



Developing a range-wide sampling framework for endangered species: a case study with light-footed Ridgway's rail

Bryan S. Stevens¹ · Courtney J. Conway² · Kimberly A. Sawyer³ · Lauren Kershek⁴ · Giselle Block⁵ · Sandra Hamilton⁴ · Rebecca Kolstrom⁵

Received: 5 February 2024 / Revised: 29 July 2024 / Accepted: 6 August 2024

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2024

Abstract

Monitoring provides the foundation for evaluating recovery of endangered species, yet many species lack monitoring programs designed to integrate a species' unique attributes, specific monitoring objectives, and principles of statistical sampling theory. We developed a framework for monitoring and assessment of endangered light-footed Ridgway's rails (*Rallus obsoletus levipes*) across their U.S. range, relative to multi-scale recovery goals. We created spatially explicit sample units and a sampling frame covering all potential habitat to facilitate range-wide probability sampling, and also built a model of the call-broadcast process commonly used to survey marsh birds that included heterogeneity in availability for detection and conditional detectability for each bird during each survey. We used the model to simulate 96 sampling strategies that included different levels of replication, multiple approaches for sample allocation amongst strata, and both simple random and weighted probability sampling (i.e., weights proportional to local rail abundance) of sample units within strata. Effective monitoring surveyed ≥ 20 –30% of the sampling frame on ≥ 3 occasions, with weighted sample selection and more targeted sampling (50% of units) for strata that are key to species recovery. We also tested Bayesian N-mixture models for estimating abundance and show that multiple models provide reasonable estimates. This work lays the foundation for statistical sampling and multi-scale population estimation for an endangered bird, and for refinement of abundance estimation models. Moreover, this work provides a replicable process for building customized and statistically defensible sampling frameworks to assess recovery of endangered species that can be used for other sensitive species.

Keywords Call-broadcast surveys · Population assessment · Probability sampling · Spatially explicit monitoring · Sensitive species · Species recovery

Communicated by Stephen Garnett.

Extended author information available on the last page of the article

Introduction

Monitoring is a cornerstone of applied ecology that provides a foundation for understanding and predicting ecological phenomena, as well as their implications for on-the-ground conservation (Yoccoz et al. 2001; Nichols and Williams 2006). Monitoring programs provide data for testing hypotheses and furthering our understanding of ecological processes (Lindenmayer et al. 2022), including changes in abundance (Harrity et al. 2020), occupancy and range dynamics (Kéry et al. 2013), habitat relationships (Kissel et al. 2023), and the effects of disturbance on wildlife (Kays et al. 2017). Large-scale monitoring programs are also increasingly used to develop decision-support tools explicitly for prediction, where models are built to inform future conservation and management across broad regions (e.g., Stevens and Conway 2020a). Such models are commonly used to predict the spatial distribution of organisms (Stevens and Conway 2019; Helmstetter et al. 2021), responses of wildlife to anthropogenic and environmental perturbations (Gerber et al. 2015; Stevens and Conway 2020b), and the effects of global change on organisms and habitats (Conroy et al. 2011; Stevens and Conway 2021). Moreover, recovery plans for threatened and endangered species (hereafter T&E species) often require monitoring programs to assess progress towards recovery goals.

Development of statistically rigorous monitoring frameworks for T&E species across large spatial domains is particularly challenging. Recovery planning is often limited by inadequate demographic data and insufficient funds to develop new monitoring programs (Campbell et al. 2002; Crouse et al. 2002), and range-wide population assessment is made more challenging by attributes common to T&E species. Broad geographic ranges paired with local rarity, large spatial variation and patchiness of populations, and the cryptic and difficult to detect nature of many T&E species combine to provide unique challenges to reliable monitoring (McDonald 2004; Mackenzie et al. 2005). In addition, the number of T&E species often overwhelms resources available for monitoring population recovery (Crouse et al. 2002; Lindenmayer et al. 2020). Thus, examples of rigorously designed and broad-scale monitoring programs for T&E species are rare (Evansen et al. 2021; Lindenmayer et al. 2022).

Secretive marsh birds (hereafter marsh birds) demonstrate many challenges of monitoring T&E species. Breeding populations are broadly, but also patchily distributed across North America (Stevens et al. 2022), and several species are listed as threatened or endangered in the U.S., Canada, or Mexico (U.S. Fish and Wildlife Service 1985; COSEWIC 2002; Diario Oficial de la Federacion 2002; U.S. Department of the Interior 2020). Marsh birds inhabit densely vegetated emergent wetlands that are difficult to access, and their cryptic behavior results in few visual or auditory detections during passive surveys (Eddleman et al. 1988; Lor and Malecki 2002). Consequently, call-broadcast methods that illicit marsh bird vocalizations by projecting calls into a wetland are used to increase detection (Conway et al. 2004; Conway and Gibbs 2005). Much effort has been devoted to developing field protocols for sampling marsh birds using call-broadcast techniques, and standardized guidelines for implementing surveys have provided a common methodology for data collection across North America (Conway 2011). Less attention has focused on sampling frameworks for marsh birds that are rooted in probability sampling concepts (e.g., spatially explicit sampling frames, sample units, and stratification, probabilistic sample selection, etc.). Johnson et al. (2009) provided recommendations for coordinated probability sampling

of marsh birds at a national scale across the U.S., and variations on these recommendations have been implemented in some areas (e.g., northeastern U.S.; Wiest et al. 2016). Yet T&E species often need targeted monitoring and assessment programs that accommodate species-specific attributes (e.g., local patchiness) and objectives (Lindenmayer et al. 2020).

The light-footed Ridgway's rail (*Rallus obsoletus levipes*) is a rare marsh bird endemic to coastal marshes and wetlands of southern California and Baja Mexico (Eddleman and Conway 2020). The species inhabits tidal wetlands and uses dense stands of California cordgrass (*Spartina foliosa*) for nesting and escape cover, and areas of vegetated high marsh as refugia during periods of high tide (Massey et al. 1984; Zembal et al. 1989; Zedler 1993; Barton 2016). Light-footed Ridgway's rails (hereafter rails when referencing *R. o. levipes* and marsh birds when referencing the broader group of birds, including other members of Rallidae) have been federally listed as endangered under the U.S. Endangered Species Act (ESA) since 1969 (U.S. Fish and Wildlife Service 1985), and the species' recovery plan provides recovery goals and population targets for downlisting (U.S. Fish and Wildlife Service 1985, 2019). However, existing monitoring does not permit estimation of range-wide abundance and hence cannot facilitate statistically rigorous tracking of populations relative to recovery goals. A statistical sampling framework is therefore needed to track recovery progress, and the objectives of this study were to: (1) develop a spatially explicit sampling frame for monitoring rails across their U.S. range, (2) develop a simulation model to test the efficacy of potential sampling designs, (3) use simulation to identify designs and levels of survey effort needed to characterize population status for rails at management-relevant scales, and (4) provide a preliminary test of a Bayesian N-mixture model for estimating rail abundance using data generated under our sampling framework and study design. This work represents an important first step towards development of a formal monitoring and assessment framework for estimating rail abundance range wide. Our results will inform monitoring not only for rails in California but also future efforts at large-scale population monitoring for other T&E species that are rare and patchily distributed over broad regions.

Methods

Study area

We developed a framework for monitoring breeding rail populations across their U.S. range (Fig. 1). The sampling areas included tidally influenced and freshwater wetlands known to provide habitat for breeding populations, as well as areas believed to contain potential breeding habitat for the subspecies within its range. We focused on breeding birds because recovery goals and past monitoring efforts focused on breeding rather than winter populations (U.S. Fish and Wildlife Service 1985, 2019). Wetlands that support rails are in coastal areas of southern California, where the landscape has been severely modified by human activities. Therefore, remaining habitat patches are typically adjacent to dense urban areas and housing developments, transportation infrastructure, or agriculture (Goodwin et al. 2001). Coastal wetlands of southern California have been reduced by 75% since the 19th century, and few unaltered wetlands remain (Marcus and Kondolf 1989). This loss and degradation of wetlands has led to the decline of wetland-dependent species across the region.

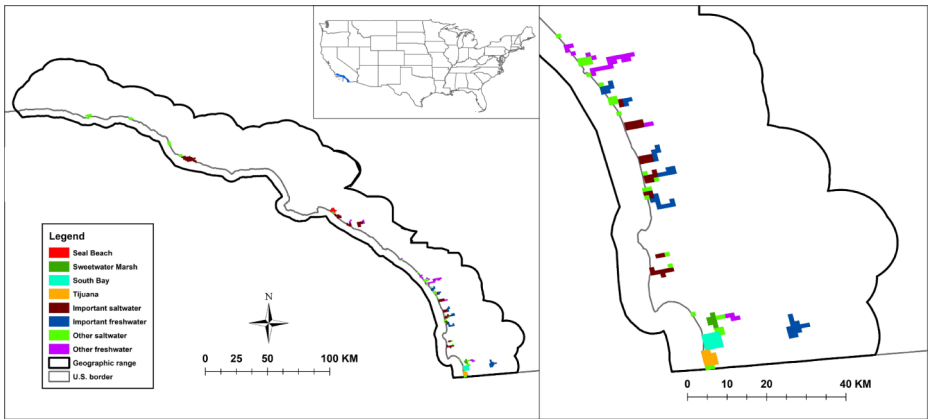


Fig. 1 Map of study area (left) where we developed a sampling frame for monitoring and assessment of light-footed Ridgway's rails in southern California, USA. The inset map (left panel) shows the geographic range (in blue) relative to the contiguous U.S., and the right panel shows only the southern portion of the species' range within the U.S. Areas covering the final sampling strata are indicated by different colors

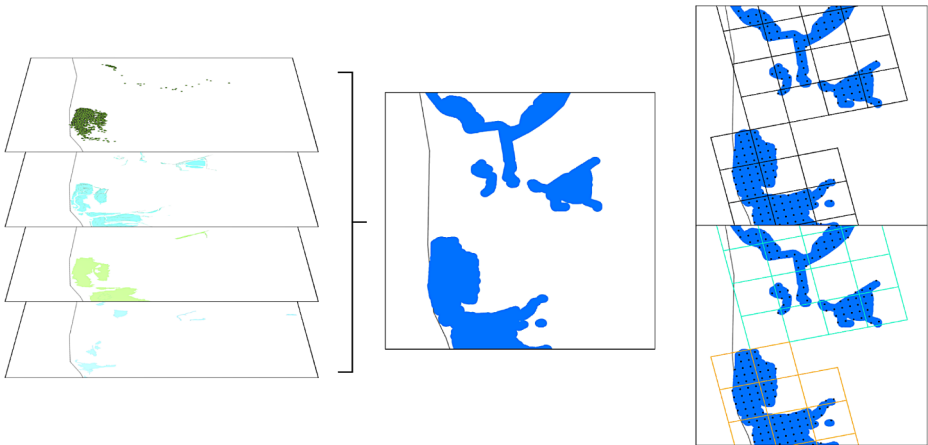


Fig. 2 Diagram depicting the process for building a sampling frame for monitoring light-footed Ridgway's rails. Rail location data (top left) were paired with spatial layers describing wetland characteristics (far left) to create a map of potential breeding habitat (blue, center). A systematic grid of survey points and 1 × 1 km grid-cell sample units were superimposed on the habitat layer (upper right), which was then stratified (e.g., different colored cells in bottom right) to accommodate local heterogeneity in rail density and the multi-scale goals of agencies responsible for monitoring and recovery

Spatially explicit sampling frame

We developed a spatially explicit sampling frame to facilitate probabilistic field surveys and assessment of rail populations. The sampling frame included potential rail habitat and was developed with the goal of facilitating population assessment across their range, as well as at specific areas of interest for recovery (e.g., national wildlife refuges). The process for developing a sampling frame involved 3 steps (Fig. 2): (1) conduct spatial analyses to

develop a simple model that maps potential habitat for rails across their range, (2) create a sampling frame that includes a grid of sample units and call-broadcast points covering all potential habitat, and (3) partition the sampling frame into spatial strata to accommodate anticipated heterogeneity in rail density and the goals of multi-scale abundance estimation.

The first step in building a sampling frame was to develop a habitat model, which resulted in the sampling universe of potential habitat available for rails across their range (Johnson et al. 2009). This process defined habitat by identifying categories of landcover and aquatic inventory variables from existing spatial data that captured as many of the known rail locations as possible, while also attempting to manage overall size of the habitat layer, and hence the feasibility of range-wide sampling, by eliminating sites that are very unlikely to be occupied. To identify wetland attribute variables to include in the habitat model, we used a geodatabase of 8,566 unique rail locations collected during annual survey efforts (1980–2021) that were conducted at important sites for the species (Zembal and Hoffman 2021). We paired rail location data with 3 publicly available data sets that characterized land cover and wetland attributes: (1) a polygon layer of existing vegetation for the coast of California created by the U.S. Forest Service (hereafter CalVeg data; Nelson et al. 2015; accessed 2/2022), (2) a 30-m resolution raster data set of regional landcover created by the National Oceanic and Atmospheric Administration as part of its Coastal Change Analysis Program (hereafter NOAA C-CAP data; NOAA Office for Coastal Management 2022; accessed 2/2022), and (3) a polygon layer of wetland attributes from the California Aquatic Resource Inventory version 0.3 (hereafter CARI data; San Francisco Estuary Institute 2017; accessed 2/2022).

We used these data to develop a spatial habitat model (i.e., an area was considered habitat or not) that was inclusive of known rail locations. However, we also managed the overall size of the sampling frame in attempt to limit the resources expended to sampling areas where breeding status is presently unknown but where extant populations are unlikely (i.e., wetland patches far from the Pacific Coast and within watersheds where the birds have not historically bred). We tested the model using location data described above and 2 auxiliary data sets: (1) incidental locations reported to the U.S. Fish and Wildlife Service under federal recovery permit requirements (hereafter incidental observations; $n=294$ observations collected 1998–2020), and (2) observations reported by citizen scientists to eBird (eBird Basic Dataset 2021, $n=11,358$ observations). The final habitat model allowed us to focus sampling to areas within watersheds where ≥ 1 rail was detected during annual survey efforts (Zembal and Hoffman 2021), and ≤ 7.75 km from the Pacific Coast, reducing the area of sampling while also containing $>98\%$ of all known rail locations across the 3 data sets. Thus, the final habitat model provided a reliable depiction of plausible breeding habitat based on known historical records, while excluding local areas within the broader geographic range (Fig. 1) where breeding populations have not historically been found. We used ArcMap 10.5.1 (ESRI, Redlands, CA) for all spatial analyses, and details of the habitat model development are provided in Supplement A.

The second step was to create a grid of call-broadcast points and sample units covering rail habitat. We generated a systematic grid with a random starting location and points spaced 200 m apart across species range and retained only those points that intersected with potential habitat. This provided a range-wide network of survey locations sufficiently close to ensure possible detection of all rails available (i.e., if all points were surveyed). To organize call-broadcast points into groups that could be sampled during a single site visit (e.g., similar to a survey route or a cluster of survey locations; Johnson et al. 2009; Conway

2011), we overlaid a 1×1 km grid of cells across the species range and retained only those cells intersecting with habitat. Under this design, each 1×1 km grid cell serves as a primary sample unit that can be randomly selected in any given year, where the range-wide sampling frame is the collection of all sample units (i.e., all grid cells) across the study region. The collection of call-broadcast points within randomly selected grid cells can then be repeatedly sampled following standardized field protocols.

The last step was to partition sample units into strata to accommodate heterogeneity in local rail density, as well as to facilitate the multi-scale estimation goals of agencies responsible for monitoring and managing recovery. The objectives for monitoring and assessment were based on the species' recovery plan and included: (1) range-wide abundance estimation, (2) local-scale abundance estimation at 4 national wildlife refuge (NWR) sites important for recovery (Seal Beach NWR, South Bay unit of San Diego Bay NWR [hereafter South Bay], Sweetwater Marsh unit of San Diego Bay NWR [hereafter Sweetwater], and Tijuana Slough NWR [hereafter Tijuana]), and (3) abundance estimation across additional protected areas outside of the NWR system that have a history of rail occurrence (hereafter additional important sites, Supplement A). We used spatial stratification to facilitate multi-scale sampling, where randomized sampling within each stratum provides a means of range-wide abundance estimation via simple addition of the estimates generated within each stratum. We started by partitioning our sample units spatially into 6 strata: (1) Seal Beach, (2) South Bay, (3) Sweetwater, (4) Tijuana, (5) additional important sites (see Appendix A), and (6) other areas not included in strata 1–5 (i.e., all remaining potential habitat).

We also anticipated rail density would differ between more inland freshwater sites (i.e., inland sites or coastal sites with infrastructure reducing influence of tidal dynamics) and coastal sites with predominantly saltwater influence. Thus, we split the additional important sites and the other habitat sites (strata 5 and 6 above) by salinity (i.e., freshwater vs. saltwater dominant) based on local expert knowledge. This resulted in 4 spatial strata covering important NWR sites (Seal Beach, South Bay, Sweetwater, and Tijuana), and 4 strata covering all remaining sample units (important saltwater, important freshwater, other saltwater, and other freshwater; Fig. 1). Lastly, each stratum and each sample unit were reviewed by local experts with knowledge of rail ecology and coastal wetlands within southern California (i.e., 2 fish and wildlife biologists, 1 refuge manager, and 1 geospatial specialist with U.S. Fish and Wildlife Service) to ensure that: (1) sample units were assigned to the correct stratum (e.g., for saltwater vs. freshwater sites), (2) sample units included rail habitat, and (3) sample units covered all potential habitat at sites known to harbor breeding rails. This resulted in several sample unit deletions, additions, and stratum reassignments to arrive at the final sampling frame (Supplement A).

Simulation model

We developed a simulation model to mimic the call-broadcast sampling process of surveying and counting rails. The simulation model included 4 general steps: (1) simulate the number of rails inhabiting each sample unit, (2) simulate the spatial location of each bird within each sample unit, (3) simulate rail counts from surveys at each survey point, and (4) replicate step 3 to simulate repeated visits within a season.

The first step was to simulate the true number of birds inhabiting each sample unit. We used data from recent annual surveys (Zembal and Hoffman 2021) to inform the simulation

and realistically capture spatial and temporal variation in rail abundance across the study area. We counted the number of rails recorded annually within each sample unit for the most recent 15 years (2007–2021) and calculated the mean and standard deviation of counts within each sample unit. We simulated the true number of rails inhabiting each cell from a truncated normal distribution by using the cell-specific means and standard deviations. To simulate rail counts under a broader range of plausible conditions for true abundance, we also simulated scenarios where true rail densities were much lower (i.e., mimicking sampling under reduced populations that would be expected under population decline) and higher (i.e., mimicking a larger population expected under population growth or if recent counts were severely underestimated) than those reflected by recent survey data (additional details below). The randomly generated number of birds was rounded to the nearest whole number to provide a discrete count and left truncated at 0 to prevent negative numbers. If the summary of recent location data resulted in a mean of zero (i.e., all recent counts equal to zero) for a given sample unit, the minimum non-zero value recorded from the other sample units within the same stratum was used to simulate truth for that cell. This ensured all sample units had potential to harbor breeding rails inside the simulation.

After generating the true number of rails within each cell, we simulated the location of each bird and calculated its distances to each survey point. We created unique shapefiles for the habitat and survey points located within each cell and used the “sf” package in R (Pebe-sma 2018) to: (1) generate random locations independently for each rail, and (2) calculate the distances between each bird and each survey point. We used the `st_sample()` function to generate locations of individuals within habitat of each sample unit from a spatial uniform distribution, with all locations equally likely. We used the `st_distance()` function to calculate distance matrices containing distances of each rail to each survey point within each cell. We treated the location of each rail as constant across a breeding season during the simulation, and therefore individual rail locations were intended to mimic a breeding season activity center for each bird. We believe this is biologically reasonable because small deviations in the location of individual birds among surveys would have negligible impact on simulated rail counts relative to survey-level variation in detectability already included in the model (described below).

Next, we simulated rail counts from call-broadcast surveys within each sample unit and replicated the process to simulate repeated visits over time (Fig. 3). An individual rail may or may not be available for detection during a survey because it may or may not respond to the call broadcast. Thus, we first simulated availability for detection for each bird during each survey (i.e., whether or not the bird responded from each survey point) as a Bernoulli random variable (1=response, 0=no response). We borrowed the baseline response probability of 0.4 from the closely related Yuma Ridgway’s rail (*R. o. yumanensis*; Conway et al. 1993), but also considered higher and lower values (described below). If a rail was not available for detection it was not counted during the simulated survey at a given point, but if the bird responded to the call broadcast it was available and therefore could be detected by a surveyor (Fig. 3A). We simulated the call type as either kek-hurrah or paired clatter (Conway 2011) when a bird was available for detection. We are unaware of data on the relative frequency of kek-hurrah or paired clatter calls in response to call broadcast, thus these call types were simulated with equal probability. Yet we simulated the call types based on local field observations because kek-hurrah and paired clatter calls have different distance-decay functions for conditional detection probability, where on average kek-hurrah calls are

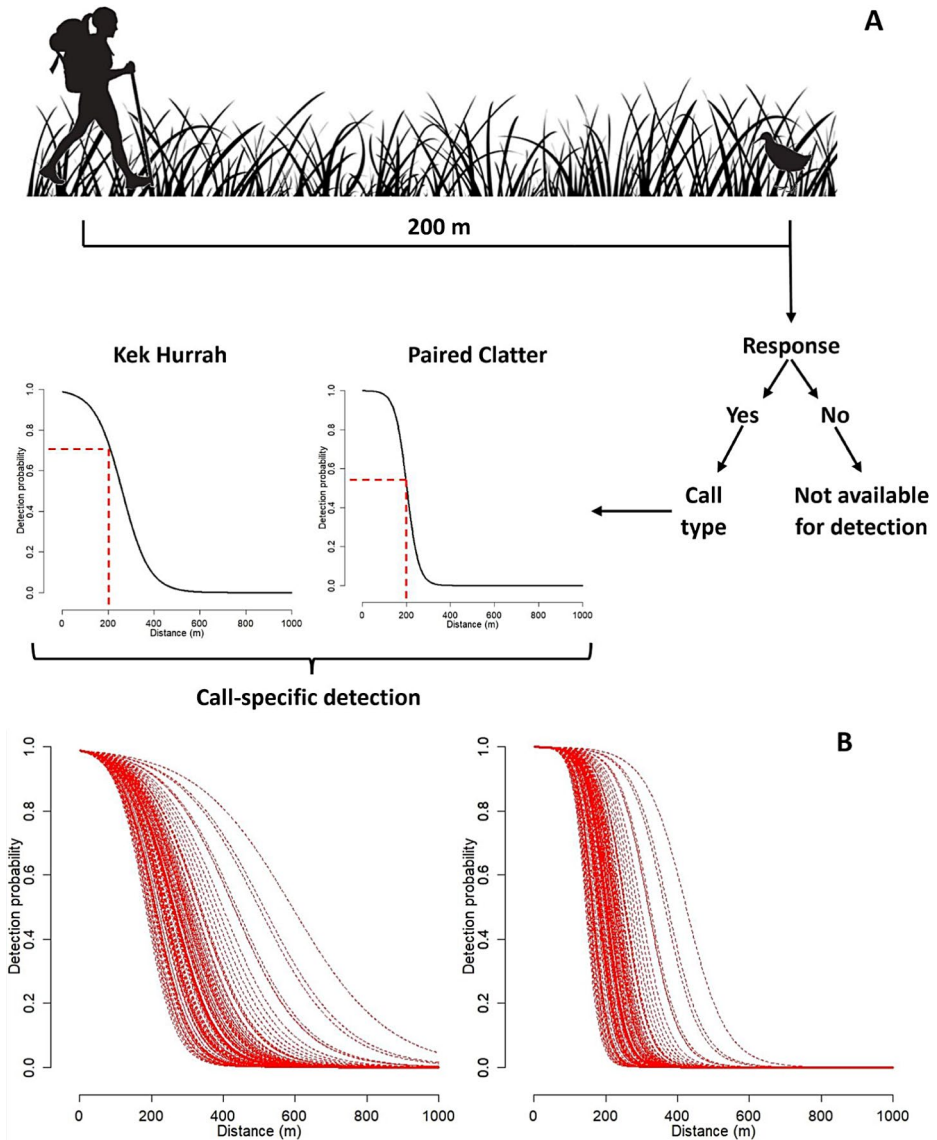


Fig. 3 Diagram of the model developed for call-broadcast sampling (**A**) and example detection functions for individual birds during surveys (**B**). For each simulated call-broadcast survey, every bird present either responded or not, and birds that responded had a detection function randomly generated for kek-hurrah (left) or paired clatter (right) calls. The distance-detection functions shown in A provide the central tendency for the model. Panel B provides an example of 100 simulated detection functions for conditional detection probabilities arising from kek-hurrah (left) and paired clatter (right) calls

more readily detected at greater distances from the observer (Fig. 3A, Supplement A). We randomly generated a conditional distance-decay relationship for each available bird during each survey from its corresponding call type (Fig. 3B, Supplement A) and used these relationships to calculate the detection probability for each bird based on its distance from

the observer. We used the individual bird- and survey-specific detection probabilities to randomly generate detection (1) or not (0) for each bird on each survey as a Bernoulli random variable. Thus, availability and conditional detection probability of each bird present within a sample unit could change for each survey point and across site visits. We also simulated multiple site visits at a survey location during the breeding season to accommodate common statistical modeling frameworks used to correct raw counts for detection error (Kéry and Royle 2015). Finally, the total count for each survey was recorded as the sum of detections over all survey points within each sample unit.

We replicated simulation of call-broadcast surveys across multiple combinations of rail density and availability for detection to understand robustness of sampling recommendations derived from our model. We considered 3 scenarios of density (low, medium, high), where the medium-density scenario used the cell-specific mean counts as described above and was therefore intended to mimic current conditions. The low- and high-density scenarios multiplied the cell-level average rail counts by 0.5 and 2.0, respectively. Similarly, we considered 3 scenarios of availability (low, medium, high), where the baseline probability of response (0.4) was considered the medium availability scenario. The low- and high-availability scenarios were 0.2 and 0.6, respectively based on the range of response rates observed for Yuma Ridgway's rail (Conway et al. 1993). Thus, we considered 9 total scenarios of truth (3 density $\times$ 3 availability). We generated 1,000 data sets for each scenario, where each data set generated the true number of birds and call-broadcast counts within each sample unit for 5 site visits. We used program R version 4.2.2 (R Core Team 2022) to program the model and to conduct all simulation analyses.

Simulated monitoring and population assessment

We simulated monitoring with multiple sampling strategies and a range of spatial and temporal replication. We simulated 4 scenarios of sampling intensity that included 20%, 30%, 40%, and 50% of all sample units from the range-wide sampling frame. We also simulated 4 scenarios of frequency that included 2, 3, 4, and 5 visits to each site during a breeding season. For each of the 16 scenarios of replication (4 spatial intensity $\times$ 4 temporal frequency), we tested 6 sample-selection strategies that included: (1) 3 approaches for allocating the total sample among the 8 strata and, (2) 2 approaches for randomly selecting sample units within each stratum. Thus, we tested performance of monitoring under 96 sampling strategies (4 space $\times$ 4 time $\times$ 3 allocation $\times$ 2 sample selection). To assess robustness of our conclusions, we replicated simulations for each of the 96 sampling strategies across all simulated data sets from the 9 truth scenarios described above (3 rail density $\times$ 3 availability). This resulted in 864,000 simulated data sets that we used to derive monitoring recommendations (96 sampling strategies $\times$ 9 truth scenarios $\times$ 1,000 data sets each).

Under stratified random sampling, a sample allocation strategy was used to determine how many samples to allocate to each stratum (Scheaffer et al. 2006). We tested proportional allocation, Neyman allocation, and a customized strategy we developed to ensure adequate samples were allocated to the NWR strata. Proportional allocation divides the sample amongst the strata according to the proportion of the sampling frame contained therein. For example, if the total sample size was 100 grid cells and stratum A contained 10% of the total number of cells in the sampling frame, then proportional allocation would allocate 10 samples to stratum A. Neyman allocation is the theoretically optimal allocation

for range-wide estimation, wherein samples are allocated proportionally to the stratum size and within stratum variance. That is, when the spatial variance of rail counts among cells within a stratum is larger (i.e., high patchiness of local density), more samples are allocated to that stratum. Given the small number of sample units contained within the 4 NWR strata (Supplement A), we were concerned that both proportional allocation and Neyman allocation would not allocate enough samples to these areas to enable accurate local-scale assessment. Consequently, we devised a customized strategy (hereafter targeted allocation) to better address the dual goals of population assessment, both range-wide and at local NWRs. For this approach we automatically sampled 50% of the grid cells within each of the NWR strata and then used Neyman allocation to allocate the remaining samples to the remaining strata.

We also tested 2 approaches for randomly selecting sample units within each stratum. First, we tested simple random sampling whereby each sample unit has an equal chance of being selected. Second, we tested weighted probability sampling, whereby the inclusion probability for each sample unit in each stratum was weighted by recent abundance data. Inclusion probabilities were calculated as a function of the average cell-level rail counts from recent surveys (2007–2021; Supplement A). Weighted sampling increases the chances of randomly selecting more populous units and therefore minimizes the risk of missing most of the birds in any one stratum due to random chance. Thus, weighted sampling can increase the efficiency of sampling when strong spatial variation in abundance exists among cells contained within the same stratum.

We simulated monitoring for each sampling strategy by randomly selecting sample units and their call-broadcast data from the simulated data sets described above. We randomly selected the appropriate number of grid cells within each stratum from each data set and retained the simulated counts corresponding to the desired temporal replication. For example, under the 2-visit per year scenario we retained call-broadcast counts from the first 2 simulated visits for all survey points within each randomly selected grid cell. We used a variety of metrics to assess the performance of sampling strategies and implications of changing rail density and availability. We assessed the efficacy of sampling by calculating the overall fraction of sample units in each data set that were truly unoccupied (because sampling is less efficient if surveyors spend too much time visiting unoccupied areas), as well as the fraction of occupied cells where ≥ 1 rail was detected. To indicate bias resulting from indices created from raw summary counts, we first calculated the maximum cell-level counts across site visits for each randomly selected sample unit (i.e., the maximum across site visits of the sum of counts over all broadcast points). We then calculated the ratio of the maximum count to the true number of rails ($\frac{\text{count}}{\text{truth}}$). Thus, we used these grid cell level metrics to indicate how efficacy and bias of count-based indices are likely to change with rail density and availability for detection.

We used a subset of our simulated data to evaluate performance of algebraic design-based abundance estimators. We focused on the medium rail density with low availability for detection scenario because medium density is most reflective of recent conditions and low availability should provide a conservative assessment of effort required (i.e., because a smaller fraction of birds is detected). The design-based estimators did not account for detection error and instead treated the maximum sum of rail counts across all call-broadcast points within a sample unit as the index of abundance within that sample unit. The computationally intensive nature of models required to estimate abundance while accurately accounting for

survey-level heterogeneity in detection prevented us from assessing model-based estimators for all simulated datasets (see below). Our assessment of design-based estimators therefore provided a straightforward approach to estimation and an understanding of how bias and precision of population indices likely change with sampling strategy and replication (Johnson et al. 2009). We calculated estimates of population size for individual strata for each simulated data set, and we calculated estimates of the range-wide total abundance as the sum of population estimates across all strata. We used the relative difference ($\frac{\text{truth}-\text{estimate}}{\text{truth}}$) and coefficient of variation (CV) of estimates to assess bias and precision of these population indices at the stratum- and range-wide levels. Mathematical details of sampling and estimation are found in Supplement A.

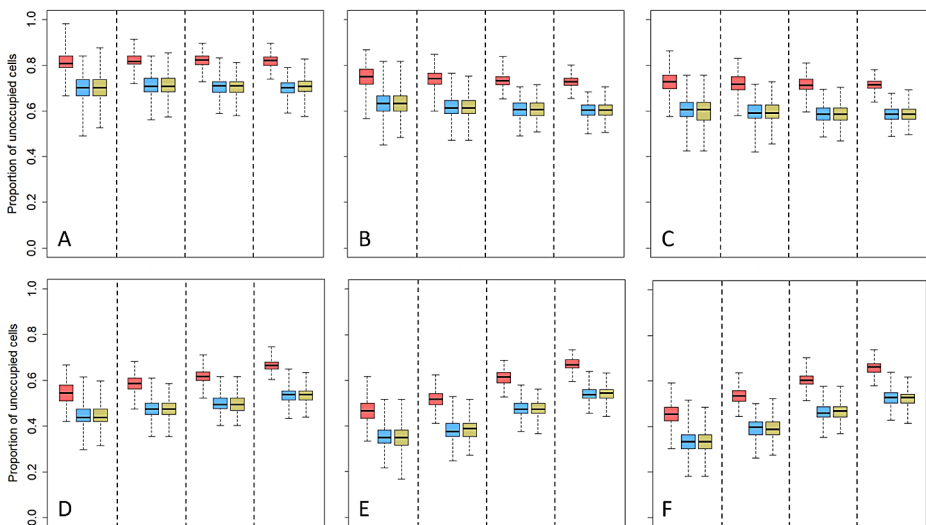
Lastly, we used Bayesian N-mixture models to test a proof-of-concept for model-based abundance estimation from rail monitoring data generated under our sampling framework. We tested 2 models, each with 2 plausible random-effects structures that can account and correct for the realistic heterogeneity in availability and detection that our sampling model simulated, resulting in 4 total parameterizations. The model parameterizations included zero-inflated versions of the Poisson-Lognormal and Poisson-Poisson mixture models, where we fit both models with site $\times$ survey and site-level random effects in the detection function (Kéry and Royle 2015). All models included stratum-level differences in zero-inflation parameters to account for differences in occupancy rates among strata, and cell-level random effects in rail density. We randomly selected one of the simulated monitoring data sets to implement model fitting, with the goal of conducting an analysis that used a conservative but realistic set of conditions consistent with monitoring recommendations deduced from analyses described above. Specifically, we selected 1 data set from the low availability and medium rail density scenario, under 20% spatial replication (i.e., 20% of the total sampling frame randomly selected and sampled), targeted sample allocation, and weighted probability sampling within each stratum. Each model was fit assuming 3, 4, and 5 site visits, respectively. We used JAGS (Plummer 2003) called from within R via the R2JAGS package (Su and Yajima 2015) for model fitting. We used 3 Markov Chain Monte Carlo (MCMC) chains to fit each model, each with a burn-in period of 250,000 samples followed by 750,000 samples that were thinned at a rate of 1:50, which resulted in 15,000 retained posterior samples per chain (45,000 total samples). We used the multivariate Gelman-Rubin statistic ($GR < 1.1$; Gelman and Rubin 1992) to assess model convergence and compared point estimates of abundance to the true simulated abundance within each stratum and range-wide. We provide complete model structure details and JAGS code in Supplement A.

Results

The final range-wide sampling frame included 268 grid cells, with stratum sizes ranging from 6 sample units in the Sweetwater stratum to 92 in the Additional Important Saltwater stratum (Table 1). The total number of survey points within the sampling frame was 2,234 and ranged from 76 points in Sweetwater to 1,010 in the Additional Important Saltwater stratum (Table 1). The final sampling frame contained 98.9% of light-footed Ridgway's rail locations (19,988 of 20,218), including 99.8% (8,553 of 8,566) of locations recorded dur-

Table 1 Final sampling strata, number of sampling units, and number of call-broadcast point locations for the sampling frame developed for range-wide monitoring of light-footed Ridgway's rails

Stratum	No. sample units	No. call-broadcast points
Seal Beach	10	128
Sweetwater	6	76
South Bay	16	109
Tijuana	11	133
Additional Important Saltwater	92	1,010
Additional Important Freshwater	44	270
Other Saltwater	56	382
Other Freshwater	33	126
Total	268	2,234

**Fig. 4** Simulated proportion of randomly selected sample units that were unoccupied by light-footed Ridgway's rails under different sampling strategies. Rows indicate sample selection (panels A–C= simple random, panels D–F=weighted probability) and columns indicate sample allocation strategies (panels A, D=proportional, panels B, E=targeted, panels C, F=Neyman). Within each panel, the vertical dashed lines separate spatial replication scenarios (20–50% from left to right), and colors show scenarios of rail density (low=red, medium=blue, high=gold). Boxplots show the interquartile range (box boundary), median (horizontal black line), and maximum and minimum values (whiskers). These plots are shown for the low availability data set

ing annual survey efforts, 92.1% (271 of 294) of incidental locations, and 98.3% (11,164 of 11,358) of locations reported to eBird.

Simulations showed that sampling efficiency as measured by the proportion of unoccupied grid cells contained within samples changed with strategies for sample allocation and randomization (Fig. 4). Targeted and Neyman allocation generally resulted in less sampling of unoccupied cells than proportional allocation of samples among strata yet differed little from each other. Weighted probability sampling also selected fewer unoccupied cells than simple random sampling (Fig. 4). For example, with targeted allocation under the medium

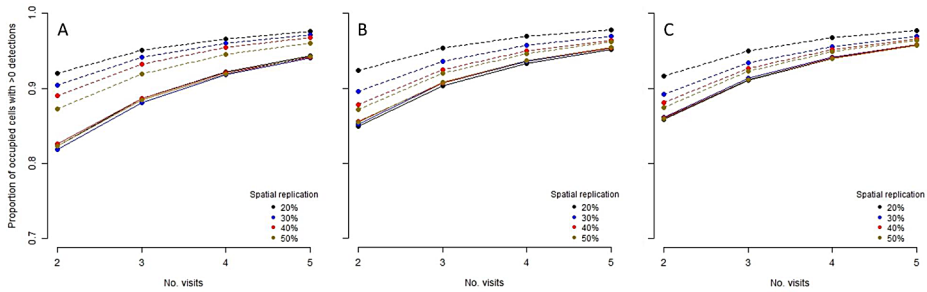


Fig. 5 Simulated proportion of randomly selected sample units where ≥ 1 rail was detected in the sample unit (occupied cells only) as a function of the number of site visits conducted annually. Dots represent the mean value across simulation iterations for the low availability scenario with medium rail density. Line types represent within-stratum sample selection strategy (solid=simple random, dashed=weighted probability sampling), and colors represent different levels of spatial replication. The panels represent different sample allocation strategies (A=proportional, B=targeted, C=Neyman)

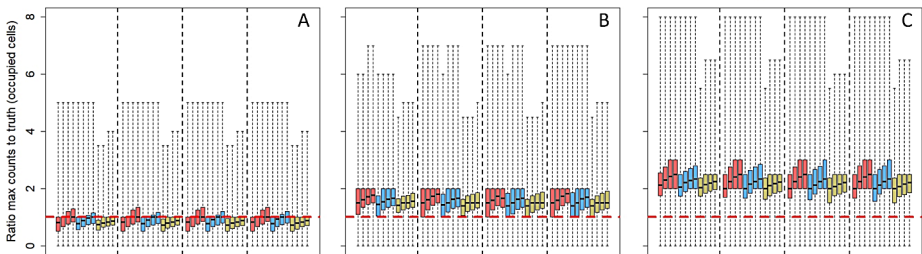


Fig. 6 Ratio of maximum count to truth (i.e., the maximum count over all surveys / true abundance at the sample-unit level) for weighted probability sampling with targeted allocation. The panels are low (A), medium (B), and high availability (A). Within each panel, the spatial replication scenarios are separated by vertical dashed lines (20%, 30%, 40%, and 50% of cells sampled, left to right). Within each spatial replication scenario, color indicates rail density (low=red, medium=blue, high=gold), and the number of surveys per year (2–5) are indicated within each color going left to right. Boxplot shows IQR (box boundaries), median (solid horizontal lines inside box), and extreme high and low values (whiskers)

density scenario, weighted sampling decreased the median proportion of unoccupied cells by 44%, 38%, 23%, and 10% relative to simple random sampling, under the 20–50% spatial replication scenarios. Weighted probability sampling was more efficient in general; under weighted sampling differences in efficiency among allocation strategies were eliminated by increasing the amount of spatial replication (Fig. 4). Weighted probability sampling was also more efficient for detecting ≥ 1 bird in a grid cell when the cell was occupied (Fig. 5). At the grid cell level, false zero counts (i.e., occupied sample units where no rails were detected during surveys) were recorded at $< 10\%$ of occupied cells when sites were visited ≥ 3 times per year, irrespective of rail density (Fig. 5, B1). In addition, the maximum sum of raw counts (i.e., summed over survey points within a cell) among all site visits provided an approximately unbiased index of local abundance, so long as availability for detection was low (Fig. 6, B2–B3). However, bias and variance of cell-level maximum counts scaled with availability for detection, resulting in systematic over-counting and larger variance of

total rail counts under medium or high availability scenarios (Fig. 6), which was consistent across all sampling strategies and rail densities (Figs. B2–B3).

Range-wide population assessment with simple design-based approaches showed that increasing the number of site visits per year decreased bias, which otherwise changed little among sampling strategies or levels of spatial replication (Fig. 7). The CV of range-wide estimates of abundance was minimized with weighted probability sampling and any of the sample allocation strategies, yet targeted and Neyman allocations showed more consistency of precision among the scenarios of spatial replication (Fig. 8). These patterns were generally shared at the stratum level, where weighted sampling typically resulted in better precision than simple random sampling (Figs. B4–B19). Moreover, while spatial replication did not meaningfully improve precision of range-wide indices, it did improve precision at the stratum level, which were less precise overall. We detected some heterogeneity in the optimal allocation strategy among individual strata, but targeted allocation was either optimal or competitive for most strata (Figs. B5–B19; Table B1). The exception was for the Other Saltwater and Other Freshwater strata, where proportional allocation of samples resulted in better precision (Figs. B17, B19). Collectively these results demonstrate that reasonably unbiased estimates with nearly optimal precision for most strata could be achieved through weighted probability sampling in tandem with targeted allocation of samples among strata and $\geq 20\%$ spatial replication, or through weighted sampling in tandem with proportional allocation and $\geq 40\%$ spatial replication (Table B1). Implementing these suggestions range-

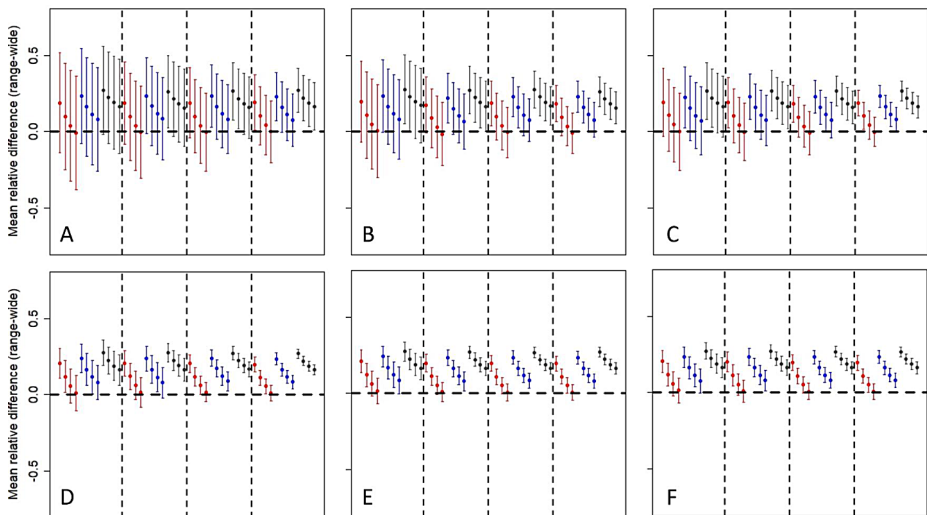


Fig. 7 Relative difference of range-wide population estimates from true abundance for simulated light-footed Ridgway's rail monitoring and design-based estimators. Dots show mean relative difference across simulation replicates and whiskers show 1 standard deviation calculated across replicates for the medium rail density and low availability scenario. Rows indicate sample selection (panels A–C=simple random, panels D–F=weighted probability) and columns indicate sample allocation strategies (panels A, D=proportional, panels B, E=targeted, panels C, F=Neyman). Within each panel, the vertical dashed lines separate spatial replication scenarios (20–50%, left to right), colors show scenarios of rail density (low=red, med=blue, high=black), and the number of site visits per year (2–5) within each density scenario is indicated (left to right). Horizontal dashed lines represent no difference between an estimate and truth

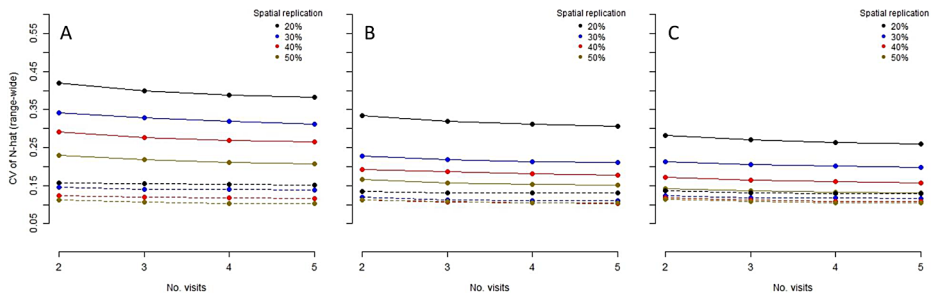


Fig. 8 Coefficient of variation (CV) of range-wide population estimates for simulated monitoring and design-based estimators. Dots represent the mean CV across simulation iterations for the low availability scenario with medium rail density. Line types represent within-stratum sample selection strategy (solid=simple random, dashed=weighted probability sampling), whereas colors represent different levels of spatial replication. The 3 panels represent different sample allocation strategies (A=proportional, B=targeted, C=Neyman)

Table 2 Estimated number of technicians required to implement range-wide sampling of light-footed Ridgway's rails at various levels of spatial (i.e., percent of sample units surveyed) and temporal replication (no. Site visits)

Spatial replication	No. site visits		
	3	4	5
20%	4 ^a	5	6
30%	5	7	9
40%	7	9	12
50%	9	12	14

^aEstimates of the number of field technicians (rounded up to the nearest whole number) required assume: (1) one person can survey 10 broadcast points per day for 40 field days in a breeding season, and (2) 8.33 survey points per 1 × 1 km grid cell sample unit (i.e., the average number of points per cell across all sample units)

wide would require approximately 4–7 technicians per year to conduct call-broadcast surveys within randomly selected sample units throughout the species range (Table 2).

All models fit to simulated data for our proof-of-concept analysis of the Bayesian N-mixture model converged ($GR < 1.1$). The accuracy of range-wide population estimates relative to the simulated true abundance was reasonable for most parameterizations and improved with the number of site visits per year (Table 3). The exception was for the Poisson-Lognormal mixture model with the site-level random effect for detection, which underestimated range-wide abundance by 23% under the best scenario. However, our analyses demonstrate that both Poisson-Poisson mixture models and the Poisson-Lognormal model with site × survey random effects in detection provided reasonable range-wide abundance estimates (Table 3). Abundance estimates at the stratum level varied considerably in accuracy, but generally results followed those of range-wide abundance estimation, with improved estimates with more site visits per year (Table 4). Again, the Poisson-Poisson mixture model provided reasonable abundance estimates for most strata (Table 4).

Table 3 Range-wide abundance estimates for light-footed Ridgway's rails using Bayesian N-mixture models. Two models were fit to simulated data: the Poisson-Lognormal and Poisson-Poisson mixture models (mathematical details provided in supplement A). Each model was fit with both site-level and site×survey level random effects in detection functions. Models were fit to a single range-wide monitoring data set generated from the low availability for detection with medium rail density scenario, under 20% spatial replication, targeted allocation of samples amongst strata, and weighted probability sampling of sample units within each stratum. Numbers represent posterior mean point estimates of range-wide abundance and estimates within 10% of true abundance ($N=434$) are in boldface

Model	Site visits/year		
	3	4	5
Poisson-Lognormal			
p(site)	334	265	230
p(site×survey)	599	503	417
Poisson-Poisson			
p(site)	508	490	456
p(site×survey)	360	352	391

Discussion

We developed the first statistical sampling framework for monitoring light-footed Ridgway's rail populations across their U.S. range to facilitate population assessment at multiple scales, including range-wide and at specific sites of interest for monitoring recovery under the ESA. This framework includes probability sampling of wetlands located within a sampling frame delineated to contain potential habitat and generates data that can be used directly to construct population indices (e.g., based on raw counts) or conduct model-based estimation that incorporates detection error, and thus provides a statistically defensible framework for monitoring rail status and recovery across their range. We also developed a simulation model of the call-broadcast process that incorporated variation in detection of individual birds based on their availability for detection, distance from the survey location, and call type. To our knowledge this is the first simulation model of the call-broadcast process that is widely used to sample marsh birds (Conway 2011). We used this model to simulate multiple sampling strategies and intensities, which allowed us to make recommendations for effective sampling. Moreover, this model will be useful to further refine sampling protocols and population assessment models used to estimate abundance.

Monitoring population status and trend over large scales are common goals of recovery programs for T&E species. Well-planned and implemented monitoring programs reliably estimate metrics that track the effectiveness of recovery at management relevant scales and are also foundational to adaptive management programs for T&E species (Campbell et al. 2002; Conroy et al. 2012). Yet, a lack of statistically defensible, coordinated, and broad-scale monitoring programs commonly obfuscates inferences about population status of T&E species (Lindenmayer et al. 2020; Evansen et al. 2021). Much research effort has focused on the important issue of model-based approaches to correct observed counts for detection error at a local level (e.g., Kery and Royle et al. 2015), whereas consideration of basic statistical sampling principles for broad-scale monitoring and assessment when not all areas can be surveyed (e.g., Johnson et al. 2009) is far less common. Consequently, there is a need for well-designed sampling frameworks that address basic questions about population status for many species, without which evidence-based conservation is difficult. Such frameworks can facilitate evaluation of ongoing recovery efforts and a reduction in uncer-

Table 4 Abundance estimates based on Bayesian N-mixture models. The models were the Poisson-Lognormal and Poisson-Poisson mixture models (supplement A). Each model was fit with both site-level and site $\times$ survey level random effects in detection functions. Models were fit to a single range-wide monitoring data set generated from the low availability for detection with medium density scenario, under 20% spatial replication, targeted allocation of samples amongst strata, and weighted probability sampling of sample units within each stratum. Numbers represent posterior mean point estimates, and bold numbers indicate estimates within 20% of true abundance. Columns with headers 3, 4, and 5 represent the number of site visits per year, and N represents the true abundance

Stratum	Poisson-Lognormal					Poisson-Poisson				
	p(site)					p(site)				
	3	4	5	5	5	3	3	4	5	5
Seal Beach (N=57)	30	24	20	45	35	73	74	67	51	52
South Bay (N=1)	0	0	0	1	0	0	0	0	0	0
Sweetwater (N=12)	7	5	4	10	8	19	18	16	13	12
Tijuana (N=62)	54	42	34	100	85	64	64	58	46	48
Important Freshwater (N=42)	66	58	51	114	110	97	50	47	37	37
Important Saltwater (N=254)	175	135	117	315	251	302	284	269	213	203
Other Freshwater (N=1)	1	0	0	1	1	1	0	0	0	0
Other Saltwater (N=5)	1	1	1	2	1	0	0	0	0	0

tainty through adaptive management, while providing data to evaluate ecological dynamics and hypotheses that are relevant for rare species conservation (e.g., range dynamics, impacts of disturbance; Lindenmayer et al. 2022).

We developed a sampling framework specific for rails in southern California, but this example is based on general principles that underlie design of statistical sampling procedures. Basic principles of sampling include the specification of a target population and sampling frame, stratification of sample units based on monitoring objectives and anticipated spatial heterogeneity, use of a probabilistic procedure to select areas to conduct surveys (i.e., when the entire region cannot be sampled), identification and use of a design that facilitates efficiency (i.e., improved precision, approximately unbiased), and implementation of protocols that facilitate estimation of detection error (e.g., repeated sampling within a season; Yoccoz et al. 2001). Unfortunately, many monitoring programs are not based on appropriate probabilistic sampling and therefore may not produce reliable estimates of population status that can be “scaled up” in any meaningful way to a larger area of interest. Consequently, monitoring for many T&E species may not track status or progress towards recovery goals in a scientifically defensible manner, and therefore fails to achieve one of its primary goals (Evansen et al. 2021; Lindenmayer et al. 2020). Previous survey efforts provided valuable data for developing the range-wide sampling framework and parameterizing our simulation model, yet these survey efforts did not include the design components described above. Hence, inferences about rail population status relative to recovery objectives can be easily challenged. Our work builds from past survey efforts and uses those data to inform sampling weights, while also using sampling principles that are more statistically defensible. Moreover, while our focus was on the U.S. portion of the light-footed Ridgway’s rail range and sampling relative to monitoring status under the ESA, the principles and approaches could be applied elsewhere (e.g., the portion of the species’ range in Mexico).

We developed a sampling framework and simulation model that provide not only a quantitative foundation for monitoring and statistical assessment but also a foundation for further refinement. The final sampling frame contained 98.9% of all known rail locations, and therefore is spatially extensive enough to capture rail habitat range wide. Weighted sampling, where sample units within each stratum are selected with probabilities proportional to local abundance, is more efficient for sampling patchy rail populations than simple random sampling. Moreover, targeted allocation of samples amongst strata, where 50% of sample units at important NWR sites are surveyed and Neyman allocation is used to allocate samples at remaining strata, was the most effective allocation for range-wide sampling. Our results also show that $\geq 20\%$ of sampling units should be surveyed ≥ 3 times during the breeding season. More than 3 surveys per year decreased bias in abundance estimates but may not appreciably increase precision (at least with ≤ 5 visits per year). At the strata level, greater spatial replication (i.e., $> 20\%$ overall replication) will likely improve precision of estimates, but a high degree of variation may remain. Lastly, we estimated that implementing these guidelines would require 4–7 dedicated field staff to conduct surveys during each breeding season. While costs are nearly always limited for monitoring T&E species (Crouse et al. 2002), this appears to be a relatively small effort when considering the task of range-wide monitoring to assess populations of an endangered species.

Previous studies developed recommendations for sampling marsh birds as a group across the continuous U.S. (Johnson et al. 2009) and implemented similar designs regionally to assess multiple species (including species other than marsh birds; Weist et al. 2016; Ladin

et al. 2020). Our approach was species-specific and focused only on the range, habitat, and sampling objectives for the target species, but followed many of the recommendations of Johnson et al. (2009), including: (1) a sampling universe built around spatial depictions of habitat, (2) use of probability sampling to select survey locations to make inference to clearly defined areas, (3) a hierarchical design that allows for aggregating estimates across strata for larger-scale assessment, (4) use of sample inclusion probabilities that are proportional to local abundance, (5) spatial aggregation of on-the-ground survey locations within larger sample units, where the number of survey points is conditional on area of habitat, (6) an assumption that field protocols will be implemented in a standardized fashion to minimize spatial-temporal variation in detection error, and (7) use of a design-based approach to generate counts as indices of abundance at scales of interest for management. Despite these similarities, there were several differences between our framework and those used or suggested previously: (1) we did not differentiate discrete from continuous wetlands, (2) we used a random systematic point grid to cover all habitat within sample units, as opposed to spatially-balanced random sampling within each sample unit, and (3) we proposed sampling all survey points within each randomly selected sample unit (i.e., to cover all habitat), as opposed to a random subset of points across the area. We believe these differences are justified and reflect differences in study objectives, scale of application, and species-specific attributes (e.g., strong spatial patchiness, endangered status, relatively small breeding range, non-migratory resident behavior). In addition, we developed a novel simulation model of the sampling process that realistically incorporated heterogeneity in detectability by individual rails and used this model to deduce sampling recommendations.

We also provide a preliminary test of Bayesian N-mixture models for estimating abundance from simulated call-broadcast data collected under our design. Obviously, we do not expect detection probability during surveys to be homogeneous in space and time, as is required for use of raw counts as abundance indices (Pollock et al. 2002), and therefore our model was designed to realistically incorporate heterogeneity in detection. Our proof-of-concept analysis showed that multiple parameterizations of N-mixture models have promise for range-wide and stratum-specific abundance estimation for rails, even under the most conservative availability scenario. Previous authors used frequentist analyses of N-mixture models for estimating abundance of clapper rails (*R. crepitans*; Wiest et al. 2016; Ladin et al. 2020), yet these analyses could not accommodate heterogeneity in detection through random effects. Recent work has called into question the robustness of N-mixture models for abundance estimation when heterogeneity in detection is present (e.g., Duarte et al. 2018). Thus, realistically incorporating detection heterogeneity in simulation models and using random effects to capture site- and site $\times$ survey level variation in detection, as we did here, are important aspects of realistically assessing model performance when estimating marsh rail abundance.

Our preliminary test showed that Bayesian N-mixture models produced reasonable estimates of abundance, but additional work is needed to determine which model and parameterization produce the most robust estimates under realistic field conditions. Specifically, more simulation is needed to understand model performance across different combinations of sampling replication, rail density, and availability of rails to sampling. Further, the N-mixture framework requires a closure assumption which could be violated by migratory birds during the breeding season (Shirkey et al. 2024). This issue is unlikely to limit application of the model to light-footed Ridgway's rail because we are sampling resident popula-

tions that are not believed to be migratory. However, the practical implication of closure violation means that some birds are not available for detection during each sampling event. Thus, our current simulation model could be easily updated to understand the implications of migration on marsh bird abundance estimation by changing availability for detection systematically across repeated site visits conducted during a breeding season. Alternative estimation models could also be tested by using our call-broadcast simulation. For example, hierarchical distance sampling estimators could be tested as an alternative to N-mixture models, assuming that rail distances from broadcast points can be accurately estimated and that birds do not move in response to broadcasting recorded calls prior to responding (Bui et al. 2015). After the best model for estimating rail abundance is established, formal power analyses can be used to test the ability of sampling and assessment procedures to estimate population trends reliably.

Our framework provides a foundation for statistical monitoring and assessment of progress towards recovery goals but can also serve as a starting point for further refinement. In addition to refining the estimation model, other components provide the opportunity for updates, including the habitat model (i.e., sampling universe) and the call-broadcast and sampling simulations. First, the spatial habitat model can be updated over time. This could occur by manually digitizing habitat boundaries of areas sampled in the field, or by using satellite imagery and remote sensing to identify wetland boundaries. These approaches would help accommodate changes in wetland area over time. For example, one approach would be to delineate habitat area annually within randomly selected cells prior to or during sampling, accounting for temporal changes in conditions.

In addition, habitat suitability models could map intensity of use for areas inside the habitat layer using rail location data and distribution models (Stevens and Conway 2020a; Helmsstetter et al. 2021). Habitat suitability could be used to further stratify sampling based on anticipated changes in density (high vs. low) within current strata. Habitat suitability could also be used to update call-broadcast simulations, which currently simulate the locations of birds with all locations equally likely to serve as an activity center. A habitat suitability model would allow for locations of individuals to be simulated proportional to suitability within each grid cell and could also allow locations of individuals to change among surveys and incorporate lack of spatial independence (e.g., territoriality) in a biologically realistic manner. Individual parameters of the simulation model could also be refined through targeted field studies. For example, we approximated availability during call-broadcast surveys by using data collected from Yuma Ridgway's rail, whereas field studies could estimate this parameter directly within the study region.

Lastly, modifications to the simulation model can be used to evaluate robustness of sampling and assessment procedures under other plausible scenarios. For example, Johnson et al. (2009) warned about the challenges of sampling wetland habitats when some areas are not directly accessible to sampling. Limited accessibility could be simulated directly by randomly omitting data from individual survey points, or clusters of points, within each sample unit and evaluating the impacts for population assessment. This would provide an understanding of the impacts of partial accessibility for inferences about marsh bird abundance, which have not been formally evaluated. Our simulations also assumed designs whereby all randomly selected sites were visited with the same intensity during a breeding season. Yet, improved sampling efficiency may be attainable by conducting repeated site visits at only a subset of locations (Pollock et al. 2002). Consequently, additional simulations could explore

the fraction of grid cells that should be visited > 1 time per season, and where these repeat visits should be conducted. Although we highlight the need for further work to optimize sampling and assessment, we provide the conceptual foundation, quantitative starting point, and simulation models to make such work possible. Moreover, we provide an example of building a fit-for-purpose framework to assess recovery for a T&E species (Lindenmayer et al. 2020, 2022) based on general sampling principles, and therefore demonstrate a process that can be used as a blueprint for developing rigorous monitoring programs for other taxa.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10531-024-02919-5>.

Acknowledgements We thank the light-footed Ridgway's rail working group, J. Terp, J. Stahl, and staff of the San Diego National Wildlife Refuge Complex and California Department of Fish and Game for discussions and feedback that improved this work. We thank D. Zembal for the many years of targeted monitoring data collection for rails in southern California and E. Harrity for preliminary work quantifying detection distances that informed our simulation model. We thank D. Johnson and the anonymous reviewer for comments that improved an earlier draft of this manuscript. The findings and conclusions in this article are those of the author(s) and do not necessarily represent the views of the U.S. Fish and Wildlife Service. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Author contributions All authors contributed to the study conception and design. Material preparation and data collection were performed by Kimberly A. Sawyer, Lauren Kershek, Giselle Block, Sandra Hamilton, and Rebecca Kohstrom. Analyses were performed and the first draft of manuscript was written by Bryan Stevens, and interpretation of simulation results was led by Bryan Stevens and Courtney Conway. All authors contributed critically important intellectual content through editing and refining previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Funding for this work was provided by U.S. Fish and Wildlife Service and the Ridgway's Rail Recovery Fund at the San Diego Foundation.

Data availability Data generated from our simulation study are available from lead author upon reasonable request.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Barton ML (2016) Nesting preferences of the light-footed Ridgway's rail in two southern California wetlands. M.S. thesis, California State University, Long Beach, California, USA
- Bayraktarov E, Ehmke G, O'Connor J, Burns EL, Nguyen HA, McRae L, Possingham HP, Lindenmayer DB (2019) Do big unstructured biodiversity data mean more knowledge? *Front Ecol Evol* 6:e239

- Bui TVD, Takekawa JY, Overton CT, Schultz ER, Hull JM, Casazza ML (2015) Movements of radio-marked California Ridgway's rails during monitoring surveys: implications for population monitoring. *J Fish Wild Man* 6:227–237
- Campbell SP, Clark JA, Crampton LH, Guerry AD, Hatch LT, Hosseini PR, Lawler JJ, O'Connor RJ (2002) An assessment of monitoring efforts in endangered species recovery plans. *Ecol Appl* 12:674–681
- Conroy MJ, Runge MC, Nichols JD, Stodola KW, Cooper RJ (2011) Conservation in the face of climate change: the roles of alternative models, monitoring, and adaptation in confronting and reducing uncertainty. *Biol Cons* 144:1204–1213
- Conroy MJ, Stodola KW, Cooper RJ (2012) Effective use of data from monitoring programs and field studies for conservation decision making: predictions, designs and models working together. *J Ornithol* 152:325–338
- Conway CJ (2011) Standardized north American marsh bird monitoring protocol. *Waterbirds* 34:319–346
- Conway CJ, Gibbs JP (2005) Effectiveness of call-broadcast surveys for monitoring marsh birds. *Auk* 122:26–35
- Conway CJ, Eddleman WR, Anderson SH, Hanebury LR (1993) Seasonal changes in Yuma clapper rail vocalization rate and habitat use. *J Wildl Man* 57:282–290
- Conway CJ, Sulzman C, Raulston BA (2004) Factors affecting detection probability of California black rails. *J Wildl Man* 68:360–370
- COSEWIC (2002) Canadian species at risk. Canadian Wildlife Service, Committee on the Status of Endangered Wildlife in Canada, Ottawa, ON, Canada
- Crouse DT, Mehrhoff LA, Parkin MJ, Elam DR, Chen LY (2002) Endangered species recovery and the SCB study: a U.S. Fish and Wildlife Service perspective. *Ecol Appl* 12:719–723
- Diario Oficial de la Federacion (2002) Norma Oficial Mexicana NOM-059-ECOL-2001, Proteccion Ambiental-Especies Nativas De Mexico de flora y fauna silvestres-categoriosde riesgo y especificaciones para su inclusion, exclusion o cambio-lista de especies en riesgo. Secretaria de Medio Ambiente y Recursos Naturales
- Duarte A, Adams MJ, Peterson JT (2018) Fitting N-mixture models to count data with unmodeled heterogeneity: bias, diagnostics, and alternative approaches. *Ecol Modelling* 374:51–59
- eBird Basic Dataset (2021) Version EBD_relNov-2021. Cornell Lab of Ornithology, Ithaca, New York, USA, accessed 11/2021
- Eddleman WR, Conway CJ (2020) Ridgway's Rail (*Rallus obsoletus*), version 1.0. In *Birds of the World* (P. G. Rodewald, Editor). Cornell Lab of Ornithology, Ithaca, NY, USA. <https://doi-org.uidaho.idm.oclc.org>
- Eddleman WR, Knopf FL, Meanley B, Reid FA, Zembal R (1988) Conservation of north American rallids. *Wilson Bull* 100:458–475
- Evansen M, Carter A, Malcom J (2021) A monitoring policy framework for the United States Endangered species Act. *Environ Res Lett* 16:e031011
- Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Stat Sci* 7:457–511
- Gerber BD, Kendall WL, Hooten MB, Dubovsky JA, Drewien RC (2015) Optimal population prediction of sandhill crane recruitment based on climate-mediated habitat limitations. *J Anim Ecol* 84:1299–1310
- Goodwin P, Mehta AJ, Zedler JB (2001) Tidal wetland restoration: an introduction. *J Coastal Res* 27:1–6
- Harriy EJ, Stevens BS, Conway CJ (2020) Keeping up with the times: mapping range-wide habitat suitability for endangered species in a changing environment. *Biol Conserv* 250:e108734
- Helmstetter N, Conway CJ, Stevens BS, Goldberg AR (2021) Balancing transferability and complexity of species distribution models for rare species conservation. *Divers Distrib* 21:95–108
- Johnson DH, Gibbs JP, Herzog M, Lor S, Neimuth ND, Ribic CA, Seamans M, Shaffer TL, Shriver WG, Stehmen SV, Thompson WL (2009) A sampling design framework for monitoring secretive marshbirds. *Waterbirds* 32:203–215
- Kays R, Parsons AW, Baker MC, Kalies EL, Forrester T, Costello R, Rota CT, Millsbaugh JJ, McShea WJ (2017) Does hunting or hiking affect wildlife communities in protected areas? *J Appl Ecol* 54:242–252
- Kéry M, Royle JA (2015) Applied hierarchical modeling in ecology: analysis of distribution, abundance, and species richness in R and BUGS. Volume 1: prelude and static models. Academic Press, San Diego, California, USA
- Kéry M, Guíllera-Arroita G, Lahoz-Monfort JJ (2013) Analysing and mapping species range dynamics using occupancy models. *J Biogeogr* 40:1463–1474
- Kissel AM, Wallace B, Anderson J, Dickson BG, Van Neste K, Landau V, Averill-Murray RC, Allison LJ, Fesnock A (2023) Range-wide occupancy trends for the Mojave desert tortoise (*Gopherus agassizii*). *Ecosphere* 14:e4462
- Ladin ZS, Wiest WA, Correll MD, Tymkiw EL, Conway M, Olsen BJ, Elphick CS, Thompson WL, Shriver WG (2020) Detection of local-scale population declines through optimized tidal marsh bird monitoring design. *Global Ecol Conserv* 23:e01128

- 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 Man* 262:e110312
- Lindenmayer DB, Lavery T, Scheele BC (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
- Lor S, Malecki RA (2002) Call-response surveys to monitor marsh bird population trends. *Wildl Soc Bull* 30:1195–1201
- Mackenzie DI, Nichols JD, Sutton N, Kawanishi K, Bailey LL (2005) Improving inferences in population studies of rare species that are detected imperfectly. *Ecology* 86:1101–1113
- Marcus L, Kondolf A (1989) The coastal wetlands of San Diego County. California State Coastal Conservancy, Oakland, California, USA
- Massey BW, Zembal R, Jorgensen PD (1984) Nesting habitat of the light-footed clapper rail in southern California. *J Field Ornithol* 55:67–80
- McDonald LL (2004) Sampling rare populations. In: Thompson WL (ed) *Sampling rare or elusive species: concepts, designs, and techniques for estimating population parameters*. Island, Washington, D.C., USA, pp 11–42
- Nelson ML, Brewer CK, Solem SJ (2015) Existing vegetation classification, mapping, and inventory technical guide, version 2.0. General Technical Report WO–90. U.S. Department of Agriculture, Forest Service, Ecosystem Management Coordination Staff, Washington, DC, USA. <https://data.fs.usda.gov/geodata/edw/datasets.php>, Accessed 2/2022
- Nichols JD, Williams BK (2006) Monitoring for conservation. *Trends Ecol Evol* 21:668–673
- NOAA Office for Coastal Management (2022) NOAA's Coastal Change Analysis Program (C-CAP) 2016 Regional Land Cover Data - Coastal United States. National Oceanic and Atmospheric Administration, Office for Coastal Management, Charleston, South Carolina, USA. <https://coast.noaa.gov/digitalcoast/data/ccapregional.html>, Accessed 2/2022.
- Pebesma E (2018) Simple features for R: standardized support for spatial vector data. *R J* 10:439–446
- Plummer M (2003) JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. Proceedings of the 3rd International Workshop on Distributed Statistical Computing. Vienna, Austria
- Pollock KH, Nichols JD, Simons TR, Farnsworth GL, Bailey LL, Sauer JR (2002) Large scale wildlife monitoring studies: statistical methods for design and analysis. *Environmetrics* 13:105–119
- R Core Team (2022) R: a language and environment for statistical computing. Version 4.2.2. R Foundation for Statistical Computing, Vienna, Austria
- San Francisco Estuary Institute (2017) California Aquatic Resource Inventory (CARI) version 0.3. San Francisco Estuary Institute Aquatic Science Center, San Francisco, California, USA. <https://www.sfei.org/data/california-aquatic-resource-inventory-cari-version-03-gis-data#sthash.WzKgReBT.dpbs>, Accessed 2/2022
- Schaeffer RL, Mendenhall IIIW, Ott RL (2006) Elementary survey sampling. Sixth edition. Thompson Brooks/Cole publishing, Belmont, California, USA
- Shirkey BT, Simpson JW, Hansen JM, Hengst NM, Gates RJ, Tonra CM (2024) Vocalization behavior of resident and migrant Virginia Rails (*Rallus limicola*) and Soras (*Porzana carolina*) in Northwestern Ohio, USA. *Waterbirds* 46:277–288
- Stevens BS, Conway CJ (2019) Predicting species distributions: unifying model selection and scale optimization for multi-scale occupancy models. *Ecosphere* 10:e02748
- Stevens BS, Conway CJ (2020a) Mapping habitat suitability at range-wide scales: spatially explicit distribution models to inform conservation and research for marsh birds. *Conserv Sci Pract* 2:e178
- Stevens BS, Conway CJ (2020b) Predictive multi-scale occupancy models at range-wide extents: effects of habitat and human disturbance on distributions of wetland birds. *Divers Distrib* 26:34–48
- Stevens BS, Conway CJ (2021) Mapping habitat quality and threats for eastern black rails (*Laterallus jamaicensis jamaicensis*). *Waterbirds* 44:245–256
- Stevens BS, Conway CJ, Luke K, Weldon A, Hand C, Schwartzer A, Smith F, Watson C, Watts B (2022) Large-scale distribution models for optimal prediction of eastern black rail habitat within tidal ecosystems. *Global Ecol Conserv* 38:e02222
- Su YS, Yajima M (2015) R2jags: a package for running jags from R. R package version 0.5–7. <http://CRAN.R-project.org/package=R2jags>.
- U.S. Department of the Interior (2020) Endangered and threatened wildlife and plants; threatened species status for Eastern Black Rail with a Sect. 4(d) rule. *Fed Reg* 85:63764–63803
- U.S. Fish and Wildlife Service (1985) Recovery plan for the light-footed clapper rail. U.S. Fish and Wildlife Service, Portland, Oregon, USA
- U.S. Fish and Wildlife Service (2019) Light-footed Ridgway's rail (*Rallus Obsoletus levipes*; formerly light-footed clapper rail, *Rallus longirostris levipes*) Recovery Plan Amendment. U.S. Fish and Wildlife Service, Portland, Oregon, USA

- Wiest WA, Correll MD, Olsen BJ, Elphick CS, Hodgman TP, Curson DR, Shriver WG (2016) Population estimates for tidal marsh birds of high conservation concern in the northeastern USA from a design-based survey. *Condor* 118:274–288
- Yoccoz NG, Nichols JD, Boulinier T (2001) Monitoring of biological diversity in space and time. *Trends Ecol Evol* 16:446–453
- Zedler JB (1993) Canopy architecture of natural and planted cordgrass marshes: selecting habitat evaluation criteria. *Ecol Appl* 3:123–138
- Zemba R, Massey BW, Fancher JM (1989) Movements and activity patterns of the light-footed clapper rail. *J Wildl Man* 53:39–42

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Bryan S. Stevens¹ · Courtney J. Conway² · Kimberly A. Sawyer³ · Lauren Kershek⁴ · Giselle Block⁵ · Sandra Hamilton⁴ · Rebecca Kolstrom⁵

✉ Bryan S. Stevens
Bryan.Stevens@usda.gov

¹ U.S. Forest Service, Pacific Northwest Research Station, La Grande, Corvallis, OR 97850, USA

² U.S. Geological Survey, Idaho Cooperative Fish and Wildlife Research Unit, University of Idaho, Moscow, ID 83844-1141, USA

³ Idaho Cooperative Fish and Wildlife Research Unit, University of Idaho, Moscow, ID 83844-1141, USA

⁴ U.S. Fish and Wildlife Service, Carlsbad, CA 92008, USA

⁵ U.S. Fish and Wildlife Service, Refuges Inventory and Monitoring Program, Sacramento, CA 95825, USA