Supporting information

__Full code for the data simulation, model, and diversity profiles can be found on github under__ *.*

Model description

We adopted the hierarchical formulation of occupancy models by Royle & Dorazio (2008) extended to a community occupancy model (Dorazio & Royle, 2005; Dorazio et al., 2006). The first level of the model represents the true community states (w) of all species (i) in each landscape's (s) community (i.e., whether a species k is part of community s):

$$w_{ks} \sim Bernoulli(\Omega)$$

The second level describes the ecological process determining the occurrences (z) of species at sampling points (j), governed by the local probability of occupancy (ψ) and conditional on w:

$$z_{ij}|w_{is} \sim Bernoulli(w_{is} * \psi_{ij})$$
$$logit(\psi_{ij}) = \beta 0_{is} + \beta 1_i * habitat_j$$

The third level describes the observation process, conditional on z, governed by the detection probability (p), modeling the detection history (y) at each point j and occasion, k:

$$y_{ijk}|z_{ij} \sim Bernoulli(z_{ij} * p_{ijk})$$
$$logit(p_{ijk}) = \alpha 0_{is}$$

The species-specific models are linked by assuming that species-specific parameters come from a common underlying distribution, governed by community hyperparameters. Note that in the simulation study, effort is constant across occasions, but variation in p induced by variation in effort can be modeled by including effort as a covariate in the logit-linear predictor of p.

To analyze our simulated data, following common practice in analyzing field data, we modelled occupancy probability as having species-specific random intercepts, $\beta 0_{ks}$, with landscape specific (indicated by s indexing) hyperparameters ($\mu_{\beta 0,s}, \sigma_{\beta 0,s}$), to allow for different baseline occupancy in the reserves and among species:

$$\beta 0_{ks} \sim Normal(\mu_{\beta 0,s}, \sigma_{\beta 0,s})$$

We further modelled species-specific effects on occupancy of the simulated habitat "disturbance" covariate ($\beta 1$). Detection probability included a species-specific random intercept with landscape specific hyperparameters, to allow for differences in baseline detection among reserves. In our case, the different landscapes had different abundances of animals, which leads to differences in species-level detection (Royle & Nichols, 2003). The formal model description can be found in the supporting information. We implemented the model in a Bayesian framework using JAGS (Plummer, 2003) accessed via the R packages

rjags (Plummer, 2019). We ran three parallel Markov chains with 250,000 iterations, of which we discarded 50,000 as burn-in, and we thinned the remaining iterations by 20 to make the output more manageable. We assessed chain convergence using the Gelman-Rubin statistic (Gelman et al., 2004). Values under 1.1 indicate convergence, and all parameters in our model had a Gelman-Rubin statistic <1.1. We tested whether the model adequately fit the data by calculating a Bayesian p-value (Gelman et al., 1996).

We used a similar model structure for the analysis of the empirical dataset. The full community occupancy model had the following parameterization:

$$w_{is} \sim Bernoulli(\Omega)$$
$$z_{ij} \sim Bernoulli(w_{is} * \psi_{ij})$$
$$logit(\psi_{ij}) = \alpha_{i,site[j]} + \beta1_i AGCD_j + \beta2_i FC_j + \beta3_i RRW_j + \varepsilon_{block[j]}$$
$$\alpha_{i,site[j]} \sim Normal(\mu_{\alpha,site[j]}, \sigma_{\alpha,site[j]})$$
$$\beta1_i \sim Normal(\mu_{\beta1}, \sigma_{\beta1})$$
$$\beta2_i \sim Normal(\mu_{\beta2}, \sigma_{\beta2})$$
$$\beta3_i \sim Normal(\mu_{\beta3}, \sigma_{\beta3})$$
$$\varepsilon_{block[j]} \sim Normal(\mu_\varepsilon, \sigma_\varepsilon)$$
$$y_{ijk} \sim Bernoulli(p_{ijk})$$
$$logit(p_{ijk}) = \alpha.p_{i,reserve[j]} + \beta.t_i time_{jk} + \beta.d_i date_{jk}$$
$$\alpha.p_{i,reserve} \sim Normal(\mu.p_{\alpha.p,reserve}, \sigma.p_{\alpha.p,reserve})$$
$$\beta.t_i \sim Normal(\mu.p_{\beta.t}, \sigma.p_{\beta.t})$$
$$\beta.d_i \sim Normal(\mu.p_{\beta.d}, \sigma.p_{\beta.d})$$

In the above formulas: z_{ij} is the true occupancy state (0 or 1) of species *i* at site *j*; ψ_{ij} is the respective occupancy probability; α is the intercept of the logit-linear predictor of occupancy probability, indexed by species and reserve, $\beta1$- $\beta3$ are the coefficients for above ground carbon density, forest cover and riparian width, respectively. Species specific intercepts and coefficients come from a Normal distribution with community means ($\mu_{\alpha,1}$ - $\mu_{\alpha,3}$ and $\mu_{\beta1}$ - $\mu_{\beta3}$ for coefficients) and standard deviations ($\sigma_{\alpha,1}$ - $\sigma_{\alpha,3}$ and $\sigma_{\beta1}$ - $\sigma_{\beta3}$ for coefficients). Following, y_{ijk} are the observations (0 or 1) of species *i* at site *j* at occasion *k*; p_{ijk} are the respective detection probabilities; α.p is the intercept of the logit-linear predictor of detection probability, indexed by species and reserve; *β.t* is the effect of time of day on detection probability, and *β.d* is the effect of date on detection probability. Species specific detection intercepts and *β.d* and *β.t* come from a Normal distribution with community means ($\mu.p_{\alpha.p,1}$ − $\mu.p_{\alpha.p,3}$, and $\mu.p_{\beta.t}$ and $\mu.p_{\beta.d}$ and for *β.t* and *β.d*, respectively) and standard deviations ($\sigma.p_{\alpha.p,1}$ − $\sigma.p_{\alpha.p,3}$, and $\sigma.p_{\beta.t}$, and $\sigma.p_{\beta.d}$ for *β.t* and *β.d*, respectively).

Rank order of site diversity

To evaluate how well occupancy-based diversity profiles were able to order landscapes by site diversity rank, we compared them to the diversity ranking in the true abundance-based profiles. Since not all communities were different in the true abundance-based profiles (e.g. no disturbance and patchy low disturbance landscapes both had an average species richness of 40 (Table S1) and could, therefore, not be distinguished in the true-abundance based profiles), we first determined how any sites we can reliably distinguish. To do so we used the results of the true abundance-based simulations and checked which landscapes could be distinguished in 95% of the 100 simulations for R, H', and D. We were

able to distinguish 3, 4, and 3 of the 5 sites in the true abundance profiles for R, H', and D, respectively (Table S1).

Table S1: Percentage of the true abundance-based simulations where pairwise diversity could be distinguished for R, H', and D. We were able to distinguish between 3, 4, and 3 of the 5 sites in the true abundance profiles for R, H', and D, respectively.

q = 0 (R)

	Site 1	Site 2	Site 3	Site 4
Site 2	**0%**	-	-	-
Site 3	100%	100%	-	-
Site 4	100%	100%	**87%**	-
Site 5	100%	100%	100%	100%

q = 1 (H')

	Site 1	Site 2	Site 3	Site 4
Site 2	97%	-	-	-
Site 3	100%	100%	-	-
Site 4	100%	100%	**72%**	-
Site 5	100%	100%	95%	97%

q = 2 (D)

	Site 1	Site 2	Site 3	Site 4
Site 2	98%	-	-	-
Site 3	99%	100%	-	-
Site 4	100%	100%	**44%**	-
Site 5	100%	100%	**82%**	**78%**

Construction of taxonomy, diet, and phylogeny similarity matrices
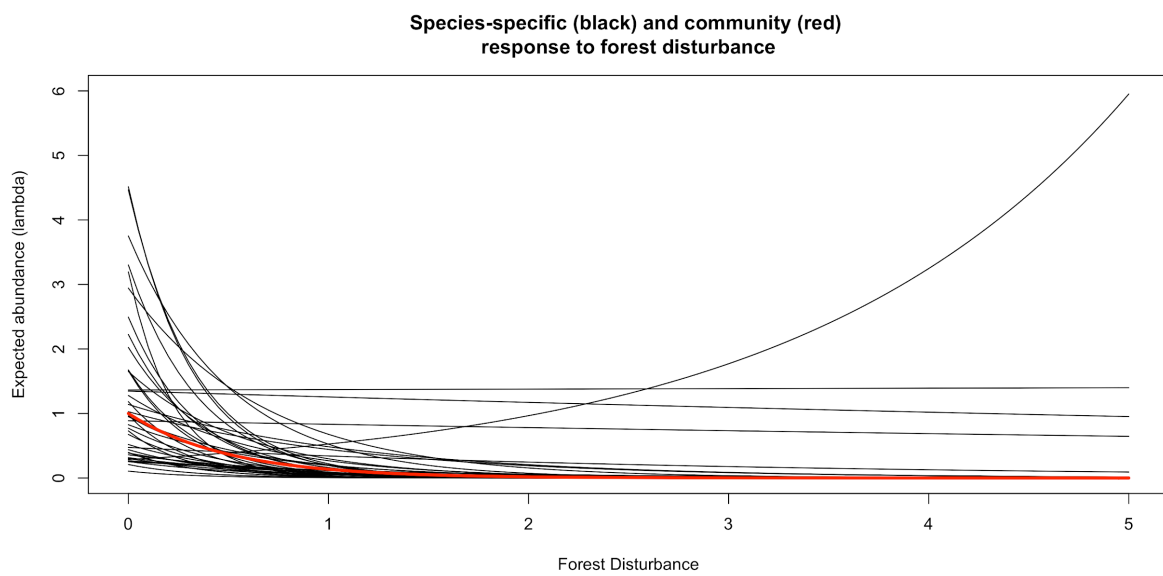
Following Leinster and Cobbold (2012) we constructed the taxonomy-based similarity matrix by defining a similarity matrix **Z** by

$$Z_{ih} = \begin{cases} 0 \text{ if the } i\text{th and } h\text{th species are of different genera} \\ 0.5 \text{ if the } i\text{th and } h\text{th species are different but congeneric} \\ 1 \text{ if } i = h \end{cases}$$

The diet similarity matrix was constructed using a pairwise comparison of species dietary preferences. The similarity matrix for phylogeny was constructed using the phylogenetic distance tree for the 167 species obtained from https://birdtree.org/. The values were then linearly rescaled to a minimum of 0 and a maximum of 1.

Distance relationship
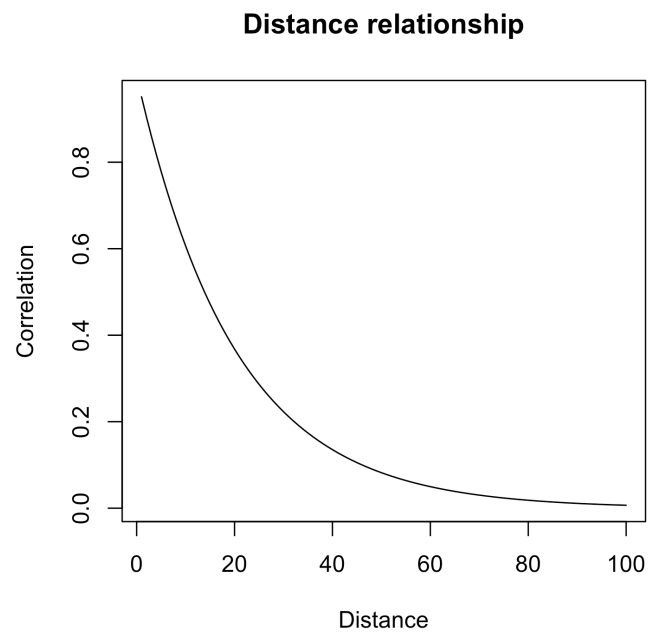
Figure S1. *Distance function used to simulate virtual landscape.*

Species-specific (black) and community (red)
response to forest disturbance

Figure S2. *Simulated responses to the habitat covariate for 40 species (black lines) according to the community hyperparameter (red line).*

Error Rate verses Threshold
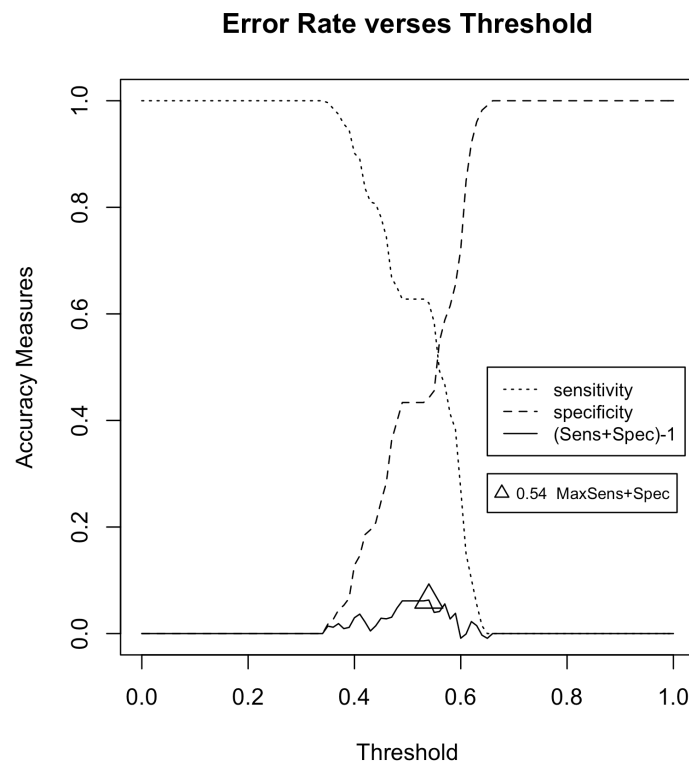
Figure S3. Example of threshold determination for one species using the max_{SSS} method that maximizes the sum of sensitivity and specificity. This was run as a Monte Carlo simulation to produce a distribution.

Table S2. *Summary of occupancy-based diversity index estimates for 5 simulated landscapes. Parentheses indicate the relative bias (bold) and 95% BCI coverage (italics) averaged over the 100 simulations when compared to the respective true abundance-based index. For the true and sample station abundance-based indices, parentheses represent the standard deviation over the 100 simulations.*

Landscape	Index	True abundance (sd)	Sample station abundance (sd)	Landscape-wide occupancy	Landscape-wide occupancy with threshold	Sample-station occupancy	Sample-station occupancy with threshold
No disturbance	**Species Richness (R, *q*=0)**	40.0 (±0)	39.8 (±0.4)	39.9 (**-0.1%**, *100%*)	38.1 (**-4.7%**, *56%*)	39.9 (**-0.1%**, *100%*)	37.5 (**-6.3%**, *54%*)
	Shannon diversity (H', *q*=1)	21.2 (±4.2)	22.0 (±4.3)	31.0 (**45.7%**, *0%*)	25.9 (**21.3%**, *39%*)	31.0 (**45.7%**, *0%*)	24.5 (**14.6%**, *61%*)
	Simpson's index (D, *q*=2)	14.8 (±4.9)	14.8 (±5.0)	26.6 (**104.9%**, *0%*)	20.5 (**56.7%**, *32%*)	26.6 (**104.8%**, *0%*)	18.5 (**42.2%**, *42%*)
Patchy low disturbance	**Species Richness (R, *q*=0)**	40.0 (±0)	38.6 (±1.1)	39.6 (**-0.9%**, *100%*)	37.1 (**-7.2%**, *16%*)	39.6 (**-0.9%**, *100%*)	36.2 (**-9.4%**, *13%*)
	Shannon diversity (H', *q*=1)	17.4 (±5.0)	16.8 (±4.9)	25.0 (**58.8%**, *0%*)	22.7 (**42.8%**, *26%*)	24.9 (**57.8%**, *0%*)	21.6 (**35.8%**, *40%*)
	Simpson's index (D, *q*=2)	10.9 (±4.7)	10.6 (±4.6)	19.3 (**118.5%**, *0%*)	16.8 (**86.5%**, *16%*)	19.2 (**116.7%**, *0%*)	15.6 (**74.6%**, *40%*)
Low disturbance across entire area	**Species Richness (R, *q*=0)**	35.1 (±1.9)	20.9 (±3.1)	32.5 (**-7.2%**, *98%*)	12.8 (**-63.5%**, *0%*)	32.5 (**-7.2%**, *98%*)	12.4 (**-64.5%**, *0%*)
	Shannon diversity (H', *q*=1)	6.4 (±3.3)	6.1 (±3.1)	9.7 (**87.2%**, *16%*)	6.0 (**14.6%**, *51%*)	9.7 (**85.9%**, *17%*)	5.9 (**12.4%**, *50%*)
	Simpson's index (D, *q*=2)	4.1 (±2.3)	4.1 (±2.2)	6.6 (**95.1%**, *13%*)	4.5 (**32.7%**, *61%*)	6.6 (**93.7%**, *13%*)	4.3 (**27.0%**, *64%*)
Patchy high disturbance	**Species Richness (R, *q*=0)**	33.0 (±2.2)	17.9 (±3.0)	30.8 (**-6.4%**, *99%*)	12.5 (**-62.1%**, *0%*)	30.8 (**-6.4%**, *99%*)	14.1 (**-57.1%**, *0%*)
	Shannon diversity (H', *q*=1)	5.5 (±3.0)	5.2 (±2.7)	8.3 (**87.6%**, *25%*)	5.8 (**30.0%**, *46%*)	8.3 (**88.9%**, *23%*)	6.1 (**36.8%**, *52%*)
	Simpson's index (D, *q*=2)	3.6 (±2.0)	3.6 (±2.0)	5.7 (**90.8%**, *18%*)	4.3 (**43.7%**, *49%*)	5.7 (**91.8%**, *15%*)	4.4 (**48.8%**, *45%*)
High disturbance	**Species Richness (R, *q*=0)**	17.2 (±2.8)	8.2 (±2.2)	22.6 (**33.8%**, *97%*)	5.1 (**-69.6%**, *1%*)	22.6 (**33.8%**, *97%*)	4.7 (**-71.9%**, *0%*)

|							
across entire area	**Shannon diversity (H', q=1)**	2.9 (±1.6)	2.7 (±1.4)	4.7 (**93.5%**, *27%*)	2.8 (**22.1%**, *50%*)	4.6 (**92.5%**, *27%*)	2.7 (**19.6%**, *49%*)
	Simpson's index (D, q=2)	2.2 (±1.2)	2.1 (±1.1)	3.6 (**90.4%**, *26%*)	2.5 (**38.9%**, *60%*)	3.5 (**89.5%**, *27%*)	2.4 (**39.0%**, *63%*)
Average	**Species Richness (R, q=0)**	33.0	25.1	33.1 (**3.8%**, *99%*)	21.1 (**-41.4%**, *15%*)	33.1 (**3.8%**, *99%*)	21.0 (**-41.8%**, *13%*)
	Shannon diversity (H', q=1)	10.9	10.6	15.7 (**74.6%**, *14%*)	12.6 (**26.1%**, *42%*)	15.7 (**74.1%**, *13%*)	12.2 (**23.9%**, *50%*)
	Simpson's index (D, q=2)	7.1	7.0	12.4 (**99.9%**, *11%*)	9.7 (**51.8%**, *44%*)	12.3 (**99.3%**, *11%*)	9.1 (**46.3%**, *48%*)
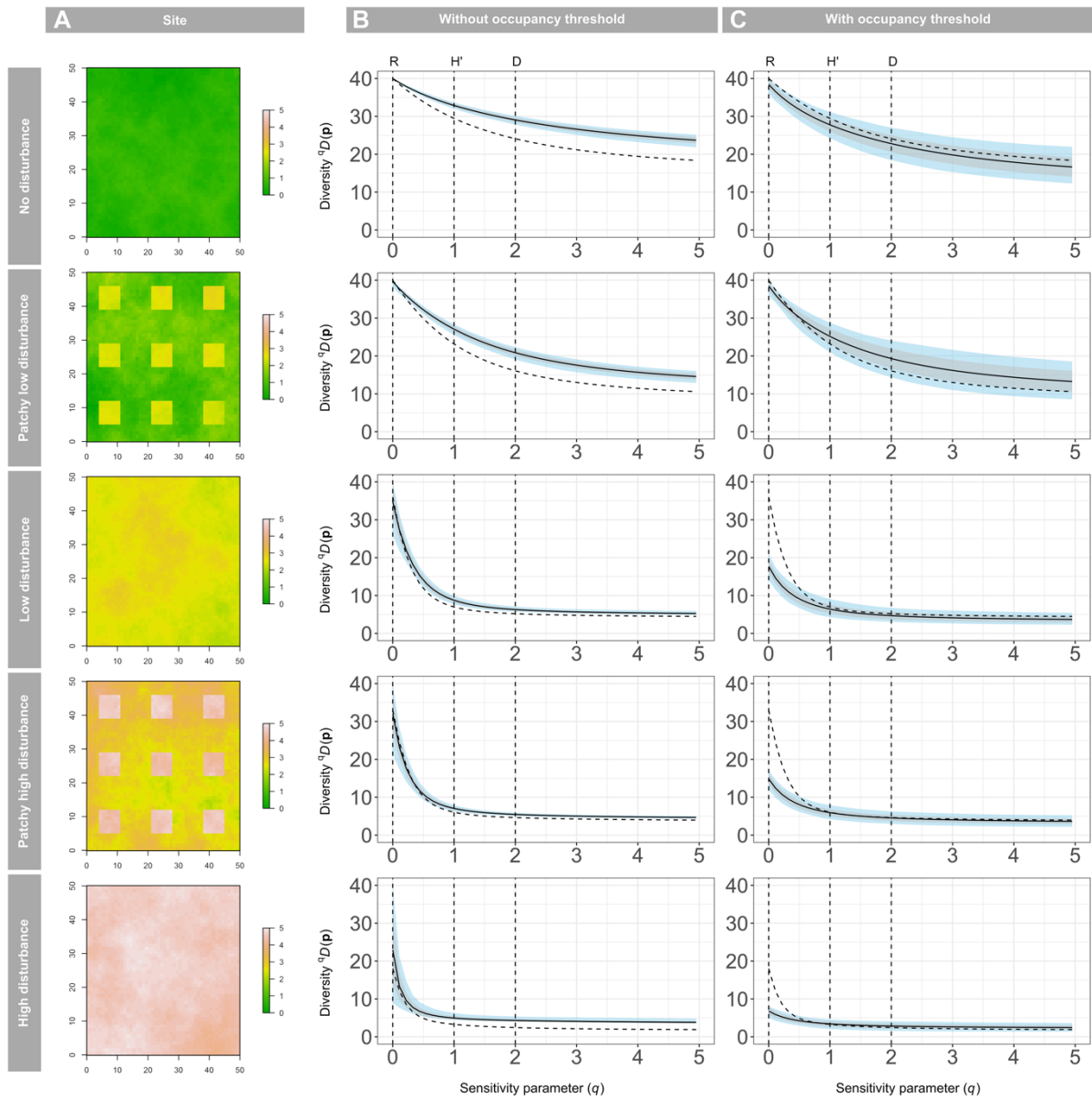
Figure S4. *Results from the simulation study. (A) Simulated "forest quality" habitat covariate for five virtual landscapes. Diversity profiles for simulated data generated using the community occupancy predictions for the entire landscape (B) without thresholding and (C) with thresholding using the max$_{SSS}$ method for the entire study landscape (solid line, SD grey shading, 95% CI blue shading) compared to the diversity profiles generated using the true abundance across the whole landscape (dashed line).*
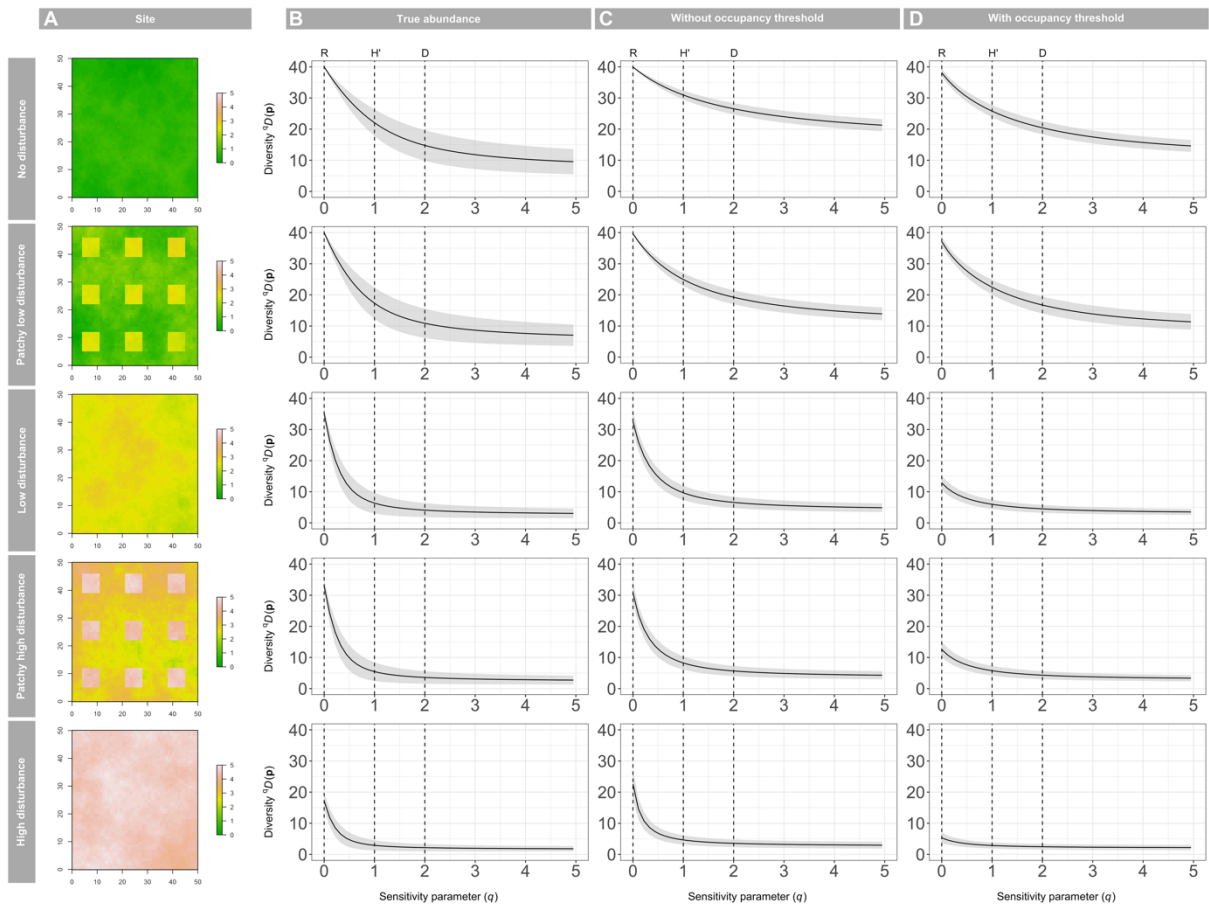
Figure S5. *Results from the simulation study. **(A)** Simulated "forest quality" habitat covariate for five virtual landscapes. Average diversity profiles (solid line) and standard deviation (grey shading) for simulated data generated using **(B)** the simulated true abundance, and the community occupancy predictions for the entire landscape **(C)** without thresholding and **(D)** with thresholding using the max$_{SSS}$ method for the entire study landscape.*
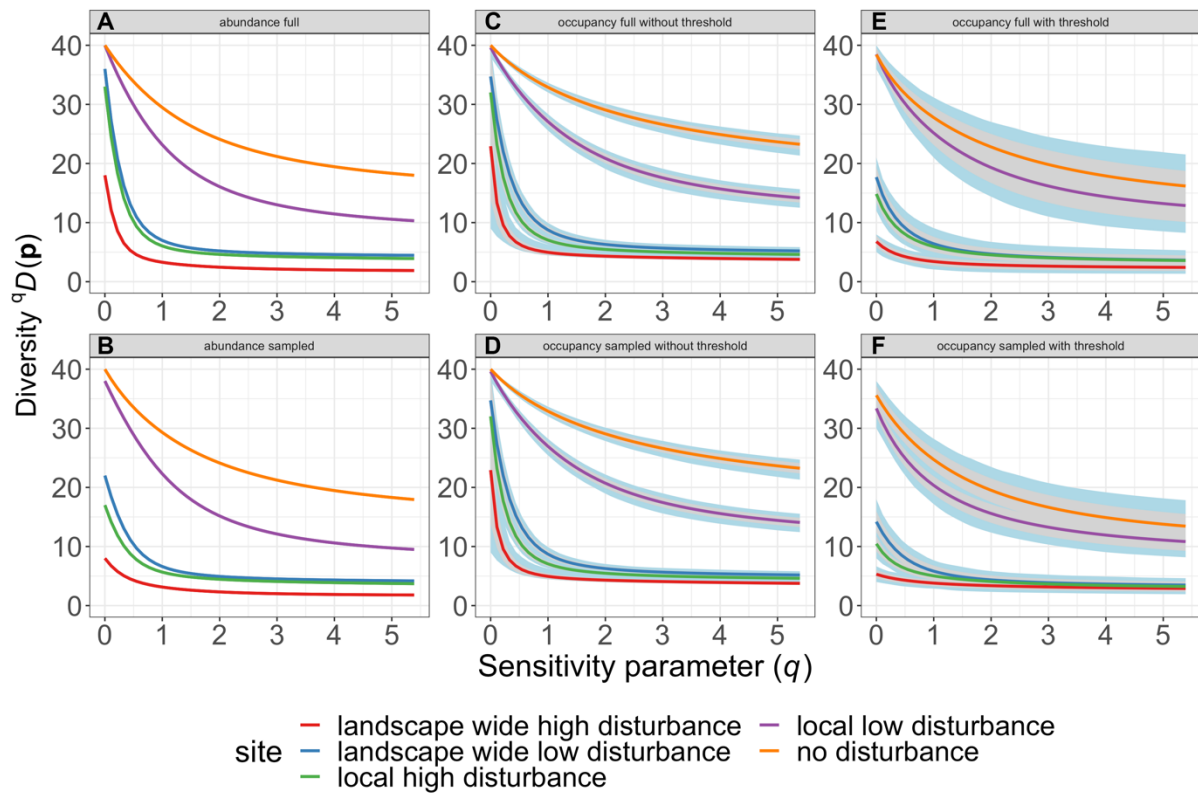
Figure S6. *Comparison among landscapes of diversity profiles generated using **(A)** the true abundance across the whole landscape, **(B)** the true abundance at the 100 sample stations in each landscape, **(C)** occupancy based predictions across the whole landscape without thresholding, **(D)** occupancy based predictions at the 100 sample stations in each landscape without thresholding, **(E)** occupancy based predictions across the whole landscape with thresholding, **(F)** occupancy based predictions at the 100 sample stations in each landscape with thresholding. The occupancy-based diversity profiles are derived quantities from the Bayesian models allowing for the calculation of uncertainty*
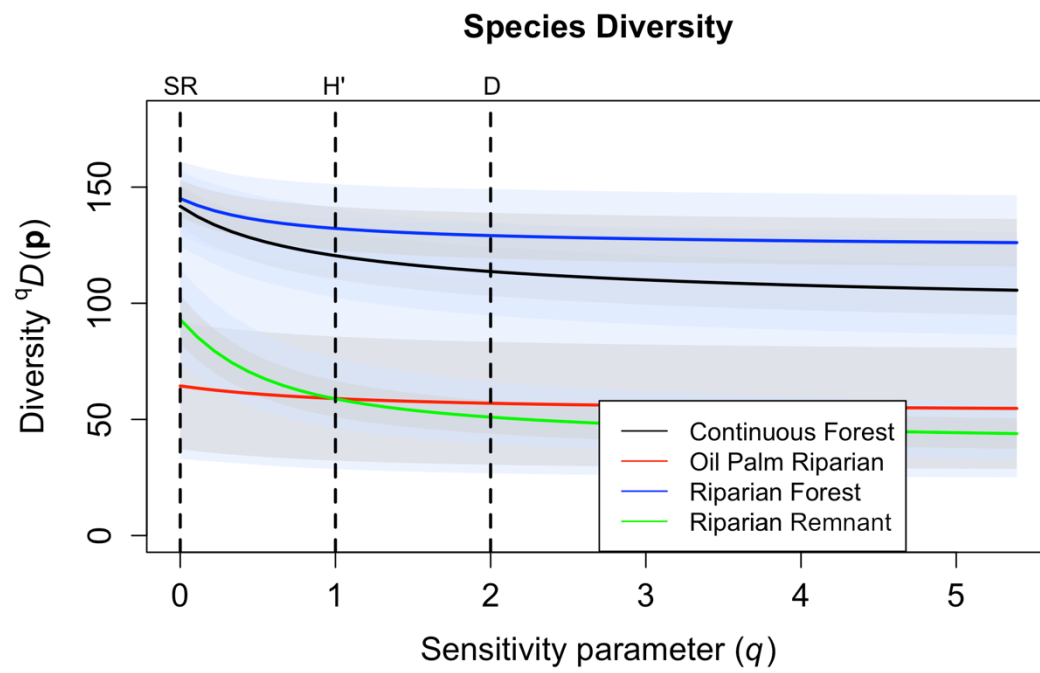
Figure S7. *Diversity profiles for the empirical bird dataset with occupancy thresholding using the max$_{SSS}$ method*