

# Model Formulation: Habitat-mediated effects of lion management on mammalian species richness across protected areas in South Africa

## The model

The modelling we used is divided in three components.

### Sub-model 1 - species richness ( $\Omega$ )

The first model component describes variation in species richness ( $N_r$ ) which is a function  $\Omega_r$ , the data augmentation parameter that defines the proportion of the zero-augmented super population ( $M_r$ ) present in region  $r$ . Inference therefore focusses on modelled variation in  $\Omega_r$  using a logit-linear model with region specific covariates and the number of species is a derived parameter:

$$\text{logit}(\Omega_r) = \beta_0^{(\Omega)} + \sum_{k=1}^{K^{(\Omega)}} \beta_k^{(\Omega)} X_{rk}$$

and

$$N_r = M_r \Omega_r.$$

Here, we consider  $K^{(\Omega)} = 5$  region-specific covariates: the log of reserve size (*Area*), human global modification index measured as the landscape average in a 5km buffer around the reserve (*HGM*), the average enhanced vegetation index measured across the reserve (*EVI*), a binary variable indicating whether lions (*Lion*) are present on the reserve (We coded the presence of lions using 1 when not detected and 0 when detected, to easily handle having to zero out the lion effect for lions), and *Landscape type* (savanna vs no-savanna or shrub-like vegetation). Because all the reserves are expected to share a common and known *expected* species pool, we set the super population ( $M$ ) to be that total species pool which consisted of “73 species that could potentially occur in each region. Part of the inference objective in multi-species modelling is to estimate the latent inclusion variable  $\omega_{si}$  that is 1 when species  $s$  is present in region  $r$  and 0 otherwise. The richness and species inclusion models are linked as follows:

$$\omega_{si} \sim \text{Bernoulli}(\Omega_i).$$

### Sub-model 2 - occupancy ( $\psi$ )

The next sub-model is for the latent species-within-region site occupancy state  $z_{isr}$ . Here  $z_{isr}$  is assumed to be a Bernoulli random variable with a corresponding occupancy probability  $\psi_{isr}$  which is conditional on being present in the region:

$$z_{irs} \sim \text{Bernoulli}(\psi_{isr}\omega_{si})$$

Here,  $\psi_{isr}$  is the probability that species  $s$  is present at camera trapping site  $i$  in region  $r$  and  $\omega_{sr}$  is the binary variable indicating whether the species is present in the region. We were interested in exploring factors that explain variation in site-level occupancy and do so using a species-specific logit-linear model with site-specific covariates.

$$\text{logit}(\psi_{isr}) = \beta_{0sr}^{(\psi)} + \sum_{k=1}^{K^{(\psi)}} \beta_{sk}^{(\psi)} X_{irk}.$$

The parameter  $\beta_{0sr}^{(\psi)}$  is the species- and region-specific intercept (see below), and  $\beta_{sk}^{(\psi)}$  is the species specific slope describing the relationship between occupancy and the  $k^{th}$  site-level covariate, noting that this is assumed constant across regions. For occupancy, we consider  $K^{(\psi)} = 5$  site-level covariates consistent of the same variables used in the richness submodel (except for Area) but calculated at the site level (i.e. Evi, Lion presence, Landscape type, GHM), GHM was maintained at the reserve level, and we also added elevation (Elev) to this submodel.

Under this adapted version of the multi-region model, where each reserve shares a common pool of potentially occurring species, species can occur in any number of reserves and we retain species identity across regions. We assume that the species-specific response to covariates is constant across regions, but accommodate within-species between-region variation using a region-level random effect as follows:

$$\beta_{0sr}^{(\psi)} \sim \text{Normal}(\beta_{0s}^{(\psi)}, \sigma_s^{(\psi)}),$$

where  $\beta_{0sr}^{(\psi)}$  are random intercepts drawn from a species-specific random effects distribution with mean  $\beta_{0s}^{(\psi)}$  and variance  $\sigma_s^{(\psi)}$ .

### Sub-model 3 - detection ( $p$ )

Finally, we have the model that accounts for imperfect detection. The data,  $y_{isr}$ , are the number of days species  $s$  was observed at camera  $i$  in reserve  $r$  out of a total of  $j_{ir}$  days that the camera was operational. These data are naturally modelled as binomial counts with detection probability  $p_{isr}$  that is conditional on the occupancy state such that detection is  $p_{isr}z_{isr} = p_{isr}$  when the species is present and  $p_{isr}z_{isr} = 0$  if not:

$$y_{isr} \sim \text{Binomial}(j_{ir}, p_{isr}z_{isr}).$$

Here,  $p_{isr}$  is the probability that species  $s$  is detected at camera trapping site  $i$  in region  $r$  given it is present, and  $z_{isr}$  is occupancy state of the species at that site. As before, we are interested in exploring factors that explain variation in site-level detectability and again do so using a species-specific logit-linear model with site-specific covariates.

$$\text{logit}(p_{isr}) = \beta_{0sr}^{(p)} + \sum_{k=1}^{K^{(p)}} \beta_{sk}^{(p)} X_{irk}.$$

The parameter  $\beta_{0sr}^{(p)}$  is the species- and region-specific intercept and  $\beta_{sk}^{(p)}$  is the species-specific slope describing the relationship between detection and the  $k^{th}$  site-level covariate, again assuming a constant effect across regions. For detection, we consider a single site-level covariate which is whether the camera had a white flash,  $X = 1$ , or not,  $X = 0$  (*Flash*).

Again we allow for within-species between-region variation using a region-level random effect as follows:

$$\beta_{0sr}^{(p)} \sim Normal(\beta_{0s}^{(p)}, \sigma_s^{(p)}),$$

where  $\beta_{0sr}^{(p)}$  are random intercepts drawn from a species-specific random effects distribution with mean  $\beta_{0s}^{(p)}$  and variance  $\sigma_s^{(p)}$ .