BIODIVERSITY RESEARCH

Using a novel model approach to assess the distribution and conservation status of the endangered Baird's tapir

Cody J. Schank[1,2] | Michael V. Cove[3] | Marcella J. Kelly[4] | Eduardo Mendoza[5] |
Georgina O'Farrill[6] | Rafael Reyna-Hurtado[7] | Ninon Meyer[7,8] |
Christopher A. Jordan[2,9,10] | Jose F. González-Maya[11] | Diego J. Lizcano[12,13] |
Ricardo Moreno[8,14] | Michael T. Dobbins[15] | Victor Montalvo[16] |
Carolina Sáenz-Bolaños[16,17] | Eduardo Carillo Jimenez[16] | Nereyda Estrada[18] |
Juan Carlos Cruz Díaz[16,17] | Joel Saenz[16] | Manuel Spínola[16] | Andrew Carver[19] |
Jessica Fort[19] | Clayton K. Nielsen[19] | Francisco Botello[20,21] | Gilberto Pozo Montuy[22] |
Marina Rivero[7,23] | Jesús Antonio de la Torre[23,24] | Esteban Brenes-Mora[25,26] |
Oscar Godínez-Gómez[5] | Margot A. Wood[27,28] | Jessica Gilbert[29] | Jennifer A. Miller[1]

[1]Department of Geography and the Environment, The University of Texas at Austin, Austin, TX, USA

[2]Global Wildlife Conservation, Austin, TX, USA

[3]Department of Applied Ecology, North Carolina State University, Raleigh, NC, USA

[4]Department of Fish and Wildlife Conservation, Virginia Tech, Blacksburg, VA, USA

[5]Instituto de Investigaciones sobre los Recursos Naturales, Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Michoacán, Mexico

[6]Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON, Canada

[7]Departamento de Conservacion de la Biodiversidad, El Colegio de la Frontera Sur, Lerma, Campeche, Mexico

[8]Fundacion Yaguara-Panama, Ciudad del Saber, Panama City, Panama

[9]Panthera, New York, NY, USA

[10]Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI, USA

[11]Proyecto de Conservación de Aguas y Tierras, ProCAT Colombia/Internacional, Bogotá, Colombia

[12]Departamento Central de Investigación, Universidad Laica Eloy Alfaro de Manabí, Manta, Ecuador

[13]The Nature Conservancy, Bogota, Colombia

[14]Smithsonian Tropical Research Institute, Balboa, Panama

[15]Department of Geography, University of Florida, Gainesville, FL, USA

[16]Instituto Internacional en Conservación y Manejo de Vida Silvestre, Universidad Nacional, Heredia, Costa Rica

[17]Department of Environmental Conservation, University of Massachusetts-Amherst, Amherst, MA, USA

[18]Escuela de Biologia, Universidad Nacional Autónoma de Honduras, Tegucigalpa, Honduras

[19]Department of Forestry and Cooperative Wildlife Research Laboratory, Southern Illinois University, Carbondale, IL, USA

[20]Departamento de Zoología, Instituto de Biología UNAM, Ciudad de México, México

[21]Departamento de Monitoreo Biológico y Planeación de Conservación, Conservación Biológica y Desarrollo Social, A. C, México City, México

[22]Conservación de la Biodiversidad del Usumacinta A.C., Emiliano Zapata, Tabasco, Mexico

[23]Bioconciencia A.C., Ciudad de México, México

[24]Laboratorio de Ecología y Conservación de Vertebrados Terrestres, Instituto de Ecología, UNAM, Ciudad Universitaria, Mexico City, Mexico

[25]Nai Conservation, San José, Costa Rica

[26]Escuela de Biología, Universidad de Costa Rica, Ciudad Universitaria, San José, Costa Rica

[27]Environmental Studies Program, Fort Lewis College, Durango, CO, USA

[28]Conservation International, Arlington, VA, USA

[29]Department of Wildlife and Fisheries Sciences, Texas A&M University, College Station, TX, USA

A Journal of Conservation Biogeography

Correspondence
Cody Schank, Department of Geography and
the Environment, The University of Texas at
Austin, Austin, TX, USA.
Email: codyschank@gmail.com

Abstract

Aim: We test a new species distribution modelling (SDM) framework, while comparing results to more common distribution modelling techniques. This framework allows for the combination of presence-only (PO) and presence-absence (PA) data and accounts for imperfect detection and spatial bias in presence data. The new framework tested here is based on a Poisson point process model, which allows for predictions of population size. We compared these estimates to those provided by experts on the species.

Species and Location: Presence data on Baird's tapir (*Tapirus bairdii*) throughout its range from southern México to northern Colombia were used in this research, primarily from the years 2000 to 2016.

Methods: Four SDM frameworks are compared as follows: (1) Maxent, (2) a presence-only (PO) SDM based on a Poisson point process model (PPM), (3) a presence-absence (PA) SDM also based on a PPM and (4) an Integrated framework which combines the previous two models. Model averaging was used to produce a single set of coefficient estimates and predictive maps for each model framework. A hotspot analysis (Gi*) was used to identify habitat cores from the predicted intensity of the Integrated model framework.

Results: Important variables to model the distribution of Baird's tapir included land cover, human pressure and topography. Accounting for spatial bias in the presence data affected which variables were important in the model. Maxent and the Integrated model produced predictive maps with similar patterns and were considered to be more in agreement with expert knowledge compared to the PO and PA models.

Main conclusions: Total abundance as predicted by the model was higher than expert opinion on the species, but local density estimates from our model were similar to available independent assessments. We suggest that these results warrant further validation and testing through collection of independent test data, development of more precise predictor layers and improvements to the model framework.

KEYWORDS

conservation, habitat cores, hotspots, maxent, occupancy model, point process model, tapir, *Tapirus bairdii*

1 | INTRODUCTION

1.1 | Conservation and SDMs

Determining current species distributions and assessing change over time are necessary for better conservation management aimed at reducing species loss in ecosystems. Species distribution models (SDMs) provide continuous spatial predictions about species habitat preferences and thus are useful tools for conservation management (Franklin, 2010). However, building useful models for rare species can be challenging for many reasons, including that data on rare species are typically scarce and vary in quality and type (Aitken et al., 2007; Wisz et al., 2008). The fact that species most in need of conservation attention are often the most difficult to model has been labelled the "rare species modelling paradox" (Lomba et al., 2010). In this study, we

compare four SDMs in terms of their abilities to generate necessary distributional information relevant for conservation of a rare, endangered species.

1.2 | Methodological background

In general, SDMs use one of two types of data as the response (i.e., dependent) variable: presence-only (PO) or presence-absence (PA). PO data contain only information about where a species has been observed, while PA data also provide information about where the species has not been found despite survey efforts.

Presence-only data are generally more available than presence-absence data, but the modelling options appropriate for use with them are limited and can lead to overestimates of distribution. Despite this, modelling methods for presence-only data have advanced, albeit

slowly. For over a decade, the most common method across all SDMs has been Maxent modelling (Phillips, Dudík, & Schapire, 2004; Vaz, Cunha, & Nabout, 2015), likely due to its more accurate predictions compared to other frameworks (Elith et al., 2006). Despite its heavy use, Maxent has been criticized for several reasons. These criticisms include its use of vague indices to determine probability of occurrence (Royle, Chandler, Yackulic, & Nichols, 2012), lack of accounting for spatial sampling bias, which is common in presence-only data (Yackulic et al., 2013), and difficulty in interpreting results (Renner & Warton, 2013).

Other methods developed to address deficiencies in SDMs include occupancy models (MacKenzie et al., 2002) and point process models (PPMs) (Warton & Shepherd, 2010). Occupancy models account for imperfect detection, or the fact that a species is not always detected despite its presence. Failing to account for imperfect detection can lead to biased parameter estimates (Dorazio, 2012; Lahoz-Monfort, Guillera-Arroita, & Wintle, 2014). In occupancy modelling, if detection probability is affected by the same set of covariates that affect occurrence probability, bias in the model is inevitable (Royle et al., 2012). To solve this issue, these models require repeated observations of either presence-absence or point-count (PC) data to estimate and account for detectability (Kéry et al., 2013).

Some researchers have attempted to work around the weaknesses of using PO data in SDMs by including pseudo-absences, or artificially compiled absences, in models. Pseudo-absences are used with methods that require PA data when only PO data are available (Zaniewski, Lehmann, & Overton, 2002). Maxent uses a similar construct termed background points. Regardless of the term, there are no clear recommendations on the number of points to use or how to distribute those points across the area of inference (but see Barbet-Massin, Jiguet, Albert, & Thuiller, 2012). PPMs solve the pseudo-absence problem through a reformulation of the basic model structure for SDMs. In PPMs, pseudo-absences are viewed as equivalent to quadrature points used to estimate integrals. Following this conceptualization, the number of background points is increased until the likelihood stabilizes (Warton & Shepherd, 2010).

An additional advantage of PPMs is that the response variable modelled is "intensity," or individuals per unit area. This is different from Maxent, which models occurrence on a per-pixel basis, and thus is dependent on the scale chosen for the analysis (Renner & Warton, 2013). The model-estimated intensity can also be summed across spatial units (e.g., a specific protected area) to obtain an estimate of population size in that area, a useful tool for conservation management.

For most rare species, PO data remain the most readily available because they are easiest to collect (Engler, Guisan, & Rechsteiner, 2004). However, over the past decade, PA data have become increasingly common due to analytical developments that have promoted their collection (i.e., MacKenzie et al., 2002) and the rapid expansion of camera-trapping, which enables this data collection (e.g., Lesmeister, Nielsen, Schauber, & Hellgren, 2015). Given that data on rare species are often scarce by nature, methods to integrate both PO and PA data on rare species into the same model have the potential to improve our

capacity to model rare species effectively. Dorazio (2014) introduced a model that combines the two frameworks of occupancy models and point process models, allowing for the use of both PO and PA data. Thus far, there have been no real-world applications of this model (see Fletcher, McCleery, Greene, & Tye, 2015 for a partial application), leaving conservation biologists unable to assess the full potential of this integrative approach of combining PO and PA data.

1.3 | Baird's tapir as a case study species

Baird's tapir (*Tapirus bairdii*) is an ideal species to test the Integrated model as there is a strong network of international scientists collaborating and sharing data across the species' range in order to fully understand its ecological and conservation needs. It is also of considerable conservation interest due to its globally endangered status, genetic uniqueness, role in ecosystem function and due to increasing threat to its habitat (Brooks, Bodmer, & Matola, 1997; García et al., 2016; Isaac, Turvey, Collen, Waterman, & Baillie, 2007).

Baird's tapir plays an important ecological role as an ecosystem engineer, primarily through seed dispersal and extensive selective browsing. Baird's tapirs are the largest terrestrial mammals in the Neotropics and are therefore the sole effective dispersers of most large seeds, especially over long distances (O'Farrill, Galetti, & Campos-Arceiz, 2013). The loss of megaherbivores like the tapir can lead to cascading detrimental effects on other species, ecological processes, as well as on overall ecosystem function (Ripple et al., 2015). Previous efforts to model the distribution of Baird's tapir include projects that used presence-only data in Maxent models (Carrillo Reyna, Weissenberger, & Reyna-Hurtado, 2015; Mendoza et al., 2013; Schank et al., 2015), and two studies that used presence-absence data in occupancy models (Cove et al., 2014; Jordan et al., 2016).

1.4 | Research objectives

Our main objective was to apply the Integrated SDM framework using both PO and PA data of Baird's tapir and compare the results to models that use each data type alone. However, given that basic knowledge about where a species occurs is imperative in designing and implementing conservation strategies, our secondary objective was to evaluate the implications of model results for Baird's tapir conservation and protection. To do this, we evaluated the accuracy and precision of model estimates, including model coefficients and population estimates created by summing intensity across spatial units. We further discuss the significance of these estimates based on our collective prior knowledge of the status of Baird's tapir populations and current threats to the species.

In addition, we address the issue of survey bias in our data. Wildlife surveys are often biased towards areas that are more accessible (e.g., close to roads), or places where the species is known or thought to occur. Hierarchical models, such as the Integrated SDM used here, separate the detection process from the process of interest: the species' actual distribution. In this hierarchical framework, the modeller can use variables in the detection process that are thought to

influence or bias the location of surveys. We tested the use of these variables and discuss their impacts on model output.

2 | METHODS

2.1 | Study area

The study area for this research ranges from Veracruz and Oaxaca states in southern Mexico to the Department of Chocó in Northern Colombia (approximate coordinates: 22°N to 3.6°S, 98° to 76°W), excluding islands. All analysis was conducted at a spatial resolution of one kilometre, in a global equal area projection.

2.2 | Model formulation

Four modelling frameworks were tested in this analysis, including (1) Maxent (Phillips et al., 2004), (2) a presence-only (PO) SDM based on a point process model (PPM), (3) a presence-absence SDM (PA), also based on a PPM and (4) the Integrated SDM framework that combines both 2 and 3 (Dorazio, 2014). Imperfect detectability is incorporated in both the PO (2) and PA (3) SDMs. In the PO model, probability of detection is modelled as a thinning of the point process describing the species distribution, using a Bernoulli distribution (Dorazio, 2014). In the PA model, detectability is estimated as in an occupancy model using a zero-inflated formulation (Royle and Dorazio, 2008). Models were fit using the "dismo" package (Hijmans, Phillips, Leathwick, & Elith, 2014) and custom code developed in R (R Core Team, 2016).

Nineteen candidate models were formulated to test differing hypotheses about the factors that affect the distribution of Baird's tapir

(Table 1). All four model frameworks were fit using these candidate models (except the null model, model 0, for Maxent, as there is no way to run the software without predictor variables). To facilitate the creation of the candidate models, predictor variables were grouped into categories related to influential processes including climate, land cover and protected status, human pressure, slope and sampling variables (a list of the data sources is found in Appendix S1). All continuous variables were scaled to have a mean of zero and standard deviation of one; thus, the magnitudes of the coefficients are comparable (except distance to/within protected areas, which was not centred, to preserve negative values as within protected areas, and positive values outside of these areas).

2.3 | Description of modelling frameworks

Maxent is a popular SDM software that minimizes the entropy between two probability distributions based on a vector, z, of environmental covariates used in modelling (Elith et al., 2011). The first distribution, $f_1(z)$, is based on presence-only (PO) locations, while the second, $f(z)$, is based on a random sample of the background environment. The goal of most SDMs is to estimate the probability of presence, or $Pr(y = 1|z)$, where $y = 1$ denotes presence ($y = 0$ denotes absence). However, without absence data, this property can only be estimated up to a constant. The missing piece of information is $Pr(y = 1)$, or the prevalence of the species in the study area. By default, the Maxent software assumes prevalence is 0.5 to create a logistic output which ranges from 0 to 1. We adjusted this setting to 0.2 to match the prevalence in our presence-absence test data. Maxent settings were also adjusted to include only linear and quadratic terms

TABLE 1 Categories of variables included in the candidate models

Model	Climate	Land cover	Human pressure	Slope	Sampling
Model0					
Model1	y				
Model2		y			
Model3			y		
Model4	y	y			
Model5	y		y		
Model6		y	y		
Model7	y	y	y		
Model8		y	y	y	
Model9	y	y	y	y	
Model10	y				y
Model11		y			y
Model12			y		y
Model13	y	y			y
Model14	y		y		y
Model15		y	y		y
Model16	y	y	y		y
Model17		y	y	y	y
Model18	y	y	y	y	y

(when included in other models). Default settings in Maxent include more complex parameters, which can lead to model overfitting, and would not provide a reasonable comparison to the other frameworks used in this research.

The Poisson point process model (PPM) is the foundation of the SDM framework used in Dorazio (2014). In this model, $\lambda(s)$ is the limiting, expected density (number of individuals per unit area) of individuals at location s, for a Poisson point process. In the context of the SDM, $\lambda(s)$ is formulated as a log-linear function of unknown parameters and location-specific regressors $x(s)$ (i.e., $log(\lambda(s)) = \beta_0 + \beta'x(s)$). The presence-only (PO) model formulated in Dorazio (2014) incorporates spatial bias in PO data through an independent thinning of the point process. This thinned point process is the product of the original point process and $p_{po}(s)$, the probability that the site is surveyed and the species is detected. $p_{po}(s)$ is formulated as a logistic function of unknown parameters and location-specific regressors $w_{po}(s)$: $logit(p_{po}(s)) = \alpha_0 + \alpha_{po}'w_{po}(s)$.

We adjusted the planned survey model from Dorazio (2014), which used point-count data, to accommodate our detection/non-detection data. The model used here follows an occupancy framework based on a zero-inflated binomial distribution. The presence or absence (i.e., occupancy) of the species at a site, C_i, follows a Bernoulli distribution. In this case, the detection histories at each site, y_i, have additional non-detections (i.e., zeros) due to imperfect detectability, the fact that an individual may go undetected even when present. This relationship is modelled as a Binomial distribution with J trials, and the probability of success (i.e., species detection) equal to the product of z_i (the occupancy state, $z_i = I(N(C_i) > 0)$ and p_{pa}, the probability of detection at the site. Occupancy (ψ_i) and intensity (λ_i) are linked following the equation below (Koshkina, Wang, Gordon, Dorazio, & Stone, 2017),

$$\varphi_i = Pr\left(N(C_i) > 0\right) = 1 - \exp\left(-\int_{C_i} \lambda(s)ds\right)$$

As with detectability in the PO model, $p_{pa}(s)$ is formulated as a logistic function of unknown parameters and location-specific regressors $w_{pa}(s)$: $logit(p_{pa}(s)) = \alpha_0 + \alpha'w_{pa}(s)$.

Though their models of detectability differ, the PO and PA frameworks share the same SDM model based on a Poisson point process. In the Integrated SDM, the PO and PA models are estimated simultaneously, such that one set of parameters for the SDM is created (i.e., the β's), while separate detectability parameters are estimated (i.e., the α's) for the PO and PA models.

2.4 | Presence data

Presence data for Baird's tapir were compiled from several sources (Appendix S1—Data Sources), including planned surveys conducted using camera traps (presence-absence data), and collections of opportunistic observations or published sightings (presence-only data), which may originate from cameras, track or signs of the animal, or direct observations of the animal. Though PA data are generally more

clustered across the study area than the PO data (Figure 2), they do provide similar coverage of environmental gradients (min–mean–max percent overlap between background/PO: 0.36-0.63-0.80, background/PA:0.46-0.60-0.75, PO/PA:0.50-0.67-0.90, Pastore, 2017).

We used presence-absence data collected by camera traps to construct 100-day long detection histories at each site, establishing ten sampling occasions that span ten days each (tapir appear to cycle through their home range every 10–12 days; Jordan, 2015). Cameras at some sites malfunctioned or were removed before the 100-day sampling period was complete. In this case, we used data only from complete sampling 10-day occasions.

Due to the nature of modelling the entire distribution of a wide-ranging species like Baird's tapir, data must be combined from several sources which are often collected during different years. In this case, the data span primarily the period from 2000 to 2015 (with few exceptions); thus, the detection histories across sites are not from the same 100-day span. For this reason, occupancy and density estimates represent the species presence during that time span (i.e., over 15 years).

2.5 | Spatial subsampling

A random spatial subsampling procedure was performed to thin the presence data prior to model fitting. The algorithm starts by randomly choosing one observation point and removing any other observation points within a given radius. The chosen observation is added to the subset, and the steps repeated until no observations are left in the original data. The effect of this procedure is to enforce a minimum distance between sampling points. A similar type of subsampling is sometimes used to remove survey bias in observation data (Beck, Böller, Erhardt, & Schwanghart, 2014); however, this grid-based approach can lead to samples that remain close in space, if they fall just across a boundary in an adjacent grid cell.

This subsampling process was used to maintain independence of observations. In occupancy models, the population at each site is assumed to be constant, with no immigration or emigration (MacKenzie et al., 2006). This assumption is known as closure, violations of which can lead to biased parameter estimates. For both types of data (PO and PA), a minimum distance of five kilometres was selected to avoid duplicate observations of the same individual. Data on home range sizes vary from 1.25 km^2 reported in Costa Rica (Foerster & Vaughan, 2002) to 8–10 km^2 in Nicaragua (C. Jordan personal observation, 2016) and 23.9 km^2 in Mexico (Reyna-Hurtado, Sanvicente-López, Pérez-Flores, Carrillo-Reyna, & Calmé, 2016). Thus, the 5-km spacing should ensure that detections are from different individuals given the tapir's home range size.

2.6 | Model averaging

Model fit was repeated 100 times for each combination of candidate model and model framework to capture the variation in the random spatial subsampling. Parameters were averaged across candidate models for each framework and point sample using Akaike weights (Symonds & Moussalli, 2011),

$$\hat{\beta} = \frac{\sum_{i=1}^{R} w_i \hat{\beta}_i}{\sum_{i=1}^{R} w_i}$$

where $\hat{\beta}$ is the parameter estimate for a given model i, and w_i is the Akaike weight of the model. The parameters were then averaged across these replicates to create one prediction and set of parameter estimates for each model framework (Table 2). In Maxent, calculation of AIC is not straightforward; thus, for this framework, we averaged across model iterations for candidate Model 9 (full model without sampling covariates). Cade (2015) questions the validity of model averaging using Akaike weights; however, for this research, only the PO and PA models are affected, because this method is not used with the Maxent results, and model 18 has all of the support in the Integrated framework (Table 2).

2.7 | Identification of habitat cores

Using the predicted intensity from the Integrated model, we calculated the Getis-Ord Gi* (i.e., local *G*) spatial statistic (Getis & Ord, 1992) to identify clustered areas both above (i.e., hotspots) and below (i.e., cold spots) the global mean (see Nelson & Boots, 2008 for use of hotspots in Landscape Ecology). The estimated intensity was then summed within each statistically significant ($\alpha = 0.05$) hotspot, and those with values above 200 were selected as habitat

TABLE 2 Mean of Akaike weights for each candidate model and model framework. Standard deviation of estimates in parentheses. Calculation of AIC is not straightforward for Maxent; thus, we focus here on the other three frameworks. Cells left blank have no support (i.e., Akaike weight = 0)

Candidate model	PO AIC weight (SD)	PA AIC weight (SD)	Integrated AIC weight (SD)
model0		0.0166 (0.0687)	
model1		0.0178 (0.053)	
model2		0.017 (0.0542)	
model3			
model4		0.0292 (0.0738)	
model5		0.0179 (0.0478)	
model6			
model7		0.022 (0.0802)	
model8			
model9	0.0354 (0.174)	0.233 (0.306)	
model10		0.093 (0.172)	
model11		0.0361 (0.0768)	
model12		0.0444 (0.112)	
model13		0.0453 (0.0871)	
model14		0.0537 (0.0945)	
model15	0.124 (0.331)	0.021 (0.0453)	
model16	0.359 (0.478)	0.0294 (0.0515)	
model17	0.0625 (0.242)	0.0679 (0.17)	
model18	0.472 (0.493)	0.238 (0.295)	0.998 (0.00421)

cores. Medici and Desbiez (2012) estimated that 200 adult individuals of *Tapirus terrestris* were required to constitute a minimum viable population in the Atlantic Forest of Brazil. We assume the same to be true for *Tapirus bairdii* although if inbreeding and genetic variability were considered, this number would likely need to be considerably higher, especially given that Baird's tapirs are known to have very low levels of heterozygosity and allelic diversity (Norton & Ashley, 2004).

3 | RESULTS AND DISCUSSION

3.1 | Coefficient estimates

The reciprocal of the condition number, the ratio of the smallest to the largest eigenvalues in the Fisher information matrix, can be used to determine whether the parameters of the SDM are identifiable (Dorazio, 2014). Following these criteria, coefficient standard errors were almost never estimated for the PO models, and thus, this framework is not included in the majority of our discussion. This trouble with identifiability could be due to very low detectability in the PO data. For comparison, PO models were estimated with standard errors in less than 1% of models, while PA models were able to estimate these measures 100% of the time, and the Integrated model 80% of the time.

Slope was the only variable in the PA model that had a 95% confidence interval that did not overlap zero (Table 3). It is likely that establishing important relationships with environmental variables was difficult in the PA model because of reduced sample size (the PA model does not include any background information, while the other three models do). This was also evident in the Akaike weights (Table 2), as there was no clear top/best model. The different random samples that resulted from the spatial subsampling of the presence data appear to have a strong effect on which model has the best fit. In contrast, for the Integrated framework, the full model (model 18) had nearly all of the support in all iterations.

In the Integrated framework, the only climate variables with 95% confidence intervals that did not overlap zero were annual precipitation (positive), and temperature seasonality (negative), both also had nonlinear relationships (i.e., quadratic terms important). Meanwhile, nearly all non-climate variables had 95% confidence intervals that did not overlap zero. The modelled relationships for these variables were mostly as expected, showing a positive relationship between Baird's tapir occurrence and forest cover, and negative relationships with roads and slope (Table 3). Among all of these, the effect of forest cover had the largest magnitude, followed by roads. This is not surprising, as deforestation and hunting are widely recognized as the two largest threats to the species (Garcìa et al., 2016; Medici et al., 2005). Interestingly, Enhanced Vegetation Index (EVI) has a negative and linear relationship with Baird's tapir occurrence in our model.

There were also several significant relationships among the sampling variables in the Integrated framework. With these, a positive relationship existed with protected status in the PO data, possibly

TABLE 3 Mean of model-averaged coefficient estimates for each model framework, with mean standard errors in parentheses. Symbols ("*") indicate estimates > 1.96 standard errors from zero (i.e., 95% confidence intervals). Standard error estimates for PO model are not provided

	PO	PA	Integrated
Beta Coefficients (Intensity)			
beta0	0.102	−1.99 (1.6)	−1.22 (0.27)*
Temp_seasonality	0.0781	0.0225 (0.534)	0.106 (0.112)
Precip_seasonality	−0.0541	−0.241 (0.553)	−0.0556 (0.118)
Max_temp_warmest_month	0.191	0.362 (0.73)	0.129 (0.131)
Annual_precip	0.404	−0.0488 (1.02)	0.497 (0.152)*
Temp_seasonality_sq	−0.254	−0.18 (0.343)	−0.284 (0.0905)*
Precip_seasonality_sq	−0.144	0.375 (0.349)	−0.115 (0.0957)
Max_temp_warmest_month_sq	0.0337	0.436 (0.292)	0.0548 (0.0278)*
Annual_precip_sq	−0.198	−0.381 (0.452)	−0.297 (0.0881)*
Treecover2000	1.39	1.63 (1.23)	1.45 (0.18)*
Distance_protected_area	−0.147	0.11 (0.359)	−0.0379 (0.118)
EVI	−0.348	−0.942 (0.729)	−0.426 (0.138)*
EVI_sq	0.0423	0.285 (0.334)	0.0206 (0.0629)
Forestloss_focal	−0.155	−0.0492 (0.156)	−0.167 (0.0561)*
Road_length_focal	−0.837	−0.499 (0.329)	−0.691 (0.142)*
Slope	−0.172	−0.834 (0.405)*	−0.358 (0.0904)*
Alpha Coefficients (Detectability)			
PO data			
alpha0.po	−1.03	–	−7.58 (0.301)*
Forest	−1.79		−0.0117 (0.208)
Protected	6.71	–	1.47 (0.203)*
Road_distance	−3.79	–	−0.752 (0.128)*
Road_distance_sq	0.673	–	0.123 (0.0301)*
PA data			
alpha0.pa	–	−1.45 (0.622)*	−1.38 (0.644)*
Treecover2000	–	0.392 (0.453)	0.24 (0.57)
Distance_protected_area	–	−0.0539 (0.182)	−0.0141 (0.273)
Road_distance	–	−0.383 (0.248)	−0.665 (0.373)
Road_distance_sq	–	0.0558 (0.109)	0.0768 (0.162)
Slope	–	0.146 (0.0762)	0.204 (0.12)

indicating the positive sampling bias towards these areas. Also with the PO data, there was a significant negative and nonlinear effect with distance to roads, which could indicate the increased difficulty in sampling far from road networks. In the PA data, none of the sampling variables had a 95% confidence interval that did not overlap zero. It is important to note that the sampling variables in the PO data are meant mostly to account for sampling bias, while those in the PA model should reveal more about what affects the detectability of the species. For this reason, we used slightly different sampling variables for the two data sets.

It is also interesting to note that distance to/within a protected area was significant in the Integrated models without sampling variables (β = −0.51, *SE* = 0.0806, *p* < .001), but was not significant in the Integrated models with sampling variables (β = −0.0379, *SE* = 0.118, *p* < .761). Thus, accounting for survey bias by including variables in the detection process reduced the effect of a variable that was expected to have a significant influence on Baird's tapir distribution.

Maxent estimates parameters are not directly comparable to the coefficients discussed above, but that can provide some context for comparison (Table 4). Forest cover, distance to/within protected areas and presence of roads were the three variables with the most importance and contribution to Maxent (in model 9). This was similar to the results we see with the Integrated model discussed above. Although it would have been possible to include the sampling variables in the Maxent model, there was no ability to differentiate between variables that affect the sampling process vs. those that affect the species distribution directly.

3.2 | Spatial predictions

Spatial predictions for probability of occurrence were strikingly similar for two of the model frameworks: Maxent and Integrated (Figure 1). These predictions match fairly well with the expected distribution for the species, although the probabilities may be too high across the

	Contribution	Importance	Lambda
Temp_seasonality	0.5	1.38	0.727
Precip_seasonality	0.0158	0.0268	−0.0628
Max_temp_warmest_month	0.0367	0.249	0.428
Annual_precip	1.06	6.86	3.42
Temp_seasonality_sq	4.45	3.67	−1.45
Precip_seasonality_sq	3.13	0.352	−0.759
Max_temp_warmest_month_sq	0.141	0.122	0.941
Annual_precip_sq	1.2	6.75	−6.16
Treecover2000	59.7	53.3	4.05
distance_protected_area	19.8	10.1	−4.36
EVI	0.531	2.78	−2.38
EVI_sq	0.00703	0.031	0.205
Forestloss_focal	1.68	1.8	−1.23
Foad_length_focal	4.74	7.96	−15.1
Slope	3.03	4.61	−2.32

TABLE 4 Mean contribution, importance and lambda values for each environmental variable used in Maxent Model 9
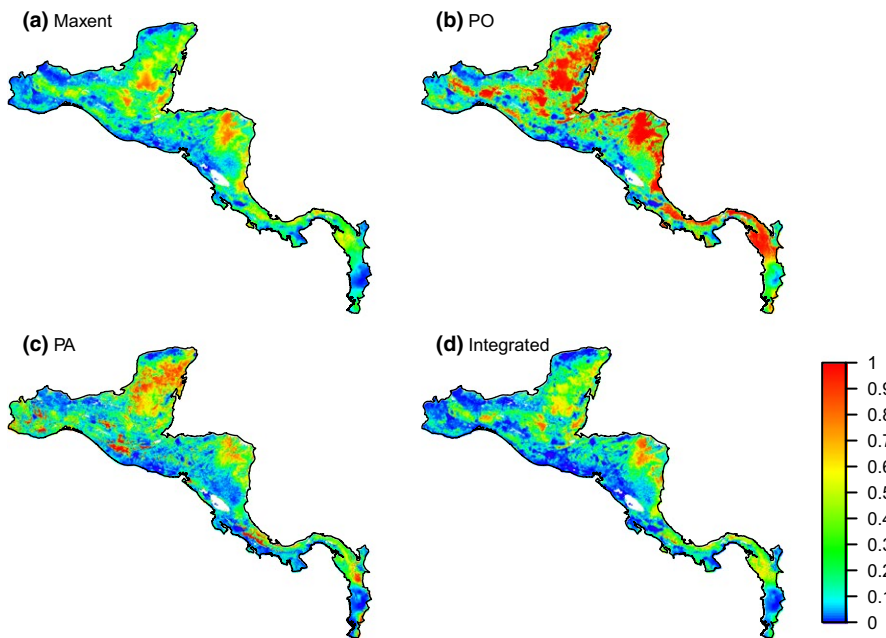

FIGURE 1 Predicted probability of presence for each model framework, (a) Maxent, (b) PO, (c) PA and (d) Integrated. [Colour figure can be viewed at wileyonlinelibrary.com]

landscape overall. Both the PO and the PA models had even higher probabilities across the study area, and the PA model had probabilities close to one in regions that have not maintained Baird's tapir presence in recent history (e.g., southern Guatemala and north-western Nicaragua). The PO model may have a similar pattern to the Integrated model, but high estimated probabilities covered too much of the distribution for the species.

Inspecting the coefficient estimates (Table 3), the difference between the PO and Integrated model was largely due to the intercept (beta0), while the difference between the PA and Integrated models could be attributed to the difference in coefficient estimates for climate variables. The advantage that the Integrated model has over Maxent is the ability to estimate intensity (i.e., abundance), which we utilize below to identify viable habitat cores.

3.3 | Assessment of habitat cores and population estimates

A hotspot analysis using Gi* (circular neighbourhood with radius of 10 km; Getis & Ord, 1992) identified 25 habitat cores in the Integrated model framework (Figure 2), with the two largest cores (the Yucatán Peninsula and the Honduras–Nicaragua border) comprising roughly 61% of core habitat for Baird's tapir. In general, the cores matched well with the expected distribution of the species. However, some small cores were predicted that have no tapir observations and likely do not harbour any individuals (e.g., La Encrucijada, Mexico; Laguna de Términos, Mexico; north and west of Lacandona, Mexico; Pixvae, Panama). It is probable that tapirs historically occurred in these areas, yet went locally extinct due to small core size and distance from larger

FIGURE 2 Map of habitat cores for the Integrated model. Purple lines represent the proposed Nicaragua canal path and the Panama Canal, black points are presence-only sample locations, red points are detection history sample locations, and the area shaded light brown is the study area. Yellow points are study areas for independent abundance estimates (1. Corcovado, 2. Triunfo, 3. Los Katios, 4. Oaxaca and 5. Talamanca). Black circles are cores overestimated by the model (a. Encrucijada, b. Laguna de Términos, c. N and W of Selva Lacandona and d. Pixvae). [Colour figure can be viewed at wileyonlinelibrary.com]
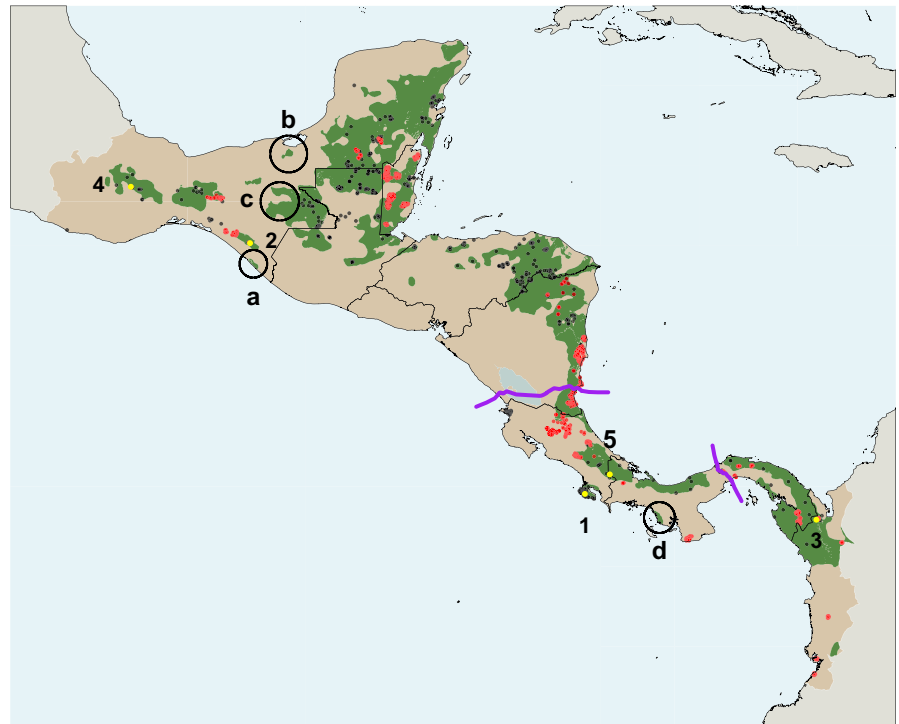
cores that act as population sources. There were also areas with confirmed tapir populations that did not appear in the prediction of viable cores (e.g., La Sepultura Biosphere Reserve and Frailescana Natural Protected Area, Mexico).

According to model estimates, there are approximately 175,000 individual tapirs spread across these 25 core areas. This is more than an order of magnitude higher than expert estimates for the species that suggest there may be fewer than 6,000 individuals left in the wild (Castellanos et al., 2008; Garcìa et al., 2016).

It is possible that model results would more closely approximate global experts' expected state of the population with the addition of more proximate or more detailed environmental variables, such as differing management regimes among protected areas, more detailed habitat types, distance to villages, and improved hunting/poaching variables. Hunting in particular has been reported as a widespread and substantial threat to the survival of Baird's tapirs due to their slow reproduction (Dunn, Estrada, & Smith, 2011; García, Leonardo, Castillo, Gómez, & García, 2010; Jordan, Galeano, & Alonzo, 2014; Koster, 2007; McCann, Wheeler, Coles, & Bruford, 2012). Although hunting does not seem to drive tapirs to local extinction in all sites where the phenomenon has been studied, hunting could maintain tapir populations at extremely low densities. The human invasion of protected areas throughout significant portions of the Baird's tapir range has increased drastically in past decades, which has likely resulted in increased access for hunters in arguably all of Baird's tapir range countries both inside and outside of protected areas. Thus, substantially lower densities than would be expected in the absence of hunting might exist throughout the species' range. However, hunting data in a format compatible with our PA and PO data on tapirs are not available on a range-wide basis for the species. Thus, to evaluate this,

hunting data would need to be collected using a uniform methodology across the range of Baird's tapir, which would entail substantial time and economic investment.

Another factor that might explain the apparent overestimation of population from our model is the detectability of the species in fragmented habitat. Cove et al. (2014) observed that tapir detection was negatively related to increasing forest cover such that tapirs were more detectable in small forest patches. The interpretation of intensity (from camera trap detections) correlating with true abundance could be misleading if tapir detection increases in small forest patches. This detection phenomenon is another factor that might explain the disparity in our global population estimate and expert opinions.

To verify the intensity estimates from this model, abundance estimations derived from other methods should be used as a comparison (i.e., via capture–recapture). Ideally, these abundance estimates should be summed up across intact cores to see how they compare to the summed intensity from this research. Interestingly, intensity values at a particular location do not appear to be outside the estimates of experts, it is only when summing the intensity across a wider region, including those areas not surveyed, that our model tends to disagree with expert opinion.

3.4 | Comparison with other studies

The maximum intensity from the Integrated model is 4.4 individuals/km^2, which is about 50% more than the highest reported estimates for the species (2.92 individuals/km^2) using capture–recapture methods (González-Maya et al., 2012). Interestingly, the location of the maximum intensity values coincide very closely with the study area of González-Maya et al. (2012) in the Talamanca Mountains on

the border between Costa Rica and Panama. Although some researchers (e.g., Oliveira-Santos, Zucco, Antunes, & Crawshaw, 2010) question the ability to use capture–recapture methods with camera traps on a species such as Baird's tapir due to lack of natural markings, others disagree with the validity of the methodology for different reasons (Tobler, Hibert, Debeir, & Richard-Hansen, 2014). Mark–resight models (McClintock, White, Antolin, & Tripp, 2009) or spatially explicit models for presence-absence data (Ramsey, Caley, & Robley, 2015) could be good alternatives for estimating density but, to our knowledge, have not yet been applied to tapir data.

Estimates from our model are also relatively consistent with other studies that estimate Baird's tapir density in other parts of Central America. Naranjo-Piñera (1995) estimated 0.6 individuals/km^2 in Corcovado National Park in Costa Rica using transects, while our model estimates a mean of 0.81 individuals/km^2 in the same protected area. Carbajal-Borges, Godínez-Gómez, and Mendoza (2014) cite 0.12 individuals/km^2 as their estimate for tapir abundance in the core area of El Triunfo Biosphere Reserve (though the estimates range up to 0.49 using different assumptions), while our model estimates a mean of 0.40 individuals/km^2 across the core area of the reserve. Mejía-Correa, Diaz-Martinez, and Molina (2014) estimate 1.02 individuals/km^2 in Los Katíos National Park of Colombia, while our model estimates a mean of 0.65 individuals/km^2 in the same protected area. Finally, Botello et al. (in press) estimate 0.32 individuals/km^2 using capture–recapture methods with camera traps in cloud forest in Oaxaca state Mexico, while our model estimates a mean of 0.63 individuals/km^2 in the same study area. None of these differences are on the order that we see between expert estimates of the population for the entire species, compared to the total population estimate across viable cores using our model estimates.

3.5 | Habitat connectivity

Regardless of the estimated population size, we found clear areas with potential obstructions to connectivity. This was partly a result of the shape of Central America, as the land mass narrows between Nicaragua and Colombia, limiting the amount of land available for Baird's tapirs, but also due to widespread forest loss in several countries, particularly in central Nicaragua and northern Honduras/southern Guatemala. The Panama Canal and the proposed Nicaragua Canal (Figure 2) also pose significant potential genetic barriers. The Panama Canal already presents a substantial barrier to connectivity between habitat cores for Baird's tapir (Meyer et al., 2015) although there are anecdotal data of tapirs crossing the canal zone (Meyer, Moreno, & Jansen, 2013). The proposed construction of the larger Nicaragua Canal would further fragment the species distribution and almost certainly pose a permanent genetic barrier because the areas with the highest density of canal-related infrastructure coincide with the only remaining areas of viable Baird's tapir habitat in the vicinity of the proposed canal zone (Jordan et al., 2016). Therefore, the proposed Nicaragua Canal would isolate populations in the two major cores to the north, from several substantial cores to the south, including the Talamanca as discussed above, and a large habitat core in the Darién

of Panama, which extends into the Chocó of Colombia. These two areas represent two of the largest cores for Baird's tapir, as the fifth and third largest cores, respectively. It is unclear to what extent this could affect the long-term survival of Baird's tapirs given their previously reported low genetic diversity (Norton & Ashley, 2004).

4 | CONCLUSION

The main advantage of the integrated SDM framework tested here is the ability to use both presence-only and presence-absence data in estimating an SDM. This is especially advantageous for wide-ranging, but relatively rare, species such as the Baird's tapir. Modelling the entire distribution of these species requires the combination of several independently collected data sets, some of which may not include absence data.

One of the most exciting possibilities of the Integrated SDM, and point process models in general, is the ability to estimate species population size by summing intensities. This can be carried out because the expected number of individuals in a given area is equal to the sum of the estimated intensities across that area (Diggle, 2013). However, care should be taken when interpreting a fitted PPM and what is actually being modelled (Renner et al. 2015). The Integrated SDM used in this research includes the observation process as a latent variable, and thus, it should estimate the true species abundance (not relative patterns in species abundance). However, our research highlights the need for further work to validate the population estimates from the model due to discrepancies between estimated intensity and expert opinion on total population size for the species.

It is possible that the seemingly high population estimates from this model for Baird's tapir are due to missing proximate (i.e., causal) environmental variables, meaning the model is closer to the potential (i.e., carrying capacity) than the actual distribution of the species (Austin, 2002). In other words, areas identified as suitable may not actually be occupied by the species due to movement constraints (e.g., dispersal barriers), competition with other species, disease, hunting pressure or other variables not included in the model.

It is also possible that duplicate records of single individuals exist in the data we used to fit the model, despite our efforts to reduce this possibility using spatial subsampling with a minimum allowable distance of 5 km. Future research using simulated data should investigate the impact of duplicate records on the modelled intensity. Although it is possible some duplicates remain in the data used to fit the model after subsampling, this number should be small, and the effect on estimated intensity minimal. Simulated data could be ideal for testing the magnitude of this effect.

A sensitivity analysis of the Integrated SDM framework using the same Baird's tapir data used in this research found that the spatial resolution of the analysis has a strong effect on the magnitude of intensity and total population estimates (Schank et al. in prep). It is suspected that this is due to the effective sample area being expanded by the individual movements of the species, and the way area is included in the model formulation. Thus, any simulation that seeks to understand the model better should incorporate animal movement.

ORCID

Cody J. Schank http://orcid.org/0000-0002-8380-5634

Michael V. Cove http://orcid.org/0000-0001-5691-0634

Eduardo Mendoza http://orcid.org/0000-0001-6292-0900

Christopher A. Jordan http://orcid.org/0000-0002-8498-6721

Jose F. González-Maya http://orcid.org/0000-0002-8942-5157

Francisco Botello http://orcid.org/0000-0002-1133-2831

Marina Rivero http://orcid.org/0000-0002-8207-4171

Jesús Antonio de la Torre http://orcid.org/0000-0003-1994-8478

Margot A. Wood http://orcid.org/0000-0001-6004-7378

REFERENCES

Aitken, M., Roberts, D.W., & Shultz, L.M. (2007). Modelling distributions of rare plantsin the great basin, Western North America. *Western North American Naturalist*, *67*, 26–38.

Austin, M. P. (2002). Spatial prediction of species distribution: An interface between ecological theory and statistical modelling. *Ecological Modelling*, *157*(2), 101–118.

Barbet-Massin, M., Jiguet, F., Albert, C. H., & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: How, where and how many? *Methods in Ecology and Evolution/British Ecological Society*, *3*(2), 327–338.

Beck, J., Böller, M., Erhardt, A., & Schwanghart, W. (2014). Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecological Informatics*, *19*, 10–15.

Botello, F., Romero-Calderón, A. G., Sánchez-Hernández, J., Hernández, O., López-Villegas, G., & Sánchez-Cordero, V. (In press) Densidad poblacional del tapir centroamericano (Tapirella bairdii) en bosque mesófilo de montaña en Totontepec Villa de Morelos, Oaxaca, México. *Revista Mexicana de Biodiversidad*.

Brooks, D. M., Bodmer, R. E., & Matola, S. (1997). *Tapirs. Status Survey and Conservation Action Plan*. Gland, Switzerland: IUCN.

Cade, B. S. (2015). Model averaging and muddled multimodel inferences. *Ecology*, *96*(9), 2370–2382.

Carbajal-Borges, J. P., Godínez-Gómez, O., & Mendoza, E. (2014). Density, abundance and activity patterns of the endangered *Tapirus bairdii* in one of its last strongholds in southern Mexico. *Tropical Conservation Science*, *7*, 100–114.

Carrillo Reyna, N. L., Weissenberger, H., & Reyna-Hurtado, R. (2015). Distribución potencial del Tapir Centroamericano en la península de Yucatán. *Therya*, *6*(3), 575–596.

Castellanos, A., Foerester, C., Lizcano, D. J., Naranjo, E., Cruz-Aldan, E., Lira-Torres, I., ... Gonzalez-Maya, J. (2008). *Tapirus bairdii*. The IUCN Red List of Threatened Species.

Cove, M. V., Vargas, L. E. P., de la Cruz, J. C., Spínola, R. M., Jackson, V. L., Saénz, J. C., & Chassot, O. (2014). Factors influencing the occurrence of the Endangered Baird's tapir *Tapirus bairdii*: Potential flagship species for a Costa Rican biological corridor. *Oryx*, *48*(03), 402–409.

Diggle, P. J. (2013). *Statistical analysis of spatial and spatio-temporal point patterns*. Boca Raton, FL: CRC Press.

Dorazio, R. M. (2012). Predicting the geographic distribution of a species from presence-only data subject to detection errors. *Biometrics*, *68*(4), 1303–1312.

Dorazio, R. M. (2014). Accounting for imperfect detection and survey bias in statistical analysis of presence-only data. *Global Ecology and Biogeography*, *23*(12), 1472–1484.

Dunn, M., Estrada, N., & Smith, D. A. (2011). The coexistence of Baird's tapir (*Tapirus bairdii*) and indigenous hunters in northeastern Honduras. *Integrative Zoology*, *7*(4), 429–438.

Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., ... Zimmermann, N. E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, *29*(2), 129–151.

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, *17*(1), 43–57.

Engler, R., Guisan, A., & Rechsteiner, L. (2004). An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Journal of Applied Ecology*, *41*(2), 263–274.

Fletcher, R. J. Jr, McCleery, R. A., Greene, D. U., & Tye, C. A. (2015). Integrated models that unite local and regional data reveal larger-scale environmental relationships and improve predictions of species distributions. *Landscape Ecology*, *31*(6), 1–14.

Foerster, C. R., & Vaughan, C. (2002). Home range, habitat use, and activity of Baird's Tapir in Costa Rica. *Biotropica*, *34*(3), 423–437.

Franklin, J. (2010). *Mapping species distributions: Spatial inference and prediction*. Cambridge, UK: Cambridge University Press.

Garcìa, M., Jordan, C., O'Farrill, G., Poot, C., Meyer, N., Estrada, N., ... Ruiz-Galeano, M. (2016). *Tapirus bairdii*. The IUCN Red List of Threatened Species 2016: e.T21471A45173340. Downloaded on 26 July 2016.

García, M., Leonardo, R., Castillo, F., Gómez, I., & García, L. (2010). Efecto de la cacería en la viabilidad poblacional del tapir (*Tapirus bairdii*) en el Parque Nacional Laguna Lachuá, Guatemala. *Revista Científica de la Facultad de Ciencias Químicas y Farmacia, Universidad de San Carlos de Guatemala*, (19), 33–36.

Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis*, *24*(3), 189–206.

González-Maya, J. F., Schipper, J., Polidoro, B., Hoepker, A., Zárrate-Charry, D., & Belant, J. L. (2012). Baird's tapir density in high elevation forests of the Talamanca region of Costa Rica. *Integrative Zoology*, *7*(4), 381–388.

Hijmans, R. J., Phillips, S., Leathwick, J., & Elith, J. (2014). dismo: Species distribution modeling. R package version 1.0-5. http://CRAN.R-project.org/package=dismo.

Isaac, N. J., Turvey, S. T., Collen, B., Waterman, C., & Baillie, J. E. (2007). Mammals on the EDGE: Conservation priorities based on threat and phylogeny. *PLoS ONE*, *2*(3), e296.

Jordan, C. A. (2015). The dynamics of wildlife and environmental knowledge in a bioculturally diverse coupled natural and human system in the Caribbean region of Nicaragua. Michigan State University.

Jordan, C. A., Galeano, M. R., & Alonzo, A. S. (2014). La cacería histórica de tapires Centroamericanos (*Tapirus bairdii*) en la RAAS, Nicaragua. *Estudios Ambientales*, *1*(1), 73–87.

Jordan, C. A., Schank, C. J., Urquhart, G. R., & Dans, A. J. (2016). Terrestrial Mammal Occupancy in the Context of Widespread Forest Loss and a Proposed Interoceanic Canal in Nicaragua's Decreasingly Remote South Caribbean Region. *PLoS ONE*, *11*(3), e0151372.

Koshkina, V., Wang, Y., Gordon, A., Dorazio, R. M., & Stone, L. (2017). Integrated species distribution models: Combining presence-background data and site-occupancy data with imperfect detection. *Methods in Ecology and Evolution*, *8*(4), 420–430.

Koster, J. M. (2007). Assessing the sustainability of Baird's tapir hunting in the Bosawas Reserve, Nicaragua. Tapir Conservation Newsletter, 15/2(20), 23–28.

Lahoz-Monfort, J. J., Guillera-Arroita, G., & Wintle, B. A. (2014). Imperfect detection impacts the performance of species distribution models. *Global Ecology and Biogeography*, *23*(4), 504–515.

Lesmeister, D. B., Nielsen, C. K., Schauber, E. M., & Hellgren, E. C. (2015). Spatial and temporal structure of a mesocarnivore guild in Midwestern North America. *Wildlife Monographs*, *191*(1), 1–61.

Lomba, A., Pellissier, L., Randin, C., Vicente, J., Moreira, F., Honrado, J., & Guisan, A. (2010). Overcoming the rare species modelling paradox: A novel hierarchical framework applied to an Iberian endemic plant. *Biological Conservation*, *143*(11), 2647–2657.

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A., & Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, *83*(8), 2248–2255.

MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L. L., & Hines, J. E. (2006). *Occupancy estimation and modeling: Inferring patterns and dynamics of species occurrence*. London, UK: Academic Press.

McCann, N., Wheeler, P. M., Coles, T., & Bruford, M. W. (2012). Rapid ongoing decline of Baird's tapir in Cusuco National Park, Honduras. *Integrative Zoology*, *7*(4), 420–428.

McClintock, B. T., White, G. C., Antolin, M. F., & Tripp, D. W. (2009). Estimating abundance using mark–resight when sampling is with replacement or the number of marked individuals is unknown. *Biometrics*, *65*(1), 237–246.

Medici, E. P., Carrillo, L., Montenegro, O. L., Miller, P. S., Carbonell, F., Chassot, O., ... Mendoza, A. (2005). Baird's Tapir (Tapirus bairdii) Conservation Workshop Population and Habitat Viability Assessment (PHVA). Baird's Tapir Conservation Workshop: Final Report. IUCN/SSC Tapir Specialist Group (TSG) & IUCN/SSC Conservation Breeding Specialist Group (CBSG).

Medici, E. P., & Desbiez, A. L. J. (2012). Population viability analysis: Using a modeling tool to assess the viability of tapir populations in fragmented landscapes. *Integrative Zoology*, *7*(4), 356–372.

Mejía-Correa, S., Diaz-Martinez, A., & Molina, R. (2014). Densidad y hábitos alimentarios de la danta Tapirus bairdii en el Parque Nacional Natural Los Katios, Colombia. *Tapir Conservation*, *23*, 16–23.

Mendoza, E., Fuller, T. L., Thomassen, H. A., Buermann, W., Ramírez-Mejía, D., & Smith, T. B. (2013). A preliminary assessment of the effectiveness of the Mesoamerican Biological Corridor for protecting potential Baird's tapir (*Tapirus bairdii*) habitat in southern Mexico. *Integrative Zoology*, *8*(1), 35–47.

Meyer, N. F. V., Esser, H. J., Moreno, R., van Langevelde, F., Liefting, Y., Oller, D. R., ... Jansen, P. A. (2015). An assessment of the terrestrial mammal communities in forests of Central Panama, using camera-trap surveys. *Journal for Nature Conservation*, *26*, 28–35.

Meyer, N., Moreno, R., & Jansen, P. (2013). Distribution and conservation status of the Baird's tapir in Panama. *Tapir Conservation*, *30*, 10–13.

Naranjo-Piñera, E. (1995). Abundancia y uso de hábitat del tapir (*Tapirus bairdii*) en un bosque tropical húmedo de Costa Rica. *Vida Silvestre Neotropical*, *4*(1), 20–31.

Nelson, T. A., & Boots, B. (2008). Detecting spatial hot spots in landscape ecology. *Ecography*, *31*(5), 556–566.

Norton, J. E., & Ashley, M. V. (2004). Genetic variability and population differentiation in captive Baird's tapirs (*Tapirus bairdii*). *Zoo Biology*, *23*(6), 521–531.

O'Farrill, G., Galetti, M., & Campos-Arceiz, A. (2013). Frugivory and seed dispersal by tapirs: An insight on their ecological role. *Integrative Zoology*, *8*(1), 4–17.

Oliveira-Santos, L. G. R., Zucco, C. A., Antunes, P. C., & Crawshaw, P. G. Jr (2010). Is it possible to individually identify mammals with no natural markings using camera-traps? A controlled case-study with lowland tapirs. *Mammalian Biology*, *75*, 375–378.

Pastore, M. (2017). overlapping: Estimation of Overlapping in Empirical Distributions. Retrieved from https://CRAN.R-project.org/package=overlapping.

Phillips, S. J., Dudík, M., & Schapire, R. E. (2004). A maximum entropy approach to species distribution modeling. In Proceedings of the twenty-first international conference on Machine learning (p. 83). ACM.

R Core Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. URL http://www.R-project.org/.

Ramsey, D. S. L., Caley, P. A., & Robley, A. (2015). Estimating population density from presence-absence data using a spatially explicit model. *Journal of Wildlife Management*, *79*, 491–499.

Renner, I. W., Elith, J., Baddeley, A., Fithian, W., Hastie, T., Phillips, S. J., Popovic, G., & Warton, D. I. (2015). Point process models for presence-only analysis. *Methods in Ecology and Evolution*, *6*(4), 366–379.

Renner, I. W., & Warton, D. I. (2013). Equivalence of MAXENT and Poisson point process models for species distribution modeling in ecology. *Biometrics*, *69*(1), 274–281.

Reyna-Hurtado, R., Sanvicente-López, M., Pérez-Flores, J., Carrillo-Reyna, N., & Calmé, S. (2016). Insights into the multiannual home range of a Baird's tapir (*Tapirus bairdii*) in the Maya Forest. *THERYA*, *7*(2), 271–276.

Ripple, W. J., Newsome, T. M., Wolf, C., Dirzo, R., Everatt, K. T., Galetti, M., ... Van Valkenburgh, B. (2015). Collapse of the world's largest herbivores. *Science Advances*, *1*, e1400103.

Royle, J. A., Chandler, R. B., Yackulic, C., & Nichols, J. D. (2012). Likelihood analysis of species occurrence probability from presence-only data for modelling species distributions. *Methods in Ecology and Evolution*, *3*(3), 545–554.

Royle, J. A., & Dorazio, R. M. (2008). *Hierarchical Modeling and Inference in Ecology: The Analysis of Data from Populations, Metapopulations and Communities. Elsevier Science*.

Schank, C., Jordan, C. A., Nielsen, C. K., Lizcano, D. J., González-Maya, J. F., Cove, M. V., & Miller, J. A. (in prep.). Sensitivity of Integrated SDMs to spatial and temporal scale.

Schank, C., Mendoza, E., Vettorazzi, M. J. G., Cove, M. V., Jordan, C. A., O'Farrill, G., ... Leonardo, R. (2015). Integrating current range-wide occurrence data with species distribution models to map the potential distribution of Baird's Tapir. *Tapir Conservation*, *24*(33), 15–25.

Symonds, M. R. E., & Moussalli, A. (2011). A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology*, *65*(1), 13–21.

Tobler, M. W., Hibert, F., Debeir, L., & Richard-Hansen, C. (2014). Estimates of density and sustainable harvest of the lowland tapir *Tapirus terrestris* in the Amazon of French Guiana using a Bayesian spatially explicit capture–recapture model. *Oryx*, *48*(03), 410–419.

Vaz, U. L., Cunha, H. F., & Nabout, J. C. (2015). Trends and biases in global scientific literature about ecological niche models. *Brazilian Journal of Biology*, *75*(4), 17–24.

Warton, D. I., & Shepherd, L. C. (2010). Poisson point process models solve the "pseudo-absence problem" for presence-only data in ecology. *The Annals of Applied Statistics*, *4*(3), 1383–1402.

Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H., & Guisan, A. (2008). Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, *14*(5), 763–773.

Yackulic, C. B., Chandler, R., Zipkin, E. F., Royle, J. A., Nichols, J. D., Campbell Grant, E. H., & Veran, S. (2013). Presence-only modelling using MAXENT: When can we trust the inferences? *Methods in Ecology and Evolution*, *4*(3), 236–243.

Zaniewski, A. E., Lehmann, A., & Overton, J. M. (2002). Predicting species spatial distributions using presence-only data: A case study of native New Zealand ferns. *Ecological Modelling*, *157*(2), 261–280.

BIOSKETCH

Cody Schank is a Ph.D. candidate at the University of Texas at Austin in the Department of Geography and the Environment, where he studies species distributions and the statistical methods used to model them. He previously worked on the Map of Life at Yale University and for the IUCN on the Global Mammal Assessment.

Author contributions: C.J.S. conceived the research; C.J.S, M.V.C., M.J.K., E.M., G.O., R.R.H., N.M., C.A.J., J.F.G.M., R.M., M.D., V.M., C.S.B., N.E., J.C.C.D., A.C., J.F., C.K.N., F.B., G.P.M., M.R., J.A.T., E.B.M., O.G.G., M.A.W. and J.G. collected the data; C.J.S. analysed the data; and C.J.S. and J.A.M. led the writing.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.