Supplement A. Supplemental description of methods for building a range-wide sampling framework to assess populations of light-footed Ridgway's rails.

Journal: Biodiversity and Conservation

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

Authors: Bryan S. Stevens[†], Courtney J. Conway, Kim Sawyer, and Lauren Kershek

[†]Pacific Northwest Research Station, U.S. Forest Service, La Grande, Oregon 97850, USA; Bryan.Stevens@usda.gov

Building a Model of Potential Habitat

Targeted breeding-season surveys for light-footed Ridgway's rails have been conducted regularly since 1980 (we had data for all years except 1994-1999 and 2020; e.g., Zembal and Hoffman 2021). These surveys were conducted in areas believed to provide key habitat for light-footed Ridgway's rails and attempted to: 1) map locations of breeding pairs, and 2) provide a census of light-footed Ridgway's rail populations within sampled wetlands. Broadcast of light-footed Ridgway's rail calls were used to locate birds, but sampling intensity and use of call-broadcast were not uniform in space or over time. These survey data nonetheless provide information on the distribution of rails within important occupied wetlands, and we linked these data (n = 8,566 locations collected from 1980-2021) with 3 publicly available spatial data sets that characterized land cover and wetland attributes to model potential habitat across the light-footed Ridgway's rail range: 1) a polygon layer of existing vegetation for the southern 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 their Coastal Change Analysis Program (hereafter NOAA C-CAP data; NOAA Office for Coastal Management 2022, accessed 2/2022), and 3) and 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 ArcMap 10.5.1 (ESRI, Redlands, CA) to conduct all spatial analyses.

Prior to modeling potential habitat, we generated a boundary layer for the U.S range of light-footed Ridgway's rails based on a compilation of range descriptions and available location data. We created the range boundary by merging the ranges from four sources: 1) U.S. Geological Survey Gap Analysis Project, which delineates the geographic limits of suitable habitat for a species (USGS Gap Analysis Project 2018, accessed 2/2021), 2) BirdLife International distribution information, which uses specimens and observations from a variety of sources to delineate the historical range of species (BirdLife International and Handbook of the Birds of the World 2020, accessed 2/2021), 3) a range map produced by California Department of Fish and Wildlife based on occurrence data (California Department of Fish and Wildlife 2021), and 4) the complete current range description provided by the U.S. Fish and Wildlife Service (U.S. Fish and Wildlife Service 2021). We adjusted this merged range boundary based on 4 light-footed Ridgway's rail location data sets: 1) all light-footed Ridgway's rail locations from annual surveys conducted from 1980-2021 (Zembal and Hoffman 2021), 2) additional locations from the recovery plan amendment for the species (U.S. Fish and Wildlife Service 2019), 3) historical species accounts from George Bird Grinell (Grinnell 1915), and 4) observed locations reported by citizen scientists through eBird (eBird Basic Dataset 2021). Specifically, we adjusted the boundaries of the merged range layer by also merging buffers that were generated around light-footed Ridgway's rail locations not included in any of the ranges

described above (2 historic locations from the 2019 amendment to the recovery plan). Finally, we added a 10 km buffer around the merged range boundary to account for movement and range shifts that may occur.

We linked attributes of the CalVeg, NOAA C-CAP, and CARI data sets with light-footed Ridgway's rail locations from the 1980-2021 surveys to characterize potential habitat for the species across their range in the U.S. We intersected light-footed Ridgway's rail locations with the 3 spatial data sets and extracted attribute values from each data set at each location. This allowed us to identify attributes from each spatial data set that described conditions at known light-footed Ridgway's rail locations, but also allowed us to eliminate non-habitat by determining which attribute categories contained few or no locations (Table A1). We then created new spatial data sets consisting of only the attribute categories that contained the most light-footed Ridgway's rail locations that were also biologically plausible habitat conditions for breeding light-footed Ridgway's rails: 2 cover categories (water and wetland) from the CalVeg data; 4 cover categories (estuarine emergent wetland, palustrine emergent wetland, water, and palustrine scrub-shrub wetland) from the NOAA C-CAP data; and 4 cover categories (tidal marsh, pond and associated vegetation, tidal flat and marsh panne, and subtidal water) from the CARI data.

We then examined combinations of the CalVeg wetland and water layer in conjunction with the important attributes from the NOAA C-CAP and CARI data sets to find a combination that contained as many light-footed Ridgway's rail locations as possible for mapping potential habitat across their range. Specifically, we considered 11 variable combination that included CalVeg wetland and water categories and important attribute categories from the NOAA C-CAP and CARI data sets (Table A2), and we evaluated the proportion of light-footed Ridgway's rail

locations contained within each of these combinations. We retained the combination of attribute variables that contained the most light-footed Ridgway's rail locations, and merged polygon layers for these attributes to create a simple model to map potential habitat in space. This model retained the CalVeg water and wetland attributes, and 3 attribute variables from the CARI data (tidal marsh, tidal flat and marsh panne, and subtidal water). Areas with this combination of CalVeg and CARI attributes contained 94.2% of the 8,566 light-footed Ridgway's rail locations and served as a starting point for further refinement.

Table A1. Top attribute categories for each of 3 spatial data sources (CalVeg, NOAA C-CAP, CARI) in terms of spatial coverage of light-footed Ridgway's rail locations collected in southern California, USA, from 1980-2021. Shown are the number of light-footed Ridgway's rail locations (No. locations), the percentage of locations (Percent locations), and the cumulative percent of locations (Cumulative percent) contained within each attribute category and the categories above it.

Data source & attribute	No. locations	Percent locations	Cumulative percent
CalVeg attributes			
Wetland	6226	72.7	72.7
Water	1270	14.8	87.5
Agriculture	356	4.2	91.7
Urban	351	4.1	95.8
valley grassland	214	2.5	98.3
montane shrubland	95	1.1	99.4
forest land	49	0.6	99.9
NOAA C-CAP attributes			
Estuarine emergent wetland	4313	50.4	50.4
Palustrine emergent wetland	2363	27.6	77.9
Water	570	6.7	84.6
Low intensity developed	462	5.4	90.0
Palustrine scrub-shrub wetland	177	2.1	92.1
Medium intensity developed	126	1.5	93.5
Developed open space	117	1.4	94.9

Grassland	109	1.3	96.2
Palustrine forested wetland	62	0.7	96.9
Bare land	57	0.7	97.6
Scrub-shrub	56	0.7	98.2
Cultivated	44	0.5	98.7
CARI legal labels			
Tidal Marsh	5997	70.0	70.0
Pond and associated vegetation	964	11.3	81.3
Tidal Flat and Marsh Panne	669	7.8	89.1
Subtidal Water	197	2.3	91.4

Table A2. Data layers considered to characterize potential habitat for light-footed Ridgway's rails across their range in southern California, USA. Variable combinations combined the top attribute categories from CARI and NOAA C-CAP with the water and wetland categories from CalVeg data. Shown are the number of light-footed Ridgway's rail locations collected from 1980-2021 that were contained within each combination of variables (No. locations), and the percentage of locations (Percent locations) covered.

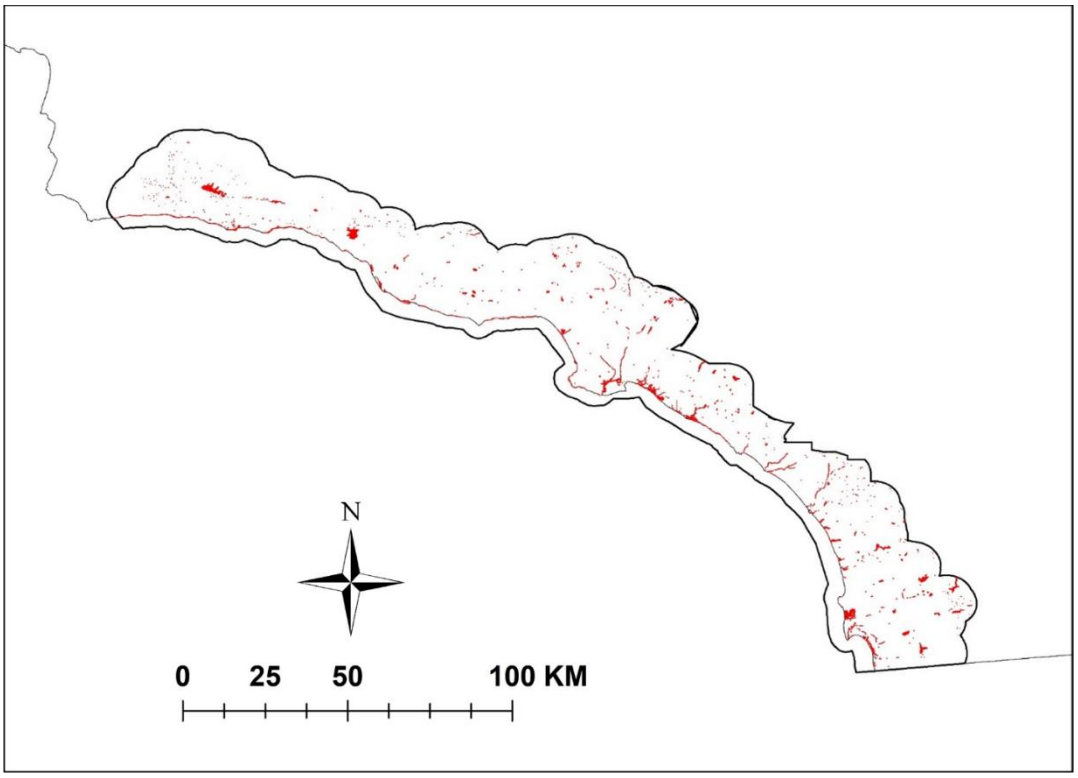
Variable combination[a]	No. locations	Percent locations
Tidal marsh[b]	7881	92.0
Pond and associated vegetation[b]	7671	89.6
Tidal flat and panne[b]	7644	89.2
Subtidal[b]	7539	88.0
Estuarine emergent wetland[c]	7701	89.9
Palustrine emergent wetland[c]	7792	91.0
Water[c]	7530	87.9
Palustrine scrub shrub wetland[c]	7546	88.1
Tidal marsh[b] + tidal flat and panne[b]	8029	93.7
Tidal marsh[b] + subtidal[b]	7924	92.5
Tidal marsh[b] + tidal flat and panne[b] + subtidal[b]	8072	94.2

[a] Variables included in addition to CalVeg water and wetland variables.
[b] Variable from CARI data set.
[c] Variable from NOAA C-CAP data set.

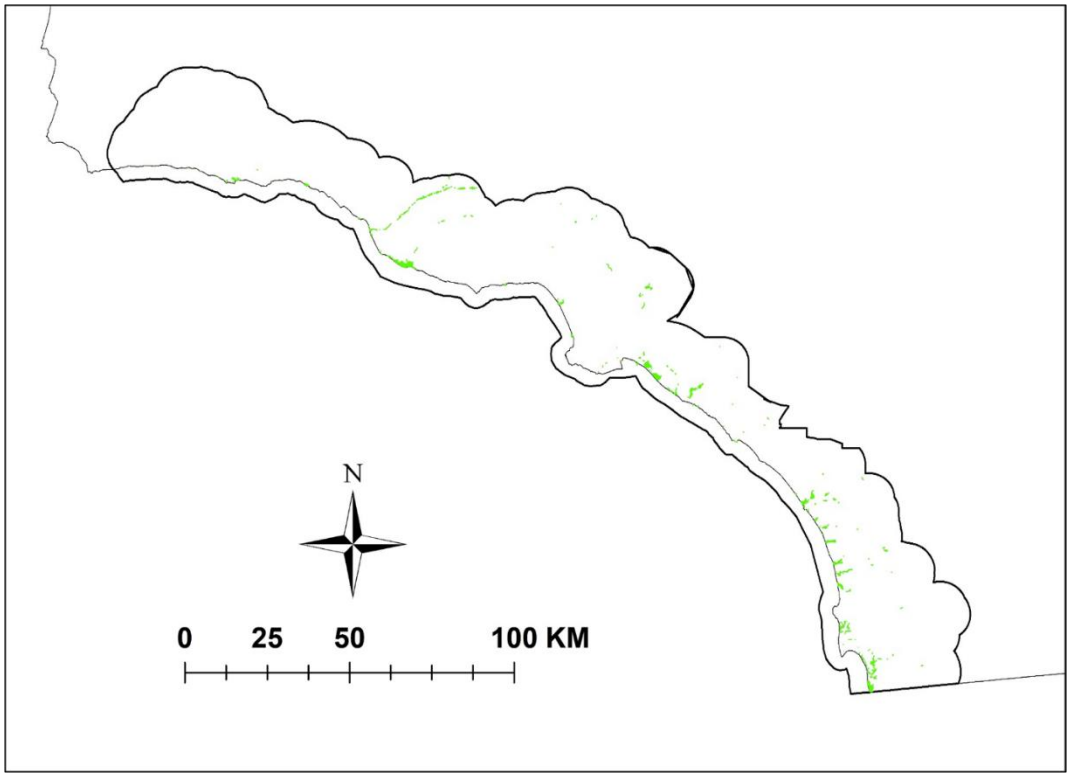
Next, we refined the potential habitat model by eliminating areas known to be non-habitat and areas unlikely to be occupied by light-footed Ridgway's rails. Secretive marsh birds generally do not use areas of open water unless they are directly adjacent to emergent vegetation. Thus, we reclassified open-water areas as non-habitat. We first created 2 mutually exclusive polygon layers from our potential habitat model (Fig. A1): 1) a polygon file to capture areas of open water by isolating the CalVeg water and CARI subtidal water areas of our potential habitat model, and 2) a polygon file that is not open water by isolating the remaining areas of potential habitat (CalVeg wetland, CARI tidal marsh, and CARI tidal flat and marsh panne polygons).

Figure A1. Open water (CalVeg water and CARI subtidal water; panel A, in red) and non-open water (CalVeg wetland, CARI tidal marsh, and CARI tidal flat and marsh panne; panel B, in green) habitat from the original model created to map potential habitat for light-footed Ridgway's rails across their range (thick black line) in California, USA. The thin black line represents the state boundary for California.

A)



B)



We calculated the distance from each light-footed Ridgway's rail location found within open water polygons to the nearest potential habitat that was not open water; nearly all distances (96.5%) were ≤100 m from potential habitat that was not open water (Figure A2; Table A3). Thus, our refined model buffered the polygon layer from areas of potential habitat that were not open water (i.e., CalVeg wetland, CARI tidal marsh, and CARI tidal plat and marsh panne pixels) by 100 m.

Figure A2. Frequency distribution for the distance from each light-footed Ridgway's rail location found in open water (Fig. A1-A) to the nearest polygon of non-open water habitat (Fig. A1-B).

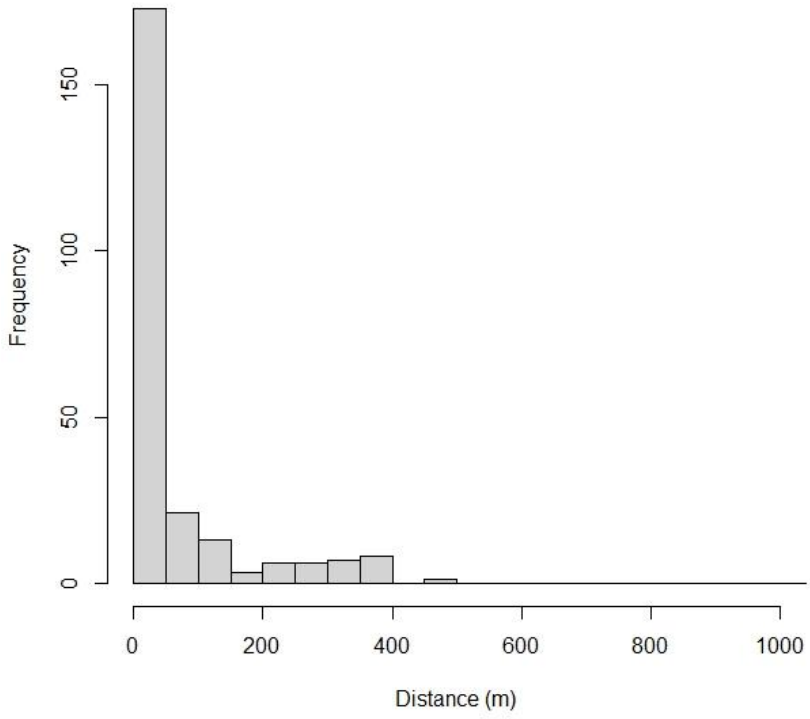


Table A3. Percentage of light-footed Ridgway's rail observations (Percent of locations) from locations within open water habitat (CalVeg water and CARI subtidal water) that were found within distance bands (Buffer distance) around non-open water habitat (CalVeg wetland, CARI tidal marsh, and CARI tidal flat and marsh panne) from the original potential habitat model created for light-footed Ridgway's rails in southern California, USA.

Buffer distance (m)	Percent of locations
50	0.95
75	0.96
100	0.97
125	0.97
150	0.98

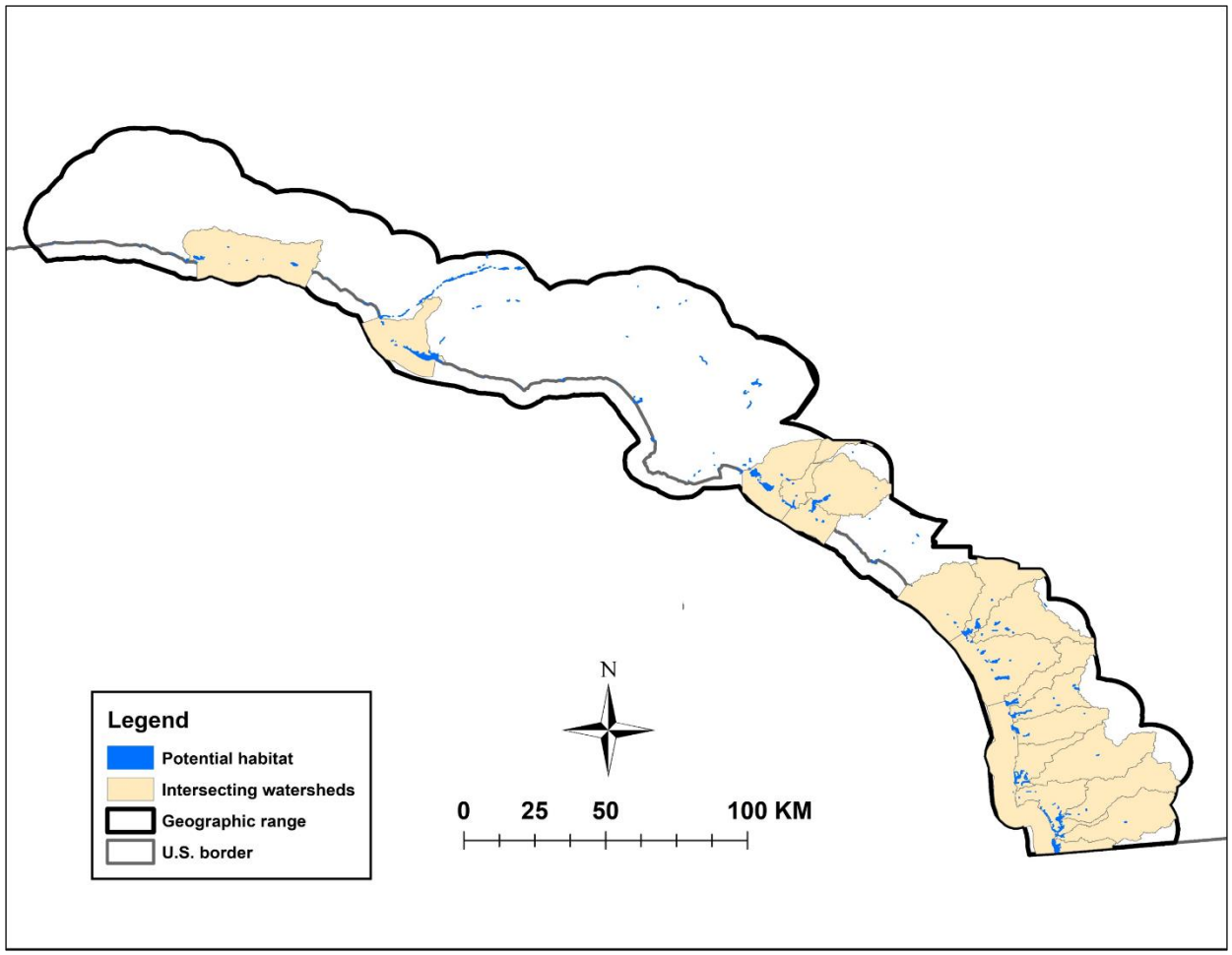
Most light-footed rail locations not contained by the original potential habitat model were <100 m from its boundary, and therefore the refined model increased the overall fraction of rail locations contained within potential habitat polygons (from 94.2% to 98.2%, n = 8,415). To further assess the potential habitat model at areas possibly occupied by breeding populations of light-footed Ridgway's rails (i.e., including areas not included in targeted sampling efforts), we tested the model using 2 auxiliary point location data sets: 1) a data set of incidental but verified light-footed Ridgway's rail locations reported to the U.S. Fish and Wildlife Service (hereafter incidental observations; n = 294 observations collected between 1998-2020), and 2) light-footed Ridgway's rail observations within the species geographic range (Fig. 1) that were reported from citizen scientists using eBird (eBird Basic Dataset 2021, n = 11,358 observations). Our refined potential habitat model captured 78.9% of incidental observations, 90.5% of eBird observations, and 93.6% of all the aggregated observations of light-footed Ridgway's rails across the data sets (18,925 of 20,218), and therefore included most areas likely to be inhabited across the species' endemic range in the U.S.

The last step to refine our potential habitat model was to eliminate areas along the California coast and further inland that are unlikely to harbor populations of light-footed Ridgway's rails. Major populations of light-footed Ridgway's rails are primarily located in tidal and estuarine wetlands on or near the coast within the southern portion of the geographic range, yet large expanses of coastal areas within the species' range are not believed to harbor breeding birds. Moreover, expansion of populations to inland sites is believed to involve inland movement up riparian corridors extending from coastal wetlands with extant populations. To eliminate unoccupied areas of coastal habitat from our potential habitat model, we used boundaries of watersheds that intersected light-footed Ridgway's rail observations from the targeted monitoring efforts. Specifically, we used HUC10 watershed boundaries from the NHDPlus version 2 watershed database (McKay et al. 2012, accessed 3/2022) to define watersheds, and retained all potential habitat that intersected watersheds containing ≥1 light-footed Ridgway's rail observation. To eliminate inland sites not likely to be occupied by light-footed Ridgway's rails, we calculated straight line distances from the California coast to each light-footed Ridgway's rail observation recorded from targeted monitoring efforts. The maximum distance from the coast was 7,743 m, and consequently we clipped our remaining potential habitat to areas within 7,750 m of the California coast. Therefore, our final potential habitat model was limited to areas within 7,750 m of the coast that were also contained within watersheds where ≥1 light-footed Ridgway's rail was previously recorded in annual survey efforts. These reductions in area of habitat resulted in no loss of coverage for existing light-footed Ridgway's rail locations, as the remaining potential habitat contained 98.2%, 78.9% and 90.4% of the targeted monitoring, incidental, and eBird detections. Furthermore, the small number of light-footed Ridgway's rail

locations that were not covered by the potential habitat model were generally close to potential habitat (median distance from points to potential habitat boundary = 98.8 m; Fig. A4).

Figure A3. Maps of potential light-footed Ridgway's rail habitat and intersecting watersheds containing ≥1 light-footed Ridgway's rail observation (A), and the final potential habitat model after removing areas outside of intersecting watersheds and >7,750 m from the Pacific Coast of California (B).

A)



B)

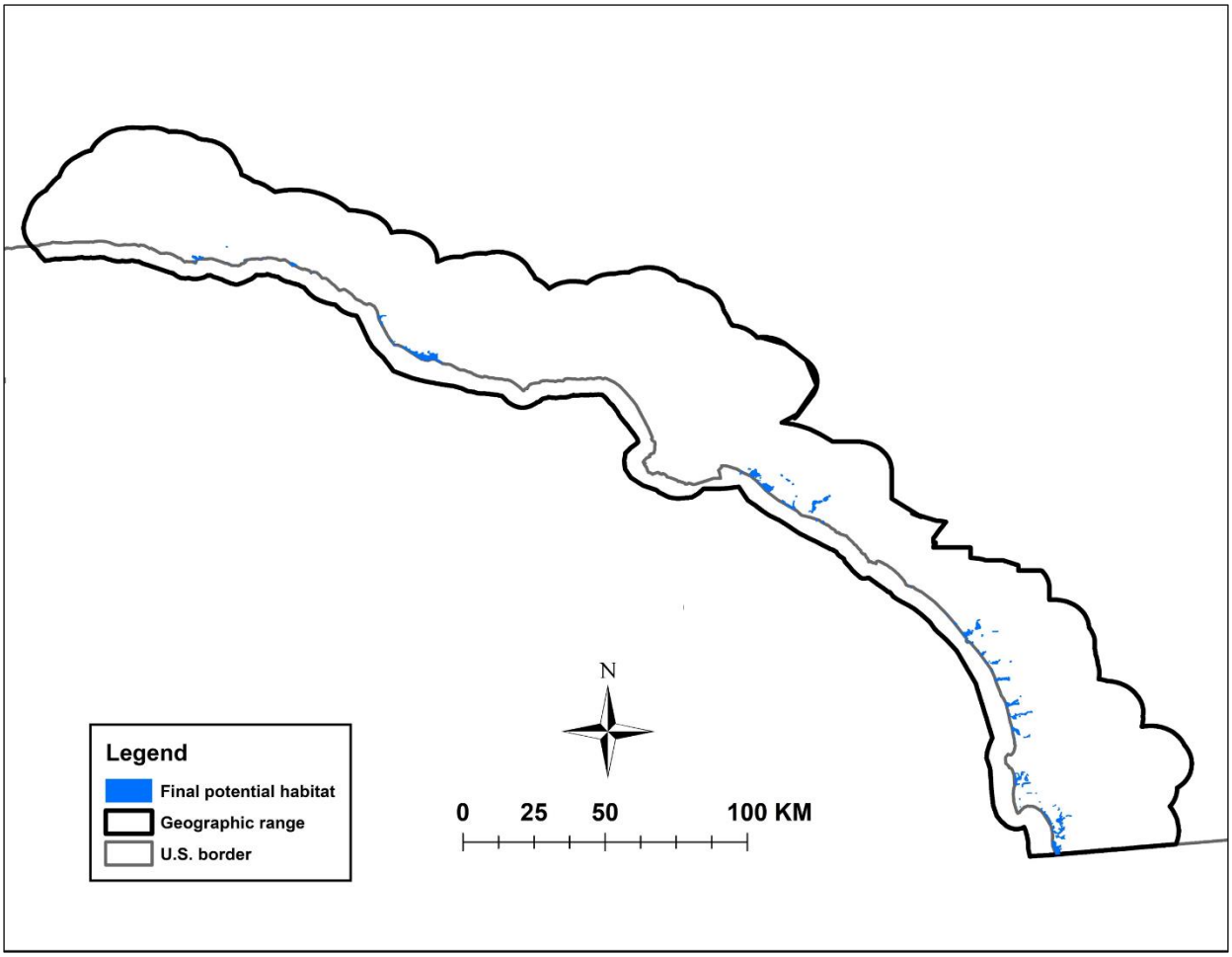
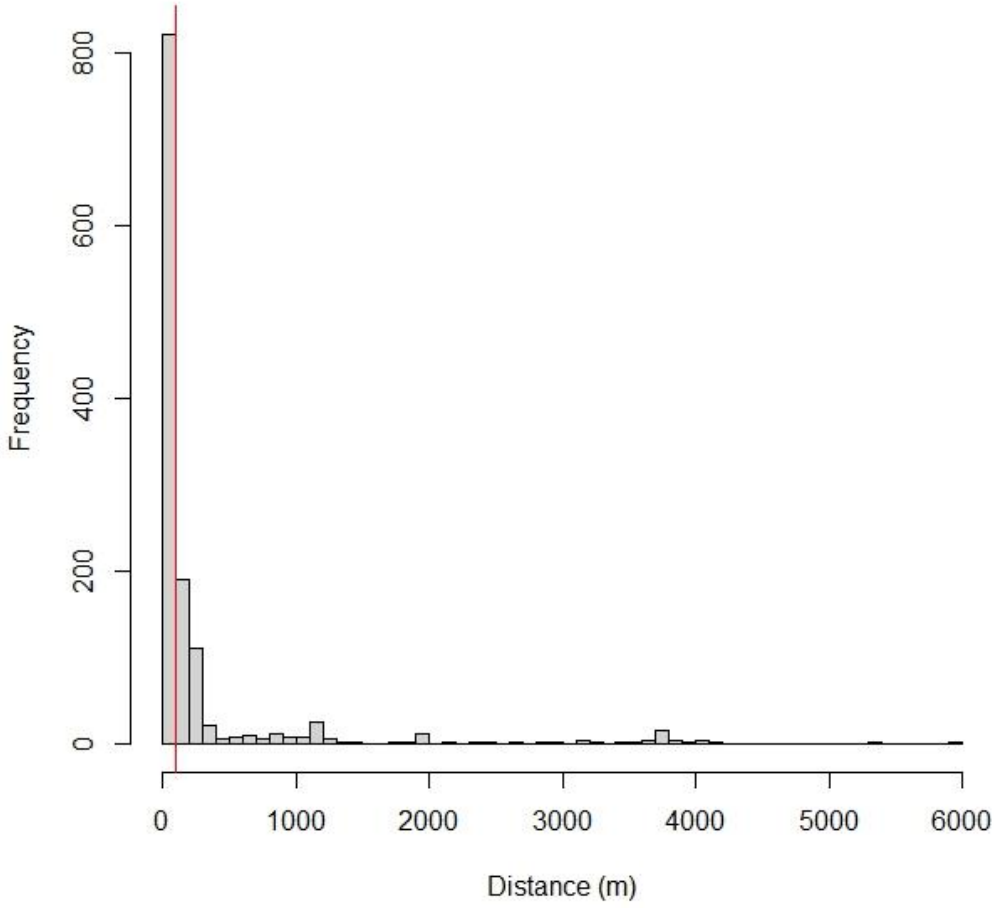

Legend
Final potential habitat
Geographic range
U.S. border
N
0
25
50
100 KM

Figure A4. Frequency distribution for the distances of light-footed Ridgway's rail locations found outside of the final potential habitat model to the nearest patch of potential habitat predicted from the model. The red vertical line indicates the median (98.8 m).



Refining the Grid of Sample Units

The preliminary sampling frame consisted of 411 1x1 km grid cells (sample units) spread across 6 sampling strata (see manuscript text for details of grid construction). Wetland complexes included in the stratum of additional important sites consisted of known areas of historical importance to light-footed Ridgway's rail conservation, which we identified in consultation with regional species experts. Locations for most of these areas (Kendall-Frost Reserve, Bataquitos

Lagoon, Los Peñasquitos Lagoon, San Elijo Lagoon, Agua Hedionda Lagoon, Buena Vista Lagoon, Upper Newport Bay) were identified using the California Protected Areas Database (CPAD; GreenInfo Network 2018, Accessed 2/2022). Cells intersecting potential habitat at these sites were assigned to the additional important stratum. Potential habitat polygons at 3 more important sites (lower San Diego River, the San Dieguito watershed, and Otay Lakes) were provided by local species experts and merged with our existing potential habitat model, and sampling cells intersecting habitat at these sites were added to the stratum of additional important sites.

Final modifications to the sampling frame were made on a case-by-case basis, where each of the 411 grid cells were carefully reviewed by local experts. This resulted in deletion of 165 grid cells that were identified to contain no light-footed Ridgway's rail habitat (e.g., cells that contained only open water, or shoreline that contained open beach, riprap, or other non-habitat), as well as the addition of 22 new 1x1 km grid cells to eliminate discontinuities in the sampling grid across light-footed Ridgway's rail habitat, or accidental omissions of areas on the edges of the first sampling grid that did indeed contain habitat. Potential habitat within these 22 new cells was digitized by the same species experts and merged with our final potential habitat model. The stratum membership of each grid cell sample unit was also evaluated, and this resulted in 33 cells originally characterized as part of the other potential habitat stratum being switched to the additional important sites stratum. Finally, the sample units in the additional important sites and other potential habitat strata were individually characterized as saltwater or freshwater dominant, resulting in 4 new strata generated from the original 2: 1) additional important saltwater, 2) additional important freshwater, 3) other saltwater, and 4) other freshwater. The final spatially explicit sampling frame thus consisted of 268 1x1 km grid cells spread across 8 sampling strata.

Simulating Conditional Detection Probabilities

We randomly generated a conditional detection function for each simulated light-footed Ridgway's rail that was available to sampling (i.e., responded during call-broadcast surveys) for either a *kek-hurrah* or *paired clatter* call type (with equal probability between call types). For each bird and during each individual survey we simulated a random distance detection function based on the following detection models:

$$P(Detection) = \frac{1}{1 + e^{(-(4.43 - \beta \times distance))}}$$

$$\beta \sim Normal(\mu = 0.017, \sigma = 0.003)$$

for the *kek-hurrah* call type, or

$$P(Detection) = \frac{1}{1 + e^{(-(7.59 - \beta \times distance))}}$$

$$\beta \sim Normal(\mu = 0.038, \sigma = 0.006)$$

for the *paired clatter* call type. We estimated the mean and standard deviation of β parameters for each of the 2 call types using mixed-effects logistic regression from detection trial experiments that we conducted for light-footed Ridgway's rails in southern California (K. Sawyer, University of Idaho, Unpublished Data). We conducted detection trial experiments by mimicking a call-broadcast survey from known distances away from wild birds (n = 118 trials), using one observer to conduct the broadcast survey and a second observer located close to the bird to determine whether or not the bird responded to the call-broadcast sequence. These data thus allowed us to directly estimate conditional detection probabilities (i.e., the probability that a rail was detected by the person conducting a survey given that the rail vocalized during the survey) and the extent to which the probabilities decayed with distance, as well the uncertainty surrounding the distance-decay function relationships (σ).

Sampling and Design-Based Estimators

We tested the performance of 3 strategies for allocating the total sample size (i.e., the total number of sample units for a given scenario of spatial replication) amongst the 8 sampling strata: 1) proportional allocation, 2) Neyman allocation, and 3) our customized allocation strategy that we developed to ensure adequate sampling of NWR strata (hereafter targeted allocation). Proportional allocation is described in the manuscript text and simply allocates the total sample size amongst strata in proportion to each stratum's fraction of sample units (i.e., relative to the entire range-wide sampling frame). Neyman allocation allocates more samples to strata with greater within-strata variance of measurements (i.e., spatial variance of light-footed Ridgway's rail counts amongst cells within the same stratum). Under Neyman allocation, the total sample is divided amongst strata based on the following formula (Scheaffer et al. 2006):

$$n_i = n\left(\frac{N_i\sigma_i}{\sum_K N_k\sigma_k}\right)$$

where n is the total sample size for a given scenario of spatial replication, n_i is the sample size allocated to stratum i, N_i is the total number of sample units (1x1 km grid cells) in stratum i, σ_i is the standard deviation of measurements amongst all N_i sample units contained within stratum i, and summation in the denominator is over all k strata (i.e., all 8 strata for range-wide sampling of light-footed Ridgway's rails). For this study we approximated σ_i for each stratum using the average cell counts from recent annual census efforts (2007-2021). That is, we first calculated the average light-footed Ridgway's rail count per 1x1 km grid cell for 2007-2021 for every sample unit in the sampling frame. We then calculated the within-strata standard deviation of average counts amongst cells to approximate σ_i for each stratum. For our targeted sample allocation strategy, we always sampled 50% of the sample units within the NWR strata (Seal Beach, South Bay, Sweetwater, and Tijuana), whereas the remaining samples for a given spatial

replication scenario were allocated to the other 4 strata (additional important saltwater, additional important freshwater, other saltwater, other freshwater,) using Neyman allocation. In this case, n represents the total number of remaining samples after accounting for the NWR samples, and k = 1,…,4. Lastly, we forced a minimum sample size of 3 samples per stratum so that a within-stratum variance could be calculated (≥3 samples needed to calculate variance).

We tested the performance of 2 strategies for randomly selecting which sample units are surveyed within each stratum: 1) simple random sampling with equal probability of selecting all sample units, and 2) weighted probability sampling with weights proportional to recent abundance. The first strategy is as simple as it sounds, where all sample units receive the same probability of selection. For example, if a sample of 3 is drawn from a stratum with 10 sample units, the selection probability for each sample unit is 1/10 on the first draw, 1/9 on the second draw, and 1/8 on the third draw. Note that sample selection probabilities always change with each selected sample when implementing sampling without replacement from a finite population (i.e., a finite sampling frame). For the weighted probability sampling, sampling weights were constructed for each cell so that within-strata weights summed to 1 for each stratum. Specifically, we calculated the average cell counts from recent annual census efforts (2007-2021, as described above) within each stratum, and calculated weights for each sample unit using the following formula:

$$w_j = \frac{\bar{x}_j}{\sum_{N_i} \bar{x}_u}$$

where w_j is the sampling weight for sample unit j, $\bar{x}_j$ is the average light-footed Ridgway's rail count for sample unit j, with u = 1,…,N_i sample units in the stratum. Thus, this formula is weighting the sampling probabilities for individual cells based on the recorded number of light-footed Ridgway's rails from recent field sampling and gives larger weight (and hence greater

probability of being selected) to 1x1 km grid cells where more light-footed Ridgway's rails have been observed. This weighting could be updated regularly as new data are collected (e.g., every 3-5 years) so that it is based on best available information, especially for grid cells that were not part of the regular survey efforts from 1980-2021.

We constructed simple design-based estimators to evaluate the accuracy (i.e., bias and precision) that can be expected from different sampling strategies and levels of spatial-temporal replication. These design-based estimators made no attempt to estimate or account for survey-level detection error in abundance estimation, and instead treated the maximum sum of individual light-footed Ridgway's rails counted across all call-broadcast points within a 1x1 km cell as the number of birds counted within that sample unit (i.e., the maximum of the sum from visit 1, the sum from visit 2, etc.). As described in the manuscript text, we also tested performance of model-based methods for estimating abundance while accounting for detection error. However, the volume of data generated by this simulation study (864,000 simulated monitoring data sets), as well as the additional work required on the front end to develop the sampling frame and the model to simulate call-broadcast sampling, prevented us from testing model-based estimation over all data sets. Thus, simple design-based estimators provided us a straightforward algebraic means of understanding how bias and precision change with sampling strategy and amount of replication, and thereby provided us the ability to make recommendations about the sampling effort required for multi-scale population estimation. We anticipate these recommendations will be updated in the future as the estimation part of the process (i.e., what we do with the monitoring data once they are in hand) gets further refined.

Stratum-level abundance estimates were generated for all sampling strategies, and then combined to produce range-wide abundance estimates. Within each stratum, design-based

estimates were generated for simple random sampling using the following formula (Scheaffer et al. 2006):

$$\hat{\tau}_l = N_l \bar{y}_l$$

where $\hat{\tau}_l$ is the estimated total abundance of rails for stratum l, N_l is the total number of sample units (1x1 km grid cells) contained within stratum l, and $\bar{y}_l$ is the average count of rails across sampled cells within stratum l. In contrast, estimates of a population total using weighted probability sampling must weight the data being obtained by the inverse of the sample inclusion probability (i.e., the overall probability that a sample unit was included in the sample) for each sample unit (Scheaffer et al. 2006, pgs 52-56):

$$\hat{\tau}_l = \sum_{i=1}^{n_l} w_{i,l} y_{i,l}$$

where $y_{i,l}$ is the count of rails at sample unit i ($i = 1,\ldots, n_l$) and $w_{i,l} = {}^{1}/_{\pi_{i,l}}$, where $\pi_{i,l}$ represents the sample inclusion probability for sample unit i in stratum l. The sample inclusion probability ($\pi_{i,l}$) changes with both the sampling weights described previously and the sample size, and therefore inclusion probabilities for each cell were specific the sample allocation strategy and amount of spatial replication. We calculated π for each sample unit under each sample allocation strategy and level of spatial replication using the *inclusionprobabilities()* function provided by the "sampling" R package (version 2.9; Tillé and Matei 2021). Finally, regardless of the sampling strategy employed, range-wide estimates of the population total were calculated via simple summation over the sampling strata:

$$\hat{\tau}_{range-wide} = \sum_{l=1}^{L} \hat{\tau}_l$$

where $\hat{\tau}_{range-wide}$ represents the range-wide total and L is the total number of strata ($L = 8$).

Model Based Abundance Estimation

We tested model-based abundance estimation that accounted for detection error using zero-inflated versions of the Poisson-Lognormal and Poisson-Poisson N-mixture models (see chapter 6 of Kéry and Royle 2015). These hierarchical models included a process model describing the spatial-temporal distribution of light-footed Ridgway's rail abundance among sample units and observation models linking true abundance back to the observed count data from call-broadcast surveys. The process model was shared between the Poisson-Lognormal and Poisson-Poisson models:

$$s_{i,s} \sim Bernoulli(\theta_s)$$

$$N_{i,s} | s_{i,s} \sim Poisson\left(s_{i,s}\lambda_{i,s}\right)$$

$$log\left(\lambda_{i,s}\right) = log\left(A_{i,s}\right) + \beta_0 + \varepsilon_{i,s}$$

$$\varepsilon_{i,s} \sim Normal(0, \sigma_\varepsilon^2)$$

where $s_{i,s}$ is the occupancy state for sample unit *i* contained within stratum *s* (1 = occupied, 0 = unoccupied), θ_s is the stratum-specific occupancy probability and dictates zero-inflation induced by the fact that some sample units are inherently unoccupied (i.e., therefore must result in zero counts), and the baseline occupancy rates change among strata. After accounting for a sample unit's occupancy status, $N_{i,s}$ is the true latent abundance of rails within sample unit *i* in stratum *s*, $\lambda_{i,s}$ is the average number of rails, $log\left(A_{i,s}\right)$ is an offset for area of potential habitat within the cell ($A_{i,s}$), β_0 is the log-scale average density of rails across all occupied cells (i.e., the intercept term), and $\varepsilon_{i,s}$ is a sample-unit level random effect that adjusts rail density within cell *i* (i.e., accounting for local factors affecting rail density). Consequently, σ_ε^2 represents the variance of the sample unit-specific random effect, and therefore the unmodeled spatial variation in density

among sample units that is not accounted for by the area of habitat or intercept term. Moreover, we estimated θ_s parameters using past census data and binomial logistic regression:

$$o_s \sim Binomial(t_s, \theta_s)$$

$$logit(\boldsymbol{\theta}) = \boldsymbol{X\zeta}$$

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \\ \theta_6 \\ \theta_7 \\ \theta_8 \end{bmatrix}$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\boldsymbol{\zeta} = \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \zeta_4 \\ \zeta_5 \\ \zeta_6 \\ \zeta_7 \\ \zeta_7 \end{bmatrix}.$$

Here o_s represents the count of sample units in stratum *s* where at least 1 light-footed Ridgway's rail was observed within annual census efforts each year (i.e., either a 1 or 0 in each cell in each year), over the past 15 years (i.e., from 2007-2021, summed across years for all cells within stratum *s*), whereas t_s represents the total number of sample units sampled over the past 15 years within stratum *s* (i.e., summed across all years for all cells within each stratum). Thus o_s is a binomial sample (each cell was occupied or not in each year) from t_s within each stratum.

Hence, past census data were used to inform estimation of zero-inflation parameters governing the probability a sample unit is occupied within each stratum ($\boldsymbol{\theta}$), while also accounting for heterogeneity in cell-level occupancy probability among strata. The matrix $\boldsymbol{X}$ is a design matrix of indicator variables representing individual strata, where logistic regression coefficients modeling strata-level changes in occupancy probability are contained within the vector $\boldsymbol{\zeta}$. A shared parameter (ζ_7) was used to model occupancy probability for the other saltwater and other freshwater strata due to the very small numbers of occupied cells observed in past data within these strata (i.e., very small values for o_s in those strata).

The observation model for the Poisson-Lognormal model linked the state variable $N_{i,s}$ to observed count data from each call-broadcast survey:

$$c_{i,s,j,v} \sim Binomial\left(p_{i,s,j,v}, N_{i,s}\right)$$

$$logit\left(p_{i,s,j,v}\right) = \alpha_0 + \gamma_{i,s,j,v}$$

$$\gamma_{i,s,j,v} \sim Normal\left(0, \sigma_\gamma^2\right)$$

where $c_{i,s,j,v}$ is light-footed Ridgway's rail count on visit *v* at call-broadcast point *j* within sample unit *i* at stratum *s*, $p_{i,s,j,v}$ is the survey-specific detection probability, α_0 is the logit-scale average detection probability across all surveys (i.e., intercept term), and $\gamma_{i,s,j,v}$ is the survey-specific random effect that adjusts detection probability for each survey. Therefore, σ_γ^2 represents the variance of the survey-specific random effect, and therefore the heterogeneity in detection probability among individual surveys. We also tested a second parameterization of the Poisson-Lognormal mixture model that replaced $\gamma_{i,s,j,v}$ with $\gamma_{i,s,j}$, and thus included a random effect that was unique to each call-broadcast point (i.e., heterogeneity in local detection around each survey point) that remained unchanged across surveys.

The observation model for the Poisson-Poisson model linked the state variable $N_{i,s}$ to the observed count data summarized over each sample unit:

$$y_{i,s,v} \sim Poisson\left(N_{i,s}\varphi_{i,s,v}\right)$$

$$logit\left(\varphi_{i,s,v}\right) = \alpha_0 + \alpha_1 \times Z_{1,i,s} + \gamma_{i,s,v}$$

$$\gamma_{i,s,v} \sim Normal\left(0, \sigma_\gamma^2\right)$$

where $y_{i,s,v}$ is light-footed Ridgway's rail count summed over all *j* broadcast points on visit *v* within sample unit *i* at stratum *s*, $\varphi_{i,s,v}$ is the rate of detection for rails within sample unit *i* (i.e., the average number of times each bird was detected across all the broadcast points within the sample unit), α_0 is intercept term and α_1 the regression coefficient for a covariate representing density of survey points within each sample unit (i.e., density of broadcast survey points per unit of potential habitat, where greater density should increase the detection rate), and $\gamma_{i,s,v}$ is the survey-specific random effect that adjusts the detection rate on each individual survey. Therefore, σ_γ^2 again represents the variance of the survey-specific random effect. We again tested a second parameterization of the model that replaced $\gamma_{i,s,v}$ with $\gamma_{i,s}$, and therefore modeled the random effect as unique to each sample unit (i.e., heterogeneity among 1x1 km cells) that remained unchanged across surveys at each point. The log transformation could also be used to model $\varphi_{i,s,v}$, instead of the logit, if double counting of individual rails across multiple broadcast points within a 1x1 km sample unit was common, as the log transformation would not bound the detection rate to be <1 (as the logit transformation does), although we did not test this alternative parameterization in our analyses.

In summary, we tested 4 Bayesian N-mixture models for estimating rail abundance while accounting for detection error, including observation models that: 1) used light-footed Ridgway's rail counts at each individual call-broadcast point as the response variable (Poisson-Lognormal),

and 2) used light-footed Ridgway's rail counts summed across call-broadcast points within a cell (uniquely for each site visit) as the response variable (Poisson-Poisson). For these models, we fit random effects structures in the detection model that included spatial (i.e., across points or cells) and temporal (i.e., across site visits) random effects, as well as models that only included spatial random effects (i.e., at the point or cell level). Lastly, we estimated stratum-specific and range-wide abundance for each iteration of our MCMC sampler while accounting for the unequal probability sampling weights, as described above for design-based estimators:

$$\widehat{N}_s = \sum_{s=1}^{n_s} w_{i,s} N_{i,s}$$

$$\widehat{N}_{range-wide} = \sum_{s=1}^{S} \widehat{N}_s$$

where $w_{i,s}$ is the inverse of the sample inclusion probability for each cell (as described above), and n_s represents the number of cells sampled in stratum *s*. This allowed us to directly sample from a posterior distribution for estimated abundance for each level and area of interest. The JAGS model statements for the most complicated versions of each model (i.e., including space and time random effects in the detection functions) are provided at the end of this appendix.

REFERENCES

BirdLife International and Handbook of the Birds of the World (2020) Bird species distribution maps of the world. Version 2020.1, http://datazone.birdlife.org/species/requestdis.

California Department of Fish and Wildlife, Biogegraphic Data Branch (2021) California Wildlife Habitat Relationship System, Version 10.1.27. Sacramento, CA. accessed: 2/2021.

eBird Basic Dataset (2021) Version EBD_relNov-2021. Cornell Lab of Ornithology, Ithaca, New York, USA, accessed 11/2021.

Grinnell, J (1915) A distributional list of the birds of California. Pacific Coast Avifauna 11.

GreenInfo Network (2018) California Protected Areas Database (CAPD). https://www.calands.org/cpad/, Accessed 2/2022.

Kéry M, and 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.

McKay L, Bondelid T, Dewald T, Johnston J, Moore R, and Reah A (2012) *NHDPlus Version 2: User Guide*. https://www.epa.gov/waterdata/get-nhdplus-national-hydrography-dataset-plus-data, Accessed 3/2022.

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.

Nelson ML, Brewer CK, and 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.

San Francisco Estuary Institute (2017) California Aquatic Resource Inventory (CARI) version 0.3. Accessed [date retrieved]. 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.

Scheaffer RL, Mendenhall III W, and Ott RL (2006) Elementary survey sampling. Sixth edition. Thompson Brooks/Cole Publishing, Belmont, California, USA.

Tillé Y, and Matei A (2021) sampling: Survey Sampling. R package version 2.9, https://CRAN.R-project.org/package=sampling.

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. 121 pp.

U.S. Fish and Wildlife Service (2021) Complete current range. https://ecos.fws.gov/ecosCommon/document/proxy/species/shapefiles/usfws_complete_species_current_range.zip

U.S. Geological Survey Gap Analysis Project (2018) U.S. Geological Survey – Gap Analysis Project Species Range Maps CONUS_2001: U.S. Geological Survey data release, https://doi.org/10.5066/F7Q81B3R.

Zembal D, and Hoffman SM (2021) Light-footed Ridgway's (Clapper) Rail in California: 2021 Season. Technical Report. Huntington Beach Wetlands Conservancy, Laguna Hills, California, USA.

Poisson-Lognormal JAGS model statement:

```
model{

###
#Priors
###

# zero-inflation/suitability
for(i in 1:7){
 zeta[i] ~ dnorm(0,1)                    #logistic regression terms for strata-level suitability model
}

# abundance
beta0 ~ dnorm(0, 0.05)                   #log(lambda) intercept
precision.epsilon <- pow(sd.epsilon, -2)
sd.epsilon ~ dunif(0, 2)                 #site-level variation in lambda

# detection
mean.p ~ dunif(0,1)
alpha0 <- logit(mean.p)
precision.gamma <- pow(sd.gamma, -2)
sd.gamma ~ dunif(0,4)                    #site-survey variation in p


########
#Likelihoods
########

##Suitability model - strata-level heterogeneity

for(s in 1:n.strata){
 past.occ[s] ~ dbin(theta[s],past.total[s])
 logit(theta[s]) <- zeta[1]*strata[s,1] + zeta[2]*strata[s,2] + zeta[3]*strata[s,3] + zeta[4]*strata[s,4] + zeta[5]*strata[s,5] +
          zeta[6]*strata[s,6] + zeta[7]*strata[s,7] + zeta[7]*strata[s,8]
}
```

```
##ZIP Abundance models - for each stratum separately because of differences in sample sizes

#Seal Beach
for(i in 1:n.cells[1]){
 occ.seal[i] ~ dbern(theta[1])
 epsilon[i] ~ dnorm(0,precision.epsilon)
 loglamb.seal[i] <- log(area[1,i]) + beta0 + epsilon[i]
 loglamb.seal.lim[i] <- min(250, max(-250, loglamb.seal[i]))
 lam.seal[i] <- exp(loglamb.seal.lim[i])
 lambda.seal[i] <- occ.seal[i]*lam.seal[i]
 N.seal[i] ~ dpois(lambda.seal[i])
}

#South Bay
for(i in 1:n.cells[2]){
 occ.sb[i] ~ dbern(theta[2])
 epsilon[i+n.cells[1]] ~ dnorm(0,precision.epsilon)
 loglamb.sb[i] <- log(area[2,i]) + beta0 + epsilon[i+n.cells[1]]
 loglamb.sb.lim[i] <- min(250, max(-250, loglamb.sb[i]))
 lam.sb[i] <- exp(loglamb.sb.lim[i])
 lambda.sb[i] <- occ.sb[i]*lam.sb[i]
 N.sb[i] ~ dpois(lambda.sb[i])
}

#Sweetwater
for(i in 1:n.cells[3]){
 occ.sweet[i] ~ dbern(theta[3])
 epsilon[i+n.cells[1]+n.cells[2]] ~ dnorm(0,precision.epsilon)
 loglamb.sweet[i] <- log(area[3,i]) + beta0 + epsilon[i+n.cells[1]+n.cells[2]]
 loglamb.sweet.lim[i] <- min(250, max(-250, loglamb.sweet[i]))
 lam.sweet[i] <- exp(loglamb.sweet.lim[i])
 lambda.sweet[i] <- occ.sweet[i]*lam.sweet[i]
 N.sweet[i] ~ dpois(lambda.sweet[i])
}

#Tijuana
for(i in 1:n.cells[4]){
 occ.tijuana[i] ~ dbern(theta[4])
 epsilon[i+n.cells[1]+n.cells[2]+n.cells[3]] ~ dnorm(0,precision.epsilon)
```

```
 loglamb.tijuana[i] <- log(area[4,i]) + beta0 + epsilon[i+n.cells[1]+n.cells[2]+n.cells[3]]
 loglamb.tijuana.lim[i] <- min(250, max(-250, loglamb.tijuana[i]))
 lam.tijuana[i] <- exp(loglamb.tijuana.lim[i])
 lambda.tijuana[i] <- occ.tijuana[i]*lam.tijuana[i]
 N.tijuana[i] ~ dpois(lambda.tijuana[i])
}

#Important fresh
for(i in 1:n.cells[5]){
 occ.imp.fresh[i] ~ dbern(theta[5])
 epsilon[i+n.cells[1]+n.cells[2]+n.cells[3]+n.cells[4]] ~ dnorm(0,precision.epsilon)
 loglamb.imp.fresh[i] <- log(area[5,i]) + beta0 + epsilon[i+n.cells[1]+n.cells[2]+n.cells[3]+n.cells[4]]
 loglamb.imp.fresh.lim[i] <- min(250, max(-250, loglamb.imp.fresh[i]))
 lam.imp.fresh[i] <- exp(loglamb.imp.fresh.lim[i])
 lambda.imp.fresh[i] <- occ.imp.fresh[i]*lam.imp.fresh[i]
 N.imp.fresh[i] ~ dpois(lambda.imp.fresh[i])
}

#Important salt
for(i in 1:n.cells[6]){
 occ.imp.salt[i] ~ dbern(theta[6])
 epsilon[i+n.cells[1]+n.cells[2]+n.cells[3]+n.cells[4]+n.cells[5]] ~ dnorm(0,precision.epsilon)
 loglamb.imp.salt[i] <- log(area[6,i]) + beta0 + epsilon[i+n.cells[1]+n.cells[2]+n.cells[3]+n.cells[4]+n.cells[5]]
 loglamb.imp.salt.lim[i] <- min(250, max(-250, loglamb.imp.salt[i]))
 lam.imp.salt[i] <- exp(loglamb.imp.salt.lim[i])
 lambda.imp.salt[i] <- occ.imp.salt[i]*lam.imp.salt[i]
 N.imp.salt[i] ~ dpois(lambda.imp.salt[i])
}

#Other fresh
for(i in 1:n.cells[7]){
 occ.o.fresh[i] ~ dbern(theta[7])
 epsilon[i+n.cells[1]+n.cells[2]+n.cells[3]+n.cells[4]+n.cells[5]+n.cells[6]] ~ dnorm(0,precision.epsilon)
 loglamb.o.fresh[i] <- log(area[7,i]) + beta0 + epsilon[i+n.cells[1]+n.cells[2]+n.cells[3]+n.cells[4]+n.cells[5]+n.cells[6]]
 loglamb.o.fresh.lim[i] <- min(250, max(-250, loglamb.o.fresh[i]))
 lam.o.fresh[i] <- exp(loglamb.o.fresh.lim[i])
 lambda.o.fresh[i] <- occ.o.fresh[i]*lam.o.fresh[i]
 N.o.fresh[i] ~ dpois(lambda.o.fresh[i])
}
```

```
#Other salt
for(i in 1:n.cells[8]){
 occ.o.salt[i] ~ dbern(theta[8])
 epsilon[i+n.cells[1]+n.cells[2]+n.cells[3]+n.cells[4]+n.cells[5]+n.cells[6]+n.cells[7]] ~ dnorm(0,precision.epsilon)
 loglamb.o.salt[i] <- log(area[8,i]) + beta0 + epsilon[i+n.cells[1]+n.cells[2]+n.cells[3]+n.cells[4]+n.cells[5]+n.cells[6]+n.cells[7]]
 loglamb.o.salt.lim[i] <- min(250, max(-250, loglamb.o.salt[i]))
 lam.o.salt[i] <- exp(loglamb.o.salt.lim[i])
 lambda.o.salt[i] <- occ.o.salt[i]*lam.o.salt[i]
 N.o.salt[i] ~ dpois(lambda.o.salt[i])
}

###
#Measurement error models
###

#Seal Beach
for(i in 1:n.cells[1]){
 for(j in 1:seal.points[i]){
  for(v in 1:visits){
   seal.count[i,j,v] ~ dbin(p.seal[i,j,v],N.seal[i])
   p.seal[i,j,v] <- 1/(1+exp(-lp.lim.seal[i,j,v]))
   lp.lim.seal[i,j,v] <- min(250, max(-250, lp.seal[i,j,v]))          # 'Stabilize' logit
   lp.seal[i,j,v] <- alpha0 + gamma.seal[i,j,v]
   gamma.seal[i,j,v] ~ dnorm(0,precision.gamma)                       # Random site-survey effects
  }
 }
}

#South Bay
for(i in 1:n.cells[2]){
 for(j in 1:sb.points[i]){
  for(v in 1:visits){
   sb.count[i,j,v] ~ dbin(p.sb[i,j,v],N.sb[i])
   p.sb[i,j,v] <- 1/(1+exp(-lp.lim.sb[i,j,v]))
   lp.lim.sb[i,j,v] <- min(250, max(-250, lp.sb[i,j,v])) # 'Stabilize' logit
   lp.sb[i,j,v] <- alpha0 + gamma.sb[i,j,v]
   gamma.sb[i,j,v] ~ dnorm(0,precision.gamma)
  }
 }
```

```
}

#Sweetwater
for(i in 1:n.cells[3]){
 for(j in 1:sweet.points[i]){
  for(v in 1:visits){
   sweet.count[i,j,v] ~ dbin(p.sweet[i,j,v],N.sweet[i])
   p.sweet[i,j,v] <- 1/(1+exp(-lp.lim.sweet[i,j,v]))
   lp.lim.sweet[i,j,v] <- min(250, max(-250, lp.sweet[i,j,v])) # 'Stabilize' logit
   lp.sweet[i,j,v] <- alpha0 + gamma.sweet[i,j,v]
   gamma.sweet[i,j,v] ~ dnorm(0,precision.gamma)
  }
 }
}

#Tijuana
for(i in 1:n.cells[4]){
 for(j in 1:tijuana.points[i]){
  for(v in 1:visits){
   tijuana.count[i,j,v] ~ dbin(p.tijuana[i,j,v],N.tijuana[i])
   p.tijuana[i,j,v] <- 1/(1+exp(-lp.lim.tijuana[i,j,v]))
   lp.lim.tijuana[i,j,v] <- min(250, max(-250, lp.tijuana[i,j,v])) # 'Stabilize' logit
   lp.tijuana[i,j,v] <- alpha0 + gamma.tijuana[i,j,v]
   gamma.tijuana[i,j,v] ~ dnorm(0,precision.gamma)
  }
 }
}

#Important fresh
for(i in 1:n.cells[5]){
 for(j in 1:imp.fresh.points[i]){
  for(v in 1:visits){
   imp.fresh.count[i,j,v] ~ dbin(p.imp.fresh[i,j,v],N.imp.fresh[i])
   p.imp.fresh[i,j,v] <- 1/(1+exp(-lp.lim.imp.fresh[i,j,v]))
   lp.lim.imp.fresh[i,j,v] <- min(250, max(-250, lp.imp.fresh[i,j,v])) # 'Stabilize' logit
   lp.imp.fresh[i,j,v] <- alpha0 + gamma.imp.fresh[i,j,v]
   gamma.imp.fresh[i,j,v] ~ dnorm(0,precision.gamma)
  }
 }
```

```
}

#Important salt
for(i in 1:n.cells[6]){
 for(j in 1:imp.salt.points[i]){
  for(v in 1:visits){
   imp.salt.count[i,j,v] ~ dbin(p.imp.salt[i,j,v],N.imp.salt[i])
   p.imp.salt[i,j,v] <- 1/(1+exp(-lp.lim.imp.salt[i,j,v]))
   lp.lim.imp.salt[i,j,v] <- min(250, max(-250, lp.imp.salt[i,j,v])) # 'Stabilize' logit
   lp.imp.salt[i,j,v] <- alpha0 + gamma.imp.salt[i,j,v]
   gamma.imp.salt[i,j,v] ~ dnorm(0,precision.gamma)
  }
 }
}

#other fresh
for(i in 1:n.cells[7]){
 for(j in 1:o.fresh.points[i]){
  for(v in 1:visits){
   o.fresh.count[i,j,v] ~ dbin(p.o.fresh[i,j,v],N.o.fresh[i])
   p.o.fresh[i,j,v] <- 1/(1+exp(-lp.lim.o.fresh[i,j,v]))
   lp.lim.o.fresh[i,j,v] <- min(250, max(-250, lp.o.fresh[i,j,v])) # 'Stabilize' logit
   lp.o.fresh[i,j,v] <- alpha0 + gamma.o.fresh[i,j,v]
   gamma.o.fresh[i,j,v] ~ dnorm(0,precision.gamma)
  }
 }
}

#other salt
for(i in 1:n.cells[8]){
 for(j in 1:o.salt.points[i]){
  for(v in 1:visits){
   o.salt.count[i,j,v] ~ dbin(p.o.salt[i,j,v],N.o.salt[i])
   p.o.salt[i,j,v] <- 1/(1+exp(-lp.lim.o.salt[i,j,v]))
   lp.lim.o.salt[i,j,v] <- min(250, max(-250, lp.o.salt[i,j,v])) # 'Stabilize' logit
   lp.o.salt[i,j,v] <- alpha0 + gamma.o.salt[i,j,v]
   gamma.o.salt[i,j,v] ~ dnorm(0,precision.gamma)
  }
 }
```

```
}

###
#Derived parameters for estimating abundance
###

N.strata.estimates[1] <- inprod(seal.weights[],N.seal[])
N.strata.estimates[2] <- inprod(sb.weights[],N.sb[])
N.strata.estimates[3] <- inprod(sweet.weights[],N.sweet[])
N.strata.estimates[4] <- inprod(tijuana.weights[],N.tijuana[])
N.strata.estimates[5] <- inprod(imp.fresh.weights[],N.imp.fresh[])
N.strata.estimates[6] <- inprod(imp.salt.weights[],N.imp.salt[])
N.strata.estimates[7] <- inprod(o.fresh.weights[],N.o.fresh[])
N.strata.estimates[8] <- inprod(o.salt.weights[],N.o.salt[])

N.rangewide.estimate <- sum(N.strata.estimates[])

}
```

Poisson-Poisson JAGS model statement:

```
model{
###
#Priors
###

# zero-inflation/suitability
for(i in 1:7){
  zeta[i] ~ dnorm(0,1)                    #logistic regression terms for strata-level suitability model
}

# abundance
beta0 ~ dnorm(0, 0.05)                    # log(lambda) intercept
precision.epsilon <- pow(sd.epsilon, -2)
sd.epsilon ~ dunif(0, 2)                  # site-level variation in lambda

# detection
mean.p ~ dunif(0,1)
alpha0 <- logit(mean.p)
alpha1 ~ dnorm(0,1)
precision.gamma <- pow(sd.gamma, -2)
sd.gamma ~ dunif(0,0.5)                   # site-survey variation n p


########
#Likelihoods
########

##Suitability model - strata-level heterogeneity

for(s in 1:n.strata){
  past.occ[s] ~ dbin(theta[s],past.total[s])
  logit(theta[s]) <- zeta[1]*strata[s,1] + zeta[2]*strata[s,2] + zeta[3]*strata[s,3] + zeta[4]*strata[s,4] + zeta[5]*strata[s,5] +
          zeta[6]*strata[s,6] + zeta[7]*strata[s,7] + zeta[7]*strata[s,8]
}

##ZIP Abundance models - for each stratum separately because of differences in sample sizes
```

```
#Seal Beach
for(i in 1:n.cells[1]){
 occ.seal[i] ~ dbern(theta[1])
 epsilon.seal[i] ~ dnorm(0,precision.epsilon)
 loglamb.seal[i] <- log(area[1,i]) + beta0 + epsilon.seal[i]
 loglamb.seal.lim[i] <- min(250, max(-250, loglamb.seal[i]))          #'Stabalize' log
 lam.seal[i] <- exp(loglamb.seal.lim[i])
 lambda.seal[i] <- occ.seal[i]*lam.seal[i]
 N.seal[i] ~ dpois(lambda.seal[i])
}

#South Bay
for(i in 1:n.cells[2]){
 occ.sb[i] ~ dbern(theta[2])
 epsilon.sb[i] ~ dnorm(0,precision.epsilon)
 loglamb.sb[i] <- log(area[2,i]) + beta0 + epsilon.sb[i]
 loglamb.sb.lim[i] <- min(250, max(-250, loglamb.sb[i]))
 lam.sb[i] <- exp(loglamb.sb.lim[i])
 lambda.sb[i] <- occ.sb[i]*lam.sb[i]
 N.sb[i] ~ dpois(lambda.sb[i])
}

#Sweetwater
for(i in 1:n.cells[3]){
 occ.sweet[i] ~ dbern(theta[3])
 epsilon.sweet[i] ~ dnorm(0,precision.epsilon)
 loglamb.sweet[i] <- log(area[3,i]) + beta0 + epsilon.sweet[i]
 loglamb.sweet.lim[i] <- min(250, max(-250, loglamb.sweet[i]))
 lam.sweet[i] <- exp(loglamb.sweet.lim[i])
 lambda.sweet[i] <- occ.sweet[i]*lam.sweet[i]
 N.sweet[i] ~ dpois(lambda.sweet[i])
}

#Tijuana
for(i in 1:n.cells[4]){
 occ.tijuana[i] ~ dbern(theta[4])
 epsilon.tijuana[i] ~ dnorm(0,precision.epsilon)
 loglamb.tijuana[i] <- log(area[4,i]) + beta0 + epsilon.tijuana[i]
 loglamb.tijuana.lim[i] <- min(250, max(-250, loglamb.tijuana[i]))
```

```
 lam.tijuana[i] <- exp(loglamb.tijuana.lim[i])
 lambda.tijuana[i] <- occ.tijuana[i]*lam.tijuana[i]
 N.tijuana[i] ~ dpois(lambda.tijuana[i])
}

#Important fresh
for(i in 1:n.cells[5]){
 occ.imp.fresh[i] ~ dbern(theta[5])
 epsilon.imp.fresh[i] ~ dnorm(0,precision.epsilon)
 loglamb.imp.fresh[i] <- log(area[5,i]) + beta0 + epsilon.imp.fresh[i]
 loglamb.imp.fresh.lim[i] <- min(250, max(-250, loglamb.imp.fresh[i]))
 lam.imp.fresh[i] <- exp(loglamb.imp.fresh.lim[i])
 lambda.imp.fresh[i] <- occ.imp.fresh[i]*lam.imp.fresh[i]
 N.imp.fresh[i] ~ dpois(lambda.imp.fresh[i])
}

#Important salt
for(i in 1:n.cells[6]){
 occ.imp.salt[i] ~ dbern(theta[6])
 epsilon.imp.salt[i] ~ dnorm(0,precision.epsilon)
 loglamb.imp.salt[i] <- log(area[6,i]) + beta0 + epsilon.imp.salt[i]
 loglamb.imp.salt.lim[i] <- min(250, max(-250, loglamb.imp.salt[i]))
 lam.imp.salt[i] <- exp(loglamb.imp.salt.lim[i])
 lambda.imp.salt[i] <- occ.imp.salt[i]*lam.imp.salt[i]
 N.imp.salt[i] ~ dpois(lambda.imp.salt[i])
}

#Other fresh
for(i in 1:n.cells[7]){
 occ.o.fresh[i] ~ dbern(theta[7])
 epsilon.o.fresh[i] ~ dnorm(0,precision.epsilon)
 loglamb.o.fresh[i] <- log(area[7,i]) + beta0 + epsilon.o.fresh[i]
 loglamb.o.fresh.lim[i] <- min(250, max(-250, loglamb.o.fresh[i]))
 lam.o.fresh[i] <- exp(loglamb.o.fresh.lim[i])
 lambda.o.fresh[i] <- occ.o.fresh[i]*lam.o.fresh[i]
 N.o.fresh[i] ~ dpois(lambda.o.fresh[i])
}
#Other salt
for(i in 1:n.cells[8]){
```

```
 occ.o.salt[i] ~ dbern(theta[8])
 epsilon.o.salt[i] ~ dnorm(0,precision.epsilon)
 loglamb.o.salt[i] <- log(area[8,i]) + beta0 + epsilon.o.salt[i]
 loglamb.o.salt.lim[i] <- min(250, max(-250, loglamb.o.salt[i]))
 lam.o.salt[i] <- exp(loglamb.o.salt.lim[i])
 lambda.o.salt[i] <- occ.o.salt[i]*lam.o.salt[i]
 N.o.salt[i] ~ dpois(lambda.o.salt[i])
}

###
#Measurement error models
###

#SealBbeach
for(i in 1:n.cells[1]){
 for(v in 1:visits){
  seal.cell.count[i,v] ~ dpois(phi.seal[i,v]*N.seal[i])
  phi.seal[i,v] <- 1/(1+exp(-lphi.lim.seal[i,v]))
  lphi.lim.seal[i,v] <- min(250, max(-250, lphi.seal[i,v]))                    # 'Stabilize' logit
  lphi.seal[i,v] <- alpha0 + alpha1*seal.point.density[i] + gamma.seal[i,v]
  gamma.seal[i,v] ~ dnorm(0,precision.gamma)                                   # Random site-survey effects
 }
}

#South Bay
for(i in 1:n.cells[2]){
 for(v in 1:visits){
  sb.cell.count[i,v] ~ dpois(phi.sb[i,v]*N.sb[i])
  phi.sb[i,v] <- 1/(1+exp(-lphi.lim.sb[i,v]))
  lphi.lim.sb[i,v] <- min(250, max(-250, lphi.sb[i,v])) # 'Stabilize' logit
  lphi.sb[i,v] <- alpha0 + alpha1*sb.point.density[i] + gamma.sb[i,v]
  gamma.sb[i,v] ~ dnorm(0,precision.gamma)                                     # Random site-survey effects
 }
}

#Sweetwater
for(i in 1:n.cells[3]){
 for(v in 1:visits){
  sweet.cell.count[i,v] ~ dpois(phi.sweet[i,v]*N.sweet[i])
```

```
   phi.sweet[i,v] <- 1/(1+exp(-lphi.lim.sweet[i,v]))
   lphi.lim.sweet[i,v] <- min(250, max(-250, lphi.sweet[i,v])) # 'Stabilize' logit
   lphi.sweet[i,v] <- alpha0 + alpha1*sweet.point.density[i] + gamma.sweet[i,v]
   gamma.sweet[i,v] ~ dnorm(0,precision.gamma)                    # Random site-survey effects
  }
}

#Tijuana
for(i in 1:n.cells[4]){
 for(v in 1:visits){
   tijuana.cell.count[i,v] ~ dpois(phi.tijuana[i,v]*N.tijuana[i])
   phi.tijuana[i,v] <- 1/(1+exp(-lphi.lim.tijuana[i,v]))
   lphi.lim.tijuana[i,v] <- min(250, max(-250, lphi.tijuana[i,v])) # 'Stabilize' logit
   lphi.tijuana[i,v] <- alpha0 + alpha1*tijuana.point.density[i] + gamma.tijuana[i,v]
   gamma.tijuana[i,v] ~ dnorm(0,precision.gamma)                  # Random site-survey effects
  }
}

#Important fresh
for(i in 1:n.cells[5]){
 for(v in 1:visits){
   imp.fresh.cell.count[i,v] ~ dpois(phi.imp.fresh[i,v]*N.imp.fresh[i])
   phi.imp.fresh[i,v] <- 1/(1+exp(-lphi.lim.imp.fresh[i,v]))
   lphi.lim.imp.fresh[i,v] <- min(250, max(-250, lphi.imp.fresh[i,v])) # 'Stabilize' logit
   lphi.imp.fresh[i,v] <- alpha0 + alpha1*imp.fresh.point.density[i] + gamma.imp.fresh[i,v]
   gamma.imp.fresh[i,v] ~ dnorm(0,precision.gamma)                # Random site-survey effects
  }
}

#Important salt
for(i in 1:n.cells[6]){
 for(v in 1:visits){
   imp.salt.cell.count[i,v] ~ dpois(phi.imp.salt[i,v]*N.imp.salt[i])
   phi.imp.salt[i,v] <- 1/(1+exp(-lphi.lim.imp.salt[i,v]))
   lphi.lim.imp.salt[i,v] <- min(250, max(-250, lphi.imp.salt[i,v])) # 'Stabilize' logit
   lphi.imp.salt[i,v] <- alpha0 + alpha1*imp.salt.point.density[i] + gamma.imp.salt[i,v]
   gamma.imp.salt[i,v] ~ dnorm(0,precision.gamma)                 # Random site-survey effects
  }
}
```

```
#Other fresh
for(i in 1:n.cells[7]){
 for(v in 1:visits){
  o.fresh.cell.count[i,v] ~ dpois(phi.o.fresh[i,v]*N.o.fresh[i])
  phi.o.fresh[i,v] <- 1/(1+exp(-lphi.lim.o.fresh[i,v]))
  lphi.lim.o.fresh[i,v] <- min(250, max(-250, lphi.o.fresh[i,v])) # 'Stabilize' logit
  lphi.o.fresh[i,v] <- alpha0 + alpha1*o.fresh.point.density[i] + gamma.o.fresh[i,v]
  gamma.o.fresh[i,v] ~ dnorm(0,precision.gamma)                    # Random site-survey effects
 }
}

#Other salt
for(i in 1:n.cells[8]){
 for(v in 1:visits){
  o.salt.cell.count[i,v] ~ dpois(phi.o.salt[i,v]*N.o.salt[i])
  phi.o.salt[i,v] <- 1/(1+exp(-lphi.lim.o.salt[i,v]))
  lphi.lim.o.salt[i,v] <- min(250, max(-250, lphi.o.salt[i,v])) # 'Stabilize' logit
  lphi.o.salt[i,v] <- alpha0 + alpha1*o.salt.point.density[i] + gamma.o.salt[i,v]
  gamma.o.salt[i,v] ~ dnorm(0,precision.gamma)                    # Random site-survey effects
 }
}

###
# Derived parameters for estimating abundance
###

N.strata.estimates[1] <- inprod(seal.weights[],N.seal[])
N.strata.estimates[2] <- inprod(sb.weights[],N.sb[])
N.strata.estimates[3] <- inprod(sweet.weights[],N.sweet[])
N.strata.estimates[4] <- inprod(tijuana.weights[],N.tijuana[])
N.strata.estimates[5] <- inprod(imp.fresh.weights[],N.imp.fresh[])
N.strata.estimates[6] <- inprod(imp.salt.weights[],N.imp.salt[])
N.strata.estimates[7] <- inprod(o.fresh.weights[],N.o.fresh[])
N.strata.estimates[8] <- inprod(o.salt.weights[],N.o.salt[])

N.rangewide.estimate <- sum(N.strata.estimates[])
}
```