Received: 2 September 2025 | Accepted: 8 January 2026

DOI: 10.1111/2041-210x.70252

Methods in Ecology and Evolution

RESEARCH ARTICLE

OccuGAMs: Non-linear occupancy and abundance modelling with imperfect detection

Johannes Maria Sassen[1] | Zachary Amir[1] | Nicholas Clark[2] | Matthew Scott Luskin[1]

[1]School of the Environment, University of Queensland, St. Lucia, Queensland, Australia

[2]School of Veterinary Science, University of Queensland, Gatton, Queensland, Australia

Correspondence
Johannes Maria Sassen
Email: joopsassen@icloud.com

Matthew Scott Luskin
Email: m.luskin@uq.edu.au

Funding information
Australian Research Council, Grant/Award Number: #DE210101440

Handling Editor: Phil J. Bouchet

Abstract

1. Hierarchical occupancy and abundance models (HOAMs) have become a leading approach for inferring wildlife population dynamics because they explicitly account for imperfect detection. HOAMs are suitable for sampling approaches that produce detection histories from repeated visits to the same sites, including direct observations (e.g. bird point counts), indirect observations (e.g. tracks, dung) and remote and passive sensors (e.g. camera traps, acoustic recorders).
2. Wildlife often exhibits non-linear temporal trends or threshold-like responses to environmental conditions. However, traditional HOAMs address non-linearity crudely using global polynomial functions, despite well-documented limitations. Generalised additive models (GAMs) provide a more flexible approach to non-linearity, allowing smooth data-driven estimation through basis functions and penalised splines. Yet, GAMs have remained sparsely adopted in hierarchical occupancy modelling, in part due to the need for custom code in Bayesian modelling languages.
3. We demonstrate the applicability of GAMs within the occupancy and abundance modelling framework (hereafter 'OccuGAMs') by comparing traditional HOAMs with polynomials to OccuGAMs. In simulations, OccuGAMs recovered non-linear relationships more accurately and more often, scoring better on energy and variogram metrics and produced more stable responses at smaller sample sizes. Polynomials performed well in some scenarios but were less generalisable, making OccuGAMs the more robust overall choice, especially when there is no a priori guidance about the functional form.
4. Limitations of OccuGAMs include interpretability of model parameters and sensitivity to the choice and number of basis functions, which can be assessed with diagnostic tools. To promote wider accessibility, we provide code for OccuGAM implementation in JAGS and Stan as well as in the R packages *flocker* and *mvgam*.

KEYWORDS
camera traps, generalised additive models, hierarchical modelling, non-linearity, tropical mammals

1 | INTRODUCTION

The ability to study animal populations in their natural habitats has been revolutionised by recent advances in wildlife monitoring technologies like camera traps and acoustic sensors (Burton et al., 2015; Lahoz-Monfort & Magrath, 2021; Rhinehart et al., 2020; Tabak et al., 2019). These tools enable the affordable collection of more observations across a range of species, often from remote or inaccessible areas (Bruce et al., 2025). Wildlife ecologists increasingly rely on advanced statistical models to analyse the large detection histories produced by cameras and eco-acoustics, such as species distribution modelling, hierarchical occupancy and abundance modelling, dynamic population monitoring, and structural equation modelling (Goldstein et al., 2024; Burton et al., 2015; Rozylowicz et al., 2024; Mendes & Luskin 2025). Hierarchical models are especially common and powerful because they explicitly separate the imperfect detection process from the ecological process that gave rise to the observation (i.e. state variables like occupancy and abundance) (MacKenzie et al., 2002; Royle, 2004). This is important because species detectability varies with environmental covariates (e.g. visibility in different vegetation types) and local behaviours, which can bias estimates of state variables and habitat associations (Gu & Swihart, 2004; Kéry & Schmidt, 2008).

Hierarchical occupancy and abundance modelling (hereafter 'HOAMs') traditionally use generalised linear models (GLMs) in the detection and state variable equations (Kéry & Schaub, 2011; MacKenzie et al., 2002). GLMs are comparatively rigid and suboptimal for capturing the shape of non-linear habitat associations or temporal trends. A common strategy to address this limitation is to incorporate a polynomial term for the covariates of interest (Briscoe et al., 2021; Dehaudt et al., 2022; Ruiz-Gutiérrez et al., 2010; Warton et al., 2017; Wearn et al., 2017). However, polynomial approximations exhibit several well-documented limitations. Polynomial terms introduce global basis functions, which constrain the response curve to exhibit smooth but inherently non-local behaviour (Harrell, 2015; Magee, 1998). This often results in implausible symmetries, boundary instability and heightened sensitivity to outliers (Aho et al., 2014; Gelman & Imbens, 2019; Guisan et al., 2002; Harrell, 2015; Magee, 1998). Moreover, polynomial models are typically unregularised, lacking mechanisms to enforce local smoothing. As a result, data in one region of the covariate space can exert a disproportionate influence on the fit in distant regions (Harrell, 2015; Magee, 1998). Despite these serious limitations, polynomials remain a widely used method for modelling non-linearity.

HOAMs with linear or polynomial effects have been widely and successfully applied across ecological systems, but more complex non-linear relationships are also likely to be common. Threshold-like responses have been observed in several systems, including in relationships between the abundance of large-bodied animals and proximity to human infrastructure (Potvin et al., 2005), American forest birds and habitat fragmentation (Morante-Filho et al., 2015; Suarez-Rubio et al., 2013) and habitat-generalist tropical mammals and their proximity to food subsidies from oil palm plantations (Luskin, Albert, & Tobler, 2017; Luskin, Brashares, et al., 2017). More complex non-linear patterns can also arise through interacting ecological and anthropogenic processes. For example, human shield effects can lead to elevated mesopredator occupancy in areas where apex predators avoid humans, producing non-monotonic or hump-shaped responses across disturbance gradients (Heit et al., 2024; Moll et al., 2018). Detecting such nuanced responses can improve ecological understanding and management outcomes but may be overlooked in some fields (Bolker et al., 2013; Heit et al., 2024; Dehaudt et al., 2025). For example, a review of 162 animal ecology studies found that only 14.2% reported a test for non-linearity, suggesting that linearity assumptions rarely receive sufficient attention when fitting GLMs (Heit et al., 2024). Here, we assess approaches for incorporating flexible, non-linear relationships into HOAMs to better represent ecological complexity.

An underutilised approach to non-linear HOAMs is the integration of generalised additive models (GAMs) (Bled et al., 2013; Kéry & Royle, 2020). Such 'OccuGAMs' extend the HOAM framework by allowing smooth, data-driven estimation of non-linear effects through penalised splines. These smooth functions are expressed as a linear combination of basis functions:

$$f(X) = \sum_{\kappa=1}^{K} \beta_\kappa b_\kappa(X) \tag{1}$$

where b_κ are evaluated basis functions, β_κ are their coefficients, K is the number of basis functions and $f(X)$ is the resulting smooth function of the covariate X. In a Bayesian framework, the coefficient vector β is often assigned a multivariate normal distribution centred at zero with a precision matrix proportional to the weighted sum of penalty matrices (Wood, 2016):

$$\beta \sim \text{MVN}\left(0, \left(\sum_\gamma S_\gamma \times \Lambda_\gamma\right)^{-1}\right) \tag{2}$$

where S_γ is the penalty matrix, Λ_γ is a smoothing parameter and β is the vector of spline coefficients. The index γ denotes distinct penalty components rather than individual basis functions. These are necessary to support adaptive or tensor product smooths and to penalise null space components, ensuring propriety of the associated Gaussian priors (Wood, 2017). The penalty matrix defines how the function is penalised for roughness, while the smoothing parameter controls the strength of this penalty. Together, the penalty matrix and smoothing parameter control the flexibility of the fitted function by penalising excessive 'wiggliness' (Wood, 2017). Alternative prior formulations are also possible; for example, a random walk prior for β can be used with a large number of knots to achieve similar smoothness control (Lee et al., 2024). Unlike polynomials, where each coefficient is estimated independently for a fixed transformation of the covariate, penalised spline coefficients are estimated jointly (Wood, 2017). As a result, the model can adapt smoothly to non-linear patterns in data while limiting

overfitting through regularisation (Wood, 2017). Moreover, because the penalised spline basis functions used in GAMs have local support (they are only active over a limited portion of the covariate range), the influence of individual data points remains localised (Wood, 2017). This is valuable when the aim is to understand the functional form of habitat or detection associations (Strebel et al., 2014), such as when an ecological threshold induces a sharp increase, decrease or a plateauing shape (Morse et al., 2003; Rhodes et al., 2008). For example, in Southeast Asia, pig-tailed macaques and wild boar are habitat generalist species that have been shown to potentially reach hyperabundance (i.e. punctuated, extreme increases in abundance) in degraded areas where they raid crops, such as near oil palm plantations (Moore et al., 2023), rendering them a compelling case for exploring potential non-linear responses.

OccuGAMs have rarely been featured in the wildlife ecology literature to date, and their use in hierarchical modelling has been largely restricted to modelling spatial autocorrelation (Bled et al., 2013; Rushing et al., 2019, 2020; Strebel et al., 2014). This may partly be due to the advanced skillset required to fit these models, namely, custom-built formulas in a Bayesian modelling framework and solved in applications such as JAGS (Plummer, 2003), Stan (Carpenter et al., 2017) or Nimble (de Valpine et al., 2017). Coding expertise in Bayesian modelling languages has been a common hurdle for ecologists more generally (Bolker et al., 2013). Importantly, recent R packages (*mvgam* and *flocker*) now make it possible to fit types of HOAMs with smooth functions using 'out-of-the-box' functionality in familiar interfaces (Clark & Wells, 2023; Socolar & Mills, 2023). Specifically, *mvgam* supports abundance (N-mixture) and *flocker* supports occupancy modelling (Table 1). However, a direct comparison between 'OccuGAMs' and their GLM counterparts is currently lacking, representing a critical gap in understanding their relative performance and ecological applicability.

Here, we compare model accuracy and interpretability of traditional HOAMs with linear and polynomial GLM component to OccuGAMs using simulated and real data sets. We begin by systematically evaluating each model's ability to recover functional forms from simulated data, where the genuine species–disturbance relationship is explicitly defined. This allows us to quantify model performance in reconstructing the original response curves across varying degrees of complexity, from linear to highly non-linear trends. Second, we use a case study to examine the applicability of these approaches to empirical data from camera trapping in Southeast Asia. This system is characterised by high rates of anthropogenic pressures (Wilcove et al., 2013) and mammal communities that exhibit substantial variation in their response to these disturbances (Amir et al., 2022; Amir, Sovie, & Luskin, 2022). We investigate the functional shape of responses to anthropogenic disturbance and assess the degree to which non-linear capabilities of OccuGAMs are warranted in this ecological context. We hypothesised that OccuGAMs, with their flexibility to model complex non-linear relationships, would outperform conventional models in accurately estimating species–disturbance relationships, particularly when those responses exhibit marked curvature or threshold-like behaviour. We further posit that such non-linear patterns are common in species' ecological responses, especially among a small set of focal group-living generalist species known to reach hyperabundance. Finally, we provide all code in both the Stan and JAGS languages, and implementation in *mvgam* (Clark & Wells, 2023) and *flocker* (Socolar & Mills, 2023), to support broader use of these methods in hierarchical modelling.

TABLE 1 Overview of current R packages and software for estimating hierarchical occupancy or N-mixture models with predictors modelled using generalised additive models (OccuGAMs). *mvgam* and *flocker* are two recently released R packages that provide out-of-the-box functionality; *mvgam* supports abundance modelling and *flocker* supports occupancy modelling but note that neither package supports both.

Software	Description	Release	Sampler	Occupancy	Abundance	Formula interface
mvgam	Fit dynamic GAMs; State-space models with highly nonlinear predictor effects	2023	Hamiltonian Monte Carlo (Stan interface)	–	✓	Yes
flocker	Fit a variety of occupancy models using *BRMS* package syntax	2024	Hamiltonian Monte Carlo (Stan interface)	✓	–	Yes
JAGS/WinBUGS	General purpose Bayesian modelling software	2003	Gibbs sampling	✓	✓	No (manual specification)
Stan	General purpose Bayesian modelling software	2017	Hamiltonian Monte Carlo	✓	✓	No (manual specification)
Nimble	General purpose Bayesian modelling software enabling enhanced sampler customisation	2017	Multiple	✓	✓	No (manual specification)

2 | METHODS

2.1 | GAM-based hierarchical occupancy and abundance models (HOAMs)

We begin by describing the general model specifications to link GAMs and HOAMs in a unified analytical framework and then detail how we applied the approach to (i) a simulation study and (ii) an empirical case study using a large camera trapping data set.

2.1.1 | General formulation: Occupancy GAMs

We model species occurrence using a hierarchical occupancy framework with a latent state model and an observation model. For each sample site i, the latent occupancy state z_i is a Bernoulli random variable with probability ψ_i.

$$z_i \sim \text{Bernoulli}(\psi_i) \tag{3}$$

In conventional occupancy models, the occurrence probability ψ_i is linked to environmental covariates through a logit link function. Most commonly, this relationship is specified using either a linear predictor:

$$\text{logit}(\psi_i) = \beta_0 + \beta_1 X_i \tag{4}$$

or a polynomial extension to accommodate simple forms of non-linearity:

$$\text{logit}(\psi_i) = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \ \ldots \ + \beta_n X_i^n \tag{5}$$

Here, β_0 denotes the intercept, $\beta_1, \ \ldots \ , \beta_n$ are regression coefficients and X_i is a site-level covariate. Polynomial terms capture curvature but impose strong global structure and often fail to represent locally non-linear relationships. To relax these constraints, smooth terms can be incorporated into the linear predictor:

$$\text{logit}(\psi_i) = \beta_0 + s(X_i) \tag{6}$$

Here, $s(X_i)$ is a penalised spline smooth, represented as a linear combination of basis functions (e.g. thin-plate regression splines). This is conceptually equivalent to replacing the GLMs in formulae 4 and 5 with a GAM. Detection probability $p_{i,j}$ for site i and survey j can similarly be modelled using a GAM rather than a GLM to account for non-linear covariate effects on detectability:

$$\text{logit}(p_{i,j}) = \alpha_0 + s(Z_{i,j}) \tag{7}$$

where α_0 is the intercept, $Z_{i,j}$ represents site- or survey-level covariates (e.g. effort or observer) and $s(Z_{i,j})$ is a penalised spline smooth. The observed detection $y_{i,j}$ is then modelled as a Bernoulli random variable conditional on the true occupancy state z_i:

$$y_{i,j} \mid \ z_i \sim \text{Bernoulli}(z_i \times p_{i,j}) \tag{8}$$

2.1.2 | General formulation: N-mixture GAMs

N-mixture models are an extension of occupancy models used to model counts rather than binary detection data (Royle, 2004). The latent state is the abundance N_i at site i, typically modelled as a Poisson random variable with mean λ_i denoting the expected mean abundance, where covariate effects can be expressed using either linear predictors or smooth terms as seen in Section 2.1.1:

$$N_i \sim \text{Poisson}(\lambda_i) \tag{9}$$

$$\log(\lambda_i) = \beta_0 + \beta_1 X_i + \ \ldots \ + \beta_n X_i^n \ \mid \ \log(\lambda_i) = \beta_0 + s(X_i) \tag{10}$$

Observed counts $y_{i,j}$ for site i and survey j are then modelled conditionally on N_i, where detection probability $p_{i,j}$ can be modelled using a GAM or a GLM:

$$y_{i,j} \mid \ N_i \sim \text{Binomial}(N_i, p_{i,j}) \tag{11}$$

$$\text{logit}(p_{i,j}) = \alpha_0 + \alpha_1 Z_i + \ \ldots \ + \alpha_n Z_i^n \ \mid \ \text{logit}(p_{i,j}) = \alpha_0 + s(Z_{i,j}) \tag{12}$$

2.2 | Simulation study

To evaluate the ability of different model formulations to detect non-linear responses in species' responses to anthropogenic disturbance, we conducted a series of simulation experiments using the *mvgam* (Clark & Wells, 2023) and *flocker* R packages (Socolar & Mills, 2023), which both employ Stan as their backend. The *mvgam* package was used to simulate and fit N-mixture models, whereas *flocker* was used for occupancy models. The simulated relationships included both simple (linear) and complex (e.g. piecewise or non-monotonic) trends selected to reflect ecologically realistic patterns (Luskin, Brashares, et al., 2017; Moore et al., 2023; Spake et al., 2022).

2.2.1 | Data

The key elements of the simulation design are visualised in Figure 1. For abundance, we simulated species count histories for four hypothetical species, differing in mean abundance ($\mu = 100, \mu = 40$ or $\mu = 2$), base detection probability ($bp0 = 0.7$ or $bp0 = 0.2$) and response to detection covariate (i.e. is the detection probability of the species linearly or non-linearly related to a detection covariate). For each site, latent abundances were drawn from a Poisson distribution with mean $\lambda_{i,k,l}$ defined as the sum of the species-specific mean abundance μ_k, a covariate effect $x_{i,l}$ and added Gaussian noise $\epsilon_{i,k,l}$ with standard deviation proportional to the mean abundance ($0.2\mu_k$) for a simulated species k, state covariate response l and site i. The covariate–abundance relationship was modelled using one of six functional forms:

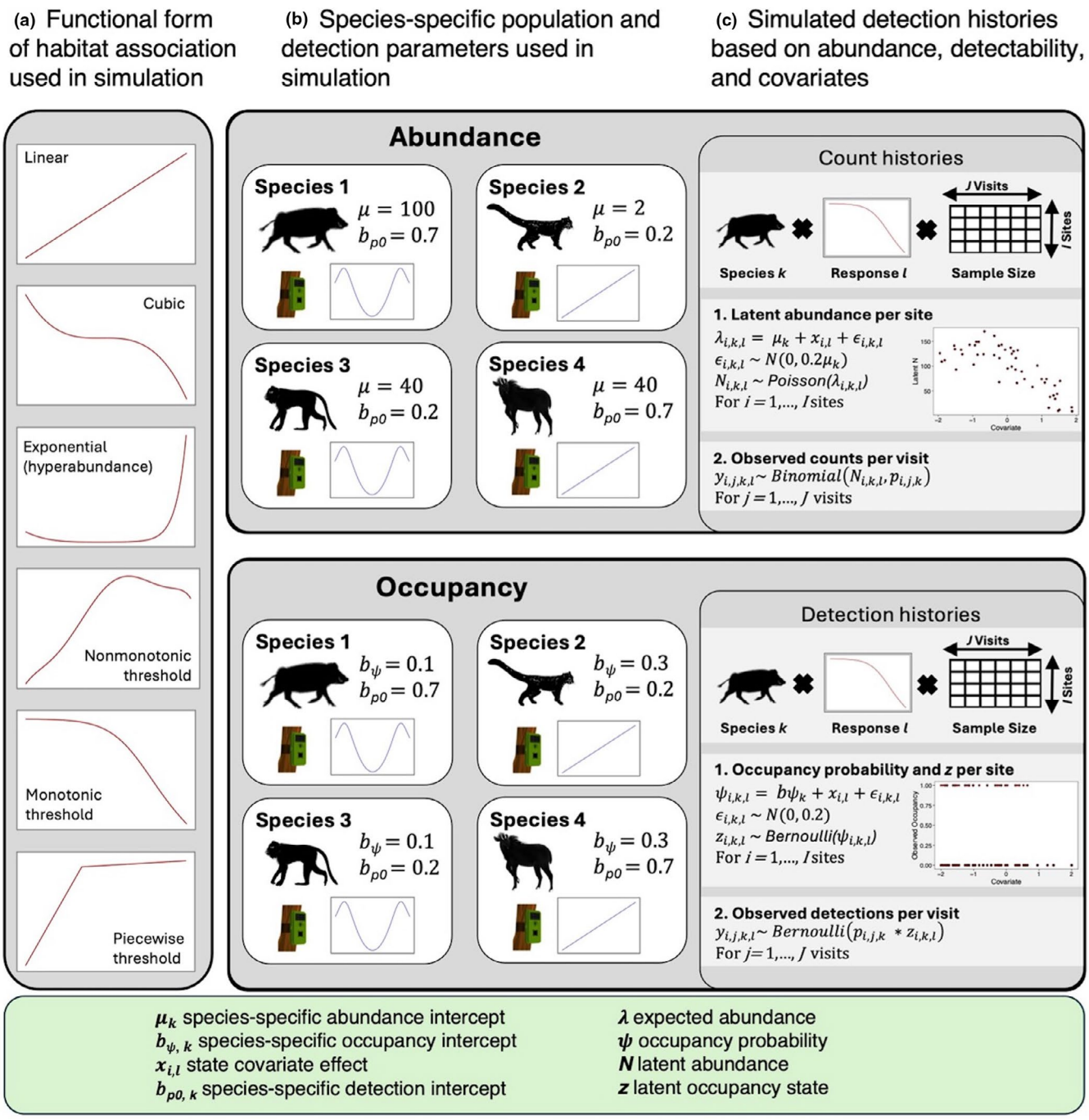

FIGURE 1 Conceptual overview of the simulation study design. (a) The six functional forms used to model the covariate–response relationship. (b) Simulated species with species-specific parameters. (c) Data generation workflow for abundance and occupancy simulations. For abundance, latent abundance $N_{i,k,l}$ at site i was drawn from a Poisson distribution with mean $\lambda_{i,k,l} = \mu_k + x_{i,l} + \varepsilon_{i,k,l}$, where μ_k is the species-specific mean abundance, $x_{i,l}$ is the covariate effect and $\varepsilon_{i,k,l} \sim N(0, 0.2\mu_k)$. Observed counts $y_{i,j,k,l}$ were generated from a binomial distribution with $N_{i,k,l}$ trials and detection probability $p_{i,j,k}$. Detection probabilities reflected a species-specific baseline detection parameter $b_{p0,\,k}$ modified by a detection covariate whose effect was applied on the logit scale, producing site- and visit-specific detection probabilities. For occupancy, latent occupancy states $z_{i,k,l}$ were drawn from a Bernoulli distribution with probability $\psi_{i,k,l} = b_{\psi,k} + x_{i,l} + \varepsilon_{i,k,l}$, where $b_{\psi,k}$ is a species-specific base occupancy probability and $\varepsilon_{i,k,l} \sim N(0, 0.2)$. Detection histories $y_{i,j,k,l}$ were generated from a Bernoulli model with success probability $p_{i,j,k,l} \times z_{i,k,l}$. We simulated 4800 data sets (96 scenarios × 50 replicates) across designs with $i = 10, 25, 50$ or 100 sites and $j = 10$ or 20 visits, for both abundance and occupancy.

(1) monotonic threshold, (2) linear, (3) cubic polynomial, (4) piecewise threshold, (5) hyperabundance and (6) non-monotonic threshold (Figure 1a). Observed counts $y_{i,j,k,l}$ for visit j at site i were generated using a binomial observation model, where each count was drawn from a binomial distribution with number of trials equal to the latent abundance $N_{i,k,l}$ and the success probability $p_{i,j,k,l}$ reflected the

species-specific baseline detection probability $b_{p0,\ k}$ modified by the effect of a detection covariate. The effect of the detection covariate could follow either a linear or non-linear relationship, depending on species, producing site- and visit-specific detection probabilities (Figure 1b). We simulated a range of study designs varying in the number of sites and replicates to represent small (10 sites × 10 visits), medium (25 sites × 10 visits), large (50 sites × 10 visits) and very large (100 sites × 20 visits) data sets. Each scenario was replicated 50 times, yielding 4800 unique count histories.

For occupancy, the simulation design was identical in terms of sites, visits and covariates, but we simulated presence–absence data instead of counts. Each species was assigned a mean occupancy probability ($b\psi$) to which the effect of the covariate and additional random noise was added, analogous to the mean abundance parameter for the count histories. Latent states $z_{i,k,l}$ were then drawn from a Bernoulli model with success probability $\psi_{i,k,l}$ and observations from a Bernoulli model with success probability $p_{i,j,k,l} \times z_{i,k,l}$. Across all scenarios and replicates, this procedure yielded 4800 unique occupancy data sets, directly comparable in structure to the abundance simulations.

2.2.2 | Models

For each simulated data set, we estimate four distinct N-mixture and occupancy model formulations that were univariate in the state formula: (1) a traditional model with linear terms; (2) a traditional model with quadratic terms; (3) a traditional model with a third-degree polynomial term; and (4) an OccuGAM with thin-plate regression splines. Each of these model types was fitted with two alternative detection sub-models: one specifying a linear effect of the detection covariate and one using a GAM. This yielded a total of eight model variants per data set, for a total of 76,800 models estimated. All N-mixture models were implemented using the *mvgam*() function from the *mvgam* R package (Clark & Wells, 2023), specifying the *nmix*() family. Occupancy models were implemented using the *flock*() function from the *flocker* R package (Socolar & Mills, 2023). We used weakly informative priors throughout. Gaussian or Student-*t* priors were placed on intercepts and regression coefficients. For models including smooth terms, additional Student-*t* priors were assigned to the corresponding smoothing parameters. In occupancy models, these priors acted directly on the standard deviation (SDs) hyperparameters controlling smoothness, whereas in N-mixture models fitted via *mvgam*, priors are set on the latent trend coefficients and associated penalty structure. We ran four chains of 1000 (N-mixture) and 5000 (occupancy) iterations, discarding the first 200 and 1000, respectively, as burn-in.

2.2.3 | Evaluation

The ability of each model formulation to recover the true shape of the response was evaluated using two proper scoring rules: the energy score (Gneiting & Raftery, 2007) and the variogram score (Scheuerer & Hamill, 2015). Importantly, both scores are computed from the full predictive distribution (i.e. the full set of posterior predictive MCMC draws) rather than from point estimates of the parameters. The energy score quantifies the overall discrepancy between the predicted distribution and true values, capturing differences in both location and spread and penalising predictions that are biased or that underestimate or overestimate uncertainty (Gneiting & Raftery, 2007). The variogram score evaluates discrepancies in pairwise distances among predictions, making it particularly informative for assessing whether a model correctly captures smoothness or correlation structure in the underlying response surface (Scheuerer & Hamill, 2015). Assessing both scores provides a comprehensive framework for evaluating the ability of the chosen statistical model to recover the true curve, where lower energy and variogram scores indicate better alignment with the underlying structure of the data. Energy and variogram scores were computed using the *es_sample*() and *vs_sample*() functions from the *scoringRules* R package (Jordan et al., 2019), respectively. To ensure all comparisons were valid, we excluded models with any parameter $\hat{R} > 1.2$.

We also evaluated the effects of alternative detection models (linear or GAM) on estimates of latent abundance N and occupancy probability ψ using the normalised root mean squared error (NRMSE) between simulated true values and model predictions. NRMSE was used in place of energy or variogram scores because it is sensitive to absolute deviations from the truth and therefore penalises systematic over- or underestimation, even when the fitted response curve has the correct shape. As a result, models that recover the correct functional form of the state process but misestimate ecological quantities due to misspecified detection are appropriately penalised with higher NRMSE values.

2.3 | Case studies: Tropical mammal responses to anthropogenic disturbances

To analyse camera trap data, we constructed hierarchical occupancy and N-mixture (i.e. abundance) models using basis-function formulations to model non-linear relationships between state variables and disturbance covariates. N-mixture models are particularly sensitive to violations of population closure assumptions, which can inflate absolute density estimates and produce to biased population inferences (Link et al., 2018), so we focus on interpreting abundance patterns only as directional responses to covariates; that is, relative abundance (Gilbert et al., 2021). We built these models in JAGS because it is the most used tool for wildlife ecologists (Kéry & Royle, 2020), but Nimble or Stan would have been equally appropriate.

2.3.1 | Data

We analysed data from 21 camera-trapping surveys conducted in 10 lowland tropical forest landscapes in Southeast Asia, including three

from Sumatra, two from Borneo, two from Peninsular Malaysia, two from Thailand and one from Singapore. Landscapes included a mix of protected areas, nearby production forests and smaller forest patches. Detailed descriptions of landscape and deployment methods relevant to this data set are provided in Amir, Moore, et al. (2022) and Amir, Sovie, and Luskin (2022) but we include a brief summary here for clarity.

Each landscape was surveyed with 22–112 unbaited infrared camera traps, positioned 20–30 cm above ground along wildlife trails. To address differences in camera spacing (typically >1 km in extensive forest blocks and <500 m in small fragments), we aggregated trap locations into 3-km^2 hexagonal sampling units. This resampling generated a spatially standardised data set, reducing spatial pseudo-replication and producing estimates of occupancy and abundance that were not confounded by variation in camera spacing (Amir, Moore, et al., 2022; Amir, Sovie, & Luskin, 2022; Rayan & Linkie, 2020). Where more than one camera fell inside a unit, environmental covariates were averaged, and detections were pooled across cameras. Consecutive records of the same species were treated as independent only if separated by at least 30 min (Rovero & Zimmermann, 2016). Data were then summarised into sampling occasions of 5 days to reduce false absences (Brodie et al., 2018). Permission to sample locations was granted by the relevant local authorities. In Singapore, permits were issued by the National Parks Board (NParks). In Indonesia, sampling permissions were granted by the Indonesian Institute of Sciences (LIPI), the Ministry of Forestry (Kemenhut) and the Ministry of Research and Technology (RISTEK) for Bukit Barisan Selatan, Kerinci Seblat and Gunung Leuser, respectively. Permissions to sample at Lambir Hills were provided by the Sarawak Forestry Corporation and the Sarawak Biodiversity Centre. In Peninsular Malaysia, permits for multiple sites were issued by the Forestry Research Institute of Malaysia (FRIM). Sampling at Danum Valley was approved by the Sabah Biodiversity Centre (SaBC), the Sabah Wildlife Department and the Danum Valley Management Committee (DVMC). In Thailand, permissions were granted by the Department of National Parks, Wildlife and Plant Conservation (DNP) and the Ministry of Natural Resources and Environment (MNRE) for Khao Yai National Park and Khao Banthat National Park.

We modelled the occupancy and abundance of pig-tailed macaques (*Macaca nemestrina*), wild boar (*Sus scrofa*), sambar deer (*Rusa unicolor*) and muntjak deer (genus *Muntiacus*) in heterogeneous environments. We produced both count history matrices for N-mixture models and detection history matrices for occupancy models (Mendes & Luskin, 2025). We included four measures of anthropogenic disturbance as covariates: forest integrity (Grantham et al., 2020), the human footprint index (Venter et al., 2016), oil palm cover (Miettinen et al., 2016) and forest cover (Hansen et al., 2013).

2.3.2 | Models

For the occupancy analysis, we estimated the probability that a species was present in grid cell i (3-km^2 hexagons) within landscape g and year t. The latent state was modelled as

$$z_i \sim \text{Bernoulli}(\psi_i) \tag{13}$$

where z_i is a Bernoulli random variable with probability ψ_i, representing the latent state in grid cell i. We model the occupancy probability ψ_i as a smooth function of a single environmental covariate, with

$$\text{logit}(\psi_i) = \beta_{0,(g,t)} + s(X_i) \tag{14}$$

where $\beta_{0,(g,t)}$ is a random intercept for each landscape g within year t and $s(X_i)$ is a thin-plate regression spline smooth. Each model includes only one disturbance covariate (i.e. percent oil palm, forest cover, forest integrity or human footprint), so X denotes the specific environmental variable used in that model. The smooth term $s(X_i)$ was represented using $K = 5$ thin-plate regression spline basis functions:

$$s(X_i) = \sum_{\kappa=1}^{K} \left(\beta_\kappa b_\kappa(X_i) \right) \tag{15}$$

where K represents the number of basis functions, $b_\kappa(*)$ denote the basis functions and β_κ are their respective coefficients. Restricting the basis dimension to $K = 5$ constrains the maximum complexity of the smooth and reduces computation time (Large et al., 2013; Samhouri et al., 2017). To evaluate alternative representations of the covariate effect, we also estimated polynomial variants of increasing order:

$$\text{logit}(\psi_i) = \beta_{0,(g,t)} + \sum_{n}^{m} \beta_n X_i^n \tag{16}$$

where $m = 1, 2, 3$ corresponds to linear, quadratic and cubic forms, respectively (Formula 16).

For the observation model, detections $y_{i,j}$ in grid cell i during sampling window j were modelled conditional on the latent state z_i and detection probability $p_{i,j}$. Detection probability incorporated camera effort and an overdispersion random effect:

$$y_{i,j} \mid z_i \sim \text{Bernoulli}(z_i \times p_{i,j}) \tag{17}$$

$$\text{logit}(p_{i,j}) = \alpha_0 + \alpha_1 \text{CameraEffort}_{i,j} + e_{i,j} \tag{18}$$

$$e_{i,j} \sim \text{Normal}(0, \tau) \tag{19}$$

where $e_{i,j}$ accounts for residual spatial and temporal heterogeneity in detection (an overdispersion random effect, ODRE) (Amir, Sovie, & Luskin, 2022; Kéry & Royle, 2016) and τ represents the standard deviation of the overdispersion random effect. Camera effort was defined as the number of active cameras in grid cell i during sampling window j. We assumed population closure during the survey period (MacKenzie et al., 2002).

For the abundance analysis, we estimated the latent abundance N_i at site i within landscape g and year t. The latent state was modelled as a Poisson random variable:

$$N_i \sim \text{Poisson}(\lambda_i) \tag{20}$$

$$\log(\lambda_i) = \beta_{0,(g,t)} + s(X_i) \tag{21}$$

where $\beta_{0(g,t)}$ is a landscape-year random intercept and $s(X_i)$ represents a thin-plate regression spline with $K = 5$ basis functions. As with occupancy models, we also estimated three polynomial formulations (linear, quadratic, cubic) for each disturbance covariate:

$$\log(\lambda_i) = \beta_{0,(g,t)} + \sum_{n}^{m} \beta_n X_i^n \quad (22)$$

where $m = 1, 2, 3$ corresponds to linear, quadratic and cubic forms, respectively (Formula 22). Finally, the observation model was identical to that of the occupancy model in that it included an effect for camera effort and the ODRE. We assumed population closure during the survey period (MacKenzie et al., 2002).

$$y_{i,j} \mid N_i \sim \text{Binomial}(p_{i,j}, N_i) \quad (23)$$

$$\text{logit}(p_{i,j}) = \alpha 0_i + \alpha 1_i \times \text{Camera Effort}_{i,j} + e_{i,j} \quad (24)$$

$$e_{i,j} \sim \text{Normal}(0, \tau) \quad (25)$$

2.3.3 | Evaluation

We fit four models for each species–covariate combination: linear, quadratic, cubic and GAM (OccuGAM), yielding 64 occupancy models and 64 N-mixture models in total. We quantified the difference between the estimated relationships of GAM and non-GAM models by calculating the NRMSE relative to the GAM. NRMSE measures the deviation between trends, with larger values reflecting more deviation. NRMSE is a normalised metric, with the range of the fitted curve used as the quantity of normalisation. Comparisons of NRMSE are therefore relative and only meaningful between the different model types within a single species–covariate combination. We inspected model goodness of fit and overdispersion through calculating Bayesian *p*-values and *C*-hat values via posterior predictive checks (Conn et al., 2018; Gelman et al., 1996). Bayesian *p*-values between 0.25 and 0.75 conventionally indicate adequate model fit, with values equalling 0.5 deemed a 'perfect fit' (Conn et al., 2018; Gelman et al., 1996; Kéry & Royle, 2016). *C*-hat values are reflective of overdispersion; we treat *C*-hat values >1.1 as indicative of model overdispersion (Kéry & Royle, 2016). In addition, we compared model predictive performance through Leave-One-Out (LOO) cross-validation (Vehtari et al., 2017). LOO is a technique for assessing pointwise out-of-sample predictive accuracy by calculating the log-likelihood at the posterior simulations of the model parameters (Vehtari et al., 2017). We computed the expected log pointwise predictive density (elpd) for each model using the *loo* R package (Vehtari et al., 2017), and then compared the four model variants (GAM, linear, quadratic, cubic) for each species–covariate combination using the *loo_compare* function from the *loo R* package (Vehtari et al., 2017), which computes differences in elpd values alongside the standard deviations (σ) with the magnitude of the mean difference relative to σ used to assess the strength of evidence. Specifically, models with $|\Delta\text{elpd}| > 2\sigma$ were considered to have meaningful differences in predictive accuracy, and models with lower values indicate better predictive performance.

3 | RESULTS

3.1 | Functional shape recovery from simulated data

OccuGAMs outperformed the linear, quadratic and cubic formulations in recovering the true abundance and occupancy relationships across the full range of sample sizes (Figure 2; Figure S7). Mean energy ranks for OccuGAMs were stable across sample sizes, whereas the polynomial models improved with increasing data, reflecting a dependence on larger samples to accurately estimate curvature (Figure 2). Rankings based on the variogram score were broadly similar (Figure S7).

Across scenarios, polynomial models sometimes captured broad features of the non-linear response but often lacked the flexibility to recover more complex or irregular shapes (Figure 3). Some specific matches did occur; for example, the quadratic model approximated the monotonic threshold relationship very well (Figures S10–S13). Polynomials performed worse when the true relationship was linear: both quadratic and cubic models produced higher energy ranks in that scenario for occupancy and abundance, indicating reduced accuracy in recovering the underlying shape (Figures S10–S13 and S15). OccuGAMs were comparatively robust. They ranked within the top two models in the linear scenarios, trailing only the linear model (Figures S10–S13), and maintained strong performance across all response shapes, with mean energy and variogram ranks rarely exceeding two (Figures S10–S13). By contrast, the performance of polynomial models was more variable. Although OccuGAM performance was largely stable across sample sizes, species' mean abundance had a clear effect. In the N-mixture analysis, OccuGAMs performed worse for species two, which had a mean abundance of two (Figure S16).

Absolute performance declined for all models as sample size decreased, indicated by reductions in raw energy scores (Figures S8 and S9), consistent with reduced ability to reconstruct functional responses under limited data. Estimation of non-linear functions was unreliable at a sample size of 10 sites, with high variance and clear deviations from the true response (Figures S8 and S9). Performance improved substantially at sample sizes of 25 and above, where most non-linear models recovered the main features of the generating functions. We note that, although OccuGAMs often had better energy and variogram scores, the absolute differences between estimated responses were often small, especially as sample sizes increased (Figures S8 and S9).

3.2 | Impact of detection sub-model on occupancy and abundance estimates

For N-mixture models, incorporating GAMs into the detection sub-model improved estimates of latent abundance when the

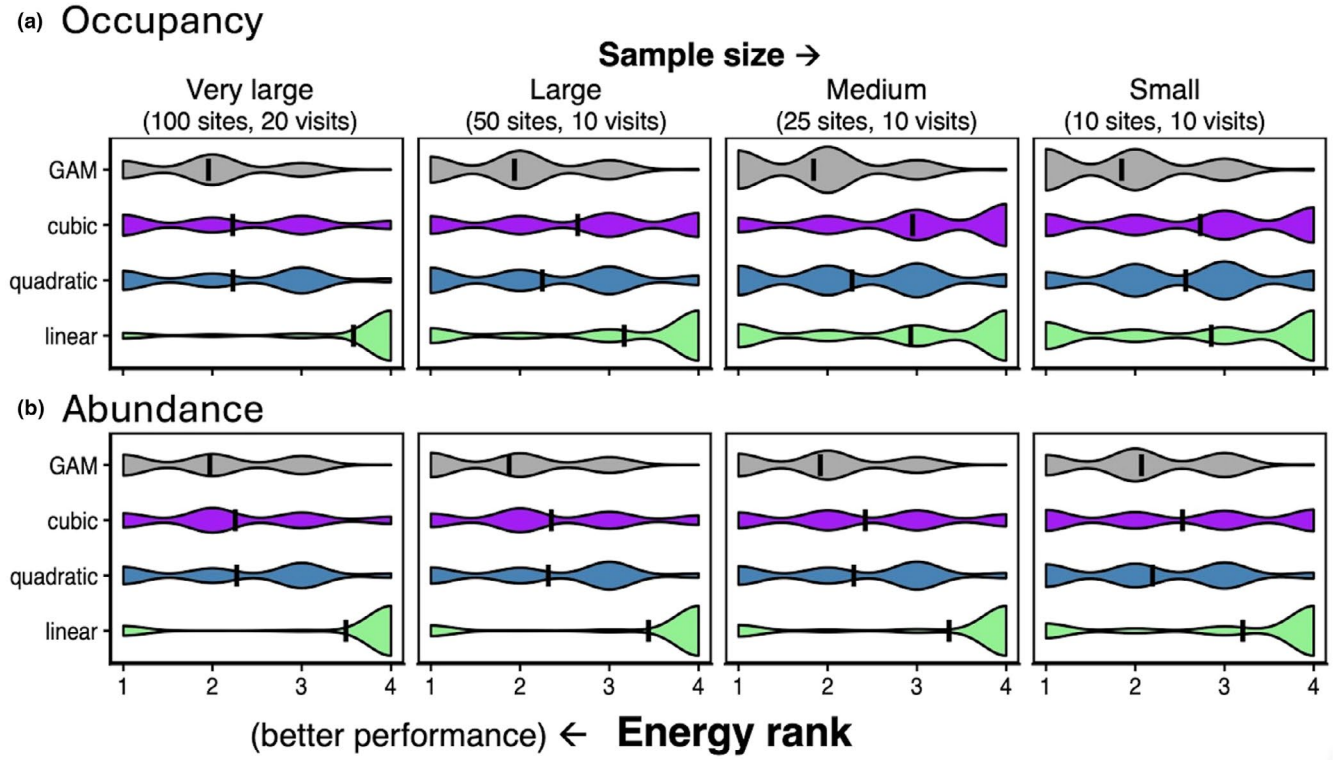

FIGURE 2 Energy rank across sample sizes for (a) occupancy and (b) abundance (N-mixture), comparing linear, quadratic, cubic and generalised additive models (GAM) hierarchical models. Panel titles indicate the sample size; very large (100 sites × 20 visits), large (50 × 10), medium (25 × 10) and small (10 × 10). Violin plots show the distribution of ranks across all simulated data set and model combinations. Models with any parameter $\widehat{R} > 1.2$ were excluded. Lower ranks indicate better performance; black vertical lines denote mean ranks.

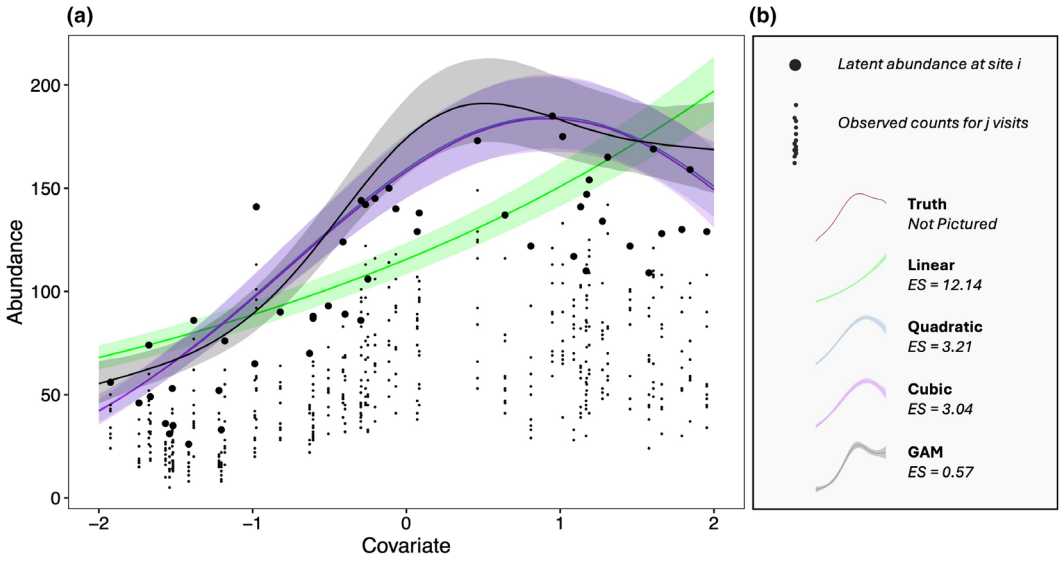

FIGURE 3 Model fits for a single simulated abundance data set from the non-monotonic scenario. This data set has a large sample size and non-linear detection covariate. The lines show posterior predictive means; shaded bands show 95% credible intervals. Latent site abundances (black points) and observed counts (small points) are shown. The energy scores for each model formulation highlight differences in ability to recover the true shape. In this example, all models contained a generalised additive models (GAM) in the detection formula.

true detection probability varied non-linearly with a detection covariate (Figure 4). Non-linear detection functions substantially affected N-mixture model estimates, with linear detection sub-models generally yielding higher NRMSE and poorer recovery of true abundance (Figure 4). Allowing the detection sub-model to match the non-linear covariate through integrating a

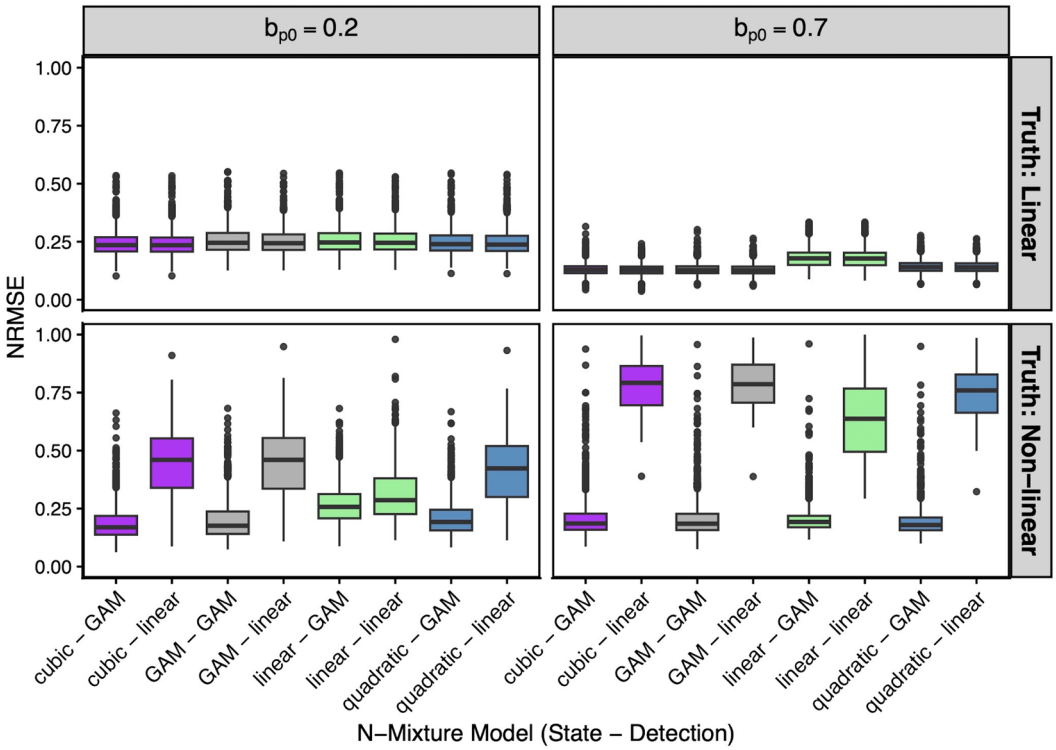

FIGURE 4 The effect of detection probability on OccuGAMs. The *y*-axis shows the discrepancy between true latent abundance *N* and N-mixture model estimates across base detection probabilities (b_{p0}), detection covariates (linear or non-linear) and detection model formulations (linear or generalised additive models [GAM]) measured by the normalised root mean squared error (NRMSE) (lower is better). NRMSE is calculated as the root of the mean squared difference between predicted and true *N*, normalised by the range of true *N*. Boxplots show NRMSE on the *y*-axis versus the eight model types on the *x*-axis, defined by crossing four state model forms with two detection model forms (linear, GAM). Models with any parameter $\hat{R} > 1.2$ were excluded. The non-linear detection covariate follows a U-shaped pattern (Figure 1; species 1 and 3).

GAM reduced NRMSE, improving fidelity of abundance estimates across state model formulations. N-mixture models using a linear detection sub-model to estimate a non-linear covariate were also more likely to produce unacceptable $\hat{R}$ values: of 38,400 models, 6320 had at least one parameter with $\hat{R} > 1.2$, and 4581 of these were linear detection models with the true detection covariate being non-linear (Table S2).

There were no clear differences between the various sub-model formulations in the occupancy model analysis. NRMSE was largely unaffected by the form of the detection covariate or detection sub-model (Figure S14). Occupancy models were less likely to have parameters $\hat{R} > 1.2$ compared to the N-mixture models, with only 581 models affected (Table S1).

3.3 | Fitting traditional HOAMs and OccuGAMs to real data

This comparison of HOAMs with linear and polynomial terms versus OccuGAMs is different since the true relationship is unknown. There was no clear difference in model goodness of fit or overdispersion between OccuGAM and polynomial models (Tables S3 and S4). All N-mixture models had Bayesian *p*-values between 0.25 and 0.75 and *C*-hat values <1.1 indicating acceptable model fit and no evidence of overdispersion (Table S4).

Most models showed high alignment across species and variables, characterised by a consistent directionality of trends among the traditional models and OccuGAMs (Figure 5; Figures S1–S3). Overall, habitat associations between occupancy and the four disturbance covariates were less pronounced than for relative abundance—the majority were flat or slightly increasing (Figures S1–S3). However, the notable exceptions were strong negative relationships for occupancy of *sambar~oil palm* and *muntjak~oil palm* (Figure 5c,d). We present the results from the N-mixture models in the main text and occupancy results in the Supporting Information.

NRMSE suggested the habitat association curve that most closely matched the one produced by the GAM varied depending on the species and covariate. No systematic bias was observed favouring simpler models or more complex ones. Crucially, there were three examples in which the GAM deviated markedly from all model types: *macaque~forest cover* (Figure S1a), *muntjak~forest cover* (Figure S1d) and *macaque~human footprint* (Figure S3a). In the first two examples, the GAM produced an initial flat relationship between relative abundance and forest cover, followed by a punctuated increase which was poorly approximated by both quadratic and cubic models (Figure S1a,d). In the latter (*macaque~human footprint*), there was a steep initial decline with

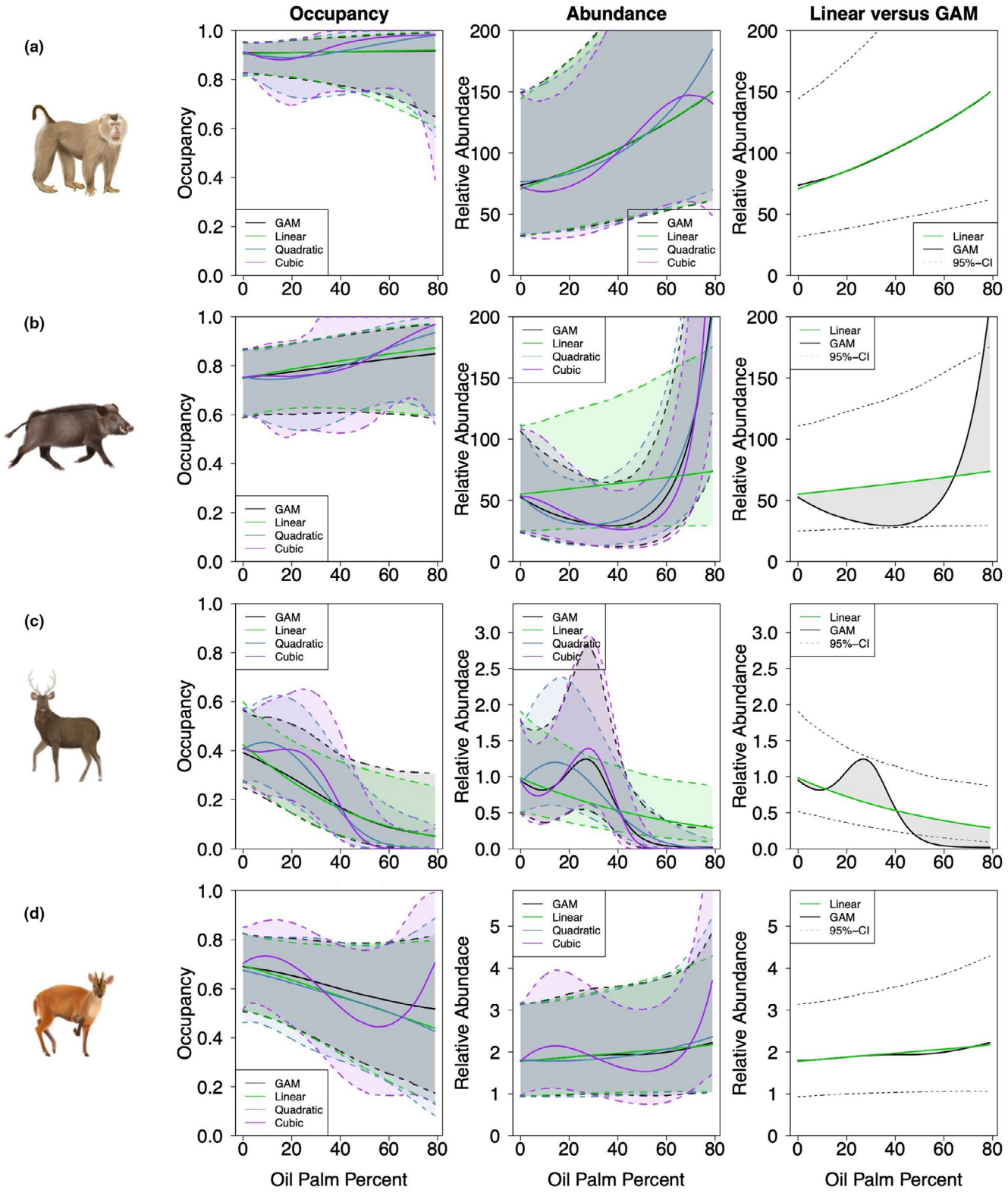

FIGURE 5 Comparison of the relationships between occupancy (left column) and relative abundance (centre column) of Southeast Asian mammals (a: macaque monkeys, *Macaca nemestrina*, b: wild boar, *Sus scrofa*, c: sambar deer, *Rusa unicolor*, d: muntjac deer, genus *Muntiacus*) and oil palm cover from camera trapping at 10 landscapes and estimated using occupancy and N-mixture models with a linear, quadratic, cubic or smooth (generalised additive models, GAM) state function. Each panel is a side-by-side comparison of the mean posterior prediction for the relevant model types. The black solid line represents the mean of the posterior predictions from the GAM, while green, blue and purple solid lines are from the linear, quadratic and cubic models, respectively. The third panel visualises the difference between the linear N-mixture model and N-mixture GAM. The dotted lines correspond to the 95% credible intervals for the traditional model in each panel.

human footprint followed by a plateau, which was poorly captured by the cubic and quadratic models (Figure S3a).

A qualitative interpretation of the relative abundance N-mixture results suggests that there were eight combinations (50%) for which habitat associations as modelled by the GAM were decidedly non-linear: *wild boar~oil palm* (Figure 5b), *sambar~oil palm* (Figure 5c), *macaque~forest cover* (Figure S1a), *muntjak~forest cover* (Figure S1d), *sambar~forest integrity* (Figure S2c), *muntjak~forest integrity* (Figure S2d), *macaque~human footprint* (Figure S3a) and *muntjak~human footprint* (Figure S3d). Sambar deer responses to oil palm cover, as modelled by GAMs, were non-linear reflecting an initial increase in relative abundance, followed by a steep decline in areas with oil palm cover >30%–40%. The sambar linear model poorly captured this trend (25.8 NRMSE relative to GAM; Figures 6c and 5c), while the predicted response from the cubic model was much closer to the GAM (8.5 NRMSE relative to GAM; Figures 6c and 5c). The quadratic model (18.6 NRMSE relative to GAM; Figure 5c) differed from the cubic and GAM models mainly in the initial steep increase in relative abundance. Finally, there were also cases where the addition of polynomial terms increased deviation from the GAM, indicating the linear fit was most appropriate. This was true for macaques and muntjac deer responses to oil palm (Figure 5a,d).

The wild boar OccuGAM showed a flat response to nearby oil cover initially, followed by a near-exponential increase in relative abundance when oil palm cover exceeded ~60% of the nearby habitat, reflecting prior work (Luskin, Brashares, et al., 2017; Moore et al., 2023). The wild boar quadratic model resembled the GAM most closely, while the inclusion of the cubic term slightly increased deviation from the GAM, but the punctuated increase was a clear feature of all non-linear model types (Figure 5b).

3.4 | N-mixture model predictive performance

N-mixture model predictive performance was similar across all species–covariate combinations, with differences rarely exceeding 2σ (Figure S6). OccuGAMs were the best performing model in 3 of 16 cases (18.75%; Figure S6). However, OccuGAMs had a Δ elpd $<2\sigma$ in 14 of 16 cases (87.50%), indicating that predictive performance was generally not meaningfully different from the best performing model. Simple linear models had the highest predictive accuracy in six cases (37.50%), quadratic models in two cases (12.50%) and cubic models in five cases (31.25%).

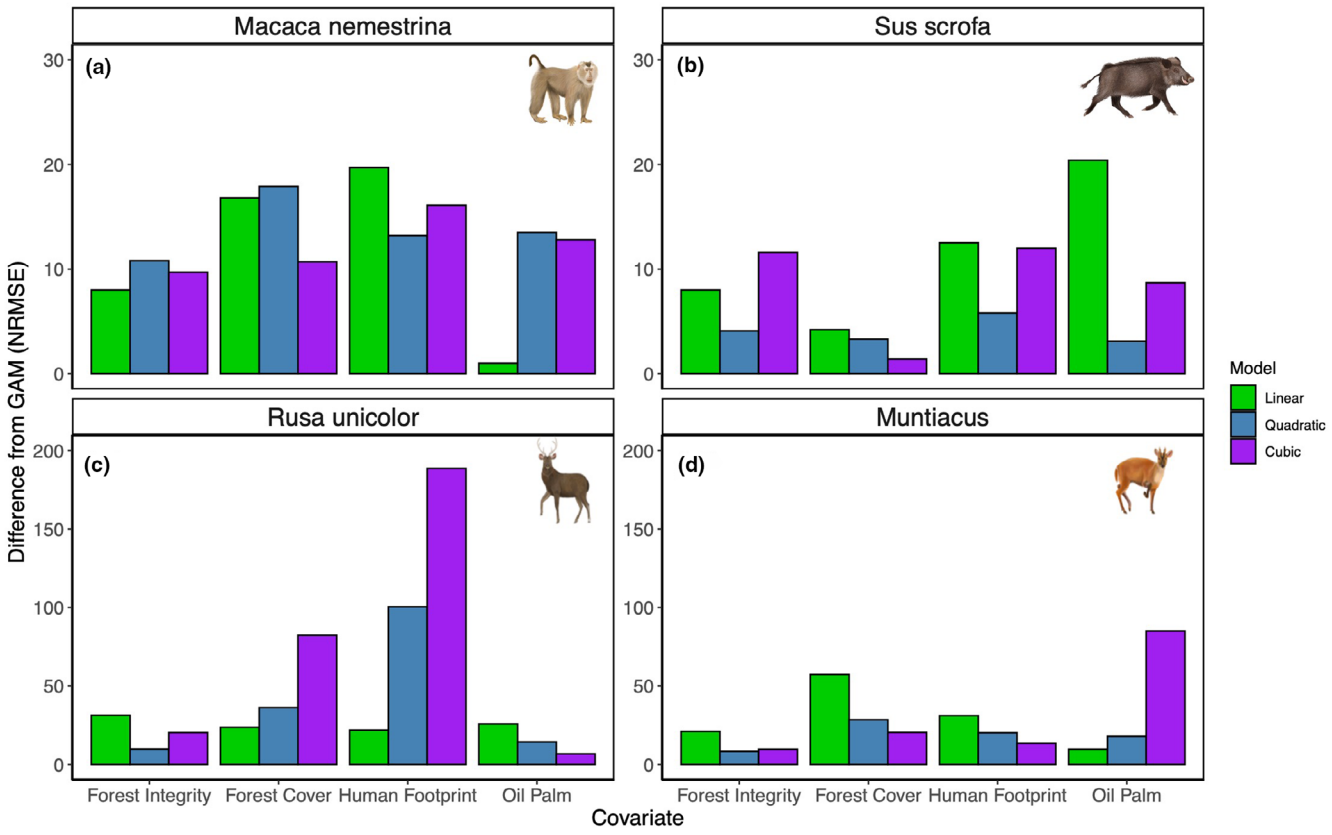

FIGURE 6 Deviation of traditional N-mixture models with polynomial versus OccuGAMs, as measured by the Normalised Root Mean Squared Error NRMSE. The NRMSE quantifies the discrepancy between curves derived from mean posterior predictive estimates of N-mixture models. The fitted curves illustrate the predicted habitat association patterns for four species of Southeast Asian mammals (a: pig-tailed macaque monkeys, *M. nemestrina*, b: wild boar, *S. scrofa*, c: sambar deer, *R. unicolor*, d: muntjac deer, genus *Muntiacus*) in response to disturbance covariates. Each bar corresponds to the NRMSE of a single species–covariate-model combination, with higher NRMSE values indicating greater divergence between the GAM and polynomial comparison model.

The difference in predictive performance of cubic models compared to the best performing model was $>2\sigma$ in five cases and the best performing model was linear or GAMs in all five of these cases (Figure S6). In the cases where cubic models performed the best, the functional shape of the GAM often suggested a highly non-linear relationship (e.g. in *macaque~forest cover*, Figure S1a; *muntjak~forest cover*, Figure S1d; *muntjak~forest integrity*, Figure S2d; *sambar~oil palm*, Figure 5c), although *macaque~oil palm* was a clear exception to this trend (Figure 5a). Unsurprisingly, GAMs were much more flexible than their cubic equivalents, generally achieving Δ elpd $<2\sigma$ when the top performing model was simple (e.g. *wild boar~forest integrity*; Figure S6d) or complex (e.g. *macaque~oil palm*; Figure 5a).

3.5 | Differences in predicted responses between GAMs and polynomial N-mixture models

An overview of the NRMSE between the fitted relationships modelled by GAMs and their polynomial counterparts across the four focal species is presented in Figure 6. We do not assume that the GAM represents the true underlying relationship; rather, we use it as a flexible reference to show how polynomial models diverge. When the NRMSE increases with the degree of the polynomial, the GAM suggests a less complex (linear) relationship (e.g. *macaque~oil palm*; Figure 6a). Alternatively, the NRMSE can decrease with the degree of the polynomial, which means the GAM suggests a more complex relationship that is not adequately captured by second-degree or even third-degree polynomials (e.g. *muntjak—forest cover*; Figure 6d).

There were 4 (25%) cases in which the NRMSE of the linear model was lower than for both the quadratic and cubic models, which means the GAM suggested a 'simple' linear relationship (Figure 6). Quadratic models most closely approximated the GAM in 6 (37.5%) cases while cubic models had the lowest NRMSE in the remaining 6 (37.5%) cases (Figure 6). Notably, three of these four cases comprised a single covariate (forest cover). There were no noticeable trends in the lowest NRMSEs among species. However, some covariate-specific patterns were evident: quadratic models had the lowest NRMSE for forest integrity in three of four species. In comparison, cubic models most closely mirrored the GAM predictions for forest cover in three of four species (Figure 6).

4 | DISCUSSION

We demonstrate the utility of incorporating GAMs into hierarchical occupancy and abundance modelling (HOAM) frameworks to flexibly model species–habitat relationships while accounting for imperfect detection using simulated and empirical data sets. We hypothesised that the greater flexibility of OccuGAMs would allow them to more accurately recover complex, non-linear species–habitat relationships compared to polynomial formulations. In line with this hypothesis, we found that OccuGAMs outperformed polynomial model formulations in their ability to recover non-linear species–habitat relationships in most simulated scenarios, showing that they can reveal complex habitat relationships that are likely to be missed by more constrained models. Moreover, in scenarios where the true relationship was linear, OccuGAMs consistently ranked second behind the linear model. This demonstrates their capacity to adapt to simpler functional forms—an advantage over polynomial formulations, which often performed worse due to spurious curvature. However, accuracy dropped for low-abundance species, reflecting limited information content in the data, which makes estimation of non-linear responses inherently unstable.

Overfitting has been a common concern when using GAMs, yet penalisation of smooth terms ensures unsupported terms are shrunk towards their null space—typically linear or constant forms (Miller, 2025). Conveniently, Bayesian software interfaces for fitting GAMs also allow for an additional penalty on the null space of each term, enabling the complete removal of unsupported terms (Marra & Wood, 2011; Miller, 2025). Indeed, we found that OccuGAMs did not exhibit the overfitting tendencies of polynomials. In contrast, complex polynomials such as the cubic performed worse at smaller sample sizes, producing variable and unrealistic predictions. This robustness could be particularly valuable in Bayesian frameworks, where model selection is not always straightforward and can be computationally intensive (Hooten & Hobbs, 2015; Tredennick et al., 2021)—positioning OccuGAMs as a practical and reliable default choice in an exploratory setting when the true shape of the response is uncertain. This is critical because the output of a linear GLM does not always indicate the need for a more flexible model even when warranted (Heit et al., 2024). In addition, we show that incorporating GAMs into the detection component of N-mixture models yields clear benefits when the true detection probability follows a non-linear trajectory. Given that detection processes are often complex in real-world systems (Strebel et al., 2014), and that fitting a large number of candidate models with alternative polynomial formulations in both the state and detection components is computationally infeasible for most researchers, the use of GAMs in the detection model should be considered a safe and practical solution.

We also showed clear evidence supporting the applicability of these models to real-world datasets. Our aim was not to assert that the GAM recovers the true functional form in every case, but rather to demonstrate that non-linear responses do arise in empirical data and that, informed by simulation results, GAMs provide a more credible means of representing such structure when it is present. Specifically, OccuGAMs produced non-linear response curves in 50% of species abundance–covariate combinations, capturing patterns that were distinctly non-linear and poorly represented by standard polynomial formulations. In contrast, the occupancy results exhibited no pronounced non-linear patterns. This difference may be partially attributed to high occupancy of some species (pigs and macaques), such that the range of responses was constrained. Another reason is that we restricted our analyses to landscapes where the species

was detected at least once (e.g. pig-tailed macaque were not present in Singapore). Expanding this framework to include landscapes from which species have recently been extirpated and extending the model structure to explicitly distinguish true absences (Amir, Sovie, & Luskin, 2022; Blasco-Moreno et al., 2019) may yield more informative insights into occupancy.

The case study OccuGAM results were consistent with the ecological findings in previous work, adding confidence. For example, wild boar hyperabundance in areas near oil palm plantations is well documented, and a comparable pattern has been observed for pig-tailed macaques (Luskin, Albert, & Tobler, 2017; Luskin, Brashares, et al., 2017; Moore et al., 2023). Our results show that while both species respond positively to oil palm presence, wild boar abundance increases sharply around 60% coverage, whereas macaque abundance increases more gradually and linearly, lacking a clear inflection point. In contrast, sambar deer abundance declines sharply to near zero around 40% oil palm coverage, though we note that the difference between the GAM and the cubic model was small in this instance. Notably, pig-tailed macaques also demonstrated distinct threshold responses to human footprint (negative) and forest cover (positive), patterns that polynomial models failed to capture. Understanding these non-linear responses can have important implications for real-world management.

GAMs employ data-driven response curves, improving the ability to detect regions of abrupt change with greater confidence. This holds relevance for studies investigating ecological thresholds, an area where empirical support remains the subject of ongoing debate (Groffman et al., 2006; Hillebrand et al., 2020; Spake et al., 2022). Indeed, traditional GAMs lacking detection–correction have been applied to identify thresholds in systems ranging from entire marine communities (Samhouri et al., 2017) to grey wolves (Potvin et al., 2005). With the advent of country- and continental-scale camera trapping data sets (Bruce et al., 2025; Mendes et al., 2024), often including repeated sampling of the same locations, OccuGAMs with camera data enable the robust investigation of ecological thresholds in wildlife communities that are traditionally difficult to monitor, such as tropical mammals inhabiting rainforests (Burns et al., 2025; Luskin, Albert, & Tobler, 2017). OccuGAMs remain underused in such contexts, presenting an opportunity to address questions such as: '*What proportion of oil palm cover triggers hyperabundance of wild boars and macaques?*'. They may also be valuable for studying biological invasions, in which a species' transition from rarity to establishment often produces non-linear responses along environmental gradients (Blackburn et al., 2011). Within HOAMs, smooth effects can identify thresholds in occupancy or local abundance, revealing the conditions that facilitate establishment and subsequent spread.

Two key limitations may impede OccuGAMs' wider adoption among applied ecologists. First, estimating complex relationships between the latent state and environmental covariates requires sufficient data support across the covariate space, which in HOAMs often translates into larger numbers of surveyed sites and increased costs (Bruce et al., 2025; Guillera-Arroita, 2017). Importantly, this challenge reflects the difficulty of resolving functional complexity rather than greater data demand of GAMs relative to polynomial alternatives. In penalised GAMs, model complexity is data-adaptive: the effective degrees of freedom are controlled by the smoothing penalty (Miller, 2025; Wood, 2017). Consequently, data requirements depend primarily on the complexity of the underlying ecological relationship, rather than on the choice of GAMs versus polynomial models per se (Wood, 2017). The key distinction is that GAMs offer greater flexibility but are more sensitive to model specification. Choices such as the number and type of basis functions can affect the fit, and poor specification may lead to over- or underfitting (Wood, 2017). Consequently, diagnostic tools such as estimated degrees of freedom (EDF) must be carefully evaluated to ensure the model appropriately captures the underlying non-linearities. By default, the widely used *mgcv* R package (Wood, 2017) implements thin-plate regression splines (TPRS) for univariate smooths, providing a flexible and well-regularised basis suitable for most environmental covariates, although alternative bases (e.g. cyclic) may be preferable where periodicity is known a priori (Wood, 2017). TPRS provide a low-rank, isotropic smoother that does not require manual selection of knot locations and is well suited to exploratory analyses where the true functional form of the response is unknown (Wood, 2017). This makes them a pragmatic default for modelling potentially complex, non-linear responses of occupancy and abundance to anthropogenic disturbance gradients. Regarding the number of basis functions, a common strategy is to specify a basis dimension K large enough to accommodate plausible ecological responses and rely on penalisation to control smoothness, rather than tuning K to optimise fit (Pedersen et al., 2019). However, when strong ecological expectations exist, such as single-threshold responses often reported for environmental drivers, constraining basis complexity can avoid small-scale oscillations and reduce computation time (Large et al., 2013; Pedersen et al., 2019; Samhouri et al., 2017). In such cases, modest K values combined with post-fit diagnostics (e.g. comparing EDF relative to K and re-running the model with higher K if EDF≈K) provide a principled balance between ecological realism and statistical flexibility.

Lastly, practitioners may be hesitant to use OccuGAMs due to the perceived difficulty in interpreting model parameters and assessing their significance (Heit et al., 2024). While this concern is valid, it also applies to polynomial models, which use multiple coefficients to capture non-linearity. Interpretation of GAM smooths should move beyond binary notions of statistical significance and instead focus on the form and magnitude of effects (Arel-Bundock et al., 2024; Pedersen et al., 2019; Simpson, 2018). In practice, this involves visualising smooth effect plots, examining EDF to gauge complexity: values near 1 indicate near-linear effects while higher EDF indicate non-linear responses, and asking targeted questions such as where slopes change or effects are strongest across the covariate range.

AUTHOR CONTRIBUTIONS

Johannes Maria Sassen led conceptualisation, method development, simulation, analysis, visualisation, writing and revision; Matthew

Scott Luskin and Zachary Amir collected the case study camera trap data; Nicholas Clark contributed to analysis of simulations; Matthew Scott Luskin, Zachary Amir and Nicholas Clark contributed to conceptualisation, methods development and revision. All authors reviewed the final manuscript and gave final approval for publication.

ACKNOWLEDGEMENTS

In-kind support was provided by the Fauna and Flora International—Indonesia programme, the Institution Conservation Society (ICS)-Solok Selatan Wahana Konservasi Masyarakat and the Leuser International Foundation (LIF). The Yayasan Sabah, the Sabah Forest Department, the Sabah Biodiversity Council and the Danum Valley Management Committee, Abdul Hamid, Glen Renolds, Jedediah Brodie, Jonathan Moore, Katie Doehla and Tombi Karolus facilitated and/or helped with fieldwork at Danum Valley. Patrick Jansen, Jorge Ahumada, Jonathan Moore, the Smithsonian Institute, the Tropical Ecology Assessment and Monitoring (TEAM) network, Yao Tse Leong and the Forest Research Institute Malaysia (FRIM) facilitated and/or helped with data collection at Pasoh. Mohizah Bt. Mohamad, Januarie Kulis, Jonathan Moore and the Sarawak Forestry Department, and the NTU field ecology courses facilitated and/or helped with fieldwork data at Lambir Hills. Shawn Lum, Adrian Loo, Max Khoo, Ben Lee, Jasyln Chan and Alexis Goh helped with fieldwork in Singapore and NParks granted permissions. Wirong Chantorn, Anuttara Nathalang, Sarayudh Bunyavejchewin, Ronglarp Sukmasuang, Felise Gutierez, Chris Scanlon and Jonathan Moore facilitated and/or helped with fieldwork at Khao Yai and Khao Ban Tat/Khao Chong. Wido Rizqi Albert, Matthew Linkie, Yoan Dinata, Hariyo Wibisono and HarimauKita facilitated fieldwork in Sumatra, and Edi Siarenta Sembiring, Tarmizi and Eka Ramadiyanta, Salpayanri, Iswandri Tanjung and Chris Decky assisted with fieldwork. The original artwork was provided by T. Barber from Talking Animals and is copyrighted. Open access publishing facilitated by The University of Queensland, as part of the Wiley - The University of Queensland agreement via the Council of Australian University Librarians.

CONFLICT OF INTEREST STATEMENT

Nicholas Clark is an Associate Editor at *Methods in Ecology and Evolution* but played no role in the handling or reviewing of this manuscript. All other authors have no conflicts of interest to declare.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210x.70252.

DATA AVAILABILITY STATEMENT

The data and the code for reproducing results are available through Zenodo: https://zenodo.org/records/18211784 (Sassen, 2026a). Full results for all analyses presented in this work are available through a separate Zenodo repository: https://zenodo.org/records/16965223 (Sassen, 2026b).

ORCID

Johannes Maria Sassen https://orcid.org/0009-0002-0579-7875
Zachary Amir https://orcid.org/0000-0002-8398-2059
Nicholas Clark https://orcid.org/0000-0001-7131-3301
Matthew Scott Luskin https://orcid.org/0000-0002-5236-7096

REFERENCES

Aho, K., Derryberry, D., & Peterson, T. (2014). Model selection for ecologists: The worldviews of AIC and BIC. *Ecology*, *95*, 631–636.

Amir, Z., Moore, J. H., Negret, P. J., & Luskin, M. S. (2022). Megafauna extinctions produce idiosyncratic Anthropocene assemblages. *Science Advances*, *8*, eabq2307.

Amir, Z., Sovie, A., & Luskin, M. S. (2022). Inferring predator–prey interactions from camera traps: A Bayesian co-abundance modeling approach. *Ecology and Evolution*, *12*, e9627.

Arel-Bundock, V., Greifer, N., & Heiss, A. (2024). How to interpret statistical models using marginaleffects for R and python. *Journal of Statistical Software*, *111*, 1–32.

Blackburn, T. M., Pyšek, P., Bacher, S., Carlton, J. T., Duncan, R. P., Jarošík, V., Wilson, J. R. U., & Richardson, D. M. (2011). A proposed unified framework for biological invasions. *Trends in Ecology & Evolution*, *26*, 333–339.

Blasco-Moreno, A., Pérez-Casany, M., Puig, P., Morante, M., & Castells, E. (2019). What does a zero mean? Understanding false, random and structural zeros in ecology. *Methods in Ecology and Evolution*, *10*, 949–959.

Bled, F., Nichols, J. D., & Altwegg, R. (2013). Dynamic occupancy models for analyzing species' range dynamics across large geographic scales. *Ecology and Evolution*, *3*, 4896–4909.

Bolker, B. M., Gardner, B., Maunder, M., Berg, C. W., Brooks, M., Comita, L., Crone, E., Cubaynes, S., Davies, T., de Valpine, P., Ford, J., Gimenez, O., Kéry, M., Kim, E. J., Lennert-Cody, C., Magnusson, A., Martell, S., Nash, J., Nielsen, A., … Zipkin, E. (2013). Strategies for fitting nonlinear ecological models in R, AD model builder, and BUGS. *Methods in Ecology and Evolution*, *4*, 501–512.

Briscoe, N. J., Zurell, D., Elith, J., König, C., Fandos, G., Malchow, A.-K., Kéry, M., Schmid, H., & Guillera-Arroita, G. (2021). Can dynamic occupancy models improve predictions of species' range dynamics? A test using Swiss birds. *Global Change Biology*, *27*, 4269–4282.

Brodie, J. F., Helmy, O. E., Mohd-Azlan, J., Granados, A., Bernard, H., Giordano, A. J., & Zipkin, E. (2018). Models for assessing local-scale co-abundance of animal species while accounting for differential detectability and varied responses to the environment. *Biotropica*, *50*, 5–15.

Bruce, T., Amir, Z., Allen, B. L., Alting, B. F., Amos, M., Augusteyn, J., Ballard, G.-A., Behrendorff, L. M., Bell, K., Bengsen, A. J., Bennett, A., Benshemesh, J. S., Bentley, J., Blackmore, C. J., Boscarino-Gaetano, R., Bourke, L. A., Brewster, R., Brook, B. W., Broughton, C., … Luskin, M. S. (2025). Large-scale and long-term wildlife research and monitoring using camera traps: A continental synthesis. *Biological Reviews*, *100*, 530–555.

Burns, P., Kaszta, Z., Cushman, S. A., Brodie, J. F., Hakkenberg, C. R., Jantz, P., Deith, M., Luskin, M. S., Ball, J. G. C., Mohd-Azlan, J., Burslem, D. F. R. P., Cheyne, S. M., Haidir, I., Hearn, A. J., Slade, E., Williams, P. J., Macdonald, D. W., & Goetz, S. J. (2025). The utility of dynamic forest structure from GEDI lidar fusion in tropical mammal species distribution models. *Frontiers in Remote Sensing*, *6*, 1563430.

Burton, A. C., Neilson, E., Moreira, D., Ladle, A., Steenweg, R., Fisher, J. T., Bayne, E., & Boutin, S. (2015). REVIEW: Wildlife camera trapping: A review and recommendations for linking surveys to ecological processes. *Journal of Applied Ecology*, *52*, 675–685.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017).

Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*, 1–32.

Clark, N. J., & Wells, K. (2023). Dynamic generalised additive models (DGAMs) for forecasting discrete ecological time series. *Methods in Ecology and Evolution*, *14*, 771–784.

Conn, P. B., Johnson, D. S., Williams, P. J., Melin, S. R., & Hooten, M. B. (2018). A guide to Bayesian model checking for ecologists. *Ecological Monographs*, *88*, 526–542.

de Valpine, P., Turek, D., Paciorek, C. J., Anderson-Bergman, C., Lang, D. T., & Bodik, R. (2017). Programming with models: Writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational and Graphical Statistics*, *26*, 403–413.

Dehaudt, B., Amir, Z., Decoeur, H., Gibson, L., Mendes, C., Moore, J. H., Nursamsi, I., Sovie, A., & Luskin, M. S. (2022). Common palm civets *Paradoxurus hermaphroditus* are positively associated with humans and forest degradation with implications for seed dispersal and zoonotic diseases. *Journal of Animal Ecology*, *91*, 794–804.

Dehaudt, B., Bose, R., Avoto, J. J., Brittain, S., Bruce, T., Chen, E. K., Forzi, F. A., Hardesty, B. D., Holbrook, K. M., Lamperti, A. M., Parker, V. T., Poulsen, J. R., Russo, N. J., Simpoh, E., Wang, B. C., Whitney, K. D., Smith, T. B., & Luskin, M. S. (2025). Thirty years of arboreal wildlife trends in an African rainforest under evolving threats and researchers' presence. *Biological Conservation*, *312*, 111475. https://doi.org/10.1016/j.biocon.2025.111475

Gelman, A., & Imbens, G. (2019). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics*, *37*, 447–456.

Gelman, A., Meng, X.-L., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, *6*, 733–760.

Gilbert, N. A., Clare, J. D. J., Stenglein, J. L., & Zuckerberg, B. (2021). Abundance estimation of unmarked animals based on camera-trap data. *Conservation Biology*, *35*, 88–100.

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*, 359–378.

Goldstein, B. R., Keller, A. G., Calhoun, K. L., Barker, K. J., Montealegre-Mora, F., Serota, M. W., Van Scoyoc, A., Parker-Shames, P., Andreozzi, C. L., & de Valpine, P. (2024). How do ecologists estimate occupancy in practice? *Ecography*, e07402.

Grantham, H. S., Duncan, A., Evans, T. D., Jones, K. R., Beyer, H. L., Schuster, R., Walston, J., Ray, J. C., Robinson, J. G., Callow, M., Clements, T., Costa, H. M., DeGemmis, A., Elsen, P. R., Ervin, J., Franco, P., Goldman, E., Goetz, S., Hansen, A., … Watson, J. E. M. (2020). Anthropogenic modification of forests means only 40% of remaining forests have high ecosystem integrity. *Nature Communications*, *11*, 5978.

Groffman, P. M., Baron, J. S., Blett, T., Gold, A. J., Goodman, I., Gunderson, L. H., Levinson, B. M., Palmer, M. A., Paerl, H. W., Peterson, G. D., Poff, N. L., Rejeski, D. W., Reynolds, J. F., Turner, M. G., Weathers, K. C., & Wiens, J. (2006). Ecological thresholds: The key to successful environmental management or an important concept with no practical application? *Ecosystems*, *9*, 1–13.

Gu, W., & Swihart, R. K. (2004). Absent or undetected? Effects of non-detection of species occurrence on wildlife–habitat models. *Biological Conservation*, *116*, 195–203.

Guillera-Arroita, G. (2017). Modelling of species distributions, range dynamics and communities under imperfect detection: Advances, challenges and opportunities. *Ecography*, *40*, 281–295.

Guisan, A., Edwards, T. C., & Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: Setting the scene. *Ecological Modelling*, *157*, 89–100.

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., & Loveland, T. R. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, *342*, 850–853.

Harrell, J. F. E. (2015). *Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis*. Springer International Publishing.

Heit, D. R., Ortiz-Calo, W., Poisson, M. K. P., Butler, A. R., & Moll, R. J. (2024). Generalized nonlinearity in animal ecology: Research, review, and recommendations. *Ecology and Evolution*, *14*, e11387.

Hillebrand, H., Donohue, I., Harpole, W. S., Hodapp, D., Kucera, M., Lewandowska, A. M., Merder, J., Montoya, J. M., & Freund, J. A. (2020). Thresholds for ecological responses to global change do not emerge from empirical data. *Nature Ecology & Evolution*, *4*, 1502–1509.

Hooten, M. B., & Hobbs, N. T. (2015). A guide to Bayesian model selection for ecologists. *Ecological Monographs*, *85*, 3–28.

Jordan, A., Krüger, F., & Lerch, S. (2019). Evaluating probabilistic forecasts with scoringRules. *Journal of Statistical Software*, *90*, 1–37.

Kéry, M., & Royle, J. A. (2016). *Applied hierarchical modeling in ecology: Analysis of distribution, abundance and species richness in R and BUGS*. Elsevier/AP.

Kéry, M., & Royle, J. A. (2020). *Applied hierarchical modeling in ecology: Analysis of distribution, abundance and species richness in R and BUGS: Volume 2: Dynamic and advanced models*. Academic Press.

Kéry, M., & Schaub, M. (2011). *Bayesian population analysis using WinBUGS: A hierarchical perspective*. Academic Press.

Kéry, M., & Schmidt, B. R. (2008). Imperfect detection and its consequences for monitoring for conservation. *Community Ecology*, *9*, 207–216.

Lahoz-Monfort, J. J., & Magrath, M. J. L. (2021). A comprehensive overview of technologies for species and habitat monitoring and conservation. *Bioscience*, *71*, 1038–1062.

Large, S. I., Fay, G., Friedland, K. D., & Link, J. S. (2013). Defining trends and thresholds in responses of ecological indicators to fishing and environmental pressures. *ICES Journal of Marine Science*, *70*, 755–767.

Lee, J., de Carvalho, M., Rua, A., & Avila, J. (2024). Bayesian smoothing for time-varying extremal dependence. *Journal of the Royal Statistical Society. Series C, Applied Statistics*, *73*, 581–597.

Link, W. A., Schofield, M. R., Barker, R. J., & Sauer, J. R. (2018). On the robustness of N-mixture models. *Ecology*, *99*, 1547–1551.

Luskin, M. S., Albert, W. R., & Tobler, M. W. (2017). Sumatran tiger survival threatened by deforestation despite increasing densities in parks. *Nature Communications*, *8*, 1783.

Luskin, M. S., Brashares, J. S., Ickes, K., Sun, I. F., Fletcher, C., Wright, S. J., & Potts, M. D. (2017). Cross-boundary subsidy cascades from oil palm degrade distant tropical forests. *Nature Communications*, *8*, 2231.

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

Magee, L. (1998). Nonlocal behavior in polynomial regressions. *The American Statistician*, *52*, 20–22.

Marra, G., & Wood, S. N. (2011). Practical variable selection for generalized additive models. *Computational Statistics & Data Analysis*, *55*, 2372–2387.

Mendes, C. P., Albert, W. R., Amir, Z., Ancrenaz, M., Ash, E., Azhar, B., Bernard, H., Brodie, J., Bruce, T., Carr, E., Clements, G. R., Davies, G., Deere, N. J., Dinata, Y., Donnelly, C. A., Duangchantrasiri, S., Fredriksson, G., Goossens, B., Granados, A., … Luskin, M. S. (2024). CamTrapAsia: A dataset of tropical forest vertebrate communities from 239 camera trapping studies. *Ecology*, *105*, e4299.

Mendes, C. P., & Luskin, M. S. (2025). The use, misuse and opportunities for structural equation modelling (SEM) in wildlife ecology. *Journal of Applied Ecology*, *62*(12), 3210–3226. https://doi.org/10.1111/1365-2664.70189

Miettinen, J., Shi, C., & Liew, S. C. (2016). 2015 land cover map of Southeast Asia at 250 m spatial resolution. *Remote Sensing Letters*, *7*, 701–710.

Miller, D. L. (2025). Bayesian views of generalized additive modelling. *Methods in Ecology and Evolution*, *16*, 446–455.

Moll, R. J., Cepek, J. D., Lorch, P. D., Dennis, P. M., Robison, T., Millspaugh, J. J., & Montgomery, R. A. (2018). Humans and urban development mediate the sympatry of competing carnivores. *Urban Ecosystems*, *21*, 765–778.

Moore, J. H., Gibson, L., Amir, Z., Chanthorn, W., Ahmad, A. H., Jansen, P. A., Mendes, C. P., Onuma, M., Peres, C. A., & Luskin, M. S. (2023). The rise of hyperabundant native generalists threatens both humans and nature. *Biological Reviews*, *98*, 1829–1844.

Morante-Filho, J. C., Faria, D., Mariano-Neto, E., & Rhodes, J. (2015). Birds in anthropogenic landscapes: The responses of ecological groups to forest loss in the Brazilian Atlantic Forest. *PLoS One*, *10*, e0128923.

Morse, C. C., Huryn, A. D., & Cronan, C. (2003). Impervious surface area as a predictor of the effects of urbanization on stream insect communities in Maine, U.S.A. *Environmental Monitoring and Assessment*, *89*, 95–127.

Pedersen, E. J., Miller, D. L., Simpson, G. L., & Ross, N. (2019). Hierarchical generalized additive models in ecology: An introduction with mgcv. *PeerJ*, *7*, e6876.

Plummer, M. (2003). *JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling* (pp. 1–10). Proceedings of the 3rd international workshop on distributed statistical computing.

Potvin, M. J., Drummer, T. D., Vucetich, J. A., Beyer, D. E., Jr., Peterson, R. O., & Hammill, J. H. (2005). Monitoring and habitat analysis for wolves in upper Michigan. *Journal of Wildlife Management*, *69*, 1660–1669.

Rayan, M., & Linkie, M. (2020). Managing threatened ungulates in logged-primary forest mosaics in Malaysia. *PLoS One*, *15*, e0243932.

Rhinehart, T. A., Chronister, L. M., Devlin, T., & Kitzes, J. (2020). Acoustic localization of terrestrial wildlife: Current practices and future opportunities. *Ecology and Evolution*, *10*, 6794–6818.

Rhodes, J. R., Callaghan, J. G., McAlpine, C. A., De Jong, C., Bowen, M. E., Mitchell, D. L., Lunney, D., & Possingham, H. P. (2008). Regional variation in habitat–occupancy thresholds: A warning for conservation planning. *Journal of Applied Ecology*, *45*, 549–557.

Rovero, F., & Zimmermann, F. (2016). *Camera trapping for wildlife research*. Pelagic Publishing.

Royle, J. A. (2004). N-mixture models for estimating population size from spatially replicated counts. *Biometrics*, *60*, 108–115.

Rozylowicz, L., Popescu, V. D., Manolache, S., Nita, A., Gradinaru, S. R., Mirea, M. D., & Bancila, R. I. (2024). Occupancy and N-mixture modeling applications in ecology: A bibliometric analysis. *Global Ecology and Conservation*, *50*, e02838.

Ruiz-Gutiérrez, V., Zipkin, E. F., & Dhondt, A. A. (2010). Occupancy dynamics in a tropical bird community: Unexpectedly high forest use by birds classified as non-forest species. *Journal of Applied Ecology*, *47*, 621–630.

Rushing, C. S., Royle, J. A., Ziolkowski, D. J., & Pardieck, K. L. (2019). Modeling spatially and temporally complex range dynamics when detection is imperfect. *Scientific Reports*, *9*, 12805.

Rushing, C. S., Royle, J. A., Ziolkowski, D. J., & Pardieck, K. L. (2020). Migratory behavior and winter geography drive differential range shifts of eastern birds in response to recent climate change. *Proceedings of the National Academy of Sciences*, *117*, 12897–12903.

Samhouri, J. F., Andrews, K. S., Fay, G., Harvey, C. J., Hazen, E. L., Hennessey, S. M., Holsman, K., Hunsicker, M. E., Large, S. I., & Marshall, K. N. (2017). Defining ecosystem thresholds for human activities and environmental pressures in the California current. *Ecosphere*, *8*, e01860.

Sassen, J. (2026a). Joopie-28/OccuGAM_Methods_ECL: MEE-PUB (MEE-PUB). *Zenodo*. https://zenodo.org/records/18211784

Sassen, J. (2026b). OccuGAMs: Non-linear occupancy and abundance modelling with imperfect detection - Full results. *Zenodo*. https://zenodo.org/records/16965223

Scheuerer, M., & Hamill, T. M. (2015). Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities. *Monthly Weather Review*, *143*, 1321–1334.

Simpson, G. L. (2018). Modelling palaeoecological time series using generalised additive models. *Frontiers in Ecology and Evolution*, *6*, 149.

Socolar, J. B., & Mills, S. C. (2023). Introducing flocker: An R package for flexible occupancy modeling via brms and Stan. *bioRxiv*. 2023.2010.2026.564080.

Spake, R., Barajas-Barbosa, M. P., Blowes, S. A., Bowler, D. E., Callaghan, C. T., Garbowski, M., Jurburg, S. D., van Klink, R., Korell, L., Ladouceur, E., Rozzi, R., Viana, D. S., Xu, W.-B., & Chase, J. M. (2022). Detecting thresholds of ecological change in the Anthropocene. *Annual Review of Environment and Resources*, *47*, 797–821.

Strebel, N., Kéry, M., Schaub, M., & Schmid, H. (2014). Studying phenology by flexible modelling of seasonal detectability peaks. *Methods in Ecology and Evolution*, *5*, 483–490.

Suarez-Rubio, M., Wilson, S., Leimgruber, P., & Lookingbill, T. (2013). Threshold responses of forest birds to landscape changes around exurban development. *PLoS One*, *8*, e67593.

Tabak, M. A., Norouzzadeh, M. S., Wolfson, D. W., Sweeney, S. J., Vercauteren, K. C., Snow, N. P., Halseth, J. M., Di Salvo, P. A., Lewis, J. S., White, M. D., Teton, B., Beasley, J. C., Schlichting, P. E., Boughton, R. K., Wight, B., Newkirk, E. S., Ivan, J. S., Odell, E. A., Brook, R. K., … Miller, R. S. (2019). Machine learning to classify animal species in camera trap images: Applications in ecology. *Methods in Ecology and Evolution*, *10*, 585–590.

Tredennick, A. T., Hooker, G., Ellner, S. P., & Adler, P. B. (2021). A practical guide to selecting models for exploration, inference, and prediction in ecology. *Ecology*, *102*, e03336.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*, 1413–1432.

Venter, O., Sanderson, E. W., Magrach, A., Allan, J. R., Beher, J., Jones, K. R., Possingham, H. P., Laurance, W. F., Wood, P., & Fekete, B. M. (2016). Sixteen years of change in the global terrestrial human footprint and implications for biodiversity conservation. *Nature Communications*, *7*, 12558.

Warton, D. I., Stoklosa, J., Guillera-Arroita, G., MacKenzie, D. I., & Welsh, A. H. (2017). Graphical diagnostics for occupancy models with imperfect detection. *Methods in Ecology and Evolution*, *8*, 408–419.

Wearn, O. R., Rowcliffe, J. M., Carbone, C., Pfeifer, M., Bernard, H., & Ewers, R. M. (2017). Mammalian species abundance across a gradient of tropical land-use intensity: A hierarchical multi-species modelling approach. *Biological Conservation*, *212*, 162–171.

Wilcove, D. S., Giam, X., Edwards, D. P., Fisher, B., & Koh, L. P. (2013). Navjot's nightmare revisited: Logging, agriculture, and biodiversity in Southeast Asia. *Trends in Ecology & Evolution*, *28*, 531–540.

Wood, S. N. (2016). Just another Gibbs additive modeler: Interfacing JAGS and mgcv. *Journal of Statistical Software*, *75*, 1–15.

Wood, S. N. (2017). *Generalized additive models: An introduction with R* (2nd ed.). CRC Press/Taylor & Francis Group.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Figure S1. Comparison of the relationships between occupancy (left column) and relative abundance (center column) of tropical mammals (A: macaque monkeys, *M. nemestrina*, B: wild boar, *S. scrofa*, C: sambar deer, *R. unicolor*, D: muntjac deer, genus *Muntiacus*) and forest cover in 10 South-East Asian landscapes estimated using Bayesian Occupancy and N-Mixture models with a linear, quadratic, cubic or smooth (GAM) function of the covariate.

Figure S2. Comparison of the relationships between occupancy (left

column) and relative abundance (center column) of tropical mammals (A: macaque monkeys, *M. nemestrina*, B: wild boar, *S. scrofa*, C: sambar deer, *R. unicolor*, D: muntjac deer, genus *Muntiacus*) and forest integrity in 10 South-East Asian landscapes estimated using Bayesian Occupancy and N-Mixture models with a linear, quadratic, cubic or smooth (GAM) function of the covariate.

Figure S3. Comparison of the relationships between occupancy (left column) and relative abundance (center column) of tropical mammals (A: macaque monkeys, *M. nemestrina*, B: wild boar, *S. scrofa*, C: sambar deer, *R. unicolor*, D: muntjac deer, genus *Muntiacus*) and human footprint in 10 South-East Asian landscapes estimated using Bayesian Occupancy and N-Mixture models with a linear, quadratic, cubic or smooth (GAM) function of the covariate.

Figure S4. Deviation of traditional occupancy models with polynomial versus OccuGAMs, as measured by the Normalised Root Mean Squared Error (NRMSE).

Figure S5. Relative model performance in predicting the occupancy of four tropical mammal species (*M. nemestrina*, *S. scrofa*, *R. unicolor*, *Muntiacus*) across four disturbance covariates.

Figure S6. Relative model performance in predicting the relative abundance of four tropical mammal species (*M. nemestrina*, *S. scrofa*, *R. unicolor*, *Muntiacus*) across four disturbance covariates.

Figure S7. Variogram rank across sample sizes for (a) Occupancy Models and (b) N-Mixture Models comparing linear, quadratic, cubic and GAM hierarchical models.

Figure S8. Declining ability to estimate highly non-linear occupancy trends with decreasing sample size.

Figure S9. Declining ability to estimate non-linear abundance trends with decreasing sample size.

Figure S10. Energy rank across sample sizes and response types for occupancy models comparing linear, quadratic, cubic and GAM hierarchical models.

Figure S11. Variogram rank across sample sizes and response types for occupancy models comparing linear, quadratic, cubic and GAM hierarchical models.

Figure S12. Energy rank across sample sizes and response types for N-mixture models comparing linear, quadratic, cubic and GAM hierarchical models.

Figure S13. Variogram rank across sample sizes and response types for N-mixture models comparing linear, quadratic, cubic and GAM hierarchical models.

Figure S14. Discrepancy between true occupancy probability ψ and occupancy model estimates across base detection probabilities (b_{p0}), detection covariates (linear or non-linear) and detection model formulations (linear or GAM) formulations measured by the NRMSE.

Figure S15. (a) Example of model fits for a single simulated abundance dataset (medium sample size) from the linear scenario with linear detection covariate. (b) Energy scores (ES) for each model formulation, highlighting differences in ability to recover the true shape.

Figure S16. Energy rank across species and response types for N-mixture models comparing linear, quadratic, cubic and GAM hierarchical models.

Table S1. $\hat{R}$ exceedance for occupancy model simulation analysis.

Table S2. $\hat{R}$ exceedance for N-mixture model simulation analysis.

Table S3. Posterior predictive Checks including Bayesian *p*-values (Goodness-of-Fit) and C-hat (Overdispersion) for 64 univariate occupancy models modelling the relationship between relative abundance of four tropical mammal species (*M. nemestrina*, *S. scrofa*, *R. unicolor*, *Muntiacus*) and four disturbance covariates, across 10 South-East Asian landscapes.

Table S4. Posterior predictive Checks including Bayesian *p*-values (Goodness-of-Fit) and C-hat (Overdispersion) for 64 univariate N-mixture models modelling the relationship between relative abundance of four tropical mammal species (*M. nemestrina*, *S. scrofa*, *R. unicolor*, *Muntiacus*) and four disturbance covariates, across 10 South-East Asian landscapes.

How to cite this article: Sassen, J. M., Amir, Z., Clark, N., & Luskin, M. S. (2026). OccuGAMs: Non-linear occupancy and abundance modelling with imperfect detection. *Methods in Ecology and Evolution*, *00*, 1–18. https://doi.org/10.1111/2041-210x.70252