



Modeling and estimating co-occurrence between the invasive Shiny Cowbird and its Puerto Rican hosts

Philip T. Patton · Krishna Pacifici ·
Jaime A. Collazo

Received: 13 November 2021 / Accepted: 6 May 2022 / Published online: 31 May 2022
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2022

Abstract Invasive species threaten island biodiversity globally. For example, the Shiny Cowbird (*Molothrus bonariensis*) parasitizes many of Puerto Rico's endemic species, particularly in the open forests in the island's southwest. Less is known, however, about cowbird parasitism in the agro-ecological highlands, which contain a patchwork of forests, shaded-coffee plantations, and coffee farms without shade. In this paper, we estimated co-occurrence rates, a potential indicator of parasitism rates, between the cowbird and four host species across these three land uses, hypothesizing that cowbirds would most likely co-occur with their hosts in shaded-coffee farms. We also hypothesized that the presence of host species would increase

the probability of cowbird occurrence. To investigate these hypotheses, we developed three Bayesian hierarchical occupancy models: one where the hosts and parasite occurred independently, one that used the latent host species richness as a predictor of cowbird occurrence, and one that used each latent host occurrence state as predictors. These methods addressed observation errors and appropriately propagated error to our predictions of co-occurrence rates. We selected the best performing model using WAIC, then used it to predict co-occurrence rates. While there was some evidence that host species richness increased the probability of cowbirds, the parsimonious model assumed no interaction. With this model, we found that cowbirds were more likely to overlap with certain hosts in shaded-coffee plantations. This may suggest increased parasitism at these plantations, potentially presenting challenges for managers who advocate for shade restoration to gain ecological services such as biodiversity conservation.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10530-022-02825-3>.

P. T. Patton · K. Pacifici
Program in Fisheries, Wildlife, and Conservation Biology,
Department of Forestry and Environmental Resources,
North Carolina State University, Raleigh, NC 27695, USA

P. T. Patton (✉)
Hawai'i Institute of Marine Biology, 46-007 Lilipuna Rd,
Kaneohe, HI 96744, USA
e-mail: pattonp@hawaii.edu

J. A. Collazo
US Geological Survey, North Carolina Cooperative Fish
and Wildlife Research Unit, Department of Applied
Ecology, North Carolina State University, Raleigh,
NC 27695, USA

Keywords Island biodiversity · Shade coffee ·
Bayesian occupancy model

Introduction

The isolation that makes islands so biologically interesting and unique also makes them particularly vulnerable to invasive species. Island invaders have contributed to 86% of insular extinctions, which in turn

represent 75% of all global extinctions since 1500 (Doherty et al. 2016; Holmes et al. 2019). Puerto Rico alone is home to approximately 180 introduced plant and animal species (Lugo et al. 2012), including the Black Rat (*Rattus rattus*), the Small Indian Mongoose (*Herpestes auro-punctatus*), and the Green Iguana (*Iguana iguana*). These omnivores and carnivores threaten the islands 18 endemic birds species through direct predation and nest predation.

In addition, the island's endemic birds face brood parasitism by the Shiny Cowbird (*Molothrus bonariensis*), which established itself in Puerto Rico sometime in the mid-20th century (Post and Wiley 1977). Since then, the cowbird proliferated throughout the island, parasitizing a multitude of host species (Wiley 1988). Unlike hosts in the cowbird's native range, the Puerto Rican bird community does not have natural defense mechanisms against cowbird parasitism. As a result, cowbird parasitism especially threatens these native birds. For example, studies in the lowlands of Puerto Rico indicate that parasitism depresses the reproductive rates of Puerto Rican Vireos (*Vireo latimeri*; Woodworth 1997; Irizarry et al. 2016). In fact, Woodworth (1999) found that cowbirds were the only one of three invasives whose presence causes population declines in vireos (the other species in the study were mongooses and iguanas). In addition, cowbirds are largely responsible for declines in lowland populations of the endangered Yellow-shouldered Blackbird (Cruz et al. 2005).

All of the aforementioned studies have taken place in the semi-arid forests in the lowlands of Puerto Rico, and largely focused on Puerto Rican Vireos and Yellow-shouldered Blackbirds. Less is known about these dynamics in the agricultural highlands of Puerto Rico, where the cowbird may co-occur with many endemic bird species. Indeed, the cowbird parasitizes multiple endemic bird species on the island (Wiley 1988), and it seems reasonable to assume that parasitism would affect these host populations similarly. Unfortunately, it would be incredibly costly to estimate cowbird nest parasitism rates on each of its preferred host species in a small area, much less at a landscape scale.

Instead, we propose studying co-occurrence rates as an indicator of nest parasitism at a broad scale. That these two rates would be perfectly correlated is a strong assumption. However, there are reasons to assume that they are correlated, perhaps strongly,

during the breeding season. First, by definition, cowbirds must be near a host's nest to parasitize it. Second, cowbirds exhibit clear preferences when choosing hosts and parasitize their preferred hosts at high rates where they co-occur (Wiley 1988; Cruz et al. 2005; Woodworth 1997). Finally, cowbird species tend to linger around their hosts' nests well after the parasitizing event to ensure that the hosts tend to their young (Smith et al. 2000; Hoover and Robinson 2007). Thus, we believe that co-occurrence rates could be a reasonable facsimile for nest parasitism rates at a landscape scale, i.e., the scale of the agricultural highlands.

Ecologists and land managers have noted Puerto Rico's agricultural highlands as an important region for conservation (Gould et al. 2008; Gladkikh et al. 2020). Many endemic species can be found in the region, which primarily consists of forests, coffee planted as row crops (sun coffee), and coffee farms overshadowed by large trees (shade coffee). Over the last 20 years, some farmers here have moved away from sun coffee, which is more agriculturally acute, for the benefits of shade coffee (Irizarry et al. 2018). The tree species planted in shade coffee farms reduce soil erosion for farmers (Beer et al. 1997), provide feeding and breeding habitat for birds (Irizarry et al. 2018), and retain carbon for the climate (Blaser et al. 2018). Thus, shade coffee has potential as the rare land use that can equally benefit communities and conservation (Blaser et al. 2018).

However, we hypothesized that these benefits may be mitigated by the invasion of the Shiny Cowbird. Specifically, we hypothesized that cowbirds would be more likely to co-occur with their preferred hosts in shade coffee because it resembles, and is intermediate between, the cowbirds preferred feeding habitat (grasslands) and the hosts' preferred breeding habitat (forests). Further, shade coffee plantations share some characteristics with edge habitats that cowbirds often use to find hosts (Smith et al. 2000), as well as the lowland dry forests in Puerto Rico (Woodworth 1997), where the cowbirds occur and parasitize hosts at high rates.

In this paper, we estimated co-occurrence rates between the Shiny Cowbird and its preferred hosts in the highlands of Puerto Rico. We hypothesized that cowbirds would most likely co-occur with their hosts in shade coffee. In addition, we hypothesized that cowbirds would be more likely to occur at sites

with their preferred hosts regardless of habitat, either because of the number of host species at a site (i.e., host species richness), or because each host species has a unique effect on cowbird presence.

To estimate co-occurrence rates, we use Bayesian hierarchical occupancy models trained with data from standard occupancy surveys (i.e., spatially and temporally replicated presence-absence surveys). This framework allowed us to test our hypotheses because it properly incorporates uncertainty in predictions of co-occurrence and can flexibly accommodate different mechanisms for the host-parasite interaction.

Methods

Data

To address these hypotheses, we used data from standard occupancy surveys conducted in the spring of 2015. We surveyed 120 randomly selected sites, stratified by land use type (41, 37, and 42, sites in forest, sun coffee plantations, and shade coffee plantations, respectively; Fig. 1). We spaced each sample point 500 m apart to mitigate dependence among sites. At each site, we conducted community-level bird surveys from March to June in 2015. This period is the peak of the breeding season for nearly the entire bird community (Collazo and Groom 2001). We visited each site three times (although 3 sites were visited twice), with three to four weeks between surveys. Each survey was conducted by one of two teams of observers—denoted here as observers A and B. The observers recorded the detection of any species seen or heard within a 50 m radius over a 12-min period.

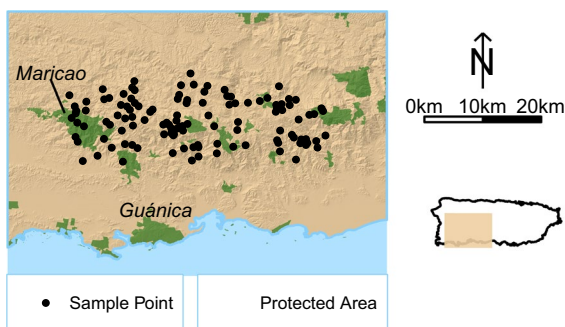


Fig. 1 Map of the study area, including sample points and overlapping protected areas

We used a 50 m radius to minimize any heterogeneity in detection probability in the sample radius. Surveys took place from 30 min before sunrise to 10 AM. To mitigate heterogeneity in detection probability due to the time of the survey, we attempted to stratify replicate surveys at a site by time of day. Further details about the study area and data can be found in (Patton 2016, pp. 14).

Modeling approach

Occupancy models (MacKenzie et al. 2002; Tyre et al. 2003) are immensely popular in ecology for their diversity of uses and return on sampling investment. Their hierarchical nature allows scientists to estimate the true probability of occurrence when individuals can go undetected during surveys (i.e., imperfect detection). The community-level variation (Dorazio and Royle 2005) expands on this approach by jointly estimating occurrence for a community of species, increasing the precision of estimates. Other variations have been used to study competition between native and invasive species (Yackulic et al. 2014), the effect of hosts on brood parasites (Péron et al. 2016), and co-occurrence between native and invasive species (Kass et al. 2020).

Estimating co-occurrence rates for interacting species pairs presents challenges. As mentioned above, the first difficulty is imperfect detection, i.e., the fact that surveyors often miss species that are truly present at a site. Second, we may need to relax the assumption that occurrence of the host species will be independent of parasite occurrence. Finally, predictions of co-occurrence must account for uncertainty in the estimates of occurrence for both species. Namely, estimates of occurrence for the parasite and host community will contain observational uncertainty (e.g., imperfect detection) and model uncertainty (i.e., imperfect description of the system).

We addressed each of these challenges in our modeling approach. Specifically, we used a Bayesian workflow to produce a full posterior predictive distribution of co-occurrence, which retained observational and model uncertainty from the host and parasite models. To address the second challenge, we developed two models with different host-parasite interaction terms, and compared their performance against a model that assumed independence. Finally, to address

imperfect detection, we used the occupancy modeling framework.

Parasite

To model occurrence for the parasite, we used a standard Bayesian static site occupancy model (MacKenzie et al. 2002; Royle and Dorazio 2008, Chapter 3). In this model, observers visit J sites K times. During each visit, the parasite is either detected, $y_{jk} = 1$, or not, $y_{jk} = 0$. The model assumes that $y_{jk} \sim \text{Bern}(f(\mathbf{u}_{jk}, \boldsymbol{\alpha}) * z_j)$, where $f(\mathbf{u}_{jk}, \boldsymbol{\alpha})$ is a function of visit specific covariates, \mathbf{u}_{jk} , that affect the detection of the parasite; $\boldsymbol{\alpha}$ are the parameters governing the effects of the covariates on detection; and z_j is a binary variable indicating whether the parasite occupies a site. Thus, the parasite is detected with probability $f(\mathbf{u}_{jk}, \boldsymbol{\alpha})$ at sites it truly occupies, and with probability zero in sites it does not occupy. A common choice of $f(\mathbf{u}_{jk}, \boldsymbol{\alpha})$, is the logit linear model, where $f(\mathbf{u}_{jk}, \boldsymbol{\alpha}) = \text{logit}^{-1}(\mathbf{u}'_{jk} \boldsymbol{\alpha})$.

The model assumes that $z_j \sim \text{Bern}(g(\mathbf{v}_j, \boldsymbol{\beta}))$, i.e., that true parasite occurrence at a site is a Bernoulli random variable with probability $g(\mathbf{v}_j, \boldsymbol{\beta})$, where $\boldsymbol{\beta}$ is a vector of coefficients representing the effect of site level covariates, \mathbf{v}_j , on the occurrence of the parasite. The full posterior distribution of the occurrence and detection parameters can be described generally as,

$$[\boldsymbol{\alpha}, \boldsymbol{\beta} | \mathbf{Y}] \propto \prod_{j=1}^J \prod_{k=1}^K \text{Bern}(y_{jk} | f(\mathbf{u}_{jk}, \boldsymbol{\alpha}) * z_j) \text{Bern}(z_j | g(\mathbf{v}_j, \boldsymbol{\beta})) [\boldsymbol{\alpha}] [\boldsymbol{\beta}], \tag{1}$$

where $[\boldsymbol{\alpha}]$ represents the prior distribution for the effects of visit-level covariates on detection, and $[\boldsymbol{\beta}]$ represents the prior distribution for the effects of site-level covariates on detection.

Host community

To estimate occurrence for the host community, we used the multispecies framework of Dorazio and Royle (2005). We assumed that there is a size N pre-defined community of host species (Pacifiçi et al. 2014). In addition, we assumed that each host species $i = 1, 2, \dots, N$ was subject to the same sampling protocol as the parasite. This model used the same

underpinnings as the parasite occurrence model (i.e., the single season occupancy model) for each of the N species. However, instead of fitting N occurrence models for each species in the host community, we fit a single model that treats each species as a random effect (Dorazio and Royle 2005). In other words, the covariate effects on detection, $\boldsymbol{\gamma}_i$, and occurrence, $\boldsymbol{\delta}_i$, for each species are derived from a common prior distribution. For example, the covariate effects on detection could come from a common multivariate normal distribution, i.e., $\boldsymbol{\gamma}_i \sim \text{Normal}(\boldsymbol{\mu}_\gamma, \boldsymbol{\Sigma}_\gamma)$, with hyperparameters $\boldsymbol{\mu}_\gamma$ and $\boldsymbol{\Sigma}_\gamma$. This formulation allows each species-level parameter to borrow strength from one another, leading to more precise estimates (Broms et al. 2016). The full posterior distribution of this model can be described as,

$$[\boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\mu}_\gamma, \boldsymbol{\Sigma}_\gamma, \boldsymbol{\mu}_\delta, \boldsymbol{\Sigma}_\delta | \mathbf{X}] \propto \prod_{i=1}^N \prod_{j=1}^J \prod_{k=1}^K \text{Bern}(x_{ijk} | h(\mathbf{o}_{jk}, \boldsymbol{\gamma}_i) * \tau_{ij}) * \text{Bern}(\tau_{ij} | \pi(\mathbf{w}_j, \boldsymbol{\delta}_i)) \text{Normal}(\boldsymbol{\gamma}_i | \boldsymbol{\mu}_\gamma, \boldsymbol{\Sigma}_\gamma) * \text{Normal}(\boldsymbol{\delta}_i | \boldsymbol{\mu}_\delta, \boldsymbol{\Sigma}_\delta) [\boldsymbol{\mu}_\gamma] [\boldsymbol{\mu}_\delta] [\boldsymbol{\Sigma}_\gamma] [\boldsymbol{\Sigma}_\delta] \tag{2}$$

where x_{ijk} is a binary variable indicating a detection of host species i at site j during visit k ; $h(\cdot)$ and $\pi(\cdot)$ are link functions; $\boldsymbol{\gamma}_i$ is a vector of parameters specifying the effect of visit specific covariates, \mathbf{o}_{jk} , on the detection of host species i ; τ_{ij} indicates the true occurrence state of host species i at site j ; $\boldsymbol{\delta}_i$ is a vector of parameters specifying the effect of site specific covariates, \mathbf{w}_j , on the occurrence of host species i ; $\boldsymbol{\mu}_\gamma$ and $\boldsymbol{\mu}_\delta$ are vectors of mean covariate effects on occurrence and detection; and $\boldsymbol{\Sigma}_\gamma$ and $\boldsymbol{\Sigma}_\delta$ are covariance matrices with the variances of covariate effects on the diagonal and zeroes on the off diagonal. The zeros represent a simplifying assumption, i.e., that there is no correlation between the covariate effects.

Modeling the host influence on parasite occurrence

We were interested in exploring the influence of host species presence or absence on the probability of occurrence for the parasite. To do so we fit three models that varied in their assumptions about this specific relationship. The first model (Model 1: Independent Model) assumes independence between the parasite and its hosts. The second (Model 2: Host Species Richness) assumes that host species richness affects parasite occurrence.

The final model (Model 3: Individual Host Effects) assumes that each host species uniquely influences parasite occurrence.

Model 1: Independent Model For the independent model, we assumed that the parasite occurs independently of the hosts. To formulate this assumption, we simultaneously fit the parasite and the host models, but did not incorporate any effect of the hosts on parasite occurrence. The full posterior distribution, written generally, is,

$$[\alpha, \beta, \gamma, \delta, \mu_\gamma, \Sigma_\gamma, \mu_\delta, \Sigma_\delta | \mathbf{Y}, \mathbf{X}] \propto \prod_{i=1}^N \prod_{j=1}^J \prod_{k=1}^K \text{Bern}(y_{jk} | f(\mathbf{u}_{jk}, \alpha) * z_j) * \text{Bern}(z_j | g(\mathbf{v}_j, \beta)) \text{Bern}(x_{ijk} | h(\mathbf{o}_{jk}, \gamma_i) * \tau_{ij}) * \text{Bern}(\tau_{ij} | \pi(\mathbf{w}_j, \delta_i)) \text{Normal}(\gamma_i | \mu_\gamma, \Sigma_\gamma) * \text{Normal}(\delta_i | \mu_\delta, \Sigma_\delta) [\alpha] [\beta] [\mu_\gamma] [\mu_\delta] [\Sigma_\gamma] [\Sigma_\delta] \tag{3}$$

Even if this model does not address our second challenge (the hosts influencing parasite occurrence), it still helps with the third challenge. That is, using the full posterior predictive distributions from both models allows us to properly propagate uncertainty in the predictions of co-occurrence.

The next two models try to address the second challenge. To be clear, the only component of Models 1, 2, and 3 that differs is $g(\cdot)$ in Eqs. 1 and 3. Aside from $g(\cdot)$, the three models are identical.

Model 2: Host Species Richness In this model, we modified $g(\cdot)$ in Eqs. 1 and 3, to incorporate a covariate, m_j , that represents the true number of host species at a site. m_j , of course, is not observed perfectly, and therefore must be modeled and estimated. To do so, we used Model 1, letting $m_j = \sum_{i=1}^N \tau_{ij}$, an unbiased estimator of host species richness (Dorazio and Royle 2005) that properly accounts for observation and process uncertainty. To incorporate this covariate into the model, we included it in the parasite occurrence function using a logit linear model, such that,

$$g(\mathbf{v}_j, m_j, \beta, \kappa) = \text{logit}^{-1}(\mathbf{v}_j \beta + m_j \kappa),$$

where κ is the effect of host species richness on parasite occurrence.

Model 3: Individual Host Effects In this model, we included the true occurrence state for every host

species, τ_j , as a covariate in the parasite occurrence model, $g(\cdot)$ using a logit linear model, such that,

$$g(\mathbf{v}_j, \tau_j, \beta, \epsilon) = \text{logit}^{-1}(\mathbf{v}_j \beta + \tau_j \epsilon), \tag{4}$$

where ϵ is a vector with N elements specifying the effects of each species on the probability of occurrence for the parasite.

One benefit of this model relative to Model 2 is that it allowed us to directly estimate the effect of each host on parasite occurrence, potentially providing a small window into the effect of host selection. That is, it allowed us to see which hosts influenced parasite occurrence and to what degree. There are a number of approaches to exploring this, including the hypothesis-testing, e.g. does the credible interval of τ_{ij} overlap zero, or the information theoretic approach, e.g., which hosts are included in the most parsimonious model. For a more general approach, i.e., one that would be useful even when N is large, we used indicator variable selection, specifically stochastic search variable selection (SVSS) (O’Hara and Sillanpää 2009).

Model fitting

Using knowledge of the system and results from previous studies (e.g., Wiley 1988), we defined the members of the host community as the Adelaide’s Warbler (ADWA), Black-whiskered Vireo (BWVI), Puerto Rican Oriole (PROR), and the Puerto Rican Vireo (PRVI; Table S1). We selected covariates that we knew affected the occurrence and detection of the host species and the cowbird, while excluding covariates found to be irrelevant by exploratory analysis (Patton 2016). For the host community, these included a site-level covariate \mathbf{w}_j indicating the land use at a site (three levels: forest, sun coffee, and shade coffee), and visit-level a covariate \mathbf{o}_{jk} indicating which set of observers conducted the survey (two levels: A and B). For the cowbird, we included a site-level covariate \mathbf{v}_j that indicated the land use at a site (two levels: forest or coffee). We only used two levels because we found that the model with all three levels overfit the data, evidenced by the credible intervals for β being markedly wider than in the two-level model. This may suggest that the model couldn’t disentangle the effect of sun vs. shade, or that there is variation in cowbird occurrence within coffee farms

that the sun/shade split could not explain. That cowbirds would occupy sun and shade equally may be ecologically plausible, if cowbirds use sun coffee for feeding habitat, and shaded coffee for breeding and feeding habitat. See *Discussion* for a discussion of gradients between coffee types, and how shade maturity might affect cowbird occurrence. Finally, we also included a visit-level covariate u_{jk} that indicated which set of observers conducted the survey.

For each model, we used semi-informative priors to enhance MCMC efficiency. $[\alpha]$ and $[\beta]$ were given Normal(0,5) priors, while $[\mu_\gamma]$ and $[\mu_\delta]$ were given Normal(0,10) priors, and the diagonal elements of $[\Sigma_\gamma]$ and $[\Sigma_\delta]$ were given truncated Cauchy(0, 2.25²) priors (Gelman et al. 2013; Broms et al. 2016). $[\kappa]$, the parameter describing the effect of host species richness on cowbird occurrence, was given a Normal(0,5) prior.

For model 3, to determine which host significantly affected cowbird occurrence, we used indicator variable selection—specifically, stochastic search variable selection (SVSS; O’Hara and Sillanpää 2009). In indicator variable selection, an individual covariate b_i is either included or excluded from the model at each MCMC iteration via a binary variable, $\omega_i b_i$, where $\omega_i = 1$ when the covariate is in the model, and $\omega_i = 0$ otherwise. As a result, the posterior mean of ω_i can be interpreted as the posterior probability that the covariate is in the model, allowing us to look at the relative effects of individual host species. SVSS is a type of indicator variable selection, whereby each b_i is given a mixture prior: $b_i \sim (1 - \omega_i)\text{Normal}(0, t^2) + \omega_i\text{Normal}(0, c^2 t^2)$ (O’Hara and Sillanpää 2009), where the variances in the Normals are tuning parameters. After exploring many options, we set $t^2 = 0.0001$ and $c^2 t^2 = 5$. We gave ω Bernoulli(ρ) priors. Each ρ_i was given a Beta(1,1) prior. This hierarchical approach alleviates the problem of ω drifting to zero (Hobbs and Hooten 2015, Section 9.2).

We fit each model using JAGS (v.4.3) and package rjags (v.4.12) in R (v.4.1.1) (Plummer 2016; R Core Team 2016). We used two chains, adapted the MCMC sampler for 40,000 iterations, discarded the first 80,000 iterations post adaptation, and then ran the model for 80,000 iterations keeping every 2nd draw, for a total of 40,000 iterations kept. We assessed convergence via traceplots and the Gelman-Rubin statistic (Gelman and Rubin 1992).

After fitting each model, we conducted posterior predictive checks to evaluate model fit (Gelman et al. 2013; Chambert et al. 2014), using the observed number of sites where each host co-occurred with the parasite as a test statistic. To evaluate the parsimony of each model, we used WAIC (p. 218 Gelman et al. 2014; Hobbs and Hooten 2015). WAIC is an information criterion (similar in construction to AIC) for Bayesian models that uses the full posterior distribution to estimate predictive accuracy (estimated using the computed log pointwise predictive density) and the effective number of parameters (estimated using the computed posterior variance of the log likelihood, summed across all data points). Using the model that fit the data and had the lowest WAIC score, we predicted the probability of co-occurrence between the parasite and each host in the three land uses, as well as the number of sites where each host co-occurred with a parasite, and the number of sites where each host occurred without the parasite.

Results

All parameters reported in Tables S2, S3, and S4 passed visual inspection of convergence, and the Gelman-Rubin statistic indicated convergence (i.e., all $\hat{R} < 1.1$). Each of the three models demonstrated adequate model fit (Fig. S1). Each model performed similarly in terms of WAIC (Table 1). While WAIC rewarded Models 1 and 2 for greater predictive ability, the metric penalized them in kind for increased model complexity. Assuming that all else being equal, simpler is better, we chose to use Model 1 for inference and prediction (Figs. 2, 3, 4).

Cowbirds were more likely to occur in coffee plantations (Posterior Mean = 0.85; 95% credible interval 0.65, 0.99) than in forests (0.24; 95% CI 0.09, 0.45).

Table 1 Values of WAIC, the information criterion used for model selection. elppd is the computed log pointwise predictive density, a measure of predictive accuracy p_{WAIC} is the computed posterior variance of the log likelihood, a measure of the number of parameters or model complexity

	WAIC	elppd	p_{WAIC}
Model 1	1460	-707.6	22.17
Model 2	1459	-705.8	23.76
Model 3	1459	-703.7	25.62

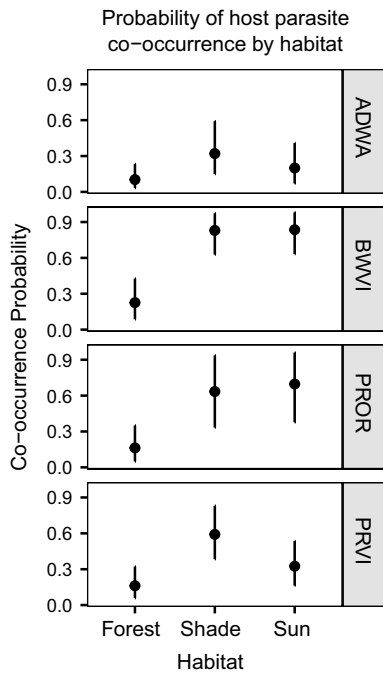


Fig. 2 Predictions from Model 1 of the probability of co-occurrence between each host and the Shiny Cowbird among the three habitats. Dots represent the posterior mean of the prediction, and the error bars represent the 95% credible intervals

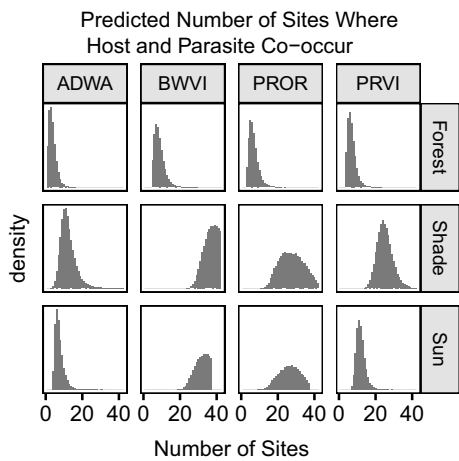


Fig. 3 Posterior distributions generated from Model 1 of the number of sites in each habitat where each host co-occurred with a cowbird

For the hosts, the difference in occurrence among land uses depended on the species (Table S2). In general, the hosts were more likely to occur in shaded

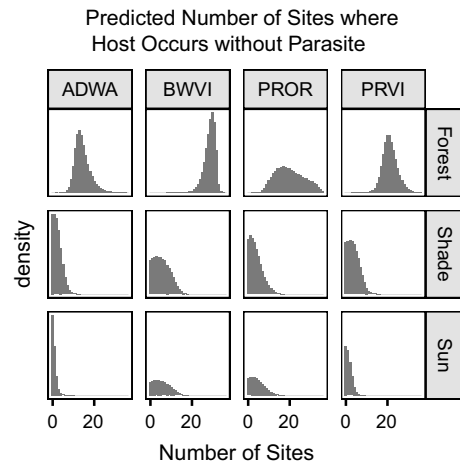


Fig. 4 Posterior distributions generated from Model 1 of the number of sites in each habitat where each host occurred without a cowbird

coffee plantations and forested sites. As a result, the hosts were generally more likely to co-occur with cowbirds in coffee plantations of either type (Fig. 2). Two endemic hosts, the Puerto Rican Vireos and Adelaide’s Warbler, were more likely to co-occur with cowbirds in shaded coffee plantations than in forests or sun coffee plantations (Figs. 3 and 4). The hosts were more likely to occur in sites without cowbirds if they were forested sites rather than if they were coffee plantations (Fig. 4). Conversely, the majority of sites where the hosts occurred with cowbirds were coffee plantations (Fig. 3).

The estimate of κ from Model 2 was slightly positive (0.47; 95% CI – 0.56, 1.54), indicating that host species richness may have a positive effect on parasite occurrence (Tables S2 and S5). Including this parameter in the model dramatically decreased the estimate of cowbird occurrence in forests, indicating that the cowbird was unlikely to be found in forested habitats without hosts. That being said, we caution against deep interpretations of this parameter because its credible interval overlaps zero.

Estimates of the effect of individual host species occurrence, ϵ_i in Model 3, did not indicate that any individual host had a notable effect on cowbird occurrence, nor did the posterior means of the model weights, ω_i (Tables S3 and S5). Each estimate of ϵ_i overlapped zero, and no estimate of ω_i deviated far from 0.5. The effect of Vireo occurrence on Cowbird occurrence, ϵ_i , was the only effect

whose 80% interval did not overlap zero. However, the effect was small and the 95% credible included zero.

Discussion

In this paper, we explored patterns in co-occurrence between the invasive Shiny Cowbird and its endemic hosts in the agricultural highlands of Puerto Rico. We hypothesized that the invasive Shiny Cowbird would most likely co-occur with its Puerto Rican hosts in shade coffee plantations, and that cowbirds would be more likely to occur in sites with their hosts regardless of land use. We found some evidence that the hosts influenced cowbird occurrence (Table S5), but the best performing model by WAIC assumed independence between the cowbird and its hosts (Table 1). We found that, in general, cowbirds were least likely to occur with their hosts in forested sites, and most likely to co-occur with their hosts in shade coffee plantations (Fig. 2). For host species that frequented both types of coffee plantations, e.g., Black-whiskered Vireo, co-occurrence rates were high regardless of the type of coffee farm. For host species that preferred shaded coffee plantations to sun coffee, e.g., Adelaide's Warbler and Puerto Rican Vireo, co-occurrence rates were highest in shaded coffee plantations (Figs. 3 and 4).

We found some evidence, although not strong, that cowbirds were more likely to occur at sites with their hosts. For example, Model 2 provided evidence for a positive association between cowbird occurrence and host species richness. This finding is in line with the results of Cummings and Veech (2014), who found that host species richness is the best predictor of Brown-headed Cowbird abundance in the Prairie Pothole region of the United States. While our credible interval of κ , the effect of host species richness on cowbird occurrence, overlapped zero (Table S5), we believe this result stems from the lack of sites with cowbirds and few or no hosts (Table S6). A study with more sites and more heterogeneity in host species richness might find a stronger relationship between richness and cowbird occurrence.

It remains unclear why Model 3 failed to separate the effect of each individual host species on cowbird occurrence. One possible explanation is our use of static, rather than dynamic, site occupancy models.

We suspect that the effects of host presence on cowbird occurrence would be much more pronounced in colonization rates, which could be estimated with a dynamic models. For example, Péron et al. (2016) used such models to elucidate the relative importance of land use and host species occurrence on parasite colonization and extinction rates. These dynamic host parasite occurrence models could be particularly interesting, given that dynamic occurrence models can provide great insight into the ecology of brood parasites (e.g., host selection), as well as the effect of brood parasites on the extinction of their hosts (Yackulic et al. 2014).

Our findings present some interesting questions for endemic bird conservation in the montane regions of Puerto Rico. Shaded coffee plantations clearly provide better habitat for native birds than their sun coffee counterparts (Irizarry et al. 2018). The shade trees provide nesting and feeding habitat for forest birds; thus, this community is much more likely to be found in shaded coffee plantations. As a result, some managers have advocated for planting shade trees in coffee farms to gain ecological services (Irizarry et al. 2018). However, our results indicate that Adelaide's Warbler and the Puerto Rican Vireo were most likely to overlap with the invasive parasite in shaded coffee plantations. The findings of Woodworth (1997, 1999) indicate that co-occurrence with cowbirds can shrink endemic host populations where the species co-occur. Thus, on the one hand, our results indicate that the invasion of the cowbird might diminish the ecological value provided by shade coffee.

On the other hand, our study may have been limited by the fact that we treated all shaded coffee plantations as equal and completely distinct from their sun coffee counterparts. In reality, canopy cover dictates a gradient between sun (0% cover) and shaded coffee (roughly 35% cover). Further, Irizarry et al. (2018) demonstrated that shade trees must fully mature before they provide ecological services to forest birds, i.e., before they develop habitat characteristics similar to forests. Thus, we might expect cowbirds to occupy fully mature plantations less, lowering co-occurrence rates relative to recently restored shade. Future studies could explore relative parasitism rates as a function of restoration maturity.

If co-occurrence (or parasitism) does not depend on maturity, there may be value in mitigating the effect of cowbirds by preserving forested sites. For example, we predicted that the Puerto Rican Vireo would occur in about 20 forested sites without the parasite, compared to only 3 shaded sites (Fig. 4). Tossas (2008) found that the cowbird parasitized fewer vireo nests in Mariacao Forest (a large forest reserve) than the open habitats, e.g., agriculture, around Guanica Forest. Further, Tossas and Thomlinson (2007) found that vireos rarely occupied and did not breed in small forest fragments, suggesting that small fragments could have less conservation value than large forest tracts. Thus, large forest preservation may be another valuable tool for managers in the highlands.

Acknowledgements We thank James F. Saracco, Jamie M. Kass, and two anonymous reviewers for insight comments on drafts of this manuscript. We also thank Katy Battle, Beatriz Romero, Luis Rodrigues, Damien Hardgrove, Joshua Morel, Mel Rivera, and Ray Robles for being integral to the data collection process. In addition, we thank the Puerto Rico Department of Natural and Environmental Resources and a host of private landowners for allowing us to sample on lands they manage. This research was funded by the Puerto Rico Department of Natural and Environmental Resources, Federal Aid project W-35. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Author contributions PTP Conceptualization, Data curation, Formal analysis, Methodology, Writing—original draft. KP Conceptualization, Methodology, Funding acquisition, Supervision, Writing—review and editing. JAC Conceptualization, Funding acquisition, Project administration, Writing—review and editing.

Funding This research was funded by the Puerto Rico Department of Natural and Environmental Resources, Federal Aid project W-35.

Data availability The code and data needed to reproduce all aspects of the analysis (results, figures, tables, etc.) are currently hosted as a standalone R package on the first author's GitHub page. Please see <https://github.com/philpatton/cowbird> for a full walk-through of how to recreate the tables and figures from scratch. The dataset itself can be found at Harvard Dataverse: Patton, Philip, 2022, "Replication Data for: Modeling and estimating co-occurrence between the invasive Shiny Cowbird and its Puerto Rican hosts", <https://doi.org/10.7910/DVN/FRD2AK>, Harvard Dataverse, V1.

Declarations

Conflict of interest The authors identify no competing interests.

References

- Beer J, Muschler R, Kass D, Somarriba E (1997) Shade management in coffee and cacao plantations. *Agrofor Syst* 38:139–164
- Blaser WJ, Oppong J, Hart SP, Landolt J, Yeboah E, Six J (2018) Climate-smart sustainable agriculture in low-to-intermediate shade agroforests. *Nat Sustain* 1:234–239
- Broms KM, Hooten MB, Fitzpatrick RM (2016) Model selection and assessment for multi-species occupancy models. *Ecology* 97:1759–1770
- Chambert T, Rotella JJ, Higgs MD (2014) Use of posterior predictive checks as an inferential tool for investigating individual heterogeneity in animal population vital rates. *Ecol Evol* 4:1389–1397
- Collazo JA, Groom MJ (2001) Avian conservation in north central forested habitats in Puerto Rico. Puerto Rican Department of Natural Resources, San Juan
- Cruz A, López-Ortiz R, Ventosa-Febles EA, Wiley JW (2005) Ecology and management of Shiny Cowbirds (*Molothrus bonariensis*) and endangered Yellow-shouldered Blackbirds (*Agelaius xanthomus*) in Puerto Rico. *Ornithol Monogr* 2005:38–44
- Cummings K, Veech JA (2014) Assessing the influence of geography, land cover and host species on the local abundance of a generalist brood parasite, the brown-headed cowbird. *Divers Distrib* 20:396–404
- Doherty TS, Glen AS, Nimmo DG, Ritchie EG, Dickman CR (2016) Invasive predators and global biodiversity loss. *Proc Natl Acad Sci* 113:11261–11265
- Dorazio RM, Royle JA (2005) Estimating size and composition of biological communities by modeling the occurrence of species. *J Am Stat Assoc* 100:389–398
- Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Stat Sci* 7:457–472
- Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013) Bayesian data analysis, 3rd edn. Chapman and Hall/CRC, London
- Gelman A, Hwang J, Vehtari A (2014) Understanding predictive information criteria for Bayesian models. *Stat Comput* 24:997–1016
- Gladkikh TM, Collazo JA, Torres-Abreu A, Reyes AM, Molina M (2020) Factors that influence participation of Puerto Rican coffee farmers in conservation programs. *Conserv Sci Pract* 2:e172
- Gould WA, Alarcon C, Fevold B, Jimenez M, Martinuzzi S, Potts G, Quinones M, Solórzano M, Ventosa E (2008) The Puerto Rico Gap Analysis Project Volume 1: land cover, vertebrate species distributions, and land stewardship. General technical report (gtr), U.S. Department of Agriculture, Forest Service
- Hobbs NT, Hooten MB (2015) Bayesian models: a statistical primer for ecologists. Princeton University Press
- Holmes ND, Spatz DR, Opper S, Tershy B, Croll DA, Keitt B, Genovesi P, Burfield IJ, Will DJ, Bond AL, et al (2019) Globally important islands where eradicating invasive mammals will benefit highly threatened vertebrates. *PLoS One* 14:e0212128

- Hoover JP, Robinson SK (2007) Retaliatory mafia behavior by a parasitic cowbird favors host acceptance of parasitic eggs. *Proc Natl Acad Sci* 104:4479–4483
- Irizarry JI, Collazo JA, Dinsmore SJ (2016) Occupancy dynamics in human-modified landscapes in a tropical island: implications for conservation design. *Divers Distrib*
- Irizarry AD, Collazo JA, Pacifici K, Reich BJ, Battle KE (2018) Avian response to shade-layer restoration in coffee plantations in Puerto Rico. *Restor Ecol* 26:1212–1220
- Kass JM, Tingley MW, Tetsuya T, Koike F (2020) Co-occurrence of invasive and native carnivorans affects occupancy patterns across environmental gradients. *Biol Invasions* 22:2251–2266
- Lugo AE, Helmer EH, Valentin ES (2012) Caribbean landscapes and their biodiversity. *Interciencia* 37:705–710
- MacKenzie DI, Nichols JD, Lachman GB, Droege S, Royle JA, Langtimm CA (2002) Estimating site occupancy rates when detection probabilities are less than one. *Ecology* 83:2248–2255
- O'Hara RB, Sillanpää MJ (2009) A review of Bayesian variable selection methods: what, how and which. *Bayesian Anal* 4:85–117
- Pacifici K, Zipkin EF, Collazo JA, Irizarry JI, DeWan A (2014) Guidelines for *a priori* grouping of species in hierarchical community models. *Ecol Evol* 4:877–888
- Patton PT (2016) Efficient use of community monitoring data: estimating patterns in co-occurrence, and novel multi-species models that address two types of sampling error. Master's thesis, North Carolina State University
- Péron G, Altwegg R, Jaime GA, Spottiswoode CN (2016) Coupled range dynamics of brood parasites and their hosts responding to climate and vegetation changes. *J Anim Ecol* 85:1191–1199
- Plummer M (2016) rjags: Bayesian Graphical Models using MCMC
- Post W, Wiley JW (1977) The shiny cowbird in the West Indies. *The Condor* 79:119–121
- R Core Team (2016) R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/edn>
- Royle JA, Dorazio RM (2008) Hierarchical modeling and inference in ecology. Academic Press, New York
- Smith JNM, Cook TL, Rothstein SI, Robinson SK, Sealy SG eds. (2000) Ecology and management of cowbirds and their hosts. University of Texas Press
- Tossas AG (2008) Reproductive success of the Puerto Rican Vireo in a montane habitat. *Wilson J Ornithol* 120:460–466
- Tossas AG, Thomlinson JR (2007) Landscape characteristics of Puerto Rican Vireo (*vireo latimeri*) nesting habitat, with source-sink implications. *Ornitol Neotropical* 18:233–242
- Tyre AJ, Tenhumberg B, Field SA, Niejalke D, Parris K, Possingham HP (2003) Improving precision and reducing bias in biological surveys: estimating false-negative error rates. *Ecol Appl* 13:1790–1801
- Wiley JW (1988) Host selection by the Shiny Cowbird. *The Condor* 90:289–303
- Woodworth BL (1997) Brood parasitism, nest predation, and season-long reproductive success of a tropical island endemic. *The Condor* 99:605–621
- Woodworth BL (1999) Modeling population dynamics of a songbird exposed to parasitism and predation and evaluating management options. *Conserv Biol* 13:67–76
- Yackulic CB, Reid JA, Nichols JD, Hines JH, Davis R, Forsman E (2014) The roles of competition and habitat in the dynamics of populations and species distributions. *Ecology* 95:265–279

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.