Combining biology and statistical hierarchicies with occpuancy models to account for imperfect detection of single and multiple species with two- and three-level occpuancy models

Richard A. Erickson, 0000-0003-4649-482X, rerickson@usgs.gov   Barb A. Bennie

Laura Pederman 0000-0001-6976-4138   Kevin's student for her data?

Steven Spear   Kevin Lafferty 0000-0001-7583-4593

18 May 2022

1 Abstract

2 Introduction

- Problem of imperfect detection

  - May miss seeing a species due to low detection
  - Occupancy models allow for imperfect detection to modeled and account for zero-inflation (MacKenzie *et al.* 2017)
  - Large application in ecological theory and practices, with text books written on the topic (e.g., Royle & Dorazio 2008; MacKenzie *et al.* 2017)

- Multi-species models allow for the co-occurrence of species to be modeled.

- Species often best predictors of other species

- Lack environmental predictor variables

- Tobler *et al.* (2019) describe deriving two methods for deriving multi-species models

  * Join species distribution models based upon an extension of the regression framework used for single species models, starting with Dorazio & Royle (2005a).
  * Multi-species models that stack together single species models by explicit capturing pair-wise correlations starting with Latimer *et al.* (2009)
  * Tobler *et al.* (2019) present a framework for estimating species co-occurrence (or correlation of occupancy) through the use of latent variables.

- we expand upon the notation of Royle & Dorazio (2008) to incorporate occupancy models into a statistical hierarchical modeling framework.

  - Bridges statistical hierarchical models such as those covered in Gelman & Hill (2006) with the biological hierarchy captured described by Royle & Dorazio (2008).
  - Directly building upon Tobler *et al.* (2019) because we directly estimate correlation without latent variables
  - Expand to include not only correlated site-level occupancy, but also detection probabilities
  - Show how model may be extended to three-level models that account for sub-sampling, following Dorazio & Erickson (2018) and their style application to eDNA.

- Motivating

  - 2-level, motivation for Correlations among detection of parasites when detection probs correlated across hosts
  - 3-level motivated by eDNA monitoring program, where we have site-revisits and care about sampling probabilities changing through time for single species
  - 3-level community analysis for metabarcoding
  - require scaleable computing, recently possibly with advances in the Stan language that allow for within chain parallel computing for HMC.

- purpose and overview of this paper

  - present models
  - numeric implementation

3 Models

3.1 Basic 2-level model

- Dorazio–Royle multispecies occupancy model (Dorazio & Royle 2005b),

- using similar notation as (Dorazio & Erickson 2018) because we generalize the model to include eDNA.

- Hierarchical model based upon Gelman & Hill (2006), described in Stan Development Team (2022) §1.13 and similar to the implementation in the `fishStan` package (Erickson *et al.* 2020; Erickson *et al.* 2022)

- Start by presenting basic models, without formal indexing until we get the final model.

- Observed unit is occupied ($Z = 1$) or not occupied ($Z = 0$)

- Probability of a unit occupied is ψ

- **Is species at site?**

$$Z \sim \text{Bernoulli}(\psi) \tag{1}$$

- **Was species detected during visit?**
- Y is detection (1) or non-detection (0)
- the lower case z is the realized value for Z
- k is number of repeated samples per site
- p is probability of detection conditional that the species was located at the site (i.e., $z = 1$).

$$Y|z \sim \text{Binomail}(k, z \times p) \tag{2}$$

- This model may be expanded to include regression coefficients.

- We use logit scale, other common choice is probit scale (e.g., Dorazio & Erickson 2018).

- Both similar, logit has slightly wider tails to the distribution (Finney 1952), and works more efficiently with Stan due to build in and optimized functions.

$$\text{logit}(\psi) = \mu_\psi \tag{3}$$

$$\text{logit}(p) = \mu_p \tag{4}$$

- Can regressors with μs

- Predictor matrices X and V as well a regression coefficients β and δ, that can include an error term:

$$\mu_\psi \sim X\beta \tag{5}$$

$$\mu_p \sim V\delta \tag{6}$$

Using $\mu_\psi \sim X\beta$ as an example, imagine that we fit an intercept for each specie's detection probability at a site. For example, the vector β looks like: $\beta_{\text{spp 1}}, \beta_{\text{spp 2}}, \ldots \beta_{\text{spp N}}$. We can then fit a regression to each site i, for example, let's say we have N sites:

$$\mu_{\psi_1} \sim X_1\beta_1 \tag{7}$$

$$\mu_{\psi_2} \sim X_2\beta_2 \tag{8}$$

$$\mu_{\psi_3} \sim X_3\beta_3 \tag{9}$$

$$\vdots \mu_{\psi_N} \sim X_N\beta_N \tag{10}$$

- We can then model the regression coefficients

- We can assume both β and δ come from a multivariate normal distribution:

- Hierarchical model, adapting notation of Stan Development Team (2022) to use use a star symbol ($\star$) for the second level hierarchy avoid confusion with too many parameters (i.e., the biology hierarchy crossed with the statistical hierarchy).

- Σ are correlated error terms

4

- Because each term coefficient is specices intercept, the correlation between these is the species' correlation across sites.

$$\beta \sim \text{multivariate normal}(X^\star \beta^\star, \Sigma_\psi) \tag{11}$$

$$\delta \sim \text{multivariate normal}(V^\star \delta^\star, \Sigma_p) \tag{12}$$

- We present two models for both computational efficiency and statistcial identifiability, specifcially one model with Σ_p and one model without Σ_p.
- The and correlation matrices Ω_ψ and Ω_p are proportional to the covariance matrices, Σ_ψ and Σ_p:

$$\Omega_{\psi_{i,j}} \propto \Sigma_{\psi_{i,j}} \text{ and} \tag{13}$$

$$\Omega_{p_{i,j}} \propto \Sigma_{p_{i,j}}. \tag{14}$$

3.2 Formal 2-level definition and notations

We base our formal nation upon Dorazio & Erickson (2018) for the occupancy model and Stan Development Team (2022) §1.13 for the hierarchical model. We also include our Stan variable names in `code format` because tracking indexing with Stan was a large challenge we faced when implementing this model. The model has units i that may a region of interest in spatial, temporal, or both where $i \in 1, 2, \ldots N_{\text{units}}$ (in Stan code, this this `n_units`). For example, multiple lakes could be visited, the same lake could be visited multiple times, or multiple lakes could be visited multiple times. The ith unit may be occupied ($Z_i = 1$) or not occupied ($Z_i = 0$) with probability ψ_i:

$$Z \sim \text{Bernoulli}(\psi). \tag{15}$$

The ith unit has k samples to the site, $k \in 1, 2, \ldots N_{\text{revisits}}$. k may be summed for each unit and then written as a vector, K for all units (`k_samples` that is the same length as the total number of observations in the data frame, $N_{\text{total}} or 'total_o bservations'). A sampling event may have a detection (Y_{i,k} = 1) or a non-detection(Y_{i,k} = 0)$. A lowercase z_i denotes the realized occupancy ($z_i = 1$) or non-occupancy ($z_i = 0$) at a

unit. The observation $Y_{i,k}$ is conditional upon z_i (denoted with the vertical bar symobol, |). Lastly, $p_{i,j}$ is the detection probability for the kth sample at the ith unit.

$$Y|z \sim \text{Binomail}(k, zp) \tag{16}$$

- This model may be expanded to include regression coefficients.
- We use logit scale, other common choice is probit scale (e.g., Dorazio & Erickson 2018).
- Both similar, logit has slightly wider tails to the distribution (Finney 1952), and works more efficiently with Stan due to build in and optimized functions.

$$\text{logit}(\psi) = \mu_\psi \tag{17}$$

$$\text{logit}(p) = \mu_p \tag{18}$$

- Can regressors with μs
- Predictor matrices X and V as well a regression coefficients β and δ, that can include an error term:

$$\mu_\psi \sim X\beta \tag{19}$$

$$\mu_p \sim V\delta \tag{20}$$

- We can assume both β and δ come from a multivariate normal distribution:
- Hierarchical model, adapting notation of Stan Development Team (2022) to use use a star symbol ($\star$) for the second level hierarchy avoid confusion with too many parameters (i.e., the biology hierarchy crossed with the statistical hierarchy).
- Σ are correlated error terms

$$\beta \sim \text{multivariate normal}(X^\star\beta^\star, \Sigma_\psi) \tag{21}$$

$$\delta \sim \text{multivariate normal}(V^\star\delta^\star, \Sigma_p) \tag{22}$$

- We present two models for both computational efficiency and statistical identifiability, specifically one model with Σ_p and one model without Σ_p.

- The and correlation matrices Ω_ψ and Ω_p are proportional to the covariance matrices, Σ_ψ and Σ_p:

$$\Omega_{\psi_{i,j}} \propto \Sigma_{\psi_{i,j}} \text{ and} \tag{23}$$

$$\Omega_{p_{i,j}} \propto \Sigma_{p_{i,j}}. \tag{24}$$

This may be extended change to the logit scale for numerical stability and includes regression coefficients.

$$\text{logit}(\psi) = \mu_\psi \tag{25}$$

$$\text{logit}(p) = \mu_p \tag{26}$$

- Sampling unit j, which can be time or repeatedly sampled through time or space for $j \in 1, 2, \ldots N_{\text{units}}$.
- indexing vector, jj that is used with loop over vectors, specifically jj_ψ and jj_p.
- Matrix of coefficients for sampling units β_ψ and δ_p
- Also, include hyper-parameters based upon syntax of Stan Development Team (2022) §1.13

$$\mu_{\psi_{jj[n]}} \sim X\beta_{jj[n]} \tag{27}$$

$$\mu_{p_{jj[n]}} \sim V\delta_{jj[n]} \tag{28}$$

The coefficients then have their own hierarchy of modeling:

$$\beta_j \sim \text{multivariate normal}(\beta^\star, \Sigma_\psi) \tag{29}$$

$$\delta_j \sim \text{multivariate normal}(\delta^\star, \Sigma_p) \tag{30}$$

$$\tag{31}$$

The covariance matrices, Σ_ψ and Σ_p are defined in terms of coefficient scales τ_ψ and τ_p and correlation matrices Ω_ψ and Ω_p. The coefficient scales are defined as $\tau_\psi = \sqrt{\Sigma_{\psi_{k,k}}}$ and $\tau_p = \sqrt{\Sigma_{p_{k,k}}}$. The correlation

matrices are defined as

$$\Omega_{\psi_{i,j}} = \frac{\Sigma_{\psi_{i,j}}}{\tau_{\psi_i}\tau_{\psi_j}} \text{ and} \tag{32}$$

$$\Omega_{p_{i,j}} = \frac{\Sigma_{p_{i,j}}}{\tau_{p_i}\tau_{p_j}}. \tag{33}$$

Both $\beta^\star$ and δ^2 are given weekly-informative priors:

$$\beta^\star \sim \text{normal}(0, 2) \text{ and} \tag{34}$$

$$\delta^\star \sim \text{normal}(0, 2). \tag{35}$$

114 Likewise, the τ parameters are given a weakly informative prior from the half-Cauchy distribution:

$$\tau_\psi \sim \text{Cauchy}(0, 2.5) \text{ contrained by } \tau_\psi > 0 \text{ and} \tag{36}$$

$$\tau_p \sim \text{Cauchy}(0, 2.5) \text{ contrained by } \tau_p > 0. \tag{37}$$

115 We used Lewandowski, Kurowick, and Joe (LKJ) priors for the for the correlation matrices as defined by
116 Lewandowski *et al.* (2009) with $\nu_p si > 1$ and $\nu_p > 1$

$$\Sigma_\psi \sim \text{LKJCrr}(\nu_\psi) \text{ and} \tag{38}$$

$$\Sigma_p \sim \text{LKJCrr}(\nu_p) \tag{39}$$

117 Notes about indexing

118 • The two-levels become confusing, often the same by coincidence.

119 • Especially for ψ-level parameters and $p^\star$-level parameters

120 • Important to think about groupings, often have problems when we were coding

3.3 Formal 3-level model

122 • Includes sub-sampling Mordecai *et al.* (2011)

- Based upon Dorazio & Erickson (2018) for syntax

- Generically, (1) is site occupied? (2) Is species present for site? (3) Did sub-sampling detect the species?

- Within context of eDNA, Three levels (see figure reprinted from Erickson *et al.* (2019)).

- For eDNA (1) Is eDNA at the site? (2) Is eDNA in sample and successfully exacted? and (3) Did molecular tool such as qPCR detect eDNA?

- We insert θ for middle-level.

Like previous section, we base our formal nation upon Dorazio & Erickson (2018) for the occupancy model and Stan Development Team (2022) §1.13 for the hierarchical model. We also include our Stan variable names in `code format` because tracking indexing with Stan was a large challenge we faced when implementing this model. The model has units i that may a region of interest in spatial, temporal, or both where $i \in 1, 2, \ldots N_{\text{units}}$ (in Stan code, this this `n_units`). For example, multiple lakes could be visited, the same lake could be visited multiple times, or multiple lakes could be visited multiple times. The ith unit may be occupied ($Z_i = 1$) or not occupied ($Z_i = 0$) with probability ψ_i:

$$Z_i \sim \text{Bernoulli}(\psi_i). \tag{40}$$

The ith unit has j_i samples taken from the site, $j_i \in 1, 2, \ldots N_{\text{revisits: } i}$ (`n_samples[unit index]`). A lowercase z_i denotes the realized occupancy ($z_i = 1$) or non-occupancy ($z_i = 0$) at a unit (`any_seen[unit index]`).

The latent, sample occurrence $a_{i,j}$ is conditional upon z_i and (denoted with the vertical bar symbol, |). Lastly, $\theta_{i,j}$ is the sample probability for the jth sample at the ith unit.

$$A_{i,j}|z_i \sim \text{Bernoulli}(z_i\theta) \tag{41}$$

A lowercase $a_{i,j}$ denotes the realized occupancy ($a_{i,j} = 1$) or non-occupancy ($a_{i,j} = 0$) at a sample (`sample_seen[unit index]`) Within each sample, there are $k_{i,j}$ sub-samples within the unit, $k_{i,j} \in 1, 2, \ldots N_{\text{subsamples: } i, j}$ (`k_samples`). Each unit may have its own revisits and each revisit to a unit may have its own number of subsamples. Hence, there are can be subscripted subscripts.

9

k may be summed for each unit and then written as a vector, K for all units (`k_samples` that is the same length as the total number of observations in the data frame, N_{total} or `total_observations`). A sampling event may have a detection ($Y\_i, j, k = 1$) or a non-detection ($Y_{i,j,k} = 0$).

The observation $Y_{i,j,k}$ is conditional upon $a_{i,j}$ (denoted with the vertical bar symbol, |). Lastly, $p_{i,j,k}$ is the detection probability for the kth sub-sample in the jth sample at the ith unit.

$$Y_{i,j,k}|a_{i,j} \sim \text{Binomail}(k_{j,k}, a_{i,j}p_{i,j,k}) \tag{42}$$

- This model may be expanded to include regression coefficients.
- We use logit scale, other common choice is probit scale (e.g., Dorazio & Erickson 2018).
- Both similar, logit has slightly wider tails to the distribution (Finney 1952), and works more efficiently with Stan due to build in and optimized functions.

$$\text{logit}(\psi) = \mu_\psi \tag{43}$$

$$\text{logit}(\theta) = \mu_\theta \tag{44}$$

$$\text{logit}(p) = \mu_p \tag{45}$$

- Can regressors with μs
- Predictor matrices X, W, and V as well a regression coefficients β, α, and δ, that can include an error term:

$$\mu_\psi \sim X\beta \tag{46}$$

$$\mu_\theta \sim W\alpha \tag{47}$$

$$\mu_p \sim V\delta \tag{48}$$

- We can assume β, α, and δ come from a multivariate normal distribution:
- Hierarchical model, adapting notation of Stan Development Team (2022) to use use a star symbol ($\star$) for the second level hierachy avoid confusion with too many parameters (i.e., the biology hierarchy

161   crossed with the statistical hierarchy).

162   • Σ are correlated error terms

$$\beta \sim \text{multivariate normal}(X^\star \beta^\star, \Sigma_\psi) \tag{49}$$

$$\delta \sim \text{multivariate normal}(V^\star \delta^\star, \Sigma_p) \tag{50}$$

163   • We present two models for both computational efficiency and statistical identifiability, specifically one

164   model with Σ_p and one model without Σ_p.

165   • The and correlation matrices Ω_ψ and Ω_p are proportional to the covariance matrices, Σ_ψ and Σ_p:

$$\Omega_{\psi_{i,j}} \propto \Sigma_{\psi_{i,j}} \text{ and} \tag{51}$$

$$\Omega_{p_{i,j}} \propto \Sigma_{p_{i,j}}. \tag{52}$$

166 This may be extended change to the logit scale for numerical stability and includes regression coefficients.

$$\text{logit}(\psi) = \mu_\psi \tag{53}$$

$$\text{logit}(\alpha) = \mu_\alpha \tag{54}$$

$$\text{logit}(p) = \mu_p \tag{55}$$

167   • Sampling unit j, which can be time or repeatedly sampled through time or space for $j \in 1, 2, \ldots N_{\text{units}}$.

168   • indexing vector, jj that is used with loop over vectors, specifically jj_ψ, jj_α, and jj_p.

169   • Matrix of coefficients for sampling units β_ψ and δ_p

170   • Also, include hyper-parameters based upon syntax of Stan Development Team (2022) §1.13

$$\mu_{\psi_{jj[n]}} \sim X\beta_{jj[n]} \tag{56}$$

$$\mu_{\alpha_{jj[n]}} \sim X\alpha jj[n] \tag{57}$$

$$\mu_{p_{jj[n]}} \sim V\delta_{jj[n]} \tag{58}$$

The coefficients then have their own hierarchy of modeling:

$$\beta_j \sim \text{multivariate normal}(\beta^{\star}, \Sigma_{\psi}) \tag{59}$$

$$\alpha j \sim \text{multivariate normal}(\alpha^{\star}, \Sigma_{\alpha}) \tag{60}$$

$$\delta_j \sim \text{multivariate normal}(\delta^{\star}, \Sigma_p) \tag{61}$$

The covariance matrices, Σ_{ψ}, Σ_{θ}, and Σ_p are defined in terms of coefficient scales τ_{ψ}, τ_{θ}, and τ_p and correlation matrices Ω_{ψ}, Ω_{θ}, and Ω_p. The coefficient scales are defined as $\tau_{\psi} = \sqrt{\Sigma_{\psi_{k,k}}}$, $\tau_{\theta} = \sqrt{\Sigma_{\theta k,k}}$, and $\tau_p = \sqrt{\Sigma_{p_{k,k}}}$. The correlation matrices are defined as

$$\Omega_{\psi_{i,j}} = \frac{\Sigma_{\psi_{i,j}}}{\tau_{\psi_i}\tau_{\psi_j}}, \tag{62}$$

$$\Omega_{\theta i,j} = \frac{\Sigma_{\theta_{i,j}}}{\tau_{\theta_i}\tau_{\theta_i}}, \text{ and} \tag{63}$$

$$\Omega_{p_{i,j}} = \frac{\Sigma_{p_{i,j}}}{\tau_{p_i}\tau_{p_j}}. \tag{64}$$

Both $\beta^{\star}$, $\alpha^{\star}$ and $\delta^s tar$ are given weekly-informative priors:

$$\beta^{\star} \sim \text{normal}(0, 2), \tag{65}$$

$$\alpha^{\star} \sim \text{normal}(0, 2), \text{ and} \tag{66}$$

$$\delta^{\star} \sim \text{normal}(0, 2). \tag{67}$$

171 Likewise, the τ parameters are given a weakly informative prior from the half-Cauchy distribution:

$$\tau_\psi \sim \text{Cauchy}(0, 2.5) \text{ contrained by } \tau_\psi > 0, \tag{68}$$

$$\tau_\theta \sim \text{Cauchy}(0, 2.5) \text{ contrained by } \theta\psi > 0, \text{ and} \tag{69}$$

$$\tau_p \sim \text{Cauchy}(0, 2.5) \text{ contrained by } \tau_p > 0. \tag{70}$$

172 We used Lewandowski, Kurowick, and Joe (LKJ) priors for the for the correlation matrices as defined by
173 Lewandowski *et al.* (2009) with $\nu_p si > 1$, $\nu_t heta > 1$, and $\nu_p > 1$

$$\Sigma_\psi \sim \text{LKJCrr}(\nu_\psi), \tag{71}$$

$$\Sigma_\theta \sim \text{LKJCrr}(\nu_\theta), \text{ and} \tag{72}$$

$$\Sigma_p \sim \text{LKJCrr}(\nu_p) \tag{73}$$

174 Notes about indexing

175 • The three-levels become confusing, second statistical-levels can often the same by coincidence.

176 • Especially for ψ-level parameters and $\theta^\star$ $p^\star$-level parameters

177 • Important to think about groupings, often have problems when we were coding

3.4 Numerical implementation

179 • Required development version of RStan as of writing, thus we used `rcmdstan`, specifically because
180 of the `reduce_sum()` function did not appear in Stan until version 2.23 and RStan version 2.21 was
181 the stable version of RStan. Likewise, `rcmdstan` allowed us to use bothin within and amoung chain
182 parallelizaiton.

183 • Onion method (Lewandowski *et al.* 2009) because default Stan option LKJ prior method does not
184 scale to large matrices

185 • Within chain parallel requiring more recent versions of Stan than currently `RStan` version 2.21 (Stan
186 Development Team 2021)

187 • use the `reduce_sum()` function, which allows parallel chains to calculate log probability distributions

188 • used Stan version 2.29 (Stan Development Team 2022), which we called through Gabry & Cesnovar
189 (2022)

• Tested using 40 cores on a local server, for test cases 10 cores per thread worked best for our models

4 Example applications of model

4.1 2-level simulated data

```
knitr::include_graphics("sim_example_figures/known_spp_matrix.jpg")

knitr::include_graphics("sim_example_figures/omega_psi_plot.jpg")
```
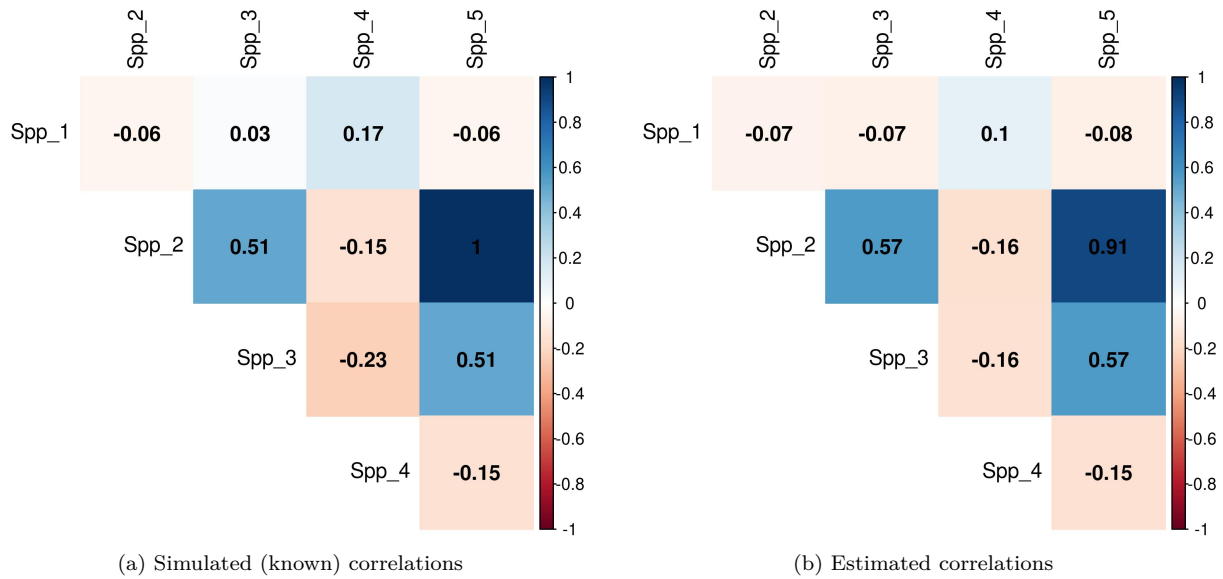

(a) Simulated (known) correlations

(b) Estimated correlations

Figure 1: Species-level occpuancy correlations .

```
knitr::include_graphics("sim_example_figures/known_detection_matrix.jpg")

knitr::include_graphics("sim_example_figures/omega_p_plot.jpg")
```
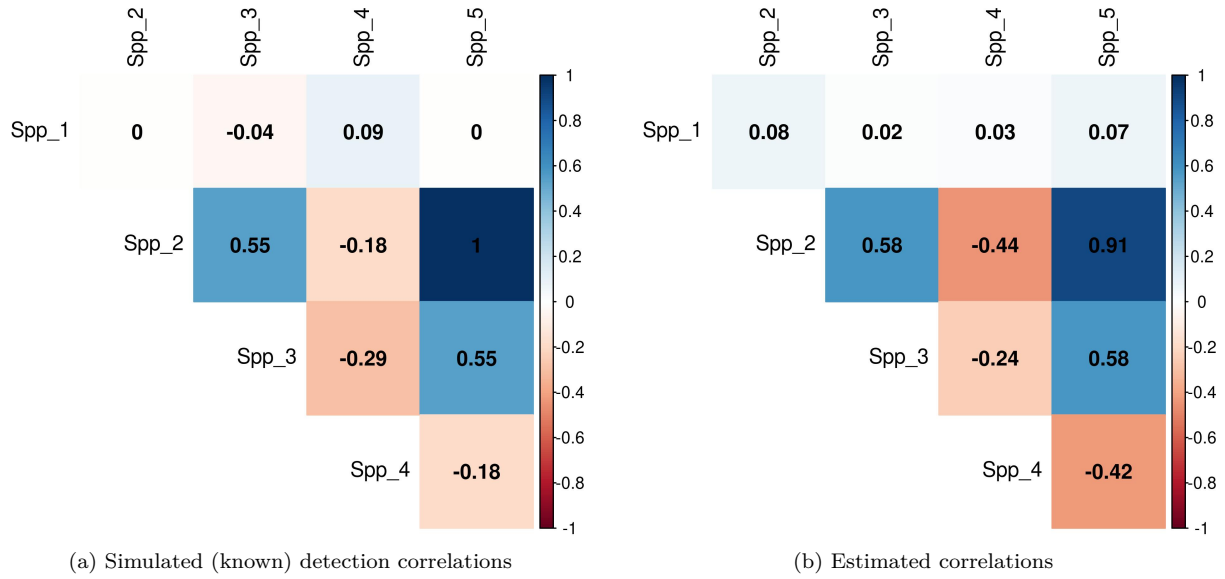
(a) Simulated (known) detection correlations

(b) Estimated correlations

Figure 2: Detection probability correlations .

4.2   2-level bird example

4.3   2-level parasite model

4.4   3-level simulated model

4.5   3-level repeated visits

4.6   3-level eDNA example

5   Discussion

1. Current application

2. Next steps

3. Future research questions

4. Implications for broader ecological literature.

6 Acknowledgments

References

Dorazio, R.M. & Erickson, R.A. (2018). ednaoccupancy: An R package for multiscale occupancy modelling of environmental DNA data. *Molecular Ecology Resources*, **18**, 368–380. Retrieved from https://doi.org/ 10.1111/1755-0998.12735

Dorazio, R.M. & Royle, J.A. (2005a). Estimating size and composition of biological communities by modeling the occurrence of species. *Journal of the American Statistical Association*, **100**, 389–398.

Dorazio, R.M. & Royle, J.A. (2005b). Estimating size and composition of biological communities by modeling the occurrence of species. *Journal of the American Statistical Association*, **100**, 389–398. Retrieved from https://doi.org/10.1198/016214505000000015

Erickson, R.A., Merkes, C.M. & Mize, E.L. (2019). Sampling designs for landscape-level eDNA monitoring programs. *Integrated Environmental Assessment and Management*, **15**, 760–771.

Erickson, R.A., Stich, D.S. & Hebert, J.L. (2022). fishStan: Hierarchical bayesian models for fisheries. *Journal of Open Source Software*, **7**, 3444.

Erickson, R.A., Stich, D.S. & Hebert, J.L. (2020). *fishStan: Hierarchical bayesian models for fisheries version 2.0*. U.S. Geological Survey software release, Reston, VA, https://doi.org/10.5066/P9TT3ILO.

Finney, D.J. (1952). *Probit analysis: A statistical treatment of the sigmoid response curve*. Cambridge university press, Cambridge.

Gabry, J. & Cesnovar, R. (2022). *Cmdstanr: R interface to 'CmdStan'*.

Gelman, A. & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

Latimer, A., Banerjee, S., Sang Jr, H., Mosher, E. & Silander Jr, J. (2009). Hierarchical models facilitate spatial analysis of large data sets: A case study on invasive plant species in the northeastern united states. *Ecology letters*, **12**, 144–154.

Lewandowski, D., Kurowicka, D. & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, **100**, 1989–2001. Retrieved from https://doi.org/10.1016/j.jmva.2009.04.008

MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L. & Hines, J.E. (2017). *Occupancy estimation and modeling: Inferring patterns and dynamics of species occurrence*. Elsevier.

Mordecai, R.S., Mattsson, B.J., Tzilkowski, C.J. & Cooper, R.J. (2011). Addressing challenges when studying mobile or episodic species: Hierarchical bayes estimation of occupancy and use. *Journal of Applied Ecology*, **48**, 56–66. Retrieved from https://doi.org/10.1111/j.1365-2664.2010.01921.x

Royle, J.A. & Dorazio, R.M. (2008). *Hierarchical modeling and inference in ecology: The analysis of data from populations, metapopulations and communities*. Elsevier.

Stan Development Team. (2021). RStan: The R interface to Stan. Retrieved from https://mc-stan.org/

Stan Development Team. (2022). *Stan Modeling Language Users Guide and Reference Manual, version 2.29*. Retrieved from https://mc-stan.org

Tobler, M.W., K'ery, M., Hui, F.K.C., Guillera-Arroita, G., Knaus, P. & Sattler, T. (2019). Joint species distribution models with species correlations and imperfect detection. *Ecology*, **100**. Retrieved from https://doi.org/10.1002/ecy.2754