Biological Conservation 286 (2023) 110260

Contents lists available at ScienceDirect

Biological Conservation

journal homepage: www.elsevier.com/locate/biocon

Strategic monitoring to minimize misclassification errors from conservation status assessments

Kylee D. Dunham[a,*,1], Patrick K. Devers[b], Abigail J. Lawson[c], James E. Lyons[a], Conor P. McGowan[d], J. Andrew Royle[a]

[a] U.S. Geological Survey, Eastern Ecological Science Center, 12100 Beech Forest Road, Laurel, MD 20708, USA
[b] U.S. Fish and Wildlife Service, Division of Migratory Bird Management, 11510 American Holly Drive, Laurel, MD 20708, USA
[c] U.S. Geological Survey, New Mexico Cooperative Fish and Wildlife Research Unit, Department of Fish, Wildlife, and Conservation Ecology, New Mexico State University, 2980 S. Espina Street, Knox Hall Room 132, Las Cruces, NM 88003, USA
[d] U.S. Geological Survey, Florida Cooperative Fish and Wildlife Research Unit, Department of Wildlife Ecology and Conservation, University of Florida, Gainesville, FL 32611, USA

ARTICLE INFO

Keywords:
Species conservation classification decisions
Endangered Species Act
Eastern black rail
Value of information
Expected value of sample information

ABSTRACT

Classifying species into risk categories is a ubiquitous process in conservation decision-making affecting regulatory procedures, conservation actions, and guiding resource allocation at global, national, and regional scales. However, monitoring programs often do not provide data required for accurate species classification decisions. Misclassification can lead to otherwise preventable species extinctions, undue regulatory burden, poor allocation of limited conservation resources, and can undermine species conservation legislation. We developed a framework that evaluates monitoring designs based on the ability to correctly inform a species classification decision, where minimizing the risk of misclassification is the central objective. We further evaluated monitoring designs by calculating the expected value of information and explored the relationship between statistical power to detect trends and misclassification. Our measure of misclassification risk, which can be tailored to the decision context, clarified the costs of over- and under-protection. High power to detect trends often corresponded to accurate species classification decisions. However, in several scenarios power to detect trends was low but the ability to correctly inform the classification decision was high. The value of information generally increased with monitoring intensity and quantified the tradeoffs between spatial and temporal replication. Our framework allows managers to assess monitoring program performance with direct implications for conservation decision-making. Our framework affords practitioners an opportunity to evaluate the effectiveness of monitoring programs a priori focusing on improving conservation decisions. We demonstrate that prioritizing monitoring to minimize misclassification errors can improve monitoring efficiency and conservation decision-making with considerable practical applications and benefits for species conservation.

1. Introduction

Conservation status assessment frameworks are used globally to classify species into risk categories and can be powerful tools for biodiversity conservation (Rodrigues et al., 2006; Smith et al., 2018). Global assessment frameworks such as the International Union for Conservation of Nature (IUCN) Red List of Threatened Species inform international agreements (e.g., Convention on International Trade in Endangered Species) and serve as the basis to monitor status of global biodiversity (e.g., Convention on Biological Diversity) (Bland et al., 2015). Assessment frameworks are used at national and regional levels to prioritize conservation actions (Partners in Flight Landbird Conservation Plan [PIF]; Rosenberg et al., 2016) or assign species and habitats legal protection (e.g., the 1973 U.S. Endangered Species Act [ESA] and the 1999 Australian Environment Protection and Biodiversity Conservation Act [EPBCA]). Species classifications often guide resource allocation affecting which species receive the limited conservation funds at global, national, and regional scales.

Conservation status classifications are typically informed by population metrics and incorporate both scientific and policy components

* Corresponding author.
E-mail address: kdunham@cornell.edu (K.D. Dunham).
[1] Present address: Cornell Lab of Ornithology, Cornell University, 159 Sapsucker Woods Road, Ithaca, NY 14850, USA.

https://doi.org/10.1016/j.biocon.2023.110260
Received 27 March 2023; Received in revised form 11 August 2023; Accepted 25 August 2023

(Cummings et al., 2018). Population metrics used to inform status differ between frameworks but often include estimates of abundance, population trends, or area occupied in relation to extinction risk. However, monitoring data may be lacking altogether or programs may fail to provide sufficient data to inform conservation decisions (Lindenmayer et al., 2013; Robinson et al., 2018). Crucially, research has demonstrated species classification decisions are sensitive to uncertainty (Connors et al., 2014; Regan et al., 2013; Rueda-Cediel et al., 2015) and are subject to two types of misclassification error (Taylor et al., 1996). Species may fail to be classified to a protected category when warranted (i.e., underprotection) or alternatively provided protection when it is not warranted (i.e., overprotection). Consequences of underprotection include an increased risk of species extinction when preventable with corrective conservation actions. Overprotection may seem preferable and even consistent with the "precautionary principle" (Myers, 1993); however, providing protection when unwarranted causes undue regulatory burden, poor allocation of limited conservation resources, and can serve to undermine species conservation legislation (Lukey et al., 2011). Ensuring sufficient data for species classification decisions to minimize misclassification errors is therefore a critical component of global species conservation status assessments.

Expected value of information (VOI) analysis quantifies the expected benefit of acquiring additional research or monitoring information to inform a decision problem (Raiffa and Schlaifer, 1961). VOI is the difference between expected management outcomes when the decision is made before acquiring new information (i.e., under uncertainty) and after acquiring new information (Runge et al., 2011; Yokota and Thompson, 2004). The expected value of perfect information (EVPI) measures difference in outcomes if one were able to resolve all uncertainty before choosing among alternative actions. EVPI provides an upper bound on the increase in performance because it assumes perfect resolution of uncertainty, which is rare in ecology. More realistically, the expected value of sample information (EVSI) measures the difference in outcomes when using less-than-perfect information (Canessa et al., 2015; Runge et al., 2011; Williams and Brown, 2020). Incorporating VOI metrics to evaluate the efficacy of monitoring designs provides critical information about the tradeoffs between differing designs. In combination with minimizing misclassification errors, metrics for EVSI have the potential to greatly improve the assessment of monitoring designs for conservation decision-making.

Traditionally, objectives for monitoring programs focus on measures of statistical precision, without formal consideration of the decision process. Assessments frequently evaluate monitoring program performance based on statistical power to detect trends or other population state summaries and assume high statistical precision equates to better decision-making. Species classification decisions, however, are often based on predictive models of extinction risk and corresponding decision thresholds and are influenced by not only parameter estimates but also value judgements associated with risk tolerance (Cummings et al., 2018). The use of power analyses to design monitoring programs to inform decisions is logical but lacks an explicit link between monitoring and the outcome of the decision-making process. That is, a core consideration in designing and allocating effort to monitoring is not only returning precise and accurate results, but also to minimize the risk of assignment to the wrong classification. Despite numerous calls for designing and evaluating monitoring programs to improve decision-making (Lindenmayer et al., 2020, 2022; Lyons et al., 2008; Nichols and Williams, 2006), to our knowledge there has not yet been a study that evaluates the efficacy of monitoring designs for classifying species to conservation risk categories.

We propose a framework to design monitoring programs for making species status classification decisions. Our framework includes two metrics to evaluate monitoring designs: 1) misclassification risk, and 2) the value of information. We used a simulation-based procedure to generate monitoring datasets with different levels of spatial and temporal replication and used these two metrics to quantify the risk of committing misclassification errors under each monitoring scenario. We also used power analysis to explore the relationship between statistical power to detect population trends and misclassification risk. We expect managers and decision-makers to find this useful to 1) determine if existing monitoring designs are sufficient for accurately informing classification decisions; and 2) design new monitoring programs with the objective to minimize classification errors and maximize the value of information. We demonstrate the utility of our approach for a hypothetical classification decision under the ESA. However, our framework is broadly applicable to species status classification decisions using other conservation status assessment frameworks at global, national, and regional scales.

2. Materials and methods

We propose a framework for evaluating monitoring decisions that characterizes outcomes of the classification process under repeated simulation of populations and sampling of these simulated populations given alternative monitoring scenarios. Our approach supposes sufficient understanding of the biological system so that a reasonable population model and parameter values can be defined for purposes of simulation and estimation. Given the population model, we define *true classification status* to be the populations conservation status determined under perfect information. That is, the conservation status of the population given the parameters of the population model. Given the model and its parameters, we can simulate a hypothetical population trajectory, subject that population to sampling by some monitoring design, and classify that population based on the monitoring data. We refer to this decision as the *predicted classification*. A correct classification is made when the predicted classification corresponds to the true classification status. Our simulation-based framework consists of 6 steps, each of which may include several components specific to the classification decision problem (Fig. 1). We describe each step here, define the decision-specific components in our case study, and describe further extensions in the discussion.

2.1. General framework

First, we defined the true biological process scenario(s) we wished to assess (e.g., increasing, declining, or stable population trends) and set initial parameter values to reflect desired population trends and corresponding correct classification status (e.g., a scenario with declining population that should be classified as endangered). Second, we simulated the true biological state and process over the designated spatiotemporal scale. These steps makeup our biological process scenarios from which we determine the true classification status, i.e., the correct classification decision given perfect information. Next, we prescribed monitoring scenarios outlining considerations such as sample size and survey duration. In step four, we simulated data collection and estimated parameters based on collected data from each monitoring program for each true biological scenario. Next, we applied our species classification criteria to the estimated population state to predict the species classification decision for each biological process–monitoring scenario. With the true classification status known for each biological process scenario and the predicted classification for each biological process–monitoring scenario, we can calculate the proportion of correct species classifications obtained under repeated simulation of the biological process–monitoring scenarios to evaluate performance of the tested monitoring scenarios. Here, our objective was to select a monitoring program that minimizes misclassification risk.

2.2. Case study

We demonstrated the utility of our framework with a hypothetical classification decision problem under the ESA based on Eastern black rails (*Laterallus jamaicensis jamaicensis*), a secretive marsh bird listed as

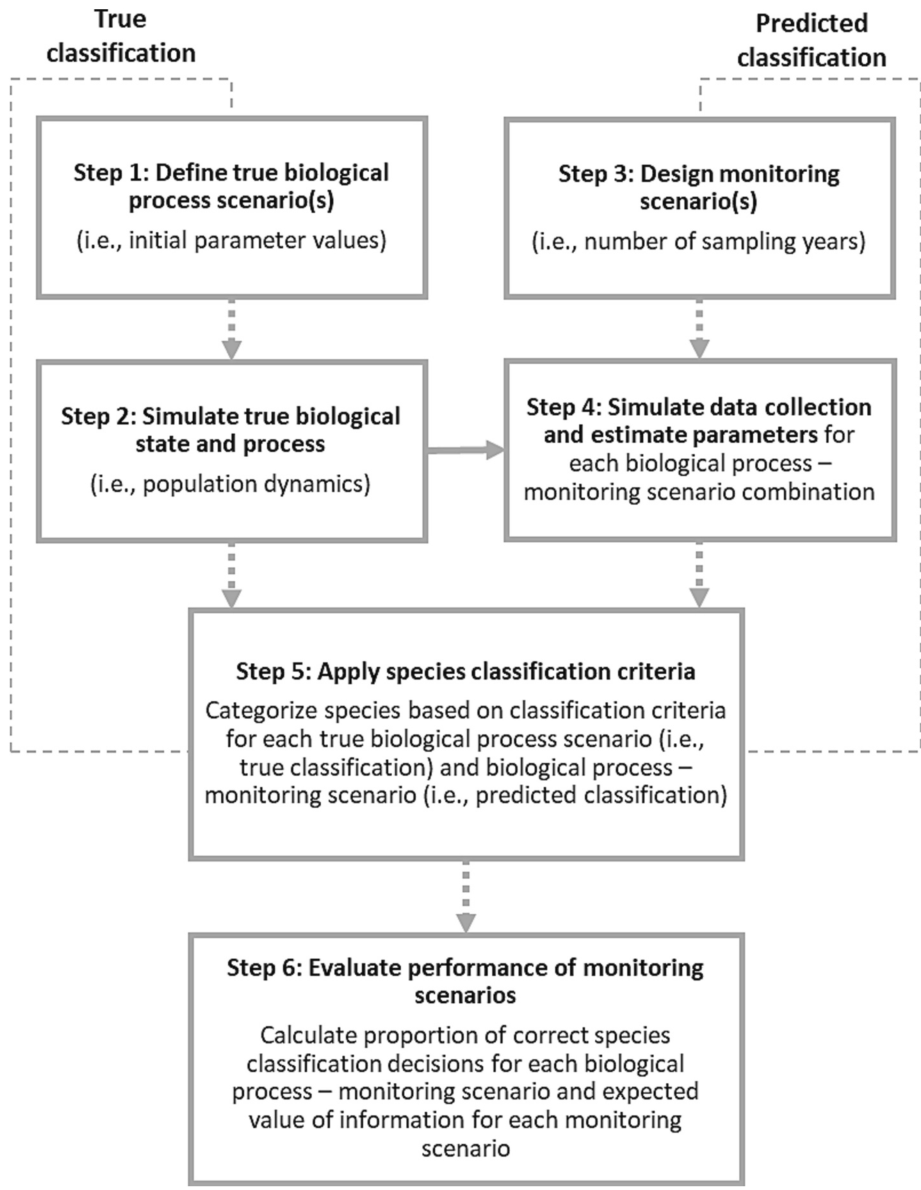

Fig. 1. General framework for evaluating monitoring programs based on the ability to correctly inform species classification decisions.

threatened under the ESA in 2020 (USFWS 2020). We assumed we are required to design and establish a monitoring program with the objective to inform a listing decision (e.g., reclassification) at some time in the future, e.g., a 5-year review. Occupancy was used as the metric to inform the original listing decision (McGowan et al., 2020) and we used percent area occupied (PAO) of the species range at year 50 as the metric for this assessment.

We framed this decision to assign the species to the correct risk category (i.e., "Putting species in the right bin" sensu Cummings et al., 2018). This decision framing requires quantitative thresholds for each risk category (or bin), which we defined as: if occupancy was >50 % of the total area available then protection would not be warranted (i.e., species could be delisted); if occupancy was ≥ 20 % and ≤ 50 % the species would remain listed as threatened; and if occupancy was <20 % the species would be listed as endangered (Table 1, Fig. 2).

Due to uncertainty in the species true biological state (e.g., PAO), we needed to define additional components to address the costs associated with misclassification (i.e., under- vs. overprotection) and risk tolerance

Table 1
Misclassification cost matrix for assigning species to the correct classification (bin) with probability values from a hypothetical example. Symmetric costs indicate that it is equally costly to underprotect and overprotect the species. Decision thresholds refer to estimates of the percent area occupied (PAO) at year 50 and correspond to specific classification decisions.

Decision thresholds	Classification	Probability	Misclassification cost matrix		
			NW	TH	EN
PAO > 50%	Not warranted (NW)	0.30	0.0	0.5	1
PAO ≥ 20% & ≤ 50%	Threatened (TH)	0.30	0.5	0.0	0.5
PAO <20%	Endangered (EN)	0.40	1.0	0.5	0.0
Expected misclassification cost			0.55	0.35	0.45

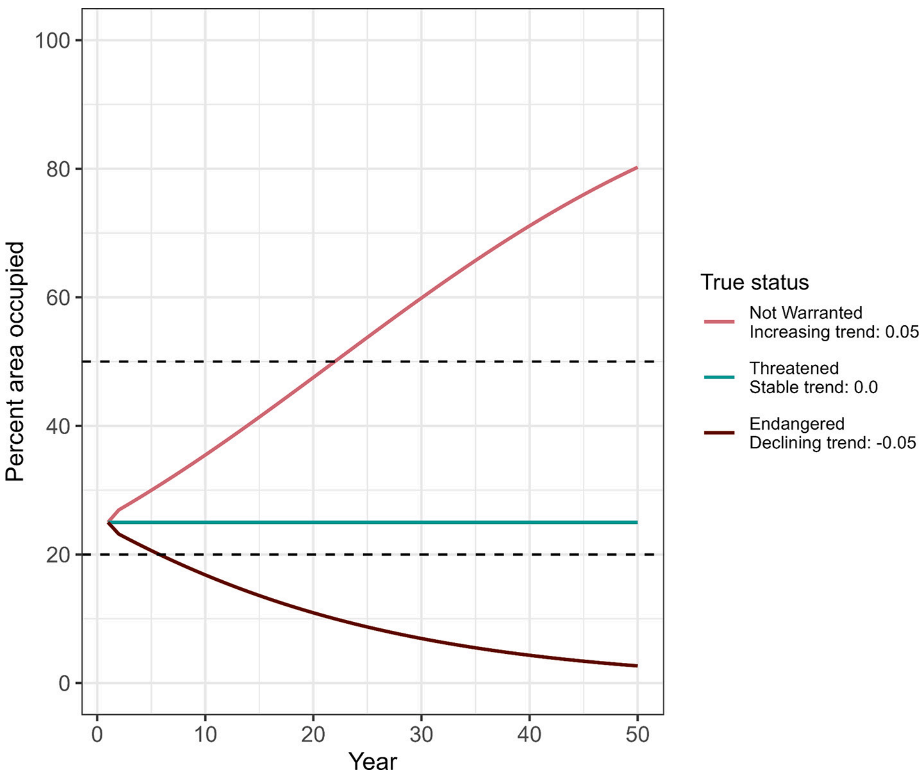

Fig. 2. Percent area occupied (PAO) over time and true status (e.g., correct classification) for each true biological process scenario (i.e., increasing trend) tested in the application of our framework. Dashed lines represent the decision thresholds separating the three species classification categories. We established the following thresholds as boundaries between the bins based on percent area occupied in year 50: if PAO was > 50 % the correct classification would be "not warranted", if PAO was ≥ 20 % & ≤ 50 % the correct classification would be "threatened", and if PAO was <20 % the correct classification would be endangered.

of decision maker(s). We established a quantitative measure of costs associated with making the incorrect classification decision (Table 1). While many options exist to quantify misclassification cost (see Cummings et al., 2018; Regan et al., 2013), we chose a symmetric misclassification cost matrix. This approach reflects a desire to balance the costs associated with under- and overprotection errors. That is, if the true status is threatened, the cost or loss associated with listing as either endangered (i.e., overprotection) or not warranted (i.e., underprotection) are equal. Further, we note that costs vary by magnitude of the error. Specifically, there is a lower cost associated with making a threatened decision (0.5) than a not warranted (NW) decision (1.0) when the correct classification is endangered (EN) (row 3, columns 1–3 of the Misclassification Cost Matrix in Table 1). When correct classifications are made, there is no cost (0) because there is no penalty associated with making the right decision. Finally, to quantify the probability or risk ("Expected misclassification cost"; Table 1) of misclassification, the misclassification cost is multiplied by the probability of PAO being in each bin (see *Step 5* below). The best or optimal classification is one that minimizes misclassification risk.

To determine which monitoring programs were sufficient for informing the species classification decision, we defined a risk tolerance level for how frequently a monitoring program returns the correct decision. Some decision makers may tolerate a scenario in which the species is misclassified relatively frequently (e.g., 40 % of the time) while others only tolerate a small misclassification rate (e.g., <10 %). We set a risk threshold of 20 % indicating we accept monitoring programs that return the correct classification a minimum of 80 % of the time and misclassifications no >20 % of the time. We then evaluated each monitoring scenario based on the expected value of sample information (EVSI) to compare the relative performance given the different spatial and temporal monitoring designs.

Step 1. Define the true biological process scenario(s)

To employ our framework, we first need to define the true biological scenario(s) of interest to the decision-maker given the decision problem (*Step 1* in Fig. 1). This depends on the expected conditions of the species and the desired range of realistic conditions the decision makers want to explore. We limited the number of biological process scenarios to a total of three possible realities. For all three initial PAO ($t = 1$) was 25 % and we set the total number of possible sites to 5000, assumed to encompass the species' range where we monitor some subset. We chose three rates of change that resulted in a PAO in year 50 that fell into the three risk categories (bins) (Fig. 2). In a realistic assessment, the rates of change may not be selected specifically to force results into each bin but would rather reflect the desired range of conditions. Through iterative testing, we identified three rates of change on the logit scale that corresponded to PAO in year 50 falling into the not warranted, threatened, and endangered bins. The "Increasing–Not warranted" scenario trend was 0.05, the "Stable–Threatened" scenario trend was 0.0, and the "Declining–Endangered" scenario trend was −0.05.

Step 2. Simulate true biological state and process

For each biological process scenario (e.g., "Increasing–Not warranted") we simulated occupancy over 50 years using a simple trend model on the logit scale (*Step 2* in Fig. 1). With the logit of the PAO in year 1 (here, 25 %) we projected logit occupancy through year 50 as

$logit\ (\psi_{t=2:T}) = logit\ (\psi_{t=1}) + \beta_1 * (t = 2 : T)$ where $\psi_{t=1}$ is initial PAO and β_1 is the true trend. We converted logit occupancy to the real scale to get PAO for all 50 years.

Step 3. Design monitoring scenarios

We created 24 monitoring scenarios (*Step 3* in Fig. 1) with two detection probabilities (0.2, 0.5), three survey period lengths (3, 5, 10 years), two levels of surveyed sites (500, 1500), and two levels of repeat surveys (i.e., visits) at each site (3, 6) (Appendix 1). Detection is generally low for black rails, but we include a "high" detection probability to reflect the possibility that we could employ actions that increase overall detection (Tolliver et al., 2019). Namely, practitioners may consider deploying autonomous recording units (ARUs) which generally increase detection for similar species (Znidersic et al., 2021). The number of sites and number of repeat surveys are loosely based on previous surveys conducted for the species along with general guidance on minimum sites and visits for estimating occupancy (Tolliver et al., 2019). We note that survey length may not correspond to timeframes required for reassessment under the ESA (e.g., 5-year reviews). Nonetheless, we wanted to evaluate performance of short (3-year and 5-year) and long term (10-year) studies for estimating trends in occupancy and for use in decision-making. In anticipation of an upcoming listing decision, management agencies may collect monitoring data that are later shared with USFWS for the Species Status Assessment. Alternatively, following a threatened or endangered species listing, the USFWS may fund short-term (1–5 years) studies for use in the status review. Therefore, these scenarios reflect plausible and relevant, real-world situations in which monitoring data could be used for the initial or periodic reassessments.

Step 4. Simulate data collection and estimate parameters

For each monitoring scenario ($n = 24$), we generated 1000 data sets from the true biological state process by simulating true occupancy for the number of sampled sites i for each year t using the true occupancy rate for each year ($\psi_{t=1:T}$, *Step 2*) as

$$z_{i,t} \sim \text{Bernoulli}(\psi_{i,t})$$

and generated site specific annual detection/non-detection data ($y_{i,t}$) as

$$y_{i,t} \sim \text{Bernoulli}(z_{i,t} * p)$$

where p is the scenario specific detection probability. We "stacked" multiple years of detection/non-detection data where the sites within each year are treated as independent sites of a single year.

We modeled occupancy (ψ) using year-stratified models in unmarked (Fiske and Chandler, 2011) within program R (R Core Team, 2021) using the same model that generated the data. Specifically, the state process for occupancy at site i was:

$$z_{i,t} \sim \text{Bernoulli}(\psi_{i,t})$$

$$\text{logit}\ (\psi_{i,t}) = \beta_0 + \beta_1 * \text{year}_t$$

where z is a latent state variable which equals 1 if site i is occupied and 0 if not, β_0 is the intercept, and β_1 is the linear effect of year. The corresponding detection process was modeled as:

$$y_{i,j,t} | z_i \sim \text{Bernoulli}(z_{i,t} p)$$

where y is the observed detection/non-detection data, p is the probability of detection given that the site is occupied, and j indexes the replicate surveys for a site. We recorded trend estimates (β_1) with uncertainty to compare to the true trend (*Step 1*). We used the predict function (Fiske and Chandler, 2011) to estimate occupancy probabilities, standard errors, and confidence intervals. Specifically, we predicted occupancy on the logit scale for a total of 50 years, inclusive of years the population was monitored.

Step 5. Apply species classification criteria

Our species classification criteria categorize the species into bins based on the percent area occupied in year 50. Specifically, when PAO is >50 % protection would not be warranted, when PAO is ≥ 20 % and ≤ 50 % the species would remain listed as threatened, and when PAO is <20 % the species would be listed as endangered. In practice, we cannot know the exact PAO at any time due to both imperfect sampling of the landscape and imperfect observation of the occupancy state at sampled sites, and thus do not know the true PAO bin. Therefore, we need to consider the risk of misclassifying the species to the incorrect risk category. We quantify the risk of misclassification by multiplying the cost of misclassification (e.g., deciding protection is not warranted when the species should be classified as endangered = 1.0) by the probability the PAO falls within each bin (Table 1). We chose a symmetric cost matrix and considered the classification with the smallest risk (expected misclassification cost) the optimal decision.

To apply the species classification criteria, we required occupancy estimates in year 50. Using predicted values of mean occupancy and standard error in year 50 (from *Step 4*) we generated 1000 replicates (r) of logit occupancy with:

$$\text{logit.Pocc}_{r,t=50} \sim \text{Normal}(\text{logit.occ.mean}_{t=50}, \text{logit.occ.se}_{t=50}).$$

We used the plogis function (R Core Team, 2021) to convert logit. Pocc to the natural scale, $Pocc_{r,t=50} = plogis(logit.Pocc_{r,t=50})$ and estimated the number of occupied sites as:

$$N_{r,t=50} \sim \text{Binomial}(\text{Tot.sites}, \text{Pocc}_{r,t=50}).$$

We calculated the percent area occupied for all replicates where:

$$\text{PAO}_{r,t=50} = N_{r,t=50} / \text{Tot.sites}$$

The metric of interest from the population projection is the probability of falling within each decision threshold and corresponding classification bin (e.g., >50 % area occupied, protection not warranted, Table 1). We calculated the probability of falling within each classification bin as the proportion of replicates in which PAO fell within the decision thresholds.

We then calculated the expected misclassification cost (i.e., misclassification risk) as the sum of the probability of being in each bin multiplied by the cost associated with making each classification. Here, the cost of listing as not warranted when the correct decision is to list as threatened is 0.5, and the cost of listing the species as not warranted when it should be listed as endangered is 1.0 ("NW" column in the cost matrix, Table 1). Thus, the cost is higher when the magnitude of the error is larger. Here we assume listing as NW when it should be EN is worse than listing it as NW when it should be threatened (TH). The decision that minimizes the risk of a misclassification error is the optimal decision. In the example in Table 1, the decision to list as TH has the lowest expected cost (i.e., risk) and can be calculated as:

$$\text{Expected cost (TH)} = 0.35 = 0.30 \times 0.5 + 0.30 \times 0.0 + 0.40 \times 0.5$$

$$\text{Expected cost (TH)} = 0.35 = Pr(\text{NW}) \times 0.5 + Pr(\text{TH}) \times 0.0 + Pr(\text{EN}) \times 0.5$$

We calculated expected costs for each of the 1000 replicate data sets resulting in one optimal decision for each data set for each biological process–monitoring scenario. We then calculated the proportion of times each decision (i.e., EN, TH, NW) was returned for each biological process–monitoring scenario for further evaluation.

Step 6. Evaluate the performance of monitoring scenarios

We evaluated monitoring scenario performance using several metrics. To measure how effective each monitoring scenario was at returning the correct decision, we calculated misclassification risk as the proportion of correct and incorrect listing decisions averaged over all data set replicates. We considered monitoring programs to be sufficient to inform our classification decision if the misclassification rate across replicates, i.e., proportion of incorrect listing decisions, was <20 % or 0.20. Additionally, we reported the type (over- or under-protection) and magnitude (e.g., EN versus TH when truly NW) of the misclassification errors.

We also evaluated the performance of each monitoring scenario with EVPI and EVSI (Runge et al., 2011; Canessa et al., 2015). Outcomes of the decision were defined by the misclassification cost matrix (Table 1), i.e., a loss function. In this case, the value of new information is expressed as the loss avoided and is equal to expected loss under uncertainty minus expected loss under certainty: $VOI = EV_{uncertainty} - EV_{certainty}$. Calculating value of information requires prior probabilities for belief in the system state (i.e., increasing, stable, or declining). We chose essentially equal prior probabilities (0.33, 0.33, and 0.34 for increasing, stable, and declining, respectively). In our case study, $EV_{uncertainty}$ is the sum of expected losses for each classification (not warranted, threatened, or endangered) given the true state of the system (increasing, stable, or declining), each weighted by the respective prior probability of being in each system state: $EV_{uncertainty} = max_a E_s[V(a,s)]$, where a represents the classification and s the system state. The optimal choice under uncertainty is the classification with the smallest expected loss. EVPI is the expected benefit from eliminating uncertainty about system state entirely, i.e., with perfect knowledge of whether the population is increasing, stable, or declining. With perfect knowledge, the decision maker would always choose the correct classification and thus incur zero loss ($EV_{certainty} = 0$); therefore, in this case, EVPI is equal to $EV_{uncertainty}$. EVPI measures the loss we could avoid with perfect knowledge, and while a useful measure of maximum improvement in outcomes, is seldom possible. EVSI is the expected benefit of imperfect information. Calculating EVSI requires a Bayesian pre-posterior analysis (Yokota and Thompson, 2004), which implies collecting data and receiving new information before making a classification. For this we need to consider all possible results of the monitoring data, and calculate how each result (increasing, stable, declining) would change our belief about system state. We used the simulated monitoring data (x) from step 4 as sample information and updated the prior probabilities for system state using Bayes theorem. For example, with monitoring data indicating a declining population (x^-), our prior probability that the population is declining $P(decline)$ is updated using:

$$\mathrm{P}(\mathrm{decline}|\mathrm{x}^-) = \frac{\mathrm{P}(\mathrm{x}^-|\mathrm{decline})\mathrm{P}(\mathrm{decline})}{\mathrm{P}(\mathrm{x}^-)}$$

where $P(x)$ is the probability of the data. After updating the priors based on new information, the expected loss under certainty is calculated for each monitoring result (increasing, stable, declining), i.e., sum of losses for each classification, weighted by the updated prior probabilities. The optimal classification under each monitoring result is the one that minimizes loss. But because we do not know which monitoring result we will find, the value for each optimal action is weighted by the probability of the data to determine expected value under certainty: $EV_{certainty} = E_x\{max_a E_{s|x}[V(a,s)]\}$. Finally, EVSI is the difference $EV_{uncertainty} - EV_{certainty}$. Given the importance of prior probabilities for system state, we used a sensitivity analysis to assess the effect of priors on EVSI.

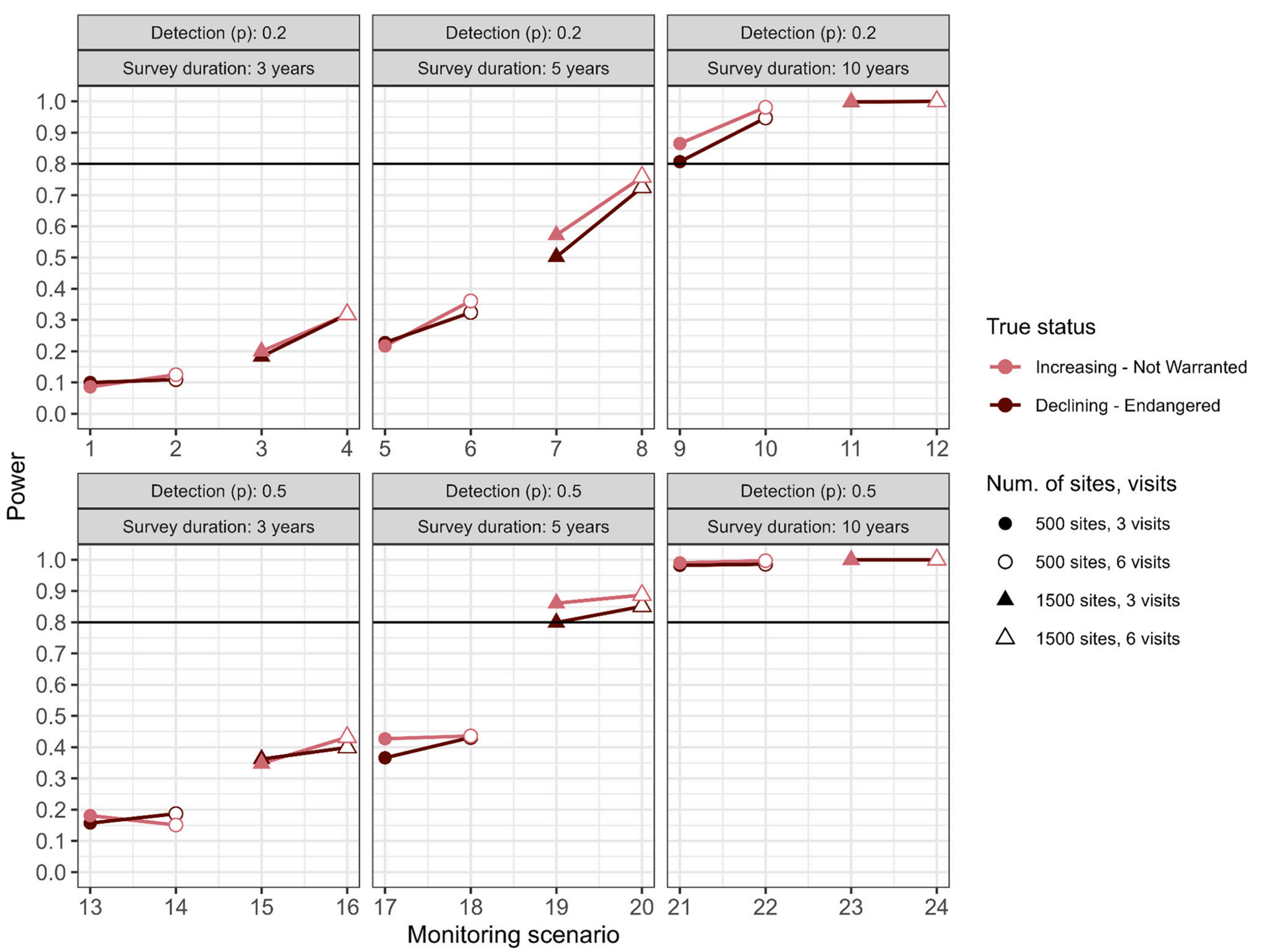

Fig. 3. Power to detect trends for increasing and declining biological process–monitoring scenarios. The solid black horizontal line represents a threshold value for the acceptable statistical power to detect trends. The biological process scenarios are represented by different color markers (True Status) and describe the population trend (increasing, declining) and true classification status (not warranted, endangered). We created 24 monitoring scenarios combining different detection probabilities, survey period lengths, number of surveyed sites, and number of repeat visits per site; monitoring scenarios are represented on the x-axis by panel label and markers. The linked points represent scenarios that only differ by the number of visits. We do not include the Stable–Threatened scenario because we could only calculate the Type 1 error rate which was low and effectively constant (approx. 0.05) across all monitoring scenarios.

Specifically, we calculated EVSI for a range of prior probabilities [0, 1] for each system state, constrained to sum to one.

2.3. Power analysis

In addition to the classification decision analysis (*Step 5* in Fig. 1), we were also interested in the statistical power to detect trends. We recorded the mean and 95 % confidence intervals of the estimated trend term from each simulated data set (Step 4) to calculate power to detect trends for each biological process–monitoring scenario. Power for the positive and negative trend biological process scenarios were calculated as the proportion of times the lower or upper 95 % confidence interval of the trend was less than or greater than zero, respectively, depending on the direction of the true trend. We computed the Type 1 error rate for the stable trend scenario as the proportion of times the 95 % confidence interval included zero (Fig. 3).

3. Results

The power to detect increasing and declining trends increased with survey effort and detection probability (Fig. 3). Power was sufficient when detection was low (0.2) only when surveys occurred for 10 years. However, when detection probability was high (0.5) power was sufficient when surveys were conducted for at least 5 years at 1500 sites and increased to nearly 100 % when conducted over 10 years. The Type 1 error rate for the Stable–Threatened scenario (not plotted) ranged from 0.049 to 0.064 across monitoring scenarios indicating a low probability that we would detect an increasing or declining trend when there is none.

Under all biological process scenarios, misclassification error declined with increasing survey effort (Figs. 4, 5). For the increasing and declining biological process scenarios, misclassification error was sufficiently low (<0.20) for nearly all monitoring scenarios starting with those including 3 years and 1500 sites monitored (monitoring scenarios 3–12 and 15–24) (Fig. 4). Alternatively, for the Stable–Threatened scenario, only two monitoring scenarios (23 and 24) met the threshold for tolerable classification error (<0.20) (Fig. 4). The proportion of overprotection errors was higher than the proportion of underprotection errors across all monitoring scenarios for the Stable–Threatened biological process scenario (Fig. 5). The magnitude of errors declined with increasing monitoring effort for both the increasing and declining biological process scenarios. For the Increasing–Not warranted scenario, the proportion of larger magnitude overprotection errors (list as endangered) was higher than the proportion of smaller magnitude overprotection errors (list as threatened) for the low monitoring effort scenarios, however, larger magnitude errors declined with increasing

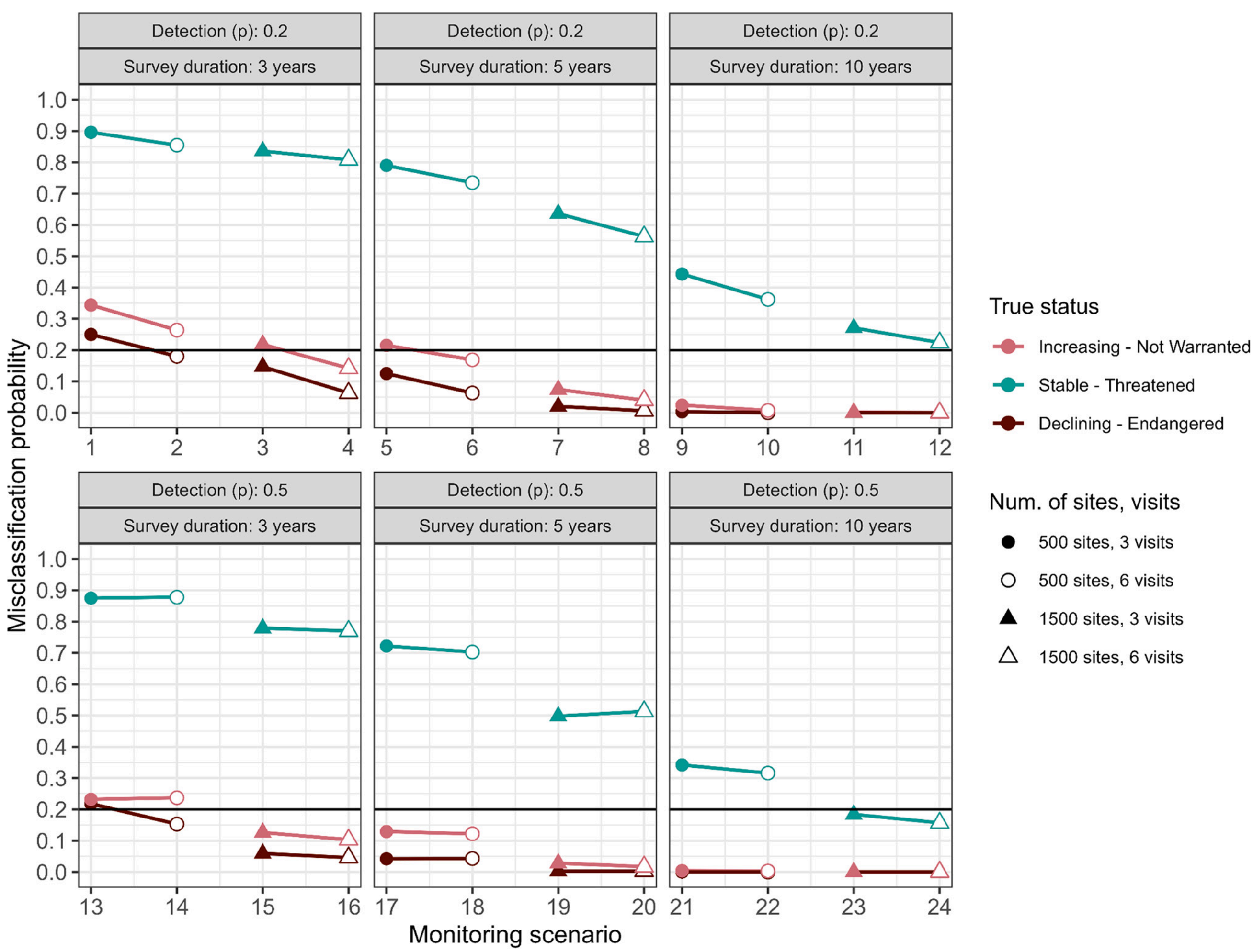

Fig. 4. Proportion of total misclassification errors (under and overprotection) for each biological process–monitoring scenario. The solid black horizontal line represents a threshold value for the acceptable misclassification error rate. We set this threshold to be 0.20, meaning that a monitoring program will only be considered sufficient for decision-making when misclassifications occur in ≤ 20 % of the replicates. The three biological process scenarios are represented by different colors (True Status) and describe the population trend (increasing, stable, declining) and true classification status (not warranted, threatened, endangered). We created 24 monitoring scenarios combining different detection probabilities, survey period lengths, number of surveyed sites, and number of repeat visits per site represented by panel labels and symbols. The linked points represent scenarios that only differ by the number of visits.

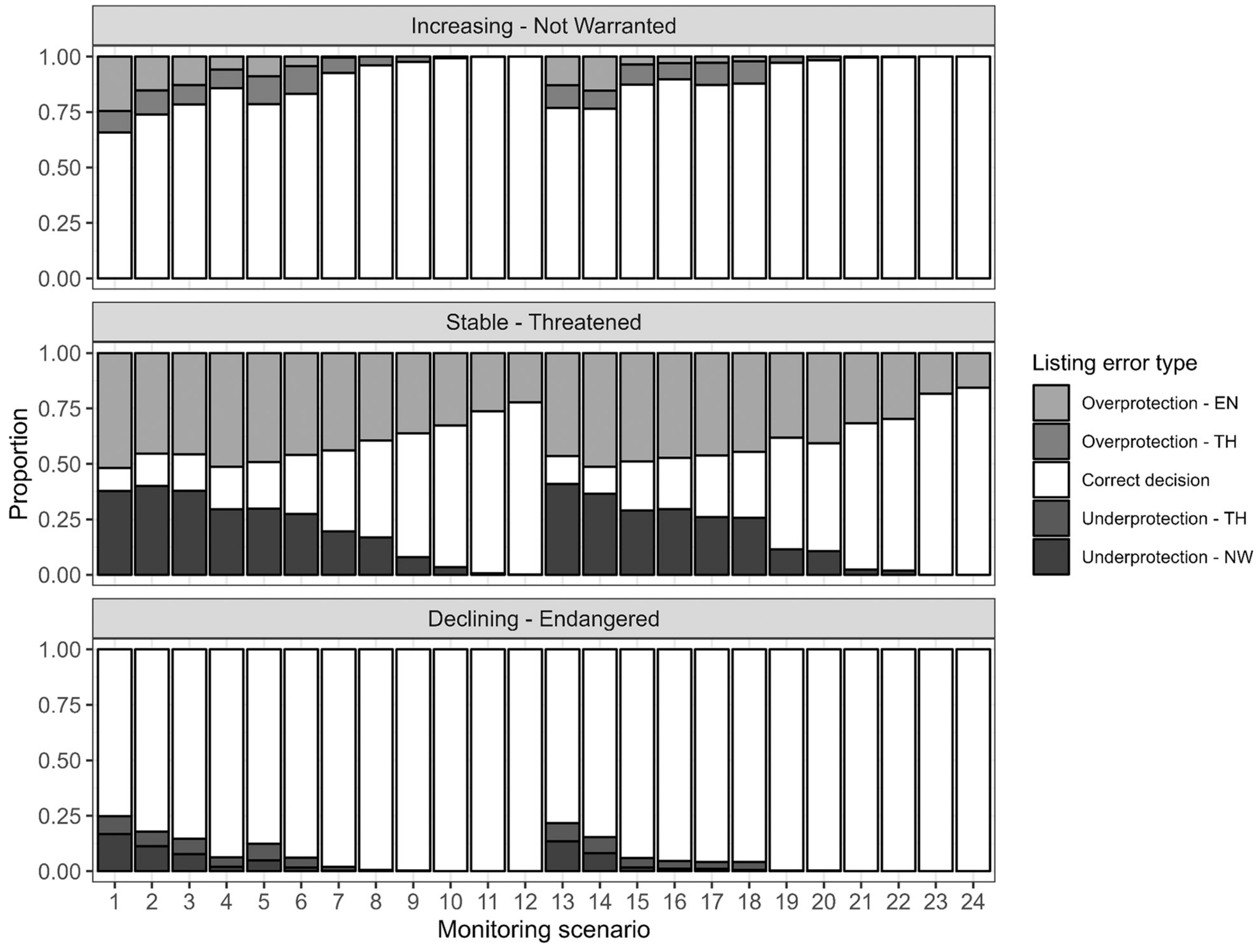

Fig. 5. Proportion and magnitude of misclassification errors for each biological process–monitoring scenario. Correct decisions are shown in white and misclassifications in gray scale.

effort (Fig. 5). We detected a similar pattern with the Declining–Endangered biological process scenario.

Classification error was sufficiently low in most monitoring scenarios to accurately classify the population to the Increasing–Not warranted and Declining–Endangered categories (monitoring scenarios 3–12 and 15–24, Fig. 4). Notably, for many of these biological process scenarios, we were able to correctly classify the population despite lacking power to detect the corresponding trend. For example, in the Increasing–Not warranted process scenario, the power to detect a positive trend for monitoring scenario 16 was roughly 0.40; however, misclassification for that scenario was 0.10, meaning we correctly classified the population 90 % of the time. Importantly, when power was sufficiently high to detect the trend (>0.80) the proportion of classification errors was sufficiently low (<0.20) to accurately classify the population in most biological process–monitoring scenario combinations.

Our misclassification cost matrix, i.e., loss function, indicates that expected loss under certainty is 0, i.e., the correct decision in all cases results in 0 penalty, so EVPI in this decision context is equal to the expected loss under uncertainty, 0.335, which is the result of maintaining status as Threatened. EVSI generally increased with monitoring effort (i.e., visits, sites, and years) but tradeoffs were evident (Fig. 6). For example, there was little difference in EVSI from monitoring a “large” number of sites (1500) for three years and a “small” number of sites (500) for five years (i.e., scenario 4 vs. 6 and 16 vs. 18, Fig. 6). To gain substantial benefit from increasing the number of years from three to five, it was necessary to increase the number of sites from 500 to 1500. That is, there was a large increase in EVSI for 1500 sites compared to 500 sites when monitored for five years. But there were diminishing returns for increased monitoring investment in the 10-year scenarios. When monitoring for 10 years, there was a relatively small increase in EVSI for scenarios with 1500 sites compared to 500 sites (Fig. 6). EVSI reflected the greater reliability of monitoring information when detection probability was “high” (0.5 compared to 0.2). EVSI was generally greater under the higher detection probability for all monitoring scenarios and the difference between three and six visits, all else being equal, was smaller when detection probability was 0.5 compared to 0.2. Sensitivity analysis for prior beliefs about system state indicated that EVSI was greatest when the priors for declining and stable were 0.5 and 0, respectively (Fig. 7). EVSI is therefore greatest when there is equal prior probability of increasing and declining.

4. Discussion

Monitoring programs are often evaluated in terms of their ability to estimate population trends or other population metrics, with the assumption that higher statistical power will lead to better decision-making. We developed a framework to evaluate monitoring programs based on the ability to correctly classify conservation status and the value of information. Further, we explored the relationship between statistical power to detect trends and the ability to correctly inform a conservation decision. Our framework allows managers to assess the performance of monitoring programs with direct implications for conservation decision-making. Our results suggest high power to detect trends often corresponds with accurate species classification decisions, however, strictly using results from the power analysis greatly limited the number of scenarios that were sufficient for making the correct

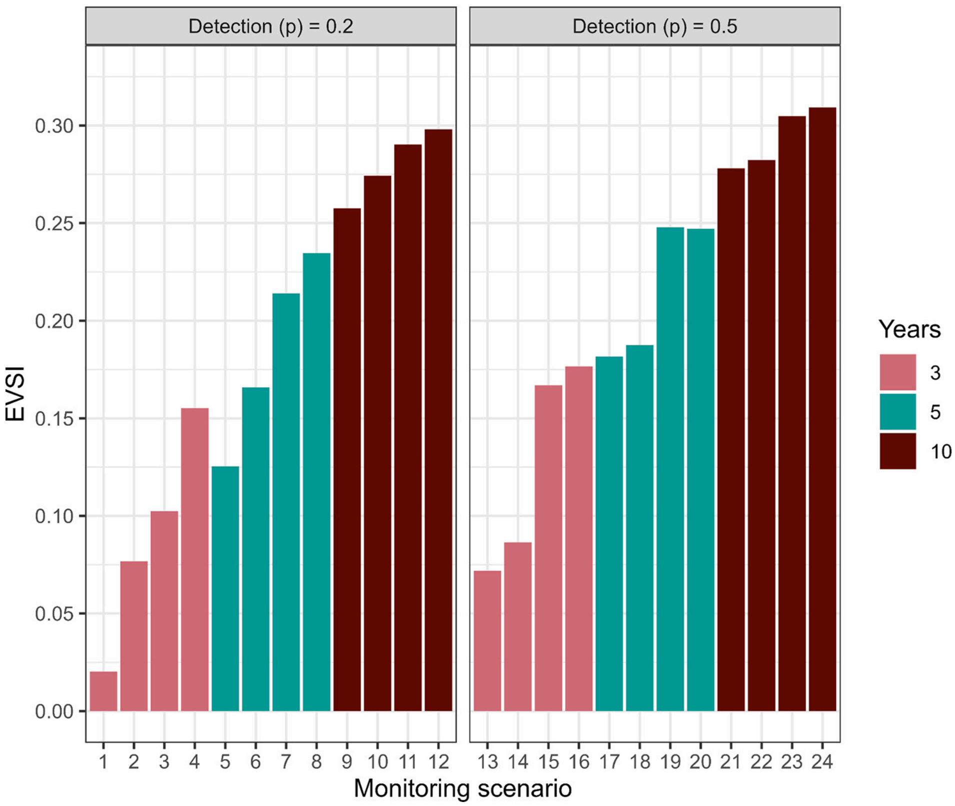

Fig. 6. Expected value of sample information (EVSI) for 24 monitoring scenarios. EVSI in this decision context is the loss avoided from an incorrect status classification (misclassification cost matrix shown in Table 1). Monitoring scenarios were either 3-, 5-, or 10-years duration. For each set of bars (3, 5, or 10 years), the first two bars are for 500 sites (3 and 6 visits, respectively); second two bars are for 1500 sites (3 and 6 visits, respectively). EVSI generally increased with monitoring effort, but interestingly in this decision context, the value of information from monitoring 1500 sites for three years was similar to 500 sites for five years (i.e., scenario 4 vs. 6 and 16 vs. 18).

classification. While power analyses are suitable for null hypothesis testing, our results demonstrate the limitations of using such analyses to inform species classification decisions. Further, our results indicate high statistical power to detect trends is not necessarily required to correctly inform classification decisions. These findings open opportunities for managers and decision-makers to consider monitoring programs with lower effort than those that meet typical statistical power thresholds, and our framework provides substantially more detail regarding the benefits and consequences of monitoring programs for making classification decisions.

Increased monitoring efforts generally reduced the risk of misclassifying a species to the incorrect extinction risk category. Declines in misclassification risk were often correlated with increases in statistical power in response to increased monitoring effort, however, we note that power is not calculable for stable population growth scenarios further limiting the utility of that approach for assessing monitoring designs for conservation decisions. Misclassification risk was initially low and the reduction in misclassification was minimal for the Increasing–Not warranted and Declining–Endangered scenarios. Alternatively, results indicate misclassification error associated with the Stable–Threatened scenario was substantially higher than for the other two biological process scenarios. We suspect this occurred for several reasons. First, misclassification risk is generally higher when the population state is near the boundary of a decision threshold (Regan et al., 2013). Under the stable process scenario, true percent area occupied was 25 %, and was close to the threshold for listing as endangered (PAO <20 %), which likely resulted in higher probability of overprotection misclassification errors particularly when uncertainty was high (e.g., low monitoring effort). Second, when estimates of occupancy and trend were highly uncertain, predicted PAO at year 50 tended to cluster near the boundaries of 0 or 1 depending on the initial sign of the mean trend estimate due to the underlying model form. This statistical artifact was particularly problematic for the stable scenario because estimates near the boundaries would end up above or below the threatened decision thresholds for all but the most intensive monitoring efforts. While we expect that the general pattern of declining misclassification risk with increased effort will hold under most biological process scenarios, we recognize the magnitude of errors and rate of decline in misclassification risk will likely vary with initial conditions of the biological process scenarios (e.g., abundance, trend), decision thresholds, and monitoring effort. In cases when population rate of change is relatively small or population metrics are near the boundary of decision thresholds, results will likely be more sensitive to precision and survey efforts. Thus, our assessment highlights the importance of evaluating monitoring programs across variable underlying (true) conditions because classification errors are sensitive to multiple components of the biological and monitoring process.

Several studies have examined effects of life history, model complexity, process error, measurement error, and decision rules on the distribution and magnitude of misclassification errors for classification decisions (Connors et al., 2014; Dunham et al., 2021; Regan et al., 2013; Rueda-Cediel et al., 2015, 2018; Taylor et al., 1996). Classification decisions varied in sensitivity to each of these components and these studies highlight the considerable consequences associated with misclassification for species conservation. Further, these studies have encouraged careful consideration and scrutiny of data used to parameterize models informing classification decisions. While several patterns have emerged from research, several corroborated here, ultimately classification decisions vary in sensitivity to multiple components. When designing monitoring programs to inform species classification decisions, we need to consider decision context (i.e., scientific and policy components), life history, data, and model types, thus, we expect this framework to be applied on a case-by-case basis. Despite considerable evidence indicating classification decisions are prone to error and the potential consequences of misclassification, to our knowledge we are the first to suggest evaluating monitoring designs with the specific objective to minimize misclassification error.

While there are many conservation status assessment frameworks, we chose to use an ESA decision as the example to apply our framework. Importantly, ESA decisions do not include explicit decision rules to

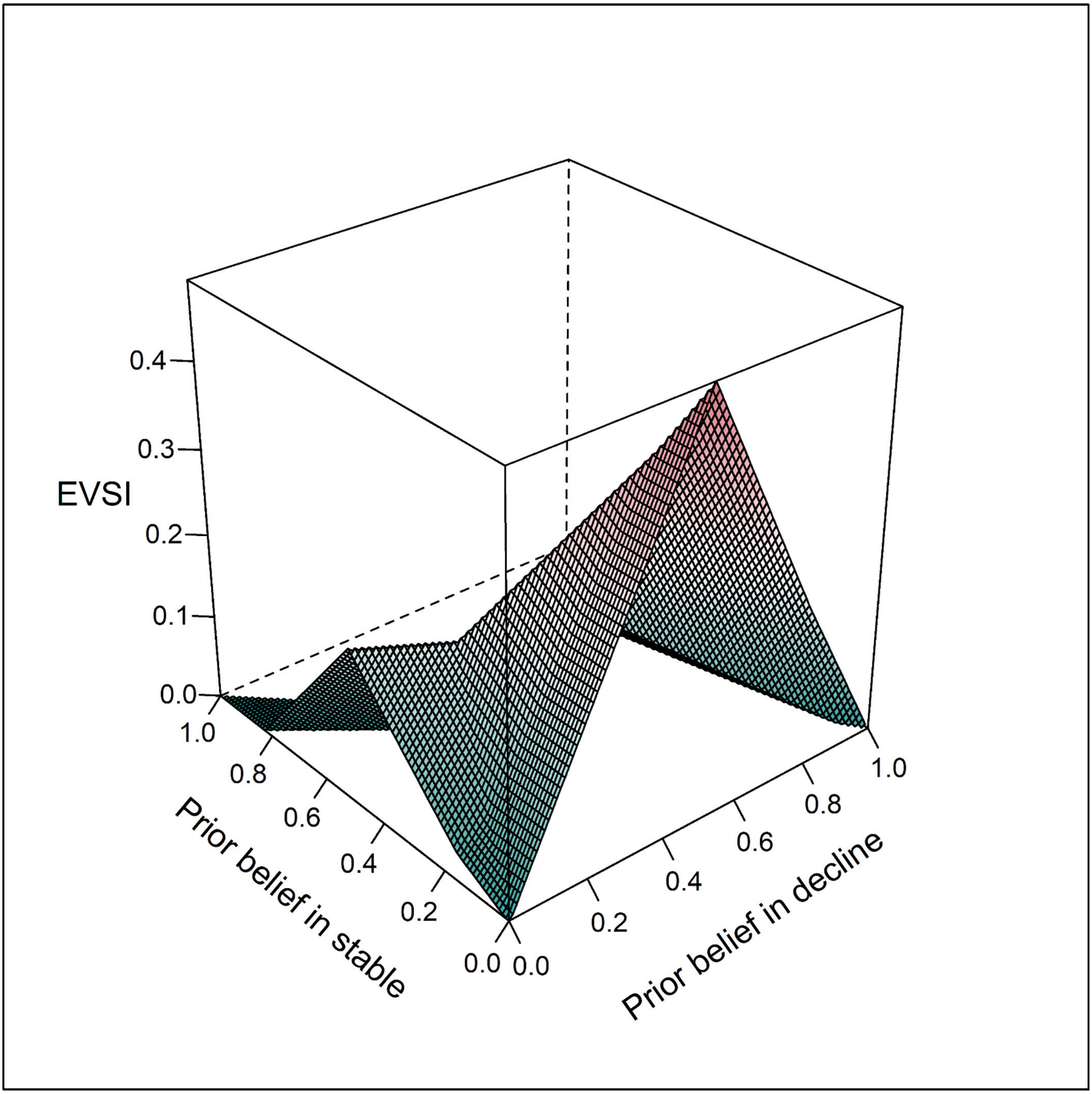

Fig. 7. Expected value of sample information (EVSI) as a function of prior belief in system state (i.e., increasing, stable, or declining). EVSI in this decision context is the loss avoided from an incorrect classification decision (loss function shown in Table 1). The results shown here are for monitoring scenario 7 (1500 sites, 3 visits, 5 years, detection = 0.2). Results are qualitatively similar for other monitoring scenarios. In this decision context, EVSI is greatest when prior probability that the population is declining is 0.5 and prior probability for stable population is 0.

categorize species to risk categories, which offers more flexibility than other frameworks such as PIF and IUCN. We believe the general framework we described here could be adapted to assess monitoring programs for many, if not all, programs that use quantitative criteria to inform classification decisions. Rueda-Cediel et al. (2018) assessed the effects of biological traits (e.g., age at first reproduction) and model choice on the accuracy of threat classifications under IUCN. Using our framework, we could readily extend their work to include defining specific monitoring scenarios and a risk tolerance threshold for evaluating the performance of monitoring scenarios to accurately classify species using IUCN Red List criteria. Such a modeling exercise may be particularly useful to design monitoring programs for species listed as data deficient, many of which are expected to be threatened (Bland et al., 2015; Parsons, 2016). For frameworks like PIF that generate scores from multiple categories of population metrics (e.g., population trend and nonbreeding distribution), practitioners could apply our framework by population metric category depending on available monitoring data or desired monitoring program(s). We expect our framework could be particularly useful for analyses requiring estimates of population trend when current scores reflect a lack of information (i.e., data deficient). In these cases, managers may need to design and assess monitoring programs to inform accurate classifications when breeding bird survey data are insufficient, a common occurrence for many species including secretive marsh birds (Carter et al., 2000). While we expect this framework to be applied on a case-by-case basis, we believe it can be useful to practitioners tasked with assessing existing programs and designing new monitoring programs across a broad range of conservation status assessment frameworks.

We selected a limited number of biological process and monitoring scenarios to demonstrate our approach and easily visualize results. Practitioners may be interested in monitoring program performance over a much wider range of biological scenarios (i.e., all possible initial occupancy rates) because the true state is unknown due to a lack of monitoring data or expert opinion on population status. We note that this step likely has important implications for the generalization of the results and the distribution and magnitude of classification errors. For example, if population state is expected to be near a decision threshold, we may need to generate many biological process scenarios aligning with potential reality to determine how intensively we need to monitor in these challenging cases. We envision these biological process scenarios being customized to the decision problem. For example, we could evaluate the true classification decision(s) using different metrics of population state (e.g., abundance) or include more realistic biological processes in which simulate truth based on species raster maps and model habitat change over time (Southwell et al., 2019). Monitoring scenarios were largely inspired by real monitoring programs employed for Eastern black rails (e.g., Tolliver et al., 2019) however, we believe our framework could be a useful tool for evaluating the performance of monitoring programs that employ human point-counts versus those using ARUs (e.g., Znidersic et al., 2021), or some combination.

We framed our case study as a single-objective problem focused on minimizing the risk of misclassification, however, our framework could be further generalized to address multiple criteria (objective) decisions (i.e., Converse, 2020). In practice, there are often multiple objectives that need to be considered, namely, financial cost. Given our results indicated monitoring scenarios with low power can return accurate classification decisions, managers may be interested in selecting options that reduce financial cost. This has potentially huge implications for resource allocation towards other species of conservation concern that could maximize cost efficiency of monitoring. We expect a particularly interesting extension for many managers would include assessing monitoring programs for multiple species that minimizes total misclassification risk over all species considered for a given budget. To address such multi-criteria decision problems, we would need to consider a different performance evaluation method in the final step that also minimizes financial costs or sets a maximum allowable budget to identify the optimal monitoring scenario.

Applications of the value of information in ecology and conservation are growing (Bolam et al., 2019). Most often, the value of information has been used to evaluate the robustness of management strategies to uncertainty, i.e., asking is it worthwhile to collect additional information to reduce uncertainty, or simply manage under uncertainty using $EV_{uncertainty}$ (Bennett et al., 2018; Johnson et al., 2014; Moore and Runge, 2012; Nicol et al., 2019). We used the value of information to evaluate prospective monitoring designs, i.e., investigate varying levels of spatial and temporal replication, which is a less common in the VOI literature (Bal et al., 2018). In one of few similar applications, Back et al. (2007) used EVSI to evaluate different sampling plans and cost-effectiveness in the context of reducing uncertainty about remediation of contaminated lands. In our case study, EVSI estimated the loss that could be avoided by increasing sampling intensity, and while EVSI generally increased with monitoring intensity, tradeoffs were evident. Our analysis gives managers information to think about in terms of information gained and costs when choosing a sampling plan. Diminishing returns on monitoring investment are not uncommon, and costs of additional monitoring are not always justified (Grantham et al., 2008; McDonald-Madden et al., 2010).

Classifying species into risk categories is a ubiquitous process in conservation decision-making with important implications for legal protection, conservation prioritization, and resource allocation (Rodrigues et al., 2006; Cummings et al., 2018). We developed this framework to propose a paradigm shift in the way monitoring programs are designed and evaluated in conservation decision-making. Managers are often tasked with assessing species conservation status and developing monitoring programs used to inform those decisions but may not assess monitoring program design with the risks of decision-making in mind (Lindenmayer et al., 2013, 2020; Lyons et al., 2008; Nichols and Williams, 2006). Our framework provides conservation practitioners opportunities to evaluate the effectiveness of monitoring programs a priori to improve conservation decisions and offers the flexibility to incorporate additional objectives depending on the decision problem (e.g., budget constraints). Our framework demonstrates that prioritizing monitoring that minimizes misclassification errors and evaluates the value of information can improve monitoring efficiency and conservation decision-making with considerable practical applications and benefits for species conservation.

CRediT authorship contribution statement

Kylee D. Dunham: Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing. **Patrick K. Devers:** Conceptualization, Writing – review & editing, Funding acquisition. **Abigail J. Lawson:** Conceptualization, Writing – review & editing. **James E. Lyons:** Conceptualization, Writing – review & editing, Formal analysis, Funding acquisition. **Conor P. McGowan:** Conceptualization, Writing – review & editing. **J. Andrew Royle:** Conceptualization, Formal analysis, Writing – review & editing, Funding acquisition, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data used in this analysis were simulated and can be replicated using the code included in the supplementary material.

Acknowledgments

We thank J. Price Tack, B. Folt, and two anonymous reviewers for their useful comments and insight that undoubtedly made this manuscript stronger. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.biocon.2023.110260.

References

Back, P.-E., Rosén, L., Norberg, T., 2007. Value of information analysis in remedial investigations. AMBIO 36, 486–493. https://doi.org/10.1579/0044-7447(2007)36[486:VOIAIR]2.0.CO;2.

Bal, P., Tulloch, A.I.T., Chadès, I., Carwardine, J., McDonald-Madden, E., Rhodes, J.R., 2018. Quantifying the value of monitoring species in multi-species, multi-threat systems. Methods Ecol. Evol. 9, 1706–1717. https://doi.org/10.1111/2041-210X.13037.

Bennett, J.R., Maxwell, S.L., Martin, A.E., Chadès, I., Fahrig, L., Gilbert, B., 2018. When to monitor and when to act: value of information theory for multiple management units and limited budgets. J. Appl. Ecol. 55, 2102–2113. https://doi.org/10.1111/1365-2664.13132.

Bland, L.M., Collen, B., Orme, C.D.L., Bielby, J., 2015. Predicting the conservation status of data-deficient species: predicting extinction risk. Conserv. Biol. 29, 250–259. https://doi.org/10.1111/cobi.12372.

Bolam, F.C., Grainger, M.J., Mengersen, K.L., Stewart, G.B., Sutherland, W.J., Runge, M.C., McGowan, P.J.K., 2019. Using the value of information to improve conservation decision making: value of information for conservation decisions. Biol. Rev. 94, 629–647. https://doi.org/10.1111/brv.12471.

Canessa, S., Guillera-Arroita, G., Lahoz-Monfort, J.J., Southwell, D.M., Armstrong, D.P., Chadès, I., Lacy, R.C., Converse, S.J., 2015. When do we need more data? A primer on calculating the value of information for applied ecologists. Methods Ecol. Evol. 6, 1219–1228. https://doi.org/10.1111/2041-210X.12423.

Carter, M.R., Hunter, W.C., Pashley, D.N., 2000. Setting conservation priorities for landbirds in the United States: the partners in flight approach. Auk 117, 541–548.

Connors, B.M., Cooper, A.B., Peterman, R.M., Dulvy, N.K., 2014. The false classification of extinction risk in noisy environments. Proc. R. Soc. B 281, 20132935. https://doi.org/10.1098/rspb.2013.2935.

Converse, S.J., 2020. Introduction to multi-criteria decision analysis. In: Structured Decision Making: Case Studies in Natural Resource Management, pp. 51–61.

Cummings, J.W., Converse, S.J., Smith, D.R., Morey, S., Runge, M.C., 2018. Implicit decision framing as an unrecognized source of confusion in endangered species classification: species classification decisions. Conserv. Biol. 32, 1246–1254. https://doi.org/10.1111/cobi.13185.

Dunham, K.D., Osnas, E.E., Frost, C.J., Fischer, J.B., Grand, J.B., 2021. Assessing recovery of spectacled eiders using a Bayesian decision analysis. PLoS One 16, e0253895. https://doi.org/10.1371/journal.pone.0253895.

Fiske, I., Chandler, R., 2011. Unmarked: an *R* package for fitting hierarchical models of wildlife occurrence and abundance. J. Stat. Softw. 43 https://doi.org/10.18637/jss.v043.i10.

Grantham, H.S., Moilanen, A., Wilson, K.A., Pressey, R.L., Rebelo, T.G., Possingham, H.P., 2008. Diminishing return on investment for biodiversity data in conservation planning: conservation value of data. Conserv. Lett. 1, 190–198. https://doi.org/10.1111/j.1755-263X.2008.00029.x.

Johnson, F.A., Jensen, G.H., Madsen, J., Williams, B.K., 2014. Uncertainty, robustness, and the value of information in managing an expanding Arctic goose population. Ecol. Model. 273, 186–199. https://doi.org/10.1016/j.ecolmodel.2013.10.031.

Lindenmayer, D.B., Piggott, M.P., Wintle, B.A., 2013. Counting the books while the library burns: why conservation monitoring programs need a plan for action. Front. Ecol. Environ. 11, 549–555. https://doi.org/10.1890/120220.

Lindenmayer, D., Woinarski, J., Legge, S., Southwell, D., Lavery, T., Robinson, N., Scheele, B., Wintle, B., 2020. A checklist of attributes for effective monitoring of threatened species and threatened ecosystems. J. Environ. Manag. 262, 110312 https://doi.org/10.1016/j.jenvman.2020.110312.

Lindenmayer, D.B., Lavery, T., Scheele, B.C., 2022. Why we need to invest in large-scale, long-term monitoring programs in landscape ecology and conservation biology. Curr. Landsc. Ecol. Rep. 7, 137–146. https://doi.org/10.1007/s40823-022-00079-2.

Lukey, J.R., Crawford, S.S., Gillis, D.J., Gillespie, M.G., 2011. Effect of ecological uncertainty on species at risk decisionmaking: COSEWIC expert opinion as a case study. Anim. Conserv. 14, 151–157.

Lyons, J.E., Runge, M.C., Laskowski, H.P., Kendall, W.L., 2008. Monitoring in the context of structured decision-making and adaptive management. J. Wildl. Manag. 72, 1683–1692. https://doi.org/10.2193/2008-141.

McDonald-Madden, E., Baxter, P.W.J., Fuller, R.A., Martin, T.G., Game, E.T., Montambault, J., Possingham, H.P., 2010. Monitoring does not always count. Trends Ecol. Evol. 25, 547–550. https://doi.org/10.1016/j.tree.2010.07.002.

McGowan, C., Angeli, N., Beisler, W., Snyder, C., Rankin, N., Woodrow, J., Wilson, J., Rivenbark, E., Schwarzer, A., Hand, C., Anthony, R., Griffin, R., Barrett, K., Haverland, A., Roach, N., Schnieder, T., Smith, A., Smith, F., Tolliver, J., Watts, B., 2020. Linking monitoring and data analysis to predictions and decisions for the range-wide eastern black rail status assessment. Endang. Species. Res. 43, 209–222. https://doi.org/10.3354/esr01063.

Moore, J.L., Runge, M.C., 2012. Combining structured decision making and value-of-information analyses to identify robust management strategies: identifying robust management strategies. Conserv. Biol. 26, 810–820. https://doi.org/10.1111/j.1523-1739.2012.01907.x.

Myers, N., 1993. Biodiversity and the precautionary principle. Ambio 22, 74–79.

Nichols, J., Williams, B., 2006. Monitoring for conservation. Trends Ecol. Evol. 21, 668–673. https://doi.org/10.1016/j.tree.2006.08.007.

Nicol, S., Brazill-Boast, J., Gorrod, E., McSorley, A., Peyrard, N., Chadès, I., 2019. Quantifying the impact of uncertainty on threat management for biodiversity. Nat. Commun. 10, 3570. https://doi.org/10.1038/s41467-019-11404-5.

Parsons, E.C.M., 2016. Why IUCN should replace "data deficient" conservation status with a precautionary "assume threatened" status—a cetacean case study. Front. Mar. Sci. 3 https://doi.org/10.3389/fmars.2016.00193.

R Core Team, 2021. R: A Language and Environmental for Statistical Computing.

Raiffa, H., Schlaifer, R., 1961. Applied statistical decision theory. Harvard University, Cambridge, MA.

Regan, T.J., Taylor, B.L., Thompson, G.G., Cochrane, J.F., Ralls, K., Runge, M.C., Merrick, R., 2013. Testing decision rules for categorizing species' extinction risk to help develop quantitative listing criteria for the U.S. Endangered Species Act: decision rules for listing species. Conserv. Biol. 27, 821–831. https://doi.org/10.1111/cobi.12055.

Robinson, N.M., Scheele, B.C., Legge, S., Southwell, D.M., Carter, O., Lintermans, M., Radford, J.Q., Skroblin, A., Dickman, C.R., Koleck, J., Wayne, A.F., Kanowski, J., Gillespie, G.R., Lindenmayer, D.B., 2018. How to ensure threatened species monitoring leads to threatened species conservation. Ecol. Manag. Restor. 19, 222–229. https://doi.org/10.1111/emr.12335.

Rodrigues, A., Pilgrim, J., Lamoreux, J., Hoffmann, M., Brooks, T., 2006. The value of the IUCN Red List for conservation. Trends Ecol. Evol. 21, 71–76. https://doi.org/10.1016/j.tree.2005.10.010.

Rosenberg, K.V., Kennedy, J.A., Dettmers, R., Ford, R.P., Reynolds, D., Alexander, J.D., Beardmore, C.J., Blancher, P.J., Bogart, R.E., Butcher, G.S., Camfield, A.F., Couturier, A., Demarest, D.W., Easton, W.E., Giocomo, J.J., Keller, R.H., Mini, A.E., Panjabi, A.O., Pashley, D.N., Rich, T.D., Ruth, J.M., Stabins, H., Stanton, J., 2016. Partners in Flight Landbird Conservation Plan: 2016 Revision for Canada and Continental United States, 119. Partners in Flight Science Committee.

Rueda-Cediel, P., Anderson, K.E., Regan, T.J., Franklin, J., Regan, H.M., 2015. Combined influences of model choice, data quality, and data quantity when estimating population trends. PLoS One 10, e0132255. https://doi.org/10.1371/journal.pone.0132255.

Rueda-Cediel, P., Anderson, K.E., Regan, T.J., Regan, H.M., 2018. Effects of uncertainty and variability on population declines and IUCN Red List classifications: assessments of IUCN Red Lists. Conserv. Biol. 32, 916–925. https://doi.org/10.1111/cobi.13081.

Runge, M.C., Converse, S.J., Lyons, J.E., 2011. Which uncertainty? Using expert elicitation and expected value of information to design an adaptive program. Biol. Conserv. 144, 1214–1223. https://doi.org/10.1016/j.biocon.2010.12.020.

Smith, D.R., Allan, N.L., McGowan, C.P., Szymanski, J.A., Oetker, S.R., Bell, H.M., 2018. Development of a Species Status Assessment Process for Decisions under the U.S. Endangered Species Act. J. Fish Wildl. Manag. 9, 302–320. https://doi.org/10.3996/052017-JFWM-041.

Southwell, D.M., Einoder, L.D., Lahoz-Monfort, J.J., Fisher, A., Gillespie, G.R., Wintle, B.A., 2019. Spatially explicit power analysis for detecting occupancy trends for multiple species. Ecol. Appl. 29 https://doi.org/10.1002/eap.1950.

Taylor, B.L., Wade, P.R., Stehn, R.A., Cochrane, J.F., 1996. A Bayesian approach to classification criteria for spectacled eiders. Ecol. Appl. 6, 1077–1089. https://doi.org/10.2307/2269592.

Tolliver, J.D.M., Moore, A.A., Green, M.C., Weckerly, F.W., 2019. Coastal Texas black rail population states and survey effort. J. Wildl. Manag. 83, 312–324. https://doi.org/10.1002/jwmg.21589.

Williams, B.K., Brown, E.D., 2020. Scenarios for valuing sample information in natural resources. Methods Ecol. Evol. 11, 1534–1549. https://doi.org/10.1111/2041-210X.13487.

Yokota, F., Thompson, K.M., 2004. Value of information analysis in environmental health risk management decisions: past, present, and future. Risk Anal. 24, 635–650. https://doi.org/10.1111/j.0272-4332.2004.00464.x.

Znidersic, E., Towsey, M.W., Hand, C., Watson, D.M., 2021. Eastern black rail detection using semi-automated analysis of long-duration acoustic recordings. ACE 16, art9. https://doi.org/10.5751/ACE-01773-160109.