynamic occupancy model is straightforward, the occupancy trend itself is not a model parameter; rather, trend in occupancy is an emergent property of the survival, colonization, and initial occupancy parameters playing out through time.

As a result, simulating data with a reliable trend magnitude that can align with policy objectives targeting trends in occupancy is nontrivial, and there is currently no standardized approach for generating a parameter-driven trend in simulated dynamic occurrence data. In previous work, the decision making surrounding which parameters will be manipulated to generate the trend and how parameters are derived to implement a specific trend magnitude are often incompletely described if they are even acknowledged (Banner et al., 2019; Popescu et al., 2012; Steenweg et al., 2016; Wood et al., 2019). More critically for conservation contexts, there has been to our knowledge no investigation of how different approaches for imposing a trend interact with biological elements to impact sample size thresholds. The lack of clarity around what may initially seem to be minor technical decisions has potentially profound implications for the ability of the resulting monitoring network to efficiently meet its goals.

Here, we develop a repeatable framework for implementing robust and realistic power analyses to detect trends in occupancy, exploring how a model's ability to detect trends is influenced by choice of trend-generation approach, power check criteria, sample size, and their interactions. We outline the important decision points for translating policy requirements to a power analysis framework, develop a repeatable approach for simulating realistic occupancy trends informed by empirical data, and optimize sampling design decisions with respect to cost. Our goals are two-fold: first, to outline key decision points in the power analysis process and their implications for final monitoring network design, and second, to identify the minimum sample requirements necessary for a monitoring program that can identify population trends following policy requirements and under different cost scenarios. Our efforts are organized around the real-world problem of designing a United States range-wide passive acoustic monitoring program for the Mexican spotted owl (*Strix occidentalis lucida*; hereafter, "MSO"), a federally threatened subspecies.

Methods

Species description

The MSO is a federally threatened subspecies emblematic of the complexities associated with monitoring design and implementation. Originally listed as threatened under the U.S. Endangered Species Act in 1993, the Mexican spotted owl's management is governed by the 2012 Mexican Spotted Owl Recovery Plan (U.S. Fish and Wildlife Service, 2012, hereafter "Recovery Plan"), which provides the criteria for delisting the subspecies and habitat protection recommendations. MSO largely inhabit mature forests and rocky canyonlands and are predominately found on public lands in Arizona, New Mexico, Colorado, Texas, and Utah, as well as in northern Mexico. Due to logistical monitoring challenges across its extensive and disjunct range, mandated monitoring of the owl is inconsistent through time and across jurisdictions and provides insufficient data for estimation of even broad-scale population trends, despite significant investments to date. Systematic range-wide monitoring with the power to detect declines or recovery of the species within budgetary constraints is therefore critical not only for optimal management of the species, but to meet statutory recovery requirements.

Recovery criteria

The Recovery Plan (U.S. Fish and Wildlife Service, 2012) outlines two criteria for recovery related to monitoring of (1) trends in owl occupancy and (2) trends in owl habitat. The monitoring plan described here focuses on criteria #1, which states:

> *"Owl occupancy rates must show a stable or increasing trend after 10 years of monitoring. The study design to verify this criterion must have a power of 90% to detect a 25% decline in occupancy rate over the 10-year period with a Type I error rate of 0.10."*

The Recovery Plan divides the owl's United States range into five distinct ecological management units (EMUs) distinguished by predominant vegetation types and other environmental factors, and outlines possible approaches to monitoring, including a requirement that 25% of samples be allocated to "marginal" habitats outside the traditional owl preferences.

Sampling frame

We limited our sampling frame to the EMU boundaries, and within the EMUs to landcover types where the owl could likely be found in primarily forested environments (Gutierrez et al. 1995). Our initial sampling design does not emphasize canyonland habitat, though many canyonlands inhabited by owls fall within primarily forested units in our sampling frame. We defined a hexagon-based sampling grid spanning the EMU area with each individual hexagonal unit (or "hex") having an area of 4 km^2, the average estimated territory size of the California spotted owl (*S. o. occidentalis*) and the best known surrogate for MSO home range size (Tempel et al. 2016; Jones et al. 2018). All hexes intersecting the EMU area were included in the sampling frame. For each hex, we identified the predominant (modal) landcover type using the LandFire Existing Vegetation Type data product (LANDFIRE 2016). We first restricted our sampling frame to hexes with tree landcover types then restricted to plausible owl habitat, either high quality or marginal, informed by the habitat requirements listed in the Recovery Plan, expert opinion, and verified by the landcover types within areas with known owl occupancy (Protected Activity Centers; PACs). As some patches of habitat were highly disjointed, we included only patches of more than five contiguous hexes in our sampling frame. All landcover types included in the final sampling frame are given in Table S1.

Intended sampling method

Traditionally, presence-absence monitoring for MSO uses a callback approach where owl presence is identified by an expert human surveyor either visually or aurally (Franklin et al. 1996). This method is time-intensive and requires highly skilled observers, making its implementation costly. Surveys are also conducted primarily at night, introducing safety concerns. Recently, passive acoustic monitoring (PAM) has proven to be a successful alternative presence-absence sampling method for identifying spotted owls and, typically as a "bonus," the broader avian community (Appel et al. 2023; Kelly et al. 2023; McGinn et al. 2025). Though the power analysis described here is agnostic to sampling method, we picked method-dependent detection probabilities for our simulations based on existing data from PAM monitoring for California spotted owls (Wood et al. 2019; Gustafson et al. 2025) and pilot PAM data for MSO (Sanderlin et al. 2025). The PAM sampling process involves deployment of 1-4 acoustic recording units (ARUs) within each site (i.e., 4 km^2 hexagon) at a sufficient distance between neighboring units to sample non-overlapping audio space. In previous work, ARUs have been programmed to record continuously on a daily basis during periods when owls are most vocally active (i.e., from before dusk until after dawn) (Reid et al. 2022) over a multi-week sampling window, with the continuous sampling window split into two discrete sampling periods to facilitate model estimation (Fig. 1).

Power analysis

We investigated the minimum sample size necessary to identify a 25% decline in occupancy over a 10-year period with 90% power at two different scales, the individual EMU level and the range-wide level, for multiple biological scenarios. We also examined the sensitivity of sample size recommendations to multiple technical decisions, including how the trend was generated in simulated data and the criteria used for considering a model successful in identifying the trend, using the Southern Rocky Mountains EMU as a case study. For all analyses, we followed the same general steps of (1) simulating many datasets, (2) fitting a model to each dataset, and (3) checking for success in identifying the trend across replicates to calculate success rate (i.e., power). We describe each of these steps in detail below. For each set of conditions, we constructed a power curve to understand how power varied as a function of sample size. We then used the power curves to produce a final sample size recommendation by calculating the sample size at which a line connecting the points before and after the threshold crossing intercepted the desired power threshold. We defined sample size as the number of hexes sampled each year, with the same hexes sampled across years.

Simulating occupancy

We simulated occupancy dynamics for each hex in the sampling frame according to a dynamic occupancy model (MacKenzie et al. 2003; Royle & Kéry 2007). The dynamic occupancy model has four key processes, each of which can be either fixed parameters or functions of other parameters: initial occupancy, ψ_{i1}, is the probability that hex *i* is occupied in the first survey year; survival, ϕ_{it}, is the probability that hex *i* is occupied during time *t* + 1 given that it was occupied in year *t*; colonization, γ_{it}, is the probability that hex *i* is occupied during year *t* + 1 given that it was not occupied in time *t*; and detectability, p_{it}, the probability that the species is detected on a single survey during year *t* in hex *i* given that the site is occupied in that year.

During simulation, the latent (true) occupancy state in hex *i* during the first year z_{i1} was given by a Bernoulli trial with a probability given by the initial occupancy probability, which had either a high (0.6 or 0.43) or low (0.03) value to account for different dynamics in the high-quality versus marginal habitat landcover types, respectively (see below). Occupancy at subsequent time steps was determined by Bernoulli trials parametrized by either annual survival, if the hex was previously occupied, or annual colonization, if the hex was not previously occupied (see below for additional details). For all approaches, we gave annual survival and colonization biologically plausible noise. Survival was drawn from a normal distribution with a mean of 0 and a standard deviation of 0.04 and noise for colonization was drawn from a normal distribution with a mean of 0 and a standard deviation of 0.01 on the probability scale (Wood et al. 2019); all draws resulted in absolute survival or colonization values between 0 and 1. We generated observed occupancy from the simulated true occupancy history based on a binomial distribution parameterized by the detection probability as the probability of success and two sampling periods as the number of trials in each year. We did not allow for false positives, so observed occupancy was always unoccupied when true occupancy was unoccupied, and observed occupancy for occupied sites was given by independent binomial draws for each sampling period and year.

Trend generation approaches

To adhere as closely as possible to the criteria of the Recovery Plan, we needed to impose a specific trend in the expected value of occupancy (i.e., the expected proportion of occupied sites) by appropriately modifying the survival and colonization parameters. We compared three different approaches for imposing a trend in occupancy using the Southern Rocky Mountains EMU as a case study. For the first two implementations we imposed the trend through changes in annual survival while holding colonization constant, and for the third we imposed changes by holding survival constant and choosing an appropriately small, constant colonization rate to generate a decline.

Equilibrium approach. The first approach was derived from the equilibrium occupancy equation (eq. 1) and follows existing approaches wherein declines are imposed by a constant multiplicative decline in survival (Banner et al., 2019; and see Wood et al., 2019 for a similar approach with an additional change in colonization). Equilibrium occupancy at a site (ψ_{equil}) with constant survival (ϕ) and colonization (γ) can be calculated as

$$\psi_{equil} = \frac{\gamma}{\gamma + 1 - \phi} \qquad \text{(eq. 1)}$$

A decay factor b can then be selected to modify annual survival as a function of time according to $\phi_t = \phi_1 b^{t-1}$, yielding a trend in expected occupancy over time. Often the decay factor is chosen by "guess and check", where a decay factor is chosen and modified until it most closely produces the desired trend (Banner et al. 2019). As the trend is a product not just of the survival decay factor but also initial occupancy, survival, and colonization, it can therefore be a substantial time investment to find the appropriate decay factor for each of many scenarios that also vary by initial occupancy and colonization.

We developed an approach to automatically select a decay factor where colonization (γ) was derived using the equilibrium occupancy equation (eq. 1), which with algebraic reorganization gives colonization as a function of initial occupancy and survival (eq. 2), and the decay factor is derived from the desired occupancy in the final time step.

$$\gamma_i = \frac{\psi_{i1}(1-\phi_i)}{1 - \psi_{i1}} \qquad \text{(eq. 2)}$$

First, we calculated a target occupancy value at *t* = 10 (ψ_{10}) based on a 25% decline in the given initial occupancy (ψ_1; therefore, $\psi_{10} = 0.75\psi_1$). Then, using the equilibrium occupancy equation (eq. 1) where occupancy was set to 75% of initial occupancy ($0.75\psi_1$) we found the necessary survival rate for that occupancy probability at equilibrium: $\phi_{10} = 1 - \frac{\gamma_{10}(1-0.75\psi_1)}{0.75\psi_1}$. We then identified the annual survival decay factor to go from initial survival $\phi_1$to final survival ϕ_{10} using the exponential decay equation $\phi_t = \phi_1 b^{t-1}$, solving for the decay factor *b* in $\phi_{10} = 1 - \frac{\gamma_{10}(1-0.75\psi_1)}{0.75\psi_1} = \phi_1 b^9$. Annual survival for the occupancy processes was then calculated by imposing the decay factor b at each time step.

Recursive approach. For the second approach, we imposed a trend on expected occupancy probability by adjusting survival in each time step based on the recursive occupancy equation (eq. 3), which describes how occupancy probability in each timestep changes as a function of occupancy in the previous timestep, colonization, and survival:

$$\psi_{t+1} = \psi_t \phi_t + (1 - \psi_t)\gamma_t \qquad \text{(eq. 3)}$$

We first calculated the annual change in occupancy necessary to get a 25% reduction in occupancy using the exponential decay formula, with the final value set to 75% of the initial occupancy value: $0.75\psi_1 = \psi_1 b^t$, where b is the decay factor and $t = 9$ is the number of transition time steps. The decay factor was therefore $b = 0.97$, and annual occupancy probability for each time step was given by $\psi_{t+1} = \psi_1 0.97^t$. Using ψ_1 and ψ_{t+1} in the recursive occupancy equation, we then solved eq. 3 for ϕ_t for each time step to generate annual survival for the occupancy process. Colonization was derived as in the equilibrium simulation using the equilibrium occupancy equation. Where the first approach implements a multiplicative decline in survival, survival in this approach declines in an unpredictable manner.

Constant survival approach. The third approach maintained constant survival based on the parametrization of initial survival and imposed a decline through selection of an also constant colonization parameter that is sufficiently small that the population necessarily declines. We selected the appropriate colonization rate by setting the equilibrium occupancy equation (eq. 1) equal to $0.75\psi_1$, parametrizing survival based on initial survival for that scenario and then solving for colonization.

We assessed the power to detect an occupancy trend for multiple biologically plausible occupancy parameter scenarios. For "high occupancy" areas, defined by their landcover type, we tested two different initial occupancy levels based on known occupancy for California spotted owls (0.43) (Wood et al. 2019), and pilot monitoring data in the Southwest (0.6) (data described in Sanderlin et al., 2025). "Low occupancy" marginal landcover types had an initial occupancy of 0.03 for all scenarios. We tested two different site survival probabilities representing typical survival (ϕ = 0.8) and a population under stress (ϕ = 0.6), and two detection levels (p = 0.4, 0.8). We reference survival ($\phi = 1 - \varepsilon$) rather than the traditional extinction parameter (ε) throughout the paper as the demographic parameter more directly relevant for conservation management (Popescu et al. 2012; Wood et al. 2019). Together, we tested eight unique scenarios across sample sizes ranging from 10 to 2000 (or the maximum number of available hexes for EMU-level models), with interim sample sizes increasing on an approximately log scale (Table 1). Curves were generated from n = 100 replicates for range-level models and n = 200 replicates for EMU-level trends to balance computational feasibility with sufficient replication to reduce variance. To assess the sample size threshold necessary to detect EMU- and range-level trends, we generated power curves based on the recursive trend generation method for reasons discussed in the results.

Model fitting

We fit single-species dynamic occupancy models (MacKenzie et al. 2003; Royle & Kéry 2007) to simulated data using a Bayesian framework in the R packages `nimble` (de Valpine et al. 2017) and `nimbleEcology` (Goldstein et al. 2024b). Model structure was dependent on both the trend generation process of the simulation and the scale at which the trend was being assessed. For equilibrium and recursive simulations where survival changed through time, models included time as a predictor of survival with a random intercept for year to account for annual heterogeneity. A random intercept for year was also used to model colonization due to annual noise. All other parameters (initial occupancy, detection probability) were estimated as intercept-only models. For the constant survival simulation, survival and colonization had only a random intercept for year and all other parameters were estimated with intercept-only models.

For EMU-level trends, we fit models for each EMU independently following the structure described above for the appropriate underlying trend generation process. For range-level trends we fit a model to range-wide simulations with a random intercept on survival for EMU. All model parameters were given uniform priors with support from 0 to 1. We fit all models with 2 chains, 1000 iterations, and a burn-in of 500; all models converged according to Gelman-Rubin statistic ($\hat{R}$ <1.1). Full model equations are provided in the Supplementary Materials and code necessary to reproduce results are archived in a Zenodo repository at http://doi.org/10.5281/zenodo.18778235.

Power check criteria

Translating the Recovery Plan requirements into concrete statistical terms of success is one of the major challenges of creating a realistic power analysis. For each fitted model, we assessed the posterior distribution of a derived parameter equal to percent change in occupancy from the first to last season following the percent change formula: $(t_{10} - t_1)/ t_1$. Most commonly, previous power analyses for monitoring plans have considered a model successful if it detects a negative trend of any kind (Test A, Fig. 2; Banner et al., 2019; Ellis et al., 2014; Wood et al., 2019). This is the least stringent possible test, using a right- or two-tailed Bayesian credible interval (CRI) with α = 0.05, and requiring only that the upper bound of the CRI be less than 0. It does not account for scenarios where the true percent decline falls outside the model's estimated CRI upper and lower bounds. A more stringent Test B uses a left-tailed CRI with $\alpha = 0.05$ and requires that the maximum value in the posterior distribution is less than zero and that the interval between the lower bound of the CRI and the maximum posterior value contains the true trend (Test B, Fig. 2). By being more stringent in the left tail of the posterior, this test does not allow for underestimation of the decline (where the slope estimated by the model is less negative than the true slope), which in conservation settings is riskier than overestimating the decline because it could potentially result in delisting a species prematurely. Test C is the most stringent, using a two-tailed CRI with $\alpha = 0.05$, and identifying success when the CRI does not contain 0 and contains the true trend (Test C, Fig. 2).

To assess sensitivity of sample size recommendations to these power check criteria, we generated power checks for Tests A, B, and C for a case study EMU, the Southern Rocky Mountains. For EMU- and range-level trend power curves we used the most stringent criteria (Test C) to assure the model and sample size's ability to detect the trend of the size required by the Recovery Plan.

Optimizing study design with respect to costs

We used a constrained optimization framework (Taha 2017) to determine optimal study design with decision variables, an objective function, and constraints in passive acoustic monitoring (Sanderlin et al., *Under Review*). Our objective function was to minimize overall project cost with respect to the constraint of required sample size to achieve power.

Our decision variables (Table 3) included combinations of biological and sampling-related factors. Biological decision variables included at what spatial scale the monitoring program would be implemented (at the EMU level or the range-wide level), as well as the eight biological scenarios described in the previous sections where initial occupancy, survival rates, or detection varied. Sampling-related decision variables included how hexes were distributed with respect to distance to road (i.e., travel costs) and vegetation cover type (i.e., different vegetation types tend to have different levels of accessibility), number of observers (10 or 12), observer types (volunteer, paid seasonal, in-kind government employee), number of ARUs per hex (2 or 4), and number of repeat deployments per year (1, 2, or 3). Number of deployments denotes the timing of ARU placement, with a single deployment indicating all hexes are sampled at the same time, and >1 deployments indicating that a smaller number of ARUs are deployed multiple times in the season to achieve full coverage of sample units. We limited maximum deployments in a year to three as a realistic possible number of deployments within the MSO breeding season.

We used estimated costs and effort from our pilot study (Sanderlin et al. 2025) to populate a series of cost equations incorporating initial project start-up costs, total annual costs, sampling unit costs, and sampling day costs that encompass our cost function (Sanderlin et al., *Under Review*). Effort and costs were modified within the cost function to best approximate the expected effort with range-wide or EMU-level monitoring for Mexican spotted owls (Table S2). We modeled sampling effort to mirror the simulation process with 25% of the samples allocated to low quality (or marginal) sites, and equal numbers of samples distributed across all EMUs for the range-wide implementation. All samples were selected using simple random sampling after the sample sizes were determined from the power threshold step above. We used minimum distance to road within a hex and vegetation cover type (mixed conifer or not) across the sampling frame as indicators of sampling time. To capture variation in the site selection process, we used 100 replicates of each set of decision variables.

Results

Our identified sampling frame covered 19% of the EMU area, with most excluded areas being either non-forest or Pinyon-Juniper landcovers that MSO do not typically inhabit. Forty-three percent of the sampling frame was considered plausibly "high" occupancy based on landcover type (Fig. 3). Sampling frame area was concentrated in the Upper Gila Mountains and Southern Rocky Mountains EMUs, with more sparse sampling areas across the other three EMUs. The sampling frame contained 95% of Protected Activity Center area designated in USDA Forest Service land in New Mexico, Colorado, and Arizona, giving a strong validation of our landcover criteria.

All three trend simulation approaches were successful in producing a true decline approximating the desired magnitude (Fig. 4), but with different attributes. Generally, the equilibrium trend generation approach was more likely to generate a smaller magnitude decline than intended and have either concave or convex non-linear attributes. Percent decline was smaller for lower initial survival values, and the trend was more likely to be insufficiently negative for lower initial occupancy values. The constant survival approach showed a strong non-linearity with most of the decline occurring in the first few years and then tapering off. Similarly to the equilibrium approach, it did not reach the full decline, particularly for low initial occupancy and high initial survival scenarios. In contrast, the recursive trend generation approach consistently resulted in a 25% decline, was not sensitive to initial occupancy or initial survival, and was strongly linear. In fact, it reproduced the trend so robustly that the two initial survival scenarios overplot on Figure 4.

Recommended sample size to meet the power threshold showed sometimes substantial sensitivity to trend generation approach (Fig. 5). While curves (Fig. 5a) were qualitatively similar across the trend approaches, direct comparison of the differences in threshold sample size across approaches showed differences of hundreds of samples for multiple scenarios (Fig. 5b). The largest difference was for scenario 5 (high initial occupancy, low survival, low detection probability), which showed a difference of over 500 samples necessary to detect the trend, with recursive requiring the most, and constant survival requiring the least. Low detection probability scenarios showed qualitatively more variation in sample size across trend approach than their high detection probability counterparts. There did not appear to be a pattern in which approach required fewer or more samples for a give biological scenario. Sample size also showed sensitivity to the power check criteria, with Test A (no check for true trend) and Test C (two tailed, check for true trend) giving similar sample size recommendations and Test B (left-tailed) requiring substantially more. Power curves for the three trend generation approaches and the three test criteria are available in the supplement (Fig. S1).

The sample size required to achieve 90% power to detect the trend varied across biological scenarios and EMUs (Fig. 6, Table 2). Higher detection probability scenarios generally required lower sample sizes across all EMUs. High initial occupancy and high detection probability scenarios all required the lowest sample size (n = ~400-500; Table 2). High detection with lower initial occupancy scenarios required for some twice as many samples to achieve required power (n = ~680-880; Table 2) as their high initial occupancy counterparts. Lower probability of detection also increased the necessary sample size by a similar magnitude. Low initial occupancy, low detection probability scenarios required the most samples, meeting power criteria between n = ~1300-1650 (Table 2). Initial survival parameterization showed small and inconsistent impacts on required sample size. Two EMUs, Basin and Range East and Basin and Range West, did not have sufficient sample area to reach the power threshold for all scenarios, indicating that identifying EMU-level trends in those areas will only be possible under specific biological scenarios.

Meeting the power requirement to detect a trend at the range-wide level, using a hierarchical model, required marginally more samples for a given scenario than any given EMU-level trend for the same scenario. The smallest required sample size was for scenario 6 (high initial occupancy, low survival, high detection probability), needing only n = 423 samples. In contrast, the highest scenario required almost three times more at n = ~1200 samples.

Optimal study designs with respect to cost differed by simulation scenario, trend scale (EMU- vs range-level), number of deployments, number of observers, and types of observers (Fig. 7, Fig. S2). Optimal designs across biological simulation scenarios mirrored sample size threshold results (Fig. S2), with higher initial occupancy, higher survival, and higher detection probability costing less than scenarios with lower initial occupancy, lower survival, and lower detection probability. Costs were similar for range-wide designs versus EMU-level designs. We used scenario 8 as a case study to visualize potential trade-offs in cost-efficiency with number of deployments and number and type of observers (Fig. 7). For most sampling and personnel scenarios, a design involving two deployments was optimal; however, for higher-cost personnel implementations (0 volunteers, 8 paid, 2 government employees; 2 volunteers, 6 paid, 2 government employee) a single deployment was optimal for most EMUs and the range-wide scale only if two ARUs were used per hex. For the lowest cost personnel scenario (8 volunteers, 2 paid, 2 government employees), three deployments were always optimal. This was also true for other low-cost personnel implementations if four rather than two ARUs were used per hex.

Discussion

We explored the implications of trend generation approach, power check criteria, and initial simulation parameterization on the minimum sample size required to achieve power, with acoustic monitoring of the Mexican spotted owl as a motivating case study. We demonstrated that decision making at each of these steps can result in variation in the recommended sample size for a monitoring program that represents significant differences in cost and optimal implementation strategy. We assessed multiple potential options for generating trends with different properties and outlined reproducible closed-form methods for selecting appropriate parameterizations for a given trend magnitude. Our results therefore illustrate not just the critical importance of intentionality and transparency in decision making throughout the power analysis process but also outline methods that facilitate that goal.

Variation in sample size due to the factors explored here translates to differences in real-world cost on a scale that could determine whether implementation of a monitoring program is financially feasible or not. For example, in the most extreme case of differences due to trend generation approach, scenario 5 showed a difference of 500 samples between the equilibrium and constant survival approaches (Fig. 5b). If a single ARU unit costs ~$450 and each sample hex requires two ARUs, selecting one trend generation approach over another could result in an additional $450,000 in hardware costs alone – and this estimate does not consider implementation costs and the accumulating costs of hardware attrition over time. This simple calculation demonstrates that selection of the right trend generation approach can have substantial implications for real-world costs.

As our results indicate that selection of the trend generation approach is nontrivial for sample size recommendation and resulting costs, how then should practitioners choose the most appropriate method? For the power analysis to be a fair test of the monitoring program's goal, the properties of the generated trend (e.g., curvature, variation in trend magnitude across replicates, etc.) should align as closely as possible with the biology of the target species, the ecological processes of interest, and the policy context. We chose the recursive method for our sample size recommendations as its ability to robustly generate a linear trend of the specified magnitude most closely aligned with the requirements of the Recovery Plan. In other settings, practitioners may choose an approach based on *a priori* knowledge of the most biologically appropriate trend shape for a target species or ecological process. For example, declines driven by changes in recruitment success by the species may be better matched to the non-linear patterns generated by the constant survival approach. We also acknowledge that the methods explored here are not a comprehensive list of possible trend generation approaches, as we focused on methods with a closed form derivation of initial parameters (rather than "guess and check"). Generation of trends with different characteristics than the ones identified here may be necessary for other applications. For example, if the goal is to detect trends responding to an acute disturbance, a discontinuous curve could be generated by a sudden or one-time drop in survival probabilities. Given that the trend properties are not necessarily intuitive, we echo other calls for visualization of the generated true trend as a standardly reported product alongside the power analysis curve and sample size recommendation to ensure properties are well matched to the goal of interest (Banner et al. 2019).

Sensitivity of sample size to the simulation's biological parameters illustrates the key role of *a priori* natural history knowledge of the target species (including pilot data, wherever possible) in producing a power analysis with sufficient biological realism to be useful. Sample sizes required to reach power for different biological scenarios ranged from ~400-1650 hexes required for EMU-level trends and ~520-1200 hexes required for a range-wide trend. While detection probability is often presumed to be the most influential factor for required sample size, we found that initial occupancy and survival had similar magnitude impacts on recommended sample size, indicating that knowledge of a target species' parameters is critical for appropriate design (Mackenzie & Royle 2005; Mckann et al. 2013; Sanderlin et al. 2014). Power analyses should therefore ideally be informed by robust pilot data on the current population parameters or be well justified by expert knowledge. Empirical data can also add useful realism to other dimensions of the power analysis. For example, landcover data used to define potential habitat availability put upper bounds on the number of samples available for each EMU, illustrating biological scenarios for which some EMUs may not have a detectable trend and thus the kind of inference that could be drawn by the program in the real world. As another example, we considered four combinations of occupancy and detection to bracket a wide range of plausible realities, which expanded our battery of simulations substantially. If pilot data can home in on fewer, more realistic values for occupancy, detection, and even temporal parameters like survival, then fewer scenarios may need to be considered. Our analysis is therefore a compelling illustration of the volume and depth of data needed to generate a power analysis that moves beyond academic exercise to being informative for implementation.

There are multiple possible routes for selecting a final sample size recommendation for implementation. One of the strengths of our approach is that it allows for flexibility in biological reality. We can therefore select an initial sample size threshold based on our best current knowledge of MSO biological parameters and either intensify or reduce sample size in subsequent years based on estimates from those data, an approach advocated for in adaptive management (Moir & Block 2001; Williams 2011; Sanderlin et al.

2019). While we included lower detection probability scenarios to explore the threshold space, pilot data (described in Sanderlin et al., 2025) show that ARUs can achieve at least 0.8 probability of detection and would therefore likely start with a recommendation from one of those scenarios. By exploring power to detect a trend across both EMU and range-wide scales, we also identified potential flexibility in how samples are allocated. For example, the sample size required to detect an EMU-level trend for scenarios 6 and 8 (high initial occupancy, high detection probability) is close to the recommendation for the range-level trend for those same scenarios. The same hardware investment could therefore be used to assess a pilot EMU and then reallocated across the range to assess the range-level trend as the implementation logistics are honed and scaled up.

Our results indicate an opportunity to optimize monitoring implementation based on the recommended sample size and landscape characteristics, as recommendations for optimal strategy varied by implementation scenarios and trend of interest (individual EMU or range-wide). Optimal sample design for a given EMU-level or range-wide trend differed by the number of acoustic recorders per hex, observer types, number of observers, and number of deployments. Given the assumptions of the cost equations, we found a trade-off between the number of deployments and number and types of observers for the optimal design. This tradeoff is likely due to relative costs of acoustic recorder units compared to observer wages, and shifts comparing between two versus four acoustic recorder units per hex. While the cost optimization analysis gives a strong first indication of which trends may be most costly to identify and how implementation should occur to minimize those costs, incorporation of further realism could improve recommendations. For example, in our simulations we assumed that additional deployments did not affect detection probability and therefore that detectability is constant throughout the breeding season. This assumption is supported by some work with California spotted owl (Winiarski et al. 2026, but see Reid et al. 2022), and would have meaningful implications for the interaction between implementation strategy and biological parameters of the underlying power simulations.

The program described here is designed to meet specific criteria for detecting a decline as described in the Recovery Plan. However, the Recovery Plan also includes more ambiguous language for meeting recovery goals regarding the programs ability to demonstrate a “stable or increasing trend”. Given that it is generally easier to identify larger magnitude changes, with a 25% decline being a substantial change, how does this power analysis ensure the ability to detect more subtle declines, occupancy stability, or increases? One of the strengths of the Bayesian approach used here is that it produces a full posterior distribution for occupancy at each time step as well as the derived percent change posterior. We can therefore say with what percent confidence a population has increased or remained stable based on how much of the posterior is greater than zero or falls within a predefined “stability” range respectively. The approach here is also easily adapted to changes in the criteria for success, with potential options including generating an increasing trend of various sizes, or stable populations, and comparing sample sizes for those scenarios to the ones examined here to determine the minimum threshold to detect all three potential occupancy trajectories.

While we incorporated the major costs associated with implementing a program of this scale, we did not factor in one of the major benefits of an acoustic monitoring program, which is the potential for collecting data on non-target species (e.g., Brunk et al., 2025; Sanderlin et al., 2025). Multi-species monitoring is increasingly the gold standard of monitoring implementation as it efficiently gathers data on multiple species of interest simultaneously. The program described here covers a substantial geographic and environmental space, making it particularly well suited for capturing data on many other forest-dwelling species and providing the opportunity to calculate community-level metrics related to ecological integrity (Miller-ter Kuile et al. 2025). Land management agencies like the USDA Forest Service are required to monitor focal species outlined in forest planning documents – many of which are vocalizing birds such as great-horned owls (*Bubo virginianus*), American goshawk (*Astur atricapillus*), and Grace's warbler (*Setophaga graciae*). While not specifically targeted, these species would be 'bycatch' in our MSO-focused monitoring program, and similar trend analysis could be implemented for these non-target species. For species that have initial occupancy, survival, and detectability within the range explored here, our analysis can also say with what likelihood we can detect a 25% decline in their populations, as the power analysis is technically species-agnostic. Of course, multi-species passive acoustic monitoring does come with some additional non-field costs, mostly relative to data storage at both the ARU-level, where the ARU may be recording for longer time periods to capture the active periods of diurnal birds, and the data processing level, where additional hours of data must be stored long-term and validated. Reaching sufficient detectability for non-MSO species may necessitate additional ARUs per sample unit, particularly for rare species or species with smaller home ranges than the hex, with additional implications for cost and optimal implementation strategy.

In addition to making recommendations for sampling and implementation decisions, we found here compelling evidence that power analysis sample size thresholds are highly sensitive to parameterization of the occupancy process and trend generation simulation approach, and that those differences impact the cost-optimal implementation strategy for a given monitoring plan. These findings imply that, despite perception of power analysis as a straightforward procedure in the study or monitoring design process, poorly justified decisions at any point in the process risk misallocation of conservation funds on the scale of hundreds of thousands of dollars. In settings where robustness of produced data may be challenged in a court of law (e.g., for the listing or delisting of a species), inappropriate or incomplete power analysis implementation is an underappreciated point of vulnerability. Future work can shore up this vulnerability by homing in on the decision points with greatest sensitivity, helping practitioners identify the *a priori* information most critical to accurately identify sample size.

References

Appel CL, Lesmeister DB, Duarte A, Davis RJ, Weldy MJ, Levi T. 2023. Using passive acoustic monitoring to estimate northern spotted owl landscape use and pair occupancy **14**:e4421.

Banner KM, Irvine KM, Rodhouse TJ, Donner D, Litt AR. 2019. Statistical power of dynamic occupancy models to identify temporal change: Informing the North American Bat Monitoring Program. Ecological Indicators **105**:166–176.

Brunk KM et al. 2025. Bioregional-scale acoustic monitoring can support fire-prone forest restoration planning. Frontiers in Ecology and the Environment **23**:e2843.

de Valpine P, Turek D, Paciorek CJ, Anderson-Bergman C, Lang DT, Bodik R. 2017. Programming With Models: Writing Statistical Algorithms for General Model Structures With NIMBLE. Journal of Computational and Graphical Statistics **26**:403–413. Taylor & Francis.

Ellis MM, Ivan JS, Schwartz MK. 2014. Spatially Explicit Power Analyses for Occupancy-Based Monitoring of Wolverine in the U.S. Rocky Mountains. Conservation Biology **28**:52–62.

Franklin AB, Anderson DR, Forsman ED, Burnham KP, Wagner FW. 1996. Methods for collecting and analyzing demographic data on the northern spotted owl. Studies in Avian Biology **17**:12–20.

Goldstein BR, Keller AG, Calhoun KL, Barker KJ, Montealegre-Mora F, Serota MW, Van Scoyoc A, Parker-Shames P, Andreozzi CL, de Valpine P. 2024a. How do ecologists estimate occupancy in practice? Ecography:e07402.

Goldstein BR, Turek D, Ponisio L, Valpine P de. 2024b. nimbleEcology: Distributions for Ecological Models in nimble. Available from https://cran.r-project.org/package=nimbleEcology.

Gonzalez A et al. 2023. A global biodiversity observing system to unite monitoring and guide action. Nature Ecology & Evolution **7**:1947–1952. Nature Publishing Group.

Griffith J et al. 2024. BON in a Box: An Open and Collaborative Platform for Biodiversity Monitoring, Indicator Calculation, and Reporting. EcoEvoRxiv. Available from https://ecoevorxiv.org/repository/view/7941/ (accessed November 5, 2024).

Gustafson ML, McGinn K, Heys JA, Sawyer SC, Wood CM. 2025. Rapid implementation and adaptive design of a large-scale monitoring program for a declining species. Biological Conservation **311**:111442.

Gutierrez RJ, Franklin AB, LaHaye WS. 1995. Spotted Owl (Strix occidentalis): Life Histories for the 21st Century. Page in Poole A, Gill F, editors. The Birds of North America No. 179. The Philadelphia Academy of Sciences and The American Ornithologists' Union, Washington, D. C.

Hughes BB et al. 2017. Long-Term Studies Contribute Disproportionately to Ecology and Policy. BioScience **67**:271–281.

Jones GM, Keane JJ, Gutiérrez RJ, Peery MZ. 2018. Declining old-forest species as a legacy of large trees lost. Diversity and Distributions **24**:341–351.

Kelly KG et al. 2023. Estimating population size for California spotted owls and barred owls across the Sierra Nevada ecosystem with bioacoustics. Ecological Indicators **154**:110851.

Leung B, Gonzalez A. 2024. Global monitoring for biodiversity: Uncertainty, risk, and power analyses to support trend change detection. Science Advances **10**:eadj1448. American Association for the Advancement of Science.

MacKenzie DI, Nichols JD, Hines JE, Knutson MG, Franklin AB. 2003. Estimating Site Occupancy, Colonization, and Local Extinction When a Species Is Detected Imperfectly. Ecology **84**:2200–2207.

MacKenzie DI, Nichols JD, Lachman GB, Droege S, Royle JA, Langtimm CA. 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology **83**:2248–2255.

Mackenzie DI, Royle JA. 2005. Designing occupancy studies: general advice and allocating survey effort. Journal of Applied Ecology **42**:1105–1114.

McGinn K et al. 2025. Frequent, heterogenous fire supports a forest owl assemblage. Ecological Applications **35**:e3080.

Mckann PC, Gray BR, Thogmartin WE. 2013. Small sample bias in dynamic occupancy models. The Journal of Wildlife Management **77**:172–180.

Miller-ter Kuile A et al. 2025. Functionalizing ecological integrity: using functional ecology to monitor animal communities. Frontiers in Ecology and the Environment **e2852**:e2852.

Moir WH, Block WM. 2001. Adaptive management on public lands in the United States: commitment or rhetoric? Environmental Management. 28(2): 141-148.:141–148.

Moussy C et al. 2022. A quantitative global review of species population monitoring. Conservation Biology **36**:e13721.

Norman KE, Poisot T. 2025. Algorithm selection for optimal ecological monitoring design. EcoEvoRxiv. Available from https://ecoevorxiv.org/repository/view/8545/ (accessed May 15, 2025).

Popescu VD, de Valpine P, Tempel D, Peery MZ. 2012. Estimating population impacts via dynamic occupancy analysis of Before–After Control–Impact studies. Ecological Applications **22**:1389–1404.

Reid DS et al. 2022. Breeding status shapes territoriality and vocalization patterns in spotted owls. Journal of Avian Biology **2022**:e02952.

Royle JA, Kéry M. 2007. A Bayesian State-Space Formulation of Dynamic Occupancy Models. Ecology **88**:1813–1823.

Sanderlin JS et al. 2025. EXPANDING BIODIVERSITY MONITORING IN THE SOUTHWESTERN UNITED STATES WITH PASSIVE ACOUSTIC MONITORING. The Southwestern Naturalist **69**:1–11. Southwestern Association of Naturalists.

Sanderlin JS, Block WM, Ganey JL. 2014. Optimizing study design for multi-species avian monitoring programmes. Journal of Applied Ecology **51**:860–870.

Sanderlin JS, Morrison ML, Block and WM. 2019. 8. Analysis of Population Monitoring Data. Pages 131–148 Quantitative Analyses in Wildlife Science. Johns Hopkins University Press. Available from https://muse.jhu.edu/pub/1/edited_volume/chapter/2383403/pdf (accessed February 18, 2026).

Taha HA. 2017. Operations research an introductionTenth edition. Pearson, Boston.

Tempel DJ et al. 2016. Meta-analysis of California Spotted Owl (Strix occidentalis occidentalis) territory occupancy in the Sierra Nevada: Habitat associations and their implications for forest management. The Condor: Ornithological Applications **118**:747–765.

U.S. Fish and Wildlife Service. 2012. Final Recovery Plan for the Mexican Spotted Owl (Strix occidentalis lucida), First Revision. U.S. Fish and Wildlife Service. Albuquerque, New Mexico, USA.

Weiser EL, Diffendorfer JE, Lopez-Hoffman L, Semmens D, Thogmartin WE. 2021. TrendPowerTool: A lookup tool for estimating the statistical power of a monitoring program to detect population trends. Conservation Science and Practice **3**:e445.

White ER. 2019. Minimum Time Required to Detect Population Trends: The Need for Long-Term Monitoring Programs. BioScience **69**:40–46.

Williams BK. 2011. Adaptive management of natural resources—framework and issues. Journal of Environmental Management **92**:1346–1353.

Williams BK, Nichols JD, Conroy MJ. 2002. Analysis and Management of Animal Populations: Modeling, Estimation and Decision Making. Academic Press. Available from https://pubs.usgs.gov/publication/5200256 (accessed February 18, 2026).

Winiarski JM et al. 2026. Passive acoustic monitoring can provide insights into occupancy dynamics and impacts of disturbance for at-risk species. Ecological Applications **36**:e70177.

Wood CM, Popescu VD, Klinck H, Keane JJ, Gutiérrez RJ, Sawyer SC, Peery MZ. 2019. Detecting small changes in populations at landscape scales: a bioacoustic site-occupancy framework. Ecological Indicators **98**:492–507.

Table 1. Occupancy parameters for each of the 8 simulation scenarios, with colonization and initial occupancy split by whether the hex is a plausibly low or high occupancy landcover type. Standard deviations are fixed, with $\boldsymbol{\gamma}_{s.d.}$ = 0.01, and $\boldsymbol{\phi}_{s.d.}$ = 0.04.

Scenario	Survival (ϕ)	Detection (p)	Colonization (γ)		Initial Occupancy (Ψ)	
			Low	*High*	*Low*	*High*
1	0.6	0.4	0.01	0.30	0.03	0.43
2	0.6	0.8	0.01	0.30	0.03	0.43
3	0.8	0.4	0.01	0.15	0.03	0.43
4	0.8	0.8	0.01	0.15	0.03	0.43
5	0.6	0.4	0.01	0.60	0.03	0.60
6	0.6	0.8	0.01	0.60	0.03	0.60
7	0.8	0.4	0.01	0.30	0.03	0.60
8	0.8	0.8	0.01	0.30	0.03	0.60

Table 2. Estimated threshold sample size to detect the trend for different biological scenarios (rows), trend scale (EMUs and range-wide model), and the three distinct trend generation approaches (implemented for the Southern Rocky Mountains EMU). Abbreviations: BRE = Basin and Range East, BRW = Basin and Range West, CP = Colorado Plateau, SRM = Southern Rocky Mountains, UGM = Upper Gila Mountains.

Scenario	BRE	BRW	CP	SRM	UGM	Range trend	Constant survival	Equilibrium	Recursive
1	-	-	1441	1476	1650	1200	1618	1400	1476
2	777	-	746	775	782	938	820	760	775
3	-	-	1304	1538	1538	-	1786	1921	1538
4	680	-	880	825	800	800	908	1000	825
5	904	-	1083	1200	987	1071	838	695	1200
6	409	412	402	412	451	586	434	448	412
7	815	-	931	973	922	1036	880	1214	973
8	442	408	448	504	408	423	495	445	504

Table 3. Decision variables for passive acoustic monitoring optimal study design.

Variable type	Component	Description	Values
Biological	*spatial scale*	Spatial scale of simulation	individual EMUs, range-wide (hierarchical')
Biological	*simulation scenario*	biological scenarios described in Table 1	scenario 1 (Ψ=0.43, φ=0.6, p=0.4), scenario 2 (Ψ=0.43, φ=0.6, p=0.8), scenario 3 (Ψ=0.43, φ=0.8, p=0.4), scenario 4 (Ψ=0.43, φ=0.8, p=0.8), scenario 5 (Ψ=0.60, φ=0.6, p=0.4), scenario 6 (Ψ=0.60, φ=0.6, p=0.8), scenario 7 (Ψ=0.60, φ=0.8, p=0.4), scenario 8 (Ψ=0.60, φ=0.8, p=0.8)
Sampling	*site selection*	Distribution of hexagonal units with simple random sample	Sample sizes determined from power thresholds and sample influences road distance and vegetation cover type
Sampling	*observers*	Number and type of observers (volunteer, paid technician, governmental employee)	10 obs (0 volunteer, 0 paid, 10 GE), 10 obs (0 volunteer, 8 paid, 2 GE), 12 obs (10 volunteer, 0 paid, 2 GE), 10 obs (2 volunteer, 6 paid, 2 GE), 10 obs (4 volunteer, 4, paid, 2 GE), 12 obs (8 volunteer, 2 paid, 2 GE)
Sampling	*n ARUs*	Number of acoustic recorder units per hexagonal unit	2, 4
Sampling	*n deployments*	number of deployments in each year	1, 2, 3

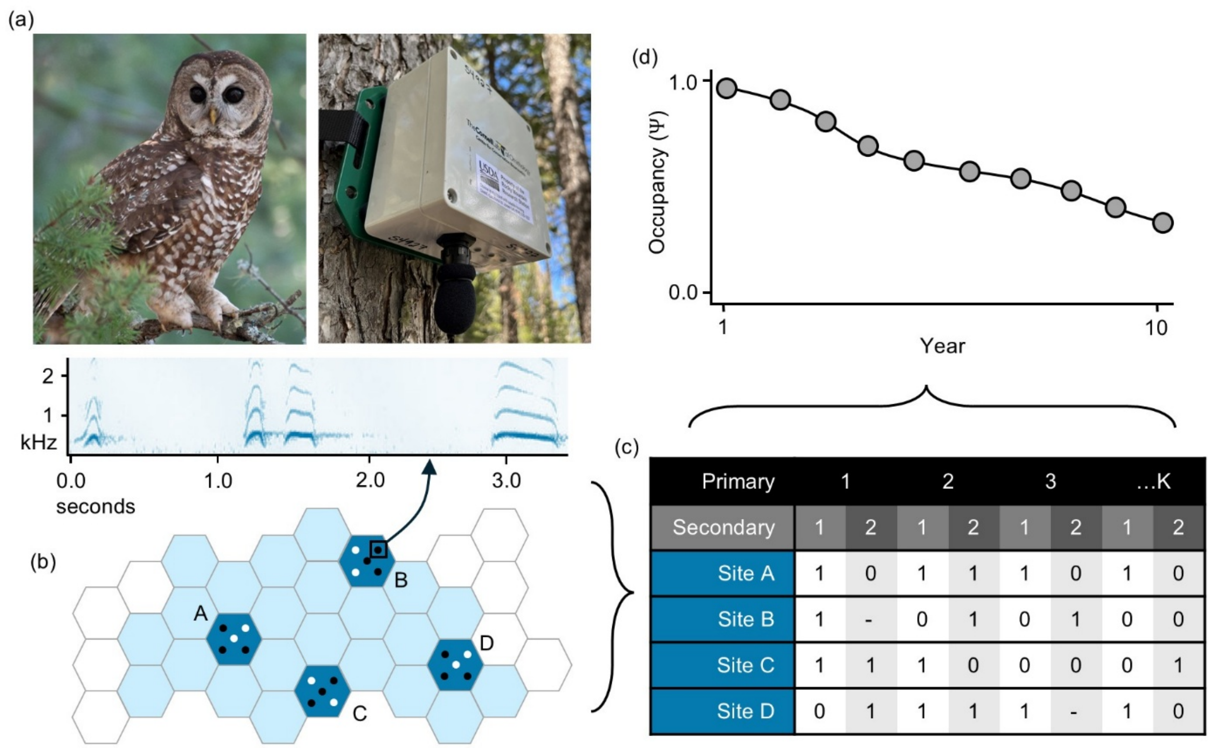

Figure 1. Conceptual diagram illustrating the passive acoustic monitoring approach. (a) Acoustic monitors detect vocalizations of Mexican spotted owls, primarily four-note territorial calls shown in the spectrogram. (b) Three acoustic monitors (black circles) out of five possible sub-sampling locations (black and white circles) are placed within selected hexagonal sampling units (dark blue hexagons). Individual acoustic monitors detect owl calls, and if any individual monitor in a hex has a positive detection, the hexagonal sampling unit is considered to be occupied. (c) Recordings are split into two discrete secondary sampling periods for each primary sampling period, and detection histories across all sampled hexes are (d) aggregated and modeled to estimate the occupancy trend across 10-year monitoring study.

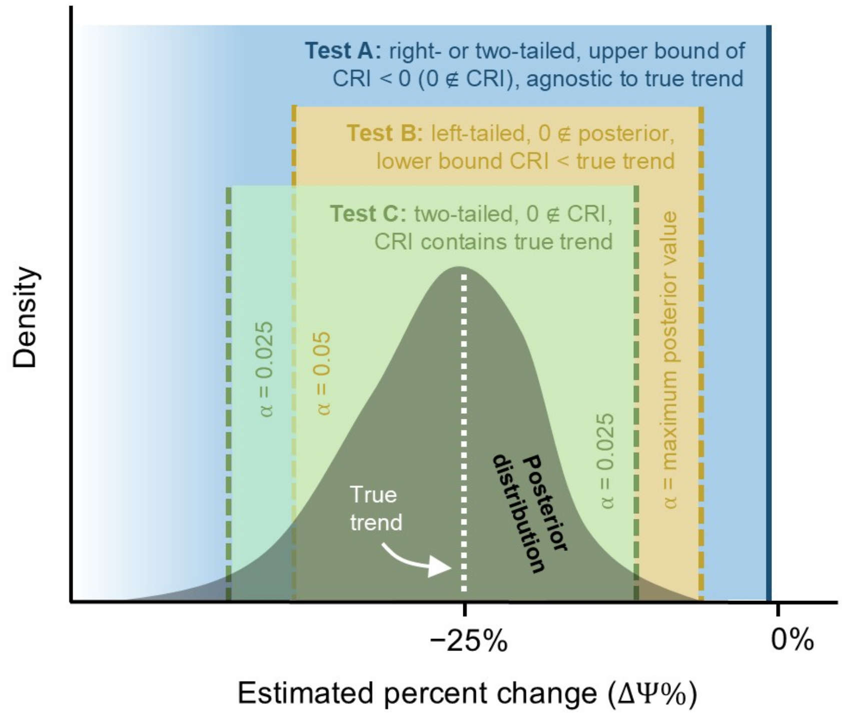

Figure 2. Conceptual figure illustrating the three possible power check tests for identifying model success in detecting the true trend (white dotted vertical line). The posterior distribution for percent change in occupancy is given by the black curve, and vertical colored dashed lines indicate the upper and lower bounds of credible intervals (CRIs) for a two tailed test in green, and a left-tailed test in yellow. For all tests the upper bound of the credible interval must be less than zero to be considered successful in detecting a negative trend. Shaded areas indicated the "region of success" where the true trend must fall for the model to be considered successful in identifying the trend for different tests, with the region bounded by CRI for test B and C (vertical colored dashed lines) and bounded by only zero for Test A (vertical solid colored line). Test A (blue) uses a right-tailed CRI with $\boldsymbol{\alpha} = \mathbf{0.05}$ and requires only that the upper bound of the CI be less than 0. Test B (gold) uses a left-tailed CRI with $\boldsymbol{\alpha} = \mathbf{0.05}$ and requires the maximum value of the posterior distribution is less than zero and that the posterior contains the true trend. Test C (green) uses a two tailed credible interval with $\boldsymbol{\alpha} = \mathbf{0.05}$, where success requires the CRI does not contain 0 and contains the true trend.

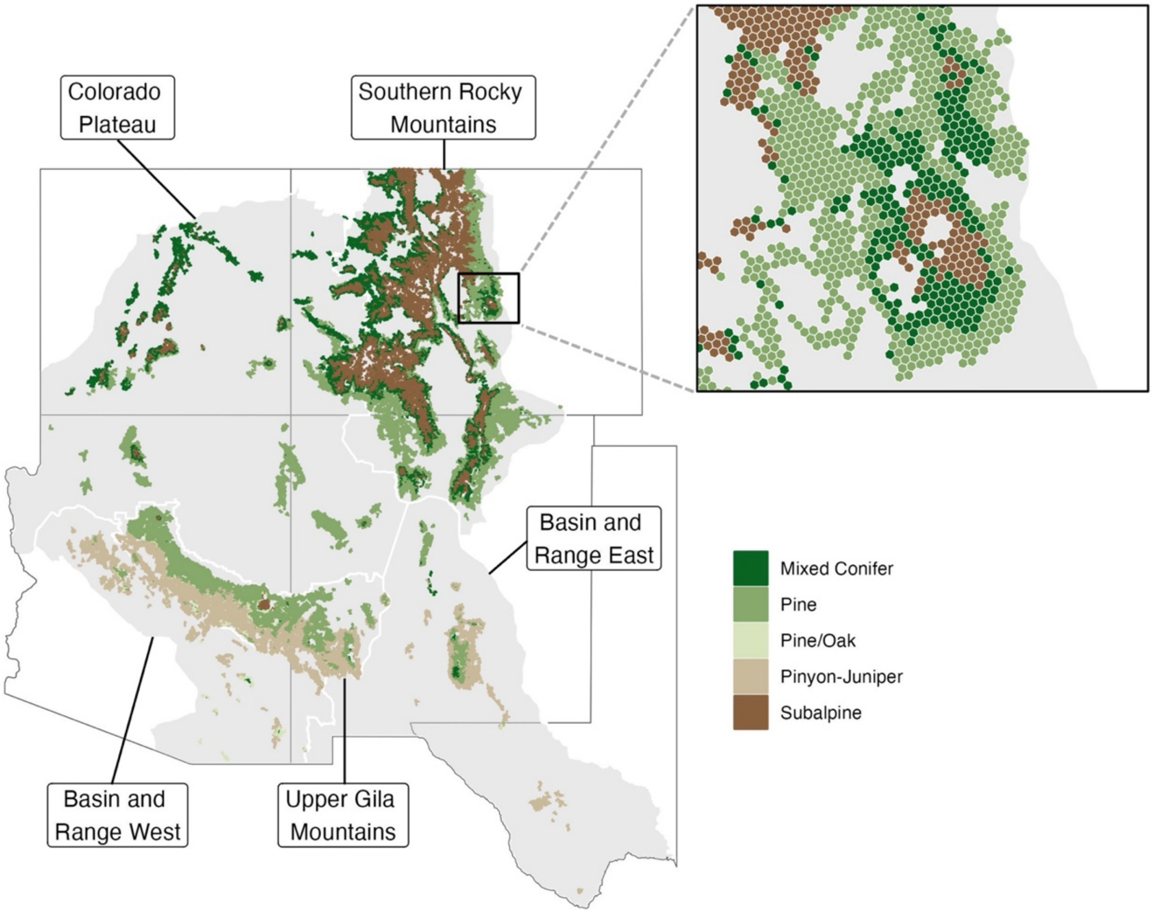

Figure 3. Map showing the five Mexican spotted owl Ecological Monitoring Units and the portion of each unit included in the sampling frame. Included hexes are colored by a simplified landcover type.

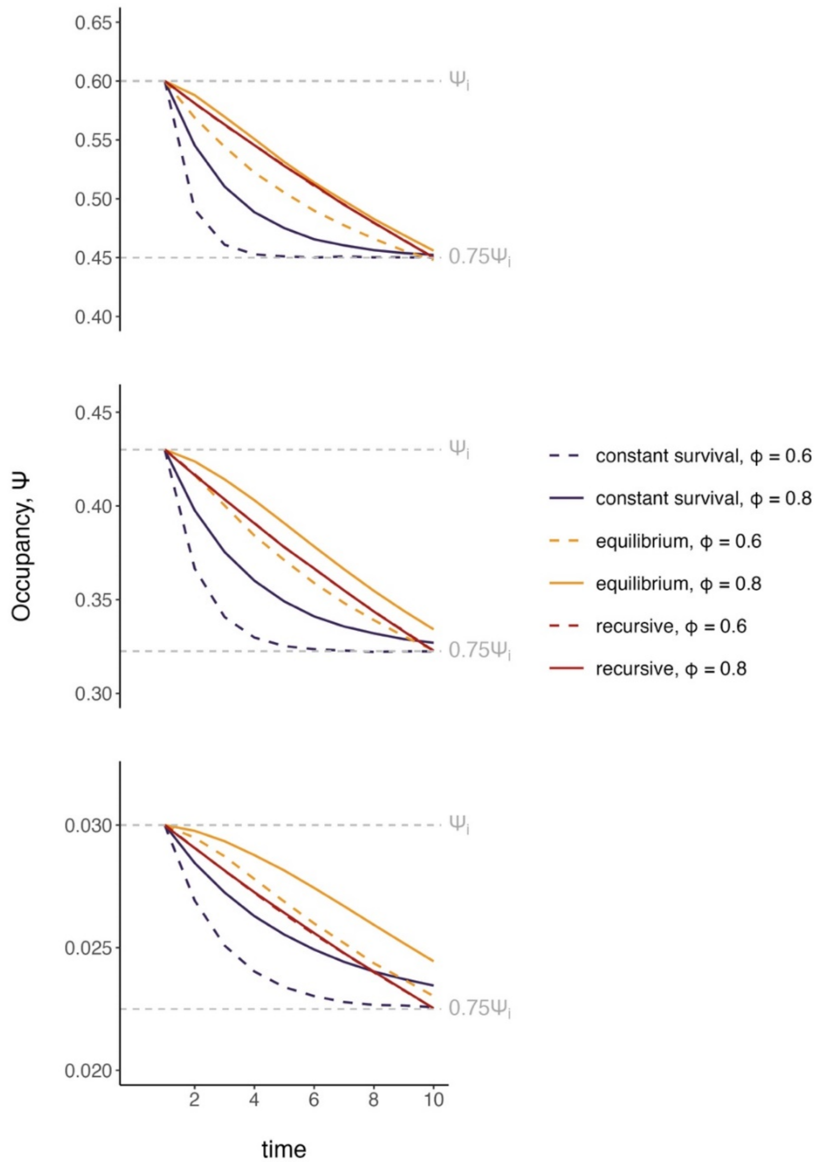

Figure 4. Simulated trends in occupancy for constant survival (purple), equilibrium (orange), and recursive (red) trend generation approaches. Trends are shown for two initial survival values (dashed for 0.6, solid for 0.8) and three different initial occupancy values (vertical panels). The horizontal dashed lines indicate initial occupancy (top line) and final occupancy (bottom line) if a true 25% decline was realized.

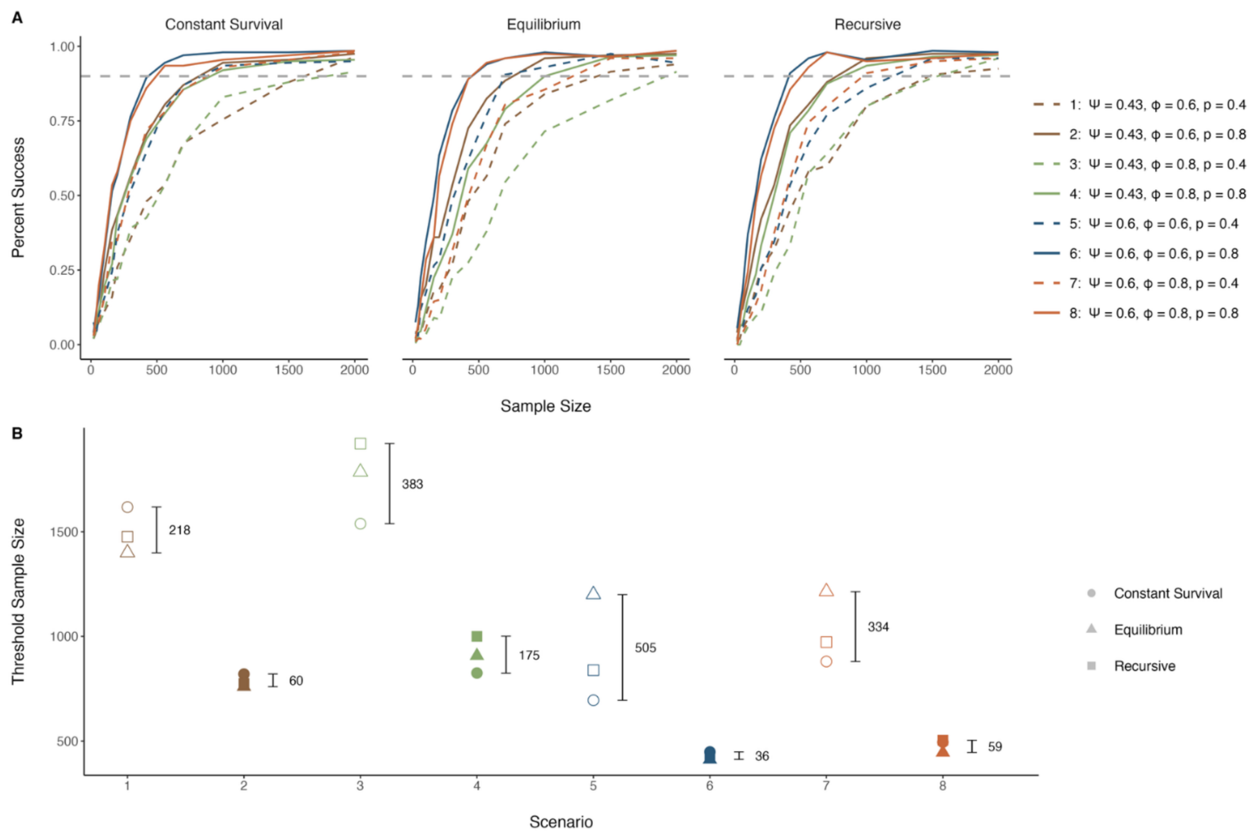

Figure 5. Comparison of power curves for three different trend generation approaches (indicated by individual plots in panel A and shape in panel B) and eight different biological scenarios (indicated by line color and type in panel A and color and shape fill in panel B). Panel A depicts the power curve for all scenarios with the target power level of 0.9 indicated by the horizontal grey dashed line. Panel B compares the sample size at the power threshold for a given scenario across trend generation approaches, with the sample size range for a scenario annotated on the plot. Dashed lines correspond to simulations with lower detection probabilities while solid lines had higher detection probabilities.

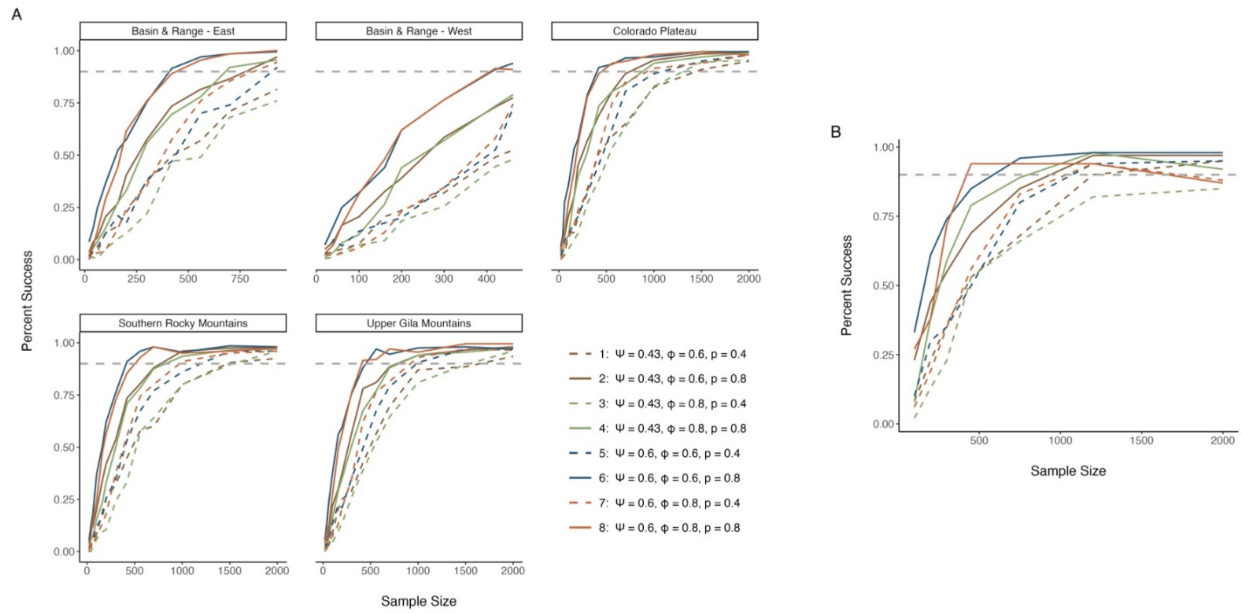

Figure 6. Power curves for trend estimation at the EMU- (panel A) and range-level (panel B) using the recursive trend generation approach for each of the 8 scenarios, with solid lines indicating higher probability of detection and dashed lines indicating lower.

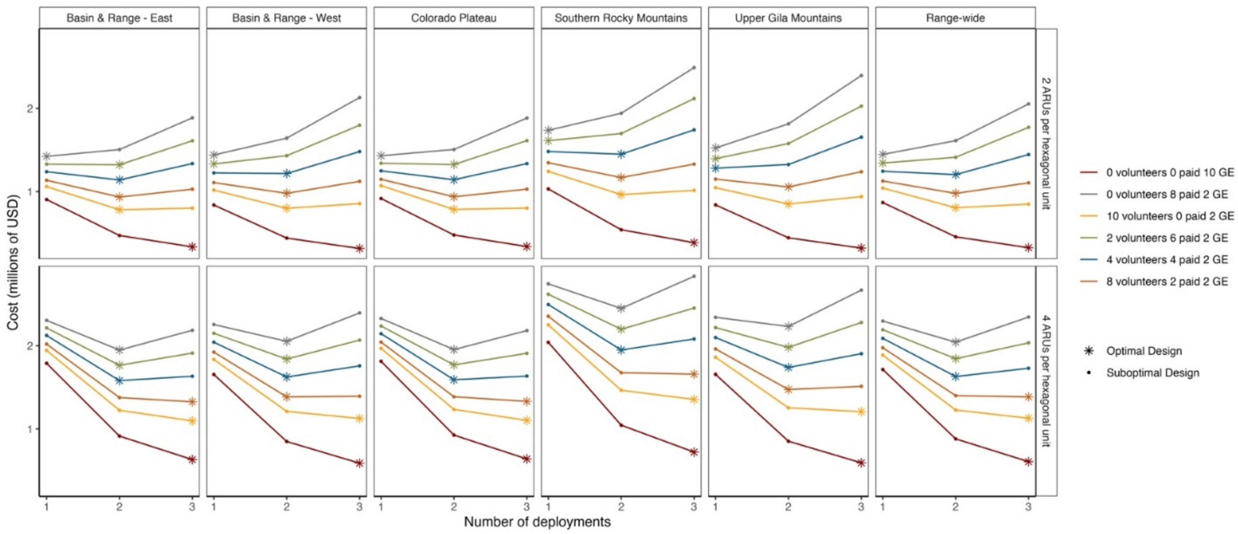

Figure 7. Mean cost of implementation for study designs for scenario 8 under different sampling and personnel implementations (volunteer, paid employee, government employee) for two or four ARUs per hexagonal unit. Each column depicts implementation for an EMU or the entire range, with differences between EMUs (or range-wide) due to both the sample size threshold and distribution of landcover types. The optimal deployment design for each personnel scenario is indicated by a star. Error in mean cost is too small to be visualized and therefore not included in the figure. GE = government employee.

Supplemental Figures and Tables for "Robust and realistic power analysis for cost-effective monitoring of occupancy trends"

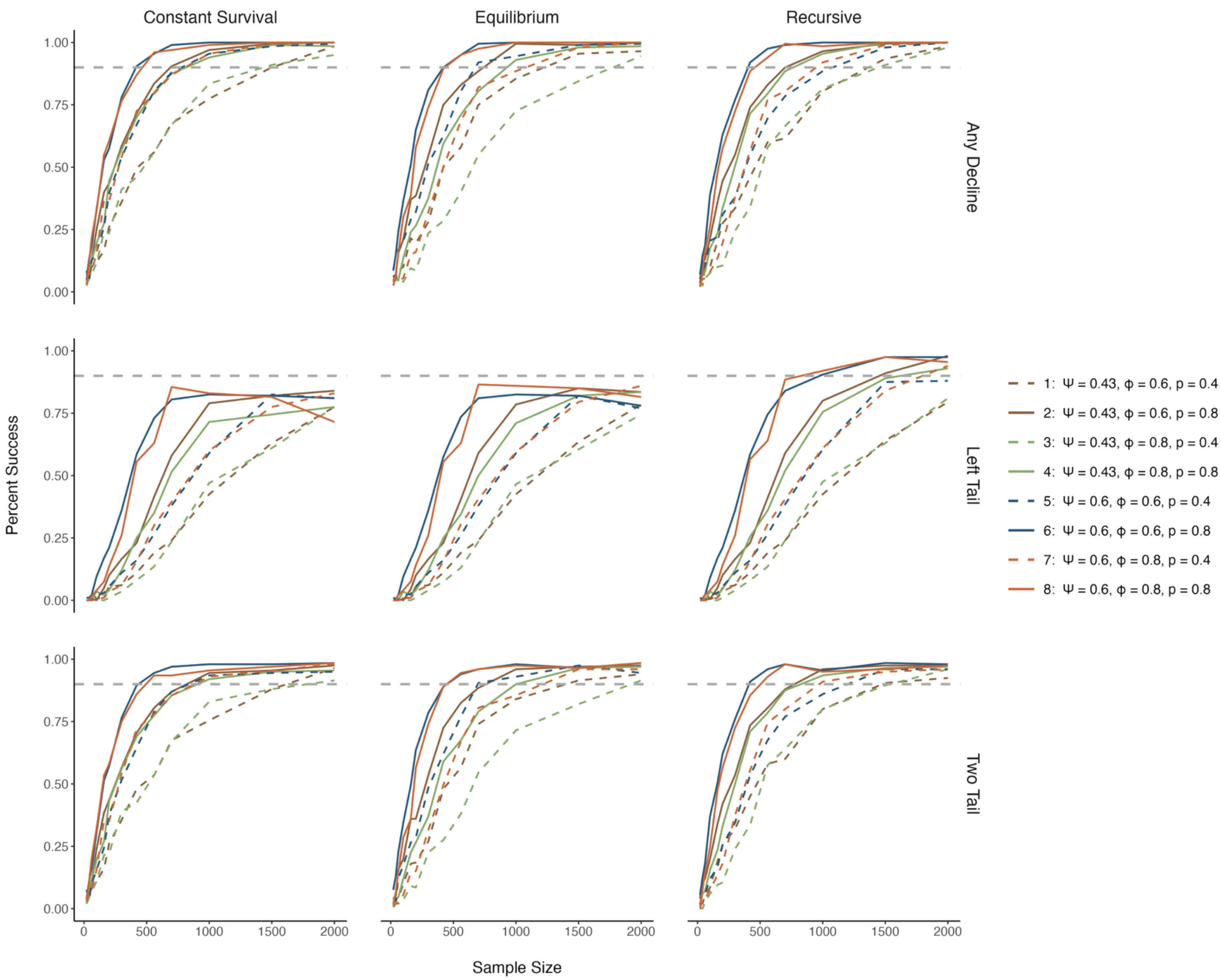

Figure S1. Power curves for each combination of trend generation approach and power check criteria, with line color indicating the biological scenario and line type indicating the detection probability.

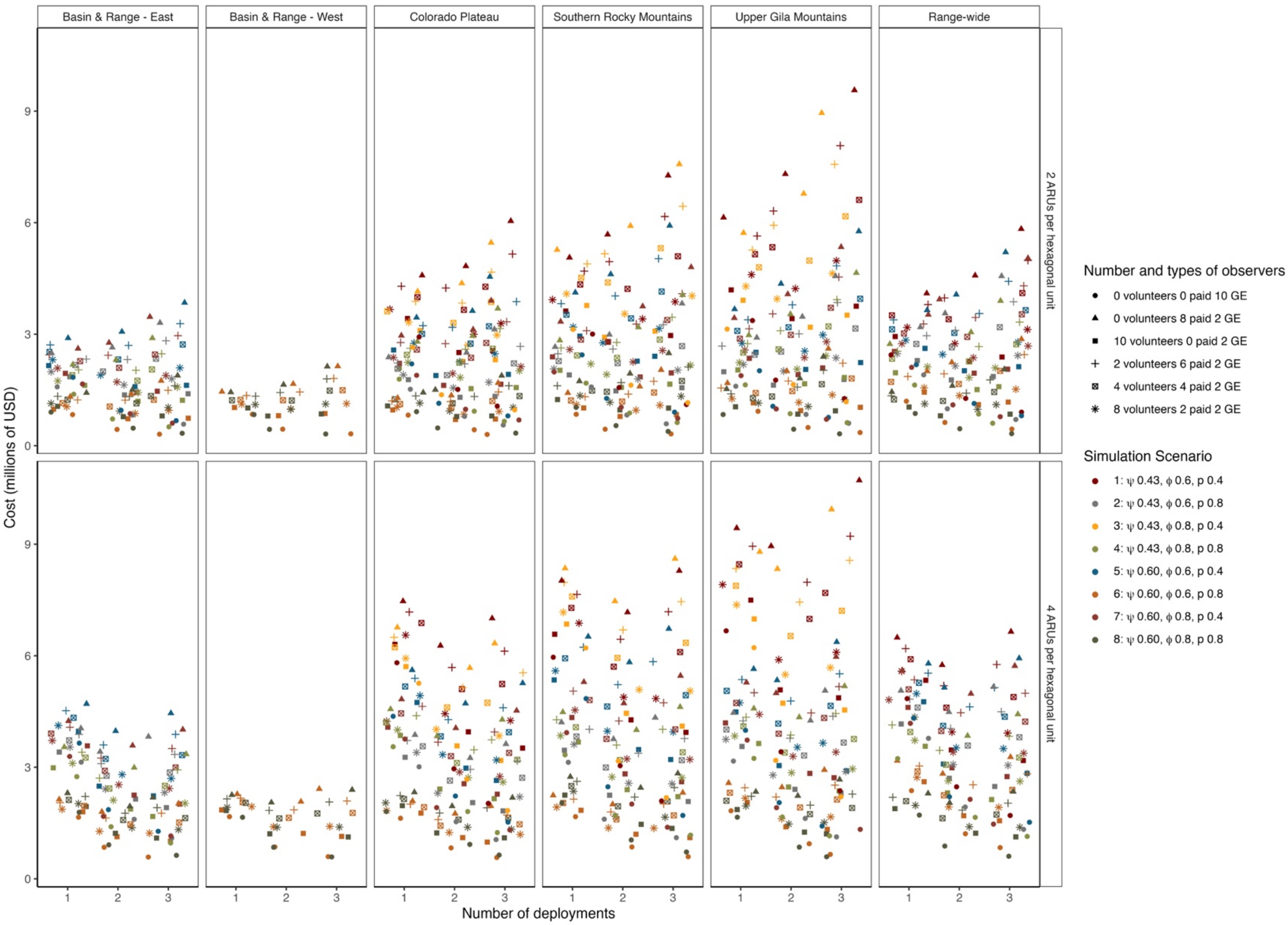

Figure S2. Mean cost of implementation for monitoring programs that meet the sample size threshold for each biological simulation scenario (given by color) under different potential observer (point shape), deployment implementations, and number of ARUs per hexagonal unit (rows). Optimal study design with respect to costs differs by number of deployments and number and types of observers across all simulation scenarios.

LandFire Existing Vegetation Type	Inclusion in sampling frame	Occupancy type
Madrean Pinyon-Juniper Woodland	yes	
Madrean Encinal	no	
Interior West Ruderal Riparian Forest	no	
Madrean Lower Montane Pine-Oak Forest and Woodland	yes	high
Western Warm Temperate Orchard	no	
Southern Rocky Mountain Dry-Mesic Montane Mixed Conifer Forest and Woodland	yes	high
North American Warm Desert Riparian Woodland	no	
North American Warm Desert Riparian Mesquite Bosque Woodland	no	
Southern Rocky Mountain Ponderosa Pine Woodland	yes	low
Madrean Upper Montane Conifer-Oak Forest and Woodland	yes	high
North American Warm Desert Lower Montane Riparian Woodland	no	
Madrean Juniper Savanna	no	
Rocky Mountain Subalpine Dry-Mesic Spruce-Fir Forest and Woodland	yes	low
Colorado Plateau Pinyon-Juniper Woodland	no	
Rocky Mountain Aspen Forest and Woodland	yes	low
Inter-Mountain Basins Aspen-Mixed Conifer Forest and Woodland	yes	low
Southern Rocky Mountain Juniper Woodland and Savanna	no	

Southern Rocky Mountain Mesic Montane Mixed Conifer Forest and Woodland	yes	high
Great Basin Pinyon-Juniper Woodland	no	
Southern Rocky Mountain Pinyon-Juniper Woodland	no	
Southern Rocky Mountain Ponderosa Pine Savanna	yes	low
Inter-Mountain Basins Subalpine Limber-Bristlecone Pine Woodland	yes	low
Inter-Mountain Basins Curl-leaf Mountain Mahogany Woodland	no	
Rocky Mountain Bigtooth Maple Ravine Woodland	yes	low
Rocky Mountain Subalpine Mesic-Wet Spruce-Fir Forest and Woodland	yes	low
Rocky Mountain Subalpine-Montane Riparian Woodland	yes	low
Western Cool Temperate Urban Evergreen Forest	no	
Rocky Mountain Lodgepole Pine Forest	yes	low
Rocky Mountain Subalpine-Montane Limber-Bristlecone Pine Woodland	yes	low
Western Cool Temperate Orchard	no	
Rocky Mountain Foothill Limber Pine-Juniper Woodland	no	
Rocky Mountain Lower Montane-Foothill Riparian Woodland	no	
Western Great Plains Riparian Woodland	no	

Table S1. All landcover types found within the EMU boundaries as classified by the LandFire existing vegetation type data product, with whether they were included in the sampling frame or not, and if so if they were high or low occupancy landcover types.

Component	Description	Data Source	Fixed values
C_e	Cost to establish study, including design, database, supplies, and QA/QC protocols	Sanderlin et al. (2025), Sanderlin et al. (*in prep*)	\$10,000
C_{u_0}	Cost per acoustic recorder unit that includes one strap, microphone, batteries, and SD cards at the start of the study	Cornell Lab of Ornithology, K. Lisa Yang Center for Conservation Bioacoustics, commercial sites for batteries and SD cards	\$430
C_{mic_0}	Cost per replacement microphone at start of study	Cornell Lab of Ornithology, K. Lisa Yang Center for Conservation Bioacoustics	\$25
C_{strap_0}	Cost per replacement strap at start of study	Commercial sites for replacement straps	\$3
p_{mics}	Average yearly proportion of replacement microphones	Sanderlin et al. (2025), Sanderlin et al. (*in prep*)	0.124
p_{straps}	Average yearly proportion of replacement straps	Sanderlin et al. (2025), Sanderlin et al. (*in prep*)	0.055

p_{units}	Average yearly proportion of replacement ARUs	Sanderlin et al. (2025), Sanderlin et al. (*in prep*)	0.140
x	Duty station distance (km) to each EMU, assuming local units are responsible for each EMU	Sanderlin et al. (2025), Sanderlin et al. (*in prep*)	20
n.years	Number of years in the study	This study	10
n.arus	Number of acoustic recorder units in each hexagonal unit	Sanderlin et al. (2025), Sanderlin et al. (*in prep*)	4
wage.vol	Volunteer per diem for participatory scientists	Sanderlin et al. (*in prep*)	$6.25/hour
wage.paid	Paid technician wage	Sanderlin et al. (*in prep*)	$21.50/hour
wage.GE	Governmental employee wage covered through other funds than project funds. No per diem because the EMU distances can be covered in a day.	This study	$0/hour

speedlimit	Average speed limit within study areas (km/hour)	Sanderlin et al. (*in prep*)	104.61 km/hour
C_v	Pro-rated vehicle rental/mileage per day	Sanderlin et al. (*in prep*)	\$50
$Time_{C_2}$	Total time per sampling unit determined from a linear model for coefficient estimates with the top model in Sanderlin et al. (*in prep*)	Sanderlin et al. (2025), Sanderlin et al. (*in prep*)	linear model estimates (β_0 = 1.67, β_{MC} = 0.32, β_{road} = 0.34) and values (μ_{road} = 134.008. σ_{road} = 180.763) for standardizing road distance with pilot data to use regression equation

Table S2. Estimates of costs and effort used in cost equations for passive acoustic monitoring optimal study design with Mexican spotted owls. We assumed that there were no cost increases with acoustic recorder units, microphones, or straps across the duration of the study. We also assumed that pairs would sample each hexagonal unit for safety and efficiency in deployments. Deployment and pick up time were equal within this study. All time estimates assumed a fixed 8-hour workday, regardless of observer type.

NIMBLE model equations for "Robust and realistic power analysis for monitoring of occupancy trends: a case study with the Mexican Spotted Owl"

1 Introduction

During the power analysis, we fit NIMBLE models to data simulated under different conditions. We defined two different models: a model for evaluating power to detect trends within a single ecological management unit (EMU; EMU-specific model), and a model for range-level trends aggregating across EMUs (range-level model). In this supplemental file, we provide the full definitions of the models used. All code necessary to reproduce our results is archived in a Zenodo repository at 10.5281/zenodo.18778235.

2 EMU-specific model

The EMU-specific model was defined by the following equations. First, the detection history at

$$\boldsymbol{y_i} \sim \mathrm{DynOcc}(\psi_0, \boldsymbol{\phi}, \boldsymbol{\gamma}, p)$$

Here, DynOcc() represents the full marginal likelihood distribution of the dynamic occupancy model which treats the full detection history in hex i across seasons and replicates as a single multivariate datum. The distribution DynOcc() provided by nimbleEcology (Goldstein et al. 2021) marginalizes over the latent occupancy state to improve mixing in some cases and also simplifies code specification (Ponisio et al. 2020). The distribution has three parameters: ψ_0 is the initial occupancy probability, which we modeled as constant across sites; $\boldsymbol{\phi}$ is a vector of season-specific survival probabilities, where ϕ_t represents the probability of a site that was occupied in time t remaining occupied in time $t+1$, modeled as varying with season but constant across hexes; $\boldsymbol{\gamma}$ is a vector of colonization probabilities, where γ_t represents the probability of a site that was unoccupied in time t becoming occupied in time $t+1$, modeled as varying with season but constant across hexes; and p is the probability of observing the species on a single visit given it is present within a hex, modeled as constant across sites, seasons, and replicates.

Colonization in each year was modeled as a separate parameter, with priors $\gamma_t \sim \mathrm{Unif}(0,1)$. Detection and initial occupancy were also given uniform priors $p \sim \mathrm{Unif}(0,1)$ and $\psi_0 \sim \mathrm{Unif}(0,1)$.

In the "recursive" and "equilibrium" simulation cases, persistence was modeled as potentially varying with time according to the equation

$$\mathrm{logit}(\phi_t) = \mathrm{logit}(\phi_0) + \mathrm{logit}(\beta) \times t$$

with priors $\phi_0 \sim \mathrm{Unif}(0,1)$ and $\beta \sim \mathrm{Unif}(0,1)$. In the "constant" case, persistence in each season was modeled as an independent parameter like colonization, with $\phi_t \sim \mathrm{Unif}(0,1)$.

3 Range-level model

The range-level model followed the same parameterization as the EMU-specific model, but differed in that it included a hierarchical representation of persistence as varying between EMUs. Detection histories in each hex were modeled as

$$\boldsymbol{y_i} \sim \mathrm{DynOcc}(\psi_0, \boldsymbol{\phi}_i, \boldsymbol{\gamma}, p)$$

following the same definitions as in the EMU-specific model but with the addition of a site-specific index on $\boldsymbol{\phi}_i$. Then, persistence within hex i was modeled as

$$\text{logit}(\phi_{i,t}) = \text{logit}(\phi_0) + \text{logit}(\beta) \times t + \alpha_{EMU(i)}$$

where $\alpha_{EMU(i)}$ is an additive random effect of EMU identity on persistence. Then, random effects were modeled as

$$\alpha_k \sim \mathcal{N}(0, \sigma_\alpha)$$

for each k in 1 to the number of EMUs. The parameter σ_α was assigned the prior $\sigma_\alpha \sim \text{Unif}(0, 10)$ and all other priors were the same as in the EMU-specific model.

References

Goldstein, Benjamin et al. (2021). "nimbleEcology: Distributions for ecological models in nimble". In: *URL: https://cran. r-project. org/package= nimbleEcology. R package version 0.4* 1.

Ponisio, Lauren C et al. (2020). "One size does not fit all: Customizing MCMC methods for hierarchical models using NIMBLE". In: *Ecology and evolution* 10.5, pp. 2385–2416.