Appendix S3

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

DOI 10.5281/zenodo.3981242

Application

This appendix provides additional details on the black-backed jackal (*Canis mesomelas*) case study.

Section S1. Study area

We used a kernel density estimator of the GPS point locations to determine the spatial extent of the study area within the Reserve. We used the upper 95% density of data values as a cutoff to define the spatial extent of the study area.

The extent of presence-only and distance sampling data differed within the study area. Presence-only data occurred throughout the study area while distance sampling occurred at a subset of the study area, which varied by month. We created indicator variables that varied by month to match the portion of the study area (i.e., portion of pixels) covered by distance sampling to the corresponding pixels across the study area (see model code).

Section S2. Opportunistic sampling: presence-only data

General information

Presence-only data are point observations in continuous space and time. To discretize space and time, we aggregated data points within pixels. We recommend that the size of pixels and the aggregation process of the presence-only data into pixels depend on both statistical and biological assumptions such that:
1) Pixels should be small enough (spatial/temporal resolution) to accurately capture heterogeneity across space and time. This allows the summation across pixels to numerically approximate the integral (within the point process model) that is used for continuous space and time analyses.
2) Pixels should also be large enough (spatial resolution) to prevent double counting of individuals across pixels (i.e., multiple observations/points of the same individual).
3) Temporal resolution of pixels should be determined based on the assumption of closure and the temporal resolution of the structured data (e.g., distance sampling data).

4) The aggregation of points within pixels should be critically considered. Depending on the temporal resolution of data and other biological factors, such as movement, data should be aggregated using the sum, mean, or max value.

Data collection

Presence-only data was collected as auxiliary information during unstructured surveying of spotted hyena behavior. Researchers drove around the Reserve looking for hyena subgroups to observer. While searching for hyenas, researchers recorded any other carnivore species and their location.

Section S3. Distance sampling: distance sampling data

General information

Distance-sampling data are counts along transects. However, if distance sampling data are not aggregated across a transect, they are simply point observations in continuous space and time. To discretize space and time, we aggregated the data within pixels. The temporal resolution of pixels are naturally defined by the timing of sampling events (e.g., monthly in our case study). The spatial resolution of pixels should be determined by the spatial variation in both observational and biological processes. Additionally, the distance of each point observation to the line is recorded to estimate imperfect detection. We used the distance from the pixel midpoint to the transect line as our spatial resolution of pixels. If pixel resolution/grain is too large, detection probability may not be estimated accurately.

Data collection

Transect surveys were conducted in the Reserve between July 2012 and March 2014 during which all observed herbivores and carnivores were recorded (Green 2015, Farr et al. 2019). To analyze the jackal data, we divided the survey transects into 10 km transect units leading to 4 transect units in the eastern region and 13 transect units in the western region. Due to terrain and off-road restrictions, straight line surveys were infeasible. However, a simulation study demonstrated that the curvy design of the transect surveys did not lead to biases in parameter estimates using a distance-sampling analysis framework (Farr et al. 2019). Replicate surveys were conducted every 4 to 6 weeks for a total of 13 replicates in the eastern region and 16 in the western region over the time frame of the study period.

Section S4. Case study covariates

We checked Pearson's correlations between covariates. We found none to be highly correlated (< 0.6).

Biological process

1) Distance to border

The distance from the midpoint of pixel g to the closest border of the Reserve. This is used as a proxy for human disturbance where pixels closer to the border are generally closer to human activity.

2) Disturbance regime

The Reserve is managed by separate entities who differ in enforcement of wildlife regulations (Green et al. 2018, Farr et al. 2019, Green et al. 2019). On the western side of the Reserve, there is active enforcement of regulations. On the eastern side of the Reserve, there is passive enforcement of regulations. The eastern and western portions of the Reserve are separated by the Mara River (Figure 2, main text). The two management regimes have caused differences in the amount of disturbance due to human activity. The

western side is relatively undisturbed and the eastern side is highly disturbed due to livestock grazing, tourism, and proximity to the urban area of Talek Town (Green et al. 2018, Green et al. 2019).

3) Interaction between distance to border and disturbance regime

We added an interaction between disturbance regime and distance to border because we hypothesized that jackals might respond to human activity differently between the two disturbance regimes.

4) Distance to water source

Distance from pixel g to the closest water source. Water sources were permanent rivers in the Reserve (e.g., Mara River, Talek River). We included this covariate to evaluate fine-scale responses of jackals to the landscape.

5) Lion density

African lions (*Panthera leo*) are the apex predator in the Reserve and may be driving jackal abundance at specific locations due to competition and avoidance. We used presence-only data of lions in the Reserve and a kernel density estimator to calculate relative density of lions within the Reserve. This covariate varied by pixel g and monthly replicate t.

6) Normalized difference vegetation index (NDVI)

NDVI is a measure of greenness used to represent the density of vegetation. We included this covariate to evaluate fine-scale responses of jackals to the landscape. This covariate varied by pixel g and monthly replicate t. NDVI data came from MODIS Terra satellite (MOD13A2) via NASA portal (Didan 2015).

Observational processes

Opportunistic sampling (presence-only) covariate: Sampling intensity

To account for biases in opportunistic sampling of the presence-only data, we added a covariate encapsulating sampling intensity. The opportunistic jackal data was collected as auxiliary data in which the primary targets were spotted hyena (*Crocuta crocuta*). Spotted hyena behavior was studied daily as part of a long term research project (Holekamp et al. 2012). Observers drove throughout territories of hyena clans to record hyena behavior. As part of the collection protocols, they also recorded the location and number of other carnivore species encountered, including black-backed jackals. The sampling intensity (i.e., the spatial locations visited monthly) depends on hyena locations, which can vary. We used behavioral session GPS locations (i.e., points; n = 16,267) over the study period and a kernel density estimator to calculate sampling intensity for each pixel g during monthly replicate t.

Distance sampling covariate:

To account for variation in detection probability during distance sampling, we added an effect of disturbance regime. We hypothesized that grass height, which differed between the two regions due to livestock grazing, could influence the detection probability of jackals.

Section S5. JAGS model code

Below is the JAGS model code. To see the full code used to run the analysis, see our GitHub repository.

```
model{

#------------#
#-LIKELIHOOD-#
#------------#

#t in TT represents monthly replicates that contain both observation processes
#Only replicate 10 does not contain distance sampling
for(t in TT){

#d in 1:Dend[cover[t]] represents the subset of 2500 m^2 pixels
#covered by the 650 m distance sampling transect half-width during replicate t
#cover[t] is a nested index to indicate pixel subsets during each montly replicate t
for(d in 1:Dend[cover[t]]){

#Observation process of distance sampling data (thinned point process)
#c3 is an identifier to match pixel d to pixel b
x[d,t] ~ dpois(pi[d] * lambda[c3[d,cover[t]],t])

}#end d

#b in 1:Bend[cover[t]] represents all the 2500 m^2 in a specific 1 km^2 pixel g
for(b in 1:Bend[cover[t]]){

#Linear predictor for biological process
#c1 is an identifier to link covariates (all at 1 km^2 resolution) to 2500 m^2 pixels
#c2 is an identifier to link 2500 m^2 to 1 km^2 pixels
log(lambda[c2[b, cover[t]],t]) <- log_lambda0 +
                                  beta1 * border[c1[b, cover[t]]] +
                                  beta2 * region[c1[b, cover[t]]] +
                                  beta3 * border[c1[b, cover[t]]] *
                                  region[c1[b, cover[t]]] +
                                  beta4 * water[c1[b, cover[t]]] +
                                  beta5 * Lion[c1[b, cover[t]],t] +
                                  beta6 * NDVI[c1[b, cover[t]],t] +
                                  psi[c1[b, cover[t]]]

}#end b

#Index g refers to 1 km^2 resolution for opportunistic sampling (presence-only)
for(g in 1:G){

#Observation process of opportunistic presence-only data (thinned point process)
y[g,t] ~ dpois(p[g,t] * intensity[g,t])

#Linear predictor for observation process of opportunistic presence-only data
logit(p[g,t]) <- logit(p0) + alpha1 * w[g,t] + z[g]

#Change-of-support between 50x50m and 1x1km
#Scale is an index to indicate if a specific pixel, g, contained distance sampling data
#Gstart[g,cover[t]]:Gend[g,cover[t]] is the subset of 2500 m^2 pixels within g
#log(400) is the area offset (i.e., 400 50x50m in 1x1km)
#to change support between distance sampling and presence only data
```

```
log(intensity[g,t]) <- ifelse(scale[g,cover[t]] < 1,
                              log_lambda0 +
                              beta1 * border[g] +
                              beta2 * region[g] +
                              beta3 * border[g] * region[g] +
                              beta4 * water[g] +
                              beta5 * Lion[g,t] +
                              beta6 * NDVI[g,t] +
                              psi[g] + log(400),
                              log(sum(lambda[Gstart[g,cover[t]]:Gend[g,cover[t]], t])))

}#end g
}#end t

for(d in 1:D){

#Linear predictor for observation process of distance sampling
sigma[d] <- exp(gamma0 + gamma1 * region[c1[c3[d,7],7]])

#Half-normal distance function
pi[d] <- exp(-dst[d]*dst[d]/(2*sigma[d]*sigma[d]))

}#end d

for(g in 1:G){

#Replicate 10 has only presence-only data
#Observation process of opportunistic presence-only data (thinned point process)
y[g,10] ~ dpois(p[g,10] * intensity[g,10])

#Linear predictor for observation process of opportunistic presence-only data
logit(p[g,10]) <- logit(p0) + alpha1 * w[g,10] + z[g]

#Change-of-support between 50x50m and 1x1km
log(intensity[g,10]) <- log_lambda0 +
                        beta1*border[g] +
                        beta2*region[g] +
                        beta3*border[g]*region[g] +
                        beta4*water[g] +
                        beta5*Lion[g,10] +
                        beta6*NDVI[g,10] +
                        psi[g] + log(400)

#-----------------#
#-DERIVED VALUES-#
#-----------------#

#Estimated density
Density[g] <- mean(intensity[g,])

#--------#
#-PRIORS-#
#--------#
```

```
#Random effects of pixel
psi[g] ~ dnorm(0, tau) #Biological process
z[g] ~ dnorm(0, tauz) #Presence only observation process

}#end g

#Biological process priors
log_lambda0 ~ dnorm(0, 0.1) #Intercept (sampled on log-scale)
beta1 ~ dnorm(0, 0.1)       #Effect of border
beta2 ~ dnorm(0, 0.1)       #Effect of distrubance region
beta3 ~ dnorm(0, 0.1)       #Interaction between border and region
beta4 ~ dnorm(0, 0.1)       #Effect of distance to water
beta5 ~ dnorm(0, 0.1)       #Effect of lion density
beta6 ~ dnorm(0, 0.1)       #Effect of NDVI
tau <- 1/(sig * sig)        #Precision of random pixel effect
sig ~ dt(0, pow(2.5,-2), 1) T(0,) #Variance of random pixel effect (half cauchy prior)

#Presence only observation process priors
p0 ~ dunif(0, 1) #Intercept
alpha1 ~ dnorm(0, 0.1) #Effect of sampling intensity
tauz <- 1/(sigz * sigz) #Precision of random pixel effect
sigz  ~ dt(0, pow(2.5,-2), 1) T(0,) #Variance of random pixel effect (half cauchy prior)

#Distance sampling observation process priors
gamma0 ~ dnorm(0, 0.1) #Intercept
gamma1 ~ dnorm(0, 0.1) #Effect of disturbance regime (grass height)

}
```

Section S6. Model results

Table S1. Mean, standard deviation, and 95% credible intervals for integrated model parameters.

	Mean	Standard Deviation	2.5% CI	97.5% CI
$log(\lambda_0)$	-9.23	0.27	-9.78	-8.75
β_1	0.84	0.18	0.50	1.19
β_2	1.75	0.33	1.17	2.46
β_3	-1.26	0.23	-1.70	-0.82
β_4	0.26	0.13	0.01	0.51
β_5	-0.04	0.04	-0.12	0.02
β_6	-0.09	0.03	-0.15	-0.02
τ	0.58	0.09	0.42	0.78
p_0	0.48	0.12	0.25	0.71
α_1	21.63	1.88	18.17	25.47
τ_z	0.07	0.02	0.04	0.11
γ_0	5.36	0.05	5.27	5.46
γ_1	-0.05	0.08	-0.20	0.09

Literature Cited

Didan, K. 2015. MOD13A2 MODIS/Terra Vegetation Indices 16-Day L3 Global 1km SIN Grid V006 [Data

set]. NASA EOSDIS Land Processes DAAC. DOI:10.5067/MODIS/MOD13A2.006.

Farr, M.T., Green, D.S., Holekamp, K.E., Roloff, G.J., & Zipkin E.F. 2019. Multispecies hierarchical modeling reveals variable responses of African carnivores to management alternatives. Ecological Applications, 29, e01845.

Green, D.S., Johnson-Ulrich, L., Couraud, H.E. & Holekamp, K.E. (2018) Anthropogenic disturbance induces opposing population trends in spotted hyenas and African lions. Biodiversity and Conservation, 27, 871–889.

Green, D.S., Zipkin, E.F., Incorvaia, D.C., & Holekamp, K.E. (2019) Long-term ecological changes influence herbivore diversity and abundance inside a protected area in the Mara-Serengeti ecosystem. Global Ecology and Conservation, 20, e00697.

Holekamp, K.E., Smith, J.E., Strelioff, C.C., Van Horn, R.C., & Watts, H.E. 2012. Society, demography and genetic structure in the spotted hyena. Molecular Ecology, 21, 613-632.