Appendix S2

Matthew T. Farr, David S. Green, Kay E. Holekamp, and Elise F. Zipkin

Integrating distance sampling and presence-only data to estimate species abundance

Ecology

Simulation study

This appendix describes the parameter values used in the simulation study.

Section S1. Parameter and covariate values:

The intercept parameter for the biological process was randomly drawn as follows: $\lambda_0 \sim uniform(0.05, 1)$. This equated to a range of 500 to 10,000 individuals for region A (assuming no covariate effects). A single effect parameter on the biological process was also assumed to come from a uniform distribution: $\beta_1 \sim uniform(-1.25, 1.25)$. The covariate values at each pixel across region A for the biological process came from a correlated multivariate normal distribution using Euclidean distances between pixels to define the variance-covariance matrix (Kéry & Royle 2016, pg. 534). We standardized the covariate values to have a mean of 0 and standard deviation of 1. Expected pixel density ranged between 0 and 85 based on the magnitude and values of the effect parameter and covariate. The intercept parameter for the observation bias of opportunistic sampling (presence-only data) was predetermined as either: $p_0 = 0.5$ or $p_0 = 0.1$, based on whether we assumed a low or high quantity of presence-only data. The covariate effect for opportunistic sampling was drawn from a random uniform distribution: $\alpha_1 \sim uniform(0, 2)$. The covariate values for the observation process of opportunistic sampling were also assumed to come from a multivariate normal distribution (Dorazio 2014). We created separate covariates for high and low quantities of presence-only data by adjusting the variance of the multivariate normal distribution, which translated to larger or smaller (high vs low) sampling intensity. The scale parameter for the distance sampling observational process was drawn from a uniform distribution: $\sigma \sim uniform(0.75, 1.25)$, which led to a detection probability ranging from approximately from 19 to 31%.

Section S2. Simulation results:

Below are the bias (estimated - truth) results for each parameter with the interquartile range and $\pm$ 1.5 the interquartile range. Each row represents a different scenario for the various data quantities. Scenarios 1-5 are for high quantities of presence-only data for each of 0, 5, 10, 15, and 20% distance sampling coverage. Scenarios 6-10 are for low quantities of presence-only data for each level of distance sampling coverage.

Table S1. Parameter Abundance (N)

	min	25	50	75	max
High PO, 0% DS	-13509.20	-6045.91	-3567.49	-1070.38	6392.90
High PO, 5% DS	-4457.62	-1716.87	-654.92	110.30	2851.05
High PO, 10% DS	-2451.25	-904.87	-331.26	126.05	1672.43
High PO, 15% DS	-1848.68	-646.47	-215.01	154.99	1357.20
High PO, 20% DS	-1400.77	-476.71	-158.86	139.33	1063.39
Low PO, 0% DS	-15099.94	-5509.04	-2263.27	884.89	10475.79
Low PO, 5% DS	-2822.29	-937.94	-276.90	318.30	2202.65
Low PO, 10% DS	-1934.85	-614.91	-165.98	265.04	1584.98
Low PO, 15% DS	-1442.93	-432.17	-114.88	241.67	1252.42
Low PO, 20% DS	-1163.49	-344.53	-59.36	201.45	1020.42

Table S2. Parameter λ_0

	min	25	50	75	max
High PO, 0% DS	-1.08	-0.48	-0.28	-0.09	0.51
High PO, 5% DS	-0.34	-0.13	-0.05	0.01	0.22
High PO, 10% DS	-0.19	-0.07	-0.03	0.01	0.13
High PO, 15% DS	-0.14	-0.05	-0.02	0.01	0.11
High PO, 20% DS	-0.11	-0.04	-0.01	0.01	0.09
Low PO, 0% DS	-1.16	-0.43	-0.18	0.07	0.80
Low PO, 5% DS	-0.22	-0.07	-0.02	0.03	0.17
Low PO, 10% DS	-0.15	-0.05	-0.01	0.02	0.13
Low PO, 15% DS	-0.12	-0.04	-0.01	0.02	0.10
Low PO, 20% DS	-0.10	-0.03	-0.01	0.02	0.09

Table S3. Parameter β_1

	min	25	50	75	max
High PO, 0% DS	-0.09	-0.02	0	0.02	0.08
High PO, 5% DS	-0.08	-0.02	0	0.02	0.08
High PO, 10% DS	-0.08	-0.02	0	0.02	0.08
High PO, 15% DS	-0.08	-0.02	0	0.02	0.08
High PO, 20% DS	-0.08	-0.02	0	0.02	0.08
Low PO, 0% DS	-0.17	-0.04	0	0.04	0.17
Low PO, 5% DS	-0.15	-0.04	0	0.04	0.15
Low PO, 10% DS	-0.13	-0.03	0	0.03	0.13
Low PO, 15% DS	-0.14	-0.04	0	0.03	0.13
Low PO, 20% DS	-0.12	-0.03	0	0.03	0.12

Table S4. Parameter σ

	min	25	50	75	max
High PO, 0% DS	NA	NA	NA	NA	NA
High PO, 5% DS	-0.27	0.00	0.06	0.18	0.44
High PO, 10% DS	-0.15	-0.01	0.03	0.08	0.21
High PO, 15% DS	-0.12	-0.01	0.02	0.06	0.16
High PO, 20% DS	-0.10	-0.02	0.01	0.04	0.13
Low PO, 0% DS	NA	NA	NA	NA	NA
Low PO, 5% DS	-0.20	-0.02	0.03	0.09	0.27
Low PO, 10% DS	-0.14	-0.02	0.01	0.06	0.18
Low PO, 15% DS	-0.12	-0.02	0.01	0.04	0.14
Low PO, 20% DS	-0.10	-0.02	0.01	0.03	0.12

Table S5. Parameter p_0

	min	25	50	75	max
High PO, 0% DS	-0.33	-0.11	-0.02	0.03	0.25
High PO, 5% DS	-0.46	-0.36	-0.33	-0.29	-0.18
High PO, 10% DS	-0.40	-0.36	-0.35	-0.34	-0.30
High PO, 15% DS	-0.39	-0.36	-0.35	-0.34	-0.32
High PO, 20% DS	-0.39	-0.36	-0.36	-0.35	-0.32
Low PO, 0% DS	-0.15	-0.02	0.01	0.06	0.19
Low PO, 5% DS	-0.08	-0.07	-0.07	-0.06	-0.05
Low PO, 10% DS	-0.08	-0.07	-0.07	-0.07	-0.06
Low PO, 15% DS	-0.08	-0.07	-0.07	-0.07	-0.06
Low PO, 20% DS	-0.08	-0.07	-0.07	-0.07	-0.06

Table S6. Parameter α_1

	min	25	50	75	max
High PO, 0% DS	-0.46	-0.11	-0.01	0.12	0.47
High PO, 5% DS	-1.17	-0.58	-0.38	-0.19	0.40
High PO, 10% DS	-1.45	-0.72	-0.46	-0.23	0.50
High PO, 15% DS	-1.57	-0.77	-0.48	-0.24	0.56
High PO, 20% DS	-1.59	-0.78	-0.49	-0.24	0.57
Low PO, 0% DS	-0.40	-0.07	0.03	0.15	0.48
Low PO, 5% DS	-0.78	-0.34	-0.16	-0.05	0.38
Low PO, 10% DS	-0.85	-0.37	-0.18	-0.06	0.42
Low PO, 15% DS	-0.86	-0.38	-0.18	-0.06	0.43
Low PO, 20% DS	-0.88	-0.39	-0.18	-0.06	0.44

Literature Cited

Dorazio, R.M. (2014) Accounting for imperfect detection and survey bias in statistical analysis of presence-only data. Global Ecology and Biogeography, 23, 1472–1484.

Kéry, M. & Royle, J.A. (2016) Applied hierarchical modeling in ecology: Analysis of distribution, abundance and species richness in R and BUGS (volume 1 – prelude and static models), Elsevier, Amsterdam.