Animal Conservation

Animal Conservation. Print ISSN 1367-9430

Multi-scale, multivariate community models improve designation of biodiversity hotspots in the Sunda Islands

L. Chiaverini[1,*], D. W. Macdonald[1,*], H. M. Bothwell[1,2], A. J. Hearn[1], S. M. Cheyne[1,3], I. Haidir[1,4], L. T. B. Hunter[5], Ż Kaszta[1], E. A. Macdonald[1], J. Ross[1] & S. A. Cushman[1,6]

1 Wildlife Conservation Research Unit, Department of Zoology, The Recanati-Kaplan Centre, University of Oxford, Tubney, UK
2 Research School of Biology, Australian National University, Canberra, ACT, Australia
3 Borneo Nature Foundation, Palangka Raya, Indonesia
4 Directorate of Conservation Area Planning, Directorate General of Natural Resources and Ecosystem Conservation, Ministry of Environment and Forestry, Jakarta, Indonesia
5 Wildlife Conservation Society, Indonesia Program, Bogor, Indonesia
6 Rocky Mountain Research Station, United States Forest Service, Flagstaff, AZ, USA

Keywords
biodiversity hotspots; Borneo; community assemblage; multi-scale; multi-species; species richness; Sumatra; Sunda.

*Joint first authors.

Correspondence
David W. Macdonald, Wildlife Conservation Research Unit, Department of Zoology, The Recanati-Kaplan Centre, University of Oxford, Tubney House, Tubney, Oxon OX13 5QL, UK.
Email: david.macdonald@zoo.ox.ac.uk

Editor: Rahel Sollmann
Associate Editor: Miguel Acevedo

Received 23 October 2020; accepted 28 January 2022

doi:10.1111/acv.12771

Abstract

Species occur in sympatric assemblages, bound together by ecological relationships and interspecific interactions. Borneo and Sumatra host some of the richest assemblages of biota worldwide. The region, however, faces the highest global deforestation rates, which seriously threaten its unique biodiversity. We used a large camera trap dataset that recorded data for 70 terrestrial species of mammals and birds, to explore the drivers of regional species richness patterns. Using a multi-scale, multivariate modelling framework which quantified the main environmental factors associated with patterns of biodiversity, while simultaneously assessing individual relationships of each species, we determined the ecological drivers of sampled biodiversity, and their contributions to community assemblages. We then mapped predicted species richness, evaluated the effectiveness of protected areas in securing biodiversity hotspots, performed gap analysis to highlight biodiverse areas lacking protection and compared our predictions with species richness maps produced by using IUCN range layers. Finally, we investigated the performance of each species as an indicator of sampled biodiversity. We demonstrate that biodiversity in Borneo and Sumatra is primarily affected by gradients of ecological and anthropogenic factors, and only marginally by topographic and spatial factors. In both islands, species are primarily associated with elevational gradients in vegetation and climate, leading to altitudinal zonation in niche separation as a major factor characterizing the islands' biodiversity. Species richness was highest in north-eastern Borneo and in western Sumatra. We found that most predicted biodiversity hotspots are not formally protected in either island; only 9.2 and 18.2% of the modelled species richness occurred within protected areas in Borneo and Sumatra, respectively. We highlighted that our prediction for Borneo performed better than, and differed drastically from, the IUCN species richness layer, while for Sumatra our modelled species richness layer and the IUCN one were similar, and both showed low predictive power. Our analysis suggests that common and generalist carnivores are the most effective indicators of sampled biodiversity and have high potential as focal, umbrella or indicator species to assist multi-species vertebrate conservation planning. Understanding existing drivers and patterns of biodiversity is critical to support the development of effective community conservation strategies in this rapidly changing region.

Introduction

Species co-occurring within biological communities are affected individualistically by habitat factors and interactively by interspecific interactions (Hutchinson, 1957). Nonetheless, the simultaneous distributions of multiple species can also be described by common factors (Davis *et al*., 2018; Haidir, Macdonald & Linkie, 2018; Hearn *et al*., 2018*b*);

co-occurrence patterns might reflect shared habitat requirements (Hearn *et al.*, 2018*b*), be driven by biotic interactions (Ovaskainen, Hottola & Siitonen, 2010), predator-prey dynamics (Coleman & Hill, 2014) or be a consequence of anthropogenic pressures (Wang, Allen & Wilmers, 2015), and these factors might not operate in a mutually exclusive way (Macdonald *et al.*, 2020). Global intensification of habitat loss necessitates a paradigmatic shift from single-species conservation and management, towards multi-species approaches (Macdonald *et al.*, 2018*a*, 2012, 2020). Therefore, a fundamental ecological question, with profound practical implications, is how biological communities are assembled and what features control co-occurrence patterns.

This question is particularly relevant to the Sunda Islands of Borneo and Sumatra. Sundaland, as this region is known, is one of the Earth's most important biodiversity hotspots, whose exceptional species richness includes 1800 vertebrates, of which 701 are endemic (Myers *et al.*, 2000). However, the biodiversity of Sundaland is seriously threatened by anthropogenic factors, including deforestation (Cushman *et al.*, 2017) and habitat conversion (Wilcove *et al.*, 2013), both directly linked with the accelerating economic growth of the region (Sodhi *et al.*, 2004). South-east Asia shows the highest rates of deforestation globally (Hansen *et al.*, 2013), and primary vegetation in Sundaland occupies only 7.8% of its original extent (Myers *et al.*, 2000). Additionally, deforestation in the region is expected to increase over the next decade (Cushman *et al.*, 2017).

Our work expands upon previous studies carried out in smaller portions of the region, and focused mainly on the felid guilds (e.g. Haidir *et al.*, 2018; Hearn *et al.*, 2018*b*), not only by assessing vaster spatial extents, but also by investigating the structure and spatial distribution of a larger, more comprehensive sample of species. A similar study was recently completed in mainland South-east Asia (Macdonald *et al.*, 2020) and a major focus of this paper is comparing the patterns and drivers of biodiversity in the Sunda Islands to those in mainland South-east Asia, using comparable datasets and methods.

Macdonald *et al.* (2020) highlighted that, in mainland South-east Asia, landscape compositional factors were the strongest drivers of species richness, and biodiversity was strongly affected by the extensiveness of natural vegetation, particularly forested areas, as well as anthropogenic factors like protected and developed areas. Additionally, these affected mainland South-east Asian biodiversity at multiple spatial scales. Here, we applied a comparable multi-scale, multivariate framework (*sensu* McGarigal *et al.*, 2016, as modified by Macdonald *et al.*, 2020) to assess the joint effect of landscape, anthropogenic, topographic and spatial factors on the simultaneous presence of sampled species to test the same hypotheses as in Macdonald *et al.* (2020) related to species co-occurrence and habitat selection in Borneo and Sumatra: (1) multi-species occurrence patterns are primarily influenced by environmental factors (e.g. forested areas, forest loss and protected areas) and only marginally by topographic and spatial factors and, (2) the aforementioned factors influence species co-occurrence at multiple spatial scales. Furthermore, Hughes (2017) and Macdonald *et al.* (2020) found that the modelled geographic distribution of terrestrial biodiversity in South-east Asia drastically differs from the predictions made by using IUCN range layers; therefore, we produced spatially explicit predictions of species richness in Borneo and Sumatra, expecting that (3) our species richness predictions for Borneo and Sumatra would differ from the IUCN maps, and would have higher predictive ability. In addition, Macdonald *et al.* (2020) highlighted that the network of protected areas in mainland South-east Asia preserved only a small portion of the modelled species richness; therefore, we performed gap analysis using the current protected areas network, hypothesizing that (4) existing protected areas are relatively poor in location and insufficient in extent to protect the main biodiversity hotspots. Additionally, building upon the finding of Macdonald *et al.* (2020), we expect that (5) species richness predictions are primarily driven by common species and only marginally affected by rare species, and that (6) carnivores are good indicators of overall biodiversity, as demonstrated in other regions (Carroll, Noss & Paquet, 2001; Dalerum *et al.*, 2008; Macdonald *et al.*, 2020).

Materials and methods

Data collection

From 2007 to 2016, our field teams conducted an extensive camera trap survey targeting wild felid populations across Borneo and Sumatra (Fig. 1; Table 1). Cameras were set 1.0–2.0 km apart, with two cameras per station set ~40 cm above ground, and were deployed along natural ridgelines and disused logging roads allowing for greater visibility (Hearn *et al.*, 2017; Macdonald *et al.*, 2018*a*). The survey provided a rich dataset giving additional insights into meso and macro regional terrestrial biodiversity. All terrestrial mammals and birds whose species was unambiguously identified were selected for the analysis. When a species could not be clearly identified, we retained the data at a broader taxonomic level (i.e. order, family or genus). Additionally, to better assess the habitat factors driving species assemblages, we also included data related to anthropogenic disturbances captured by camera traps, such as people and domestic animals (i.e. domestic cats, domestic cattle and domestic dogs). For each species, we used the number of detections per camera trap station, reducing overestimation bias by discounting records of the same species at the same camera station within 1 h, except when animals were individually recognizable (e.g. based on the pattern of spots and stripes for felids) and when sex and/or age class were unambiguous.

Habitat covariates

We selected a preliminary set of 35 covariates considered important for regional terrestrial biodiversity (Macdonald *et al.*, 2018*a*; Hearn *et al.*, 2018*b*), including 19 landscape, four anthropogenic, four topographic and eight spatial covariates (Table 2). Additionally, we included camera effort as a

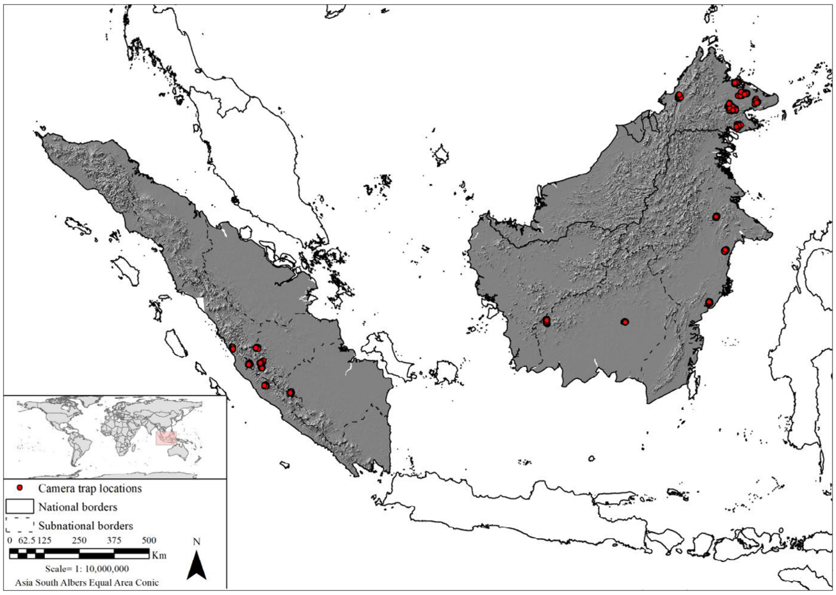

Figure 1 Map of the study area, including the islands of Borneo (right) and Sumatra (left), showing the camera trap locations. Each location contains one or more camera trap grids. Number and model of camera traps for each location are provided in Table 1.

Table 1 Sampling effort in Borneo and Sumatra. Shown are sampling locations per region, survey dates, extent of the surveyed areas, number of camera trap stations per sampling location, total number of trap nights and average number of trap nights per camera trap station, maximum and average number of species detected per sampling location, protected status of the sampling locations and camera trap models. Protected status acronyms: No (not protected), NR (Nature Reserve), PF (Production Forest), NP (National Park), Pl (plantation), CA (Conservation Area), WS (Wildlife Sanctuary), FR (Forest Reserve), WR (Wildlife Reserve). Camera trap model acronyms: CA (Cuddeback Ambush IR), Pa3 (Panthera V3), SS (Snapshot Sniper), Re5 (Reconyx HC 500), Bu (Bushnell), Re8 (Reconyx PC 800), Pa4 (Panthera V4)

Regions	Sampling locations	Survey dates	Surveyed area (km^2)	Camera trap stations	Total trap nights (mean)	Max species detected (mean)	Protected status	Camera trap model
Kalimantan	Bawan	09/09/2012–26/11/2012	32.9	63	4070 (64.60)	4 (0.94)	No	CA
	Belantikan	23/02/2014–17/06/2014	112.2	51	3846 (75.41)	8 (3.57)	NR/PF	CA
	Kutai	12/12/2012–23/03/2013	88.9	52	3269 (62.87)	9 (3.37)	NP	CA
	Lesan	15/07/2013–26/01/2014	70.8	73	10 057 (137.77)	11 (3.03)	NR	CA
	Sungai Wain	17/05/2012–08/08/2012	69.6	78	4729 (60.63)	7 (3.54)	NR	CA
Sabah	Crocker	06/10/2011–07/02/2012	149.7	35	4059 (115.97)	31 (17.77)	NP	Pa3/SS/Re5/Bu
	Danum Palm	15/03/2009–07/07/2009	7.8	23	2214 (96.26)	21 (8.00)	Pl	SS
	Danum Valley	24/03/2012–06/10/2012	157.0	79	5880 (74.43)	25 (15.13)	CA	Pa3/Re5/Re8
	IJM	26/05/2011–18/08/2011	44.0	33	1855 (56.21)	15 (3.73)	Pl	Pa3/SS/Re5/Bu
	Kinabatangan	24/07/2010–17/12/2010	359.5	68	4450 (65.44)	17 (7.81)	WS	SS/CA/Bu
	Malua	09/07/2008–12/02/2009	102.8	38	3867 (101.76)	27 (14.26)	FR	SS
	Sepilok	09/02/2011–25/05/2011	49.4	35	2067 (59.06)	17 (11.11)	FR	Pa3/SS/Re5/Bu
	Tabin	18/09/2009–22/04/2010	144.3	74	6877 (92.93)	24 (15.53)	WR	SS/CA
	Tawau	21/10/2012–30/12/2013	149.0	77	17 600 (228.57)	33 (20.12)	NP	Re5/Re8
	Ulu Segama	24/05/2007–18/10/2007	60.1	22	2847 (129.41)	26 (18.27)	FR	SS
Sumatra	Bungo	11/06/2014–11/11/2014	74.3	79	8405 (106.39)	14 (5.49)	NP	Pa4/CA
	Ipuh	06/09/2015–24/12/2015	63.6	79	6276 (79.44)	17 (6.49)	NP	CA
	Kambang	01/04/2015–08/07/2015	54.4	149	12 599 (84.56)	9 (2.81)	NP	CA
	Linggau	10/01/2016–24/03/2016	47.5	120	7269 (60.58)	11 (3.55)	NP	CA
	Muara Hemat	17/10/2014–05/02/2015	72.0	164	13 466 (82.11)	13 (4.83)	NP	CA
	RKE	25/04/2015–10/08/2015	91.4	72	5586 (77.58)	18 (4.42)	NP	CA
	Sipurak	24/11/2014–08/04/2015	80.0	80	7227 (90.34)	18 (6.25)	NP	CA
Total		24/05/2007–24/03/2016	2081.2	1544	138 515 (89.71)	33 (7.15)		

predictor variable, defined as the cumulative number of nights in which camera traps were active. Similar to Macdonald *et al*. (2020), we kept the distinct covariates' grouping (i.e. environmental, defined as the combination of landscape and anthropogenic covariates, topographic, spatial and camera effort), to evaluate the contribution of each of them to sampled species detection counts.

Species-habitat relationships are known to be scale-dependent (Cushman & McGarigal, 2002; Cushman & McGarigal, 2004*b*; McGarigal *et al*., 2016). To investigate the scales at which sampled biodiversity responded to habitat features, we calculated each covariate at eight spatial scales, by averaging the pixel values of the original raster layers using circular buffers of 250, 500, 1000, 2000, 4000, 6000, 8000 and 10 000 m radii, centred on each camera trap location. Scales were chosen to cover a wide range of behavioural processes, describing different ways in which sampled species could interact with environmental features of the landscape, from fine-scale resource selection within home ranges, to broad-scale selection of home ranges (Simcharoen *et al*., 2014; Hearn *et al*., 2018*a*).

To produce more biologically informative variables, we calculated class- and landscape-level metrics in FRAGSTATS (McGarigal *et al*., 2012) for discrete layers, and focal statistics in ArcMap v10.6.1 (ESRI, 2018) for continuous layers, using for both the software circular moving windows of radii corresponding to the aforementioned spatial scales. At the class level, defined as every patch type in the landscape (e.g. every class in a land-cover layer), we calculated PLAND, representing the percentage of the landscape occupied by a specific class, and GYRATE_AM, providing a measure of landscape continuity. At the landscape level, defined as the combination of all classes in the patch mosaic (e.g. the entire land-cover layer, comprising all cover classes simultaneously), we computed aggregation index (AI), edge density (ED), contrast-weighted edge density (CWED), patch density (PD) and largest patch index (LPI). These landscape metrics were selected following past research showing they are particularly strong indicators of species occurrence patterns in complex landscapes (Grand *et al*., 2004; Chambers *et al*., 2016; Macdonald *et al*., 2020). For continuous layers we calculated focal mean (FM) and standard deviation (SD). These resulted in a total of 67 covariates (Table 2).

Scale optimization

Since interspecific relationships among species assemblages are different between Borneo and Sumatra, we implemented the modelling workflow previously applied by Macdonald *et al*. (2020), and described in the next paragraphs, independently for the two islands. We removed all poorly sampled covariates (i.e. those occurring at <10% of camera stations) to avoid assessing unrepresentative habitat features, since zero-inflation of explanatory variables can negatively affect parameter estimation (Martin *et al*., 2005).

To identify representative scales for sampled biodiversity, we performed Canonical Correspondence Analysis (CCA; Fig. 2a) (ter Braak, 1986; ter Braak & Prentice, 1988; McGarigal, Cushman & Stafford, 2000) on randomly sampled 80% of the camera trap stations (we used the remaining 20% to evaluate the ability of the species richness surface to correctly predict observed richness – see below), independently at each scale for each covariate, using the *cca* function in the *vegan* package (Oksanen *et al*., 2018) in R v3.5.1 (R Core Team, 2018). CCA is a multivariate modelling technique that accounts for unimodal distribution of response variables, which are expected in species presence patterns across broad environmental gradients (Austin *et al*., 1994). CCA allowed us to model species-habitat relationships simultaneously for all sampled species and, at the same time, to account for the individual contribution of each species to the modelled relationships (Macdonald *et al*., 2020). For each covariate, at each scale, we performed CCA using the number of detections of the sampled species as response variables, and we identified the optimal scale by selecting the scale whose CCA showed the highest canonical *eigenvalue* (Borcard, Legendre & Drapeau, 1992). We then checked multicollinearity by calculating Pearson's correlation coefficient between all pairs of scale-optimized covariates (Fig. 2b). When two covariates were highly correlated ($|r| \geq 0.7$), we removed the covariate whose univariate CCA had the lower adjusted-R^2 (Guisan & Zimmermann, 2000). The final set of covariates was selected by performing forward selection on the previously selected 80% of camera trap stations, using the *ordistep* function in the *vegan* package (Fig. 2c), independently on each group of covariates (i.e. environmental, topographic and spatial) and, within the multivariate context of the CCA, simultaneously for all sampled species, retaining only those covariates significant at $P < 0.001$ within each group (Cushman & McGarigal, 2002; Cushman & McGarigal, 2004*a*).

We performed variance partitioning analysis (Fig. 2d) to investigate the variance explained in the multivariate detection counts of sampled species by each group of covariates (individual contribution), as well as by different combinations of groups of covariates (joint contribution) (Borcard *et al*., 1992; Cushman & McGarigal, 2004*a*, *b*), by using the number of detections of the sampled species as response variables and applying the *varpart* function in the *vegan* package.

Species richness and model performance

We performed CCA between the number of detections of the sampled species at the previously selected 80% of camera trap stations and the final set of scale-optimized covariates (Fig. 2e) to assess their contribution to the sampled species detections, and interpreted the biplot to evaluate the interactions of each species with specific habitat factors (Macdonald *et al*., 2020). CCA biplots are graphical representations of the multivariate results that simultaneously show both the sampled species and the influence of the covariates, which are shown as vectors pointing in the direction of increasing value and with length proportional to their influence on predicted species composition and detection. The vectors' lengths (i.e. the influence of the covariates on the species) are given by the *eigenvalues* calculated in the CCA. Species

Table 2 Preliminary set of covariates and corresponding representative scales for biodiversity in Borneo and Sumatra. Additionally, the description of the primary set of covariates selected for Borneo and Sumatra biodiversity models is provided. For the analyses, landscape and anthropogenic covariates were merged into the macro-group of environmental covariates. We calculated class- and landscape-level metrics in FRAGSTATS (McGarigal, Cushman & Ene, 2012) for discrete layers, and focal statistics in ArcMap v10.6.1 (ESRI, 2018) for continuous layers. PLAND, represents the percentage of the landscape occupied by a specific class, GYRATE_AM is a measure of landscape continuity. At the landscape level, we computed aggregation index (AI), edge density (ED), contrast-weighted edge density (CWED), patch density (PD) and largest patch index (LPI). For continuous layers we calculated focal mean (FM) and standard deviation (SD). All raster layers were first re-projected to Asia South Albers Equal Area Conic projection in ArcMap v10.6.1, by applying nearest neighbour re-sampling technique for discrete layers and bilinear interpolation re-sampling technique for continuous layers. Then, by applying the same re-sampling techniques, all raster layers were re-sampled to 250 m resolution

				Borneo	Sumatra
Class	Covariate	Source	Metric	Best scale (m)	Best scale (m)
Landscape	Landcover	Miettinen *et al.* (2012)	AI	4000[a]	4000[a]
			CWED	2000[a]	4000[a]
			ED	2000[a]	4000[a]
			PD	4000[a]	2000[a]
	Lowland forests	Miettinen *et al.* (2012)	PLAND	4000[a]	8000[a]
			GYRATE_AM	8000	10 000
	Lowland mosaic areas	Miettinen *et al.* (2012)	PLAND	4000[a]	6000[a]
			GYRATE_AM	1000[a]	6000[a]
	Lower-montane forests	Miettinen *et al.* (2012)	PLAND	6000[a]	4000[a]
			GYRATE_AM	6000[a]	2000[a]
	Lowland open areas	Miettinen *et al.* (2012)	PLAND	2000[a]	6000[b]
			GYRATE_AM	2000	8000
	Mangroves	Miettinen *et al.* (2012)	PLAND	10 000	NA[c]
			GYRATE_AM	4000[a]	NA[c]
	Montane mosaic areas	Miettinen *et al.* (2012)	PLAND	NA[c]	2000[a]
			GYRATE_AM	NA[c]	2000
	Montane open areas	Miettinen *et al.* (2012)	PLAND	NA[c]	8000[a]
			GYRATE_AM	NA[c]	8000[b]
	Palm plantations	Miettinen *et al.* (2012)	PLAND	2000	10 000[a]
			GYRATE_AM	4000[a]	10 000
	Peat swamp forests	Miettinen *et al.* (2012)	PLAND	NA[c]	NA[c]
			GYRATE_AM	NA[c]	NA[c]
	Plantation and regrowth	Miettinen *et al.* (2012)	PLAND	10 000[a]	2000[a]
			GYRATE_AM	500[a]	1000
	Upper-montane forests	Miettinen *et al.* (2012)	PLAND	NA[c]	4000
			GYRATE_AM	NA[c]	4000[a]
	Urban areas	Miettinen *et al.* (2012)	PLAND	NA[c]	NA[c]
			GYRATE_AM	NA[c]	NA[c]
	Water	Miettinen *et al.* (2012)	PLAND	6000[a]	NA[c]
			GYRATE_AM	4000[b]	NA[c]
	Continuous tree canopy cover	Hansen *et al.* (2013)	FM	4000[a]	4000[a]
			SD	2000[a]	4000[a]
	Tree canopy cover reclassified	Hansen *et al.* (2013)	AI	2000[a]	4000[a]
			CWED	2000[a]	4000[a]
			ED	2000[a]	4000[a]
			LPI	4000[a]	4000[a]
			PD	4000	4000[a]
	Non-forested areas (0–20%)	Hansen *et al.* (2013)	PLAND	4000[a]	2000[a]
			GYRATE_AM	1000[a]	2000[a]
	Open forests (20–40%)	Hansen *et al.* (2013)	PLAND	6000[a]	6000[a]
			GYRATE_AM	2000[b]	4000[b]
	Closed forests (>40%)	Hansen *et al.* (2013)	PLAND	4000[a]	4000[a]
			GYRATE_AM	6000[a]	6000[b]

Table 2 Continued

Class	Covariate	Source	Metric	Borneo Best scale (m)	Sumatra Best scale (m)
Anthropogenic	Human footprint	WCS CIESIN (2005)	FM	10 000	10 000[a]
			SD	2000	10 000
	Human population density	CIESIN, CIAT (2016)	FM	6000[a]	10 000[a]
			SD	6000[b]	8000[b]
	Forest loss	Hansen *et al.* (2013)	PLAND	10 000	4000
			GYRATE_AM	8000[a]	2000[a]
	Protected areas	IUCN, UNEP-WCMC (2017)	PLAND	4000	10 000[a]
			GYRATE_AM	4000[a]	8000
Topographic	Elevation	Jarvis *et al.* (2008)	FM	250	250
			SD	8000[a]	10 000[a]
	Roughness	Jarvis *et al.* (2008), Evans *et al.* (2014)	FM	4000[a]	10 000[a]
			SD	6000[a]	10 000
	Slope position	Jarvis *et al.* (2008), Evans *et al.* (2014)	FM	8000	10 000
			SD	4000[a]	10 000[a]
	Compound topographic index	Jarvis *et al.* (2008), Evans *et al.* (2014)	FM	2000[a]	500
			SD	4000	4000
Spatial	X	NA	NA	NA	NA
	Y	NA	NA	NA[a]	NA[a]
	$X \times Y$	NA	NA	NA[a]	NA[a]
	X^2	NA	NA	NA[a]	NA[a]
	Y^2	NA	NA	NA[a]	NA[a]
	$X^2 \times Y$	NA	NA	NA[a]	NA[a]
	$X \times Y^2$	NA	NA	NA[a]	NA[a]
	$X^2 \times Y^2$	NA	NA	NA	NA[a]
Camera effort	Camera trap nights	NA	NA	NA	NA

[a] Covariates excluded from the model after the multicollinearity analysis.

[b] Covariates excluded from the model after the forward selection analysis.

[c] Covariates excluded from the model because occurring in <10% of camera trap stations.

are represented in the biplot by dots, and distances from the origin along the dot-product vector of each covariate are proportional to the correlation between that species' number of detections and the value of that predictor variable.

To predict species richness, we first performed single-species generalized linear models (GLMs) by using the package *stats* in R v3.5.1, independently for each species (Fig. 2f; Supporting Information Tables S1 and S2), using the standardized final set of scale-optimized covariates from the CCA as predictor variables, and accommodating the continuous discrete response variables (i.e. species' detections) by adopting a quasi-Poisson distribution (Vincent & Haworth, 1983). We opted for a quasi-Poisson distribution given that the data for several species were overdispersed and did not meet the Poisson distribution's assumption of equal mean and variance (Supporting Information Figures S1 and S2).

We then used the parameter estimates from the single-species GLMs in combination with the covariate layers (Table 2) for Borneo and Sumatra to predict species counts across the entire extent of both islands. The single-species predictions were then converted to binary form, treating <1 value pixels as absences, and ≥1 value pixels as presences (Fig. 2g). Finally, the single-species binary surfaces were combined into a species richness layer by addition (Fig. 2h; Grand *et al*., 2004; Macdonald *et al*., 2020).

Since the camera trap surveys did not sample the full gradients of all the chosen predictor variables over the entire study area (e.g. most of the camera trap locations occurred within protected areas, surveys did not cover the most remote areas occurring at high elevations and camera traps were mainly deployed in easily accessible areas like along disused logging roads; Fig. 1), to avoid producing unreliable models by making predictions on areas that were ecologically too different from the sampling locations, we masked out the most different areas by producing a Mahalanobis distance surface. The Mahalanobis distance is the distance between two points in a multivariate space, that we defined by using the layers of the environmental factors that most likely determined a geographically biased camera trap survey: tree canopy cover, protected areas, human footprint, elevation, roughness, annual maximum temperature, annual minimum temperature and annual precipitations. The climatic covariates were derived from PRISM Climate Group (2016), while the landscape, anthropogenic and topographic covariates' sources were the same as the ones used to model species richness (Table 2). The analysis, performed by using the package *dismo* (Hijmans *et al*., 2017) in R v3.5.1, provided a continuous surface of the distances in the multidimensional space of each pixel from those where the camera traps were deployed. We then binarized the continuous surface by calculating the 20th percentile of the distribution of the

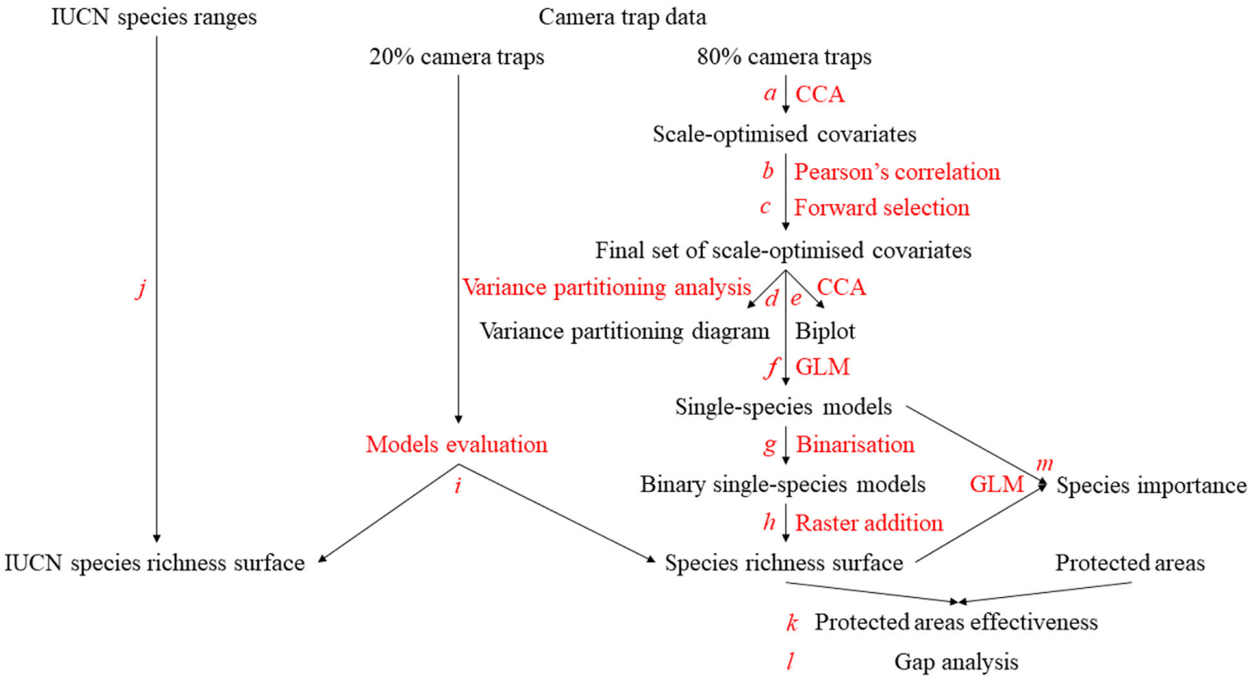

Figure 2 Workflow diagram of the methods applied to map the hotspots of terrestrial vertebrate biodiversity in Borneo and Sumatra using camera trap species detection data. Using 80% of camera traps, we first performed canonical correspondence analysis (CCA) independently for each covariate, for each scale, using the number of detections of the sampled species as response variable, to assess the most representative spatial scales for sampled biodiversity (a). Then, we calculated Pearson's correlation between all remaining covariates, and removed those highly correlated at $|r| \geq 0.7$ (b) and we obtained the final set of covariates by applying forward selection, retaining covariates at $P < 0.001$ (c). We performed variance partitioning to evaluate the variance explained by different groups of covariates (d), and performed CCA between the final set of scale-optimized covariates and the number of detections of the sampled species, producing biplots to evaluate relationships of sampled species with covariates (e). We then produced single-species generalized linear models (GLMs), independently for each species and predicted these across both islands (f), converted the predictions to binary by considering pixels with values <1 as absences, and as presences otherwise (g), and produced a species richness layer by adding the single-species binary models (h). We evaluated the performance of the species richness surface by using the 20% of camera traps held out for validation and calculating Nagelkerke-pseudo-R^2 (i). We also produced a layer of species richness by adding IUCN range layers for the sampled species (j), and evaluated the IUCN species richness surface by using the same method used to evaluate the modelled one (i). Finally, we evaluated the effectiveness of the current system of protected areas to preserve modelled species richness (k), and performed gap analysis evaluating the most biodiverse areas lacking formal protection (l). Finally, we evaluated the importance of each species as an indicator to represent the distribution of the overall sampled biodiversity (m).

Mahalanobis distances and excluded all pixels with distance values equal to or higher than that value from our predictive surfaces.

We produced the single-species models by using the previously randomly selected 80% of camera trap stations, keeping the remaining 20% to evaluate the ability of the species richness surface to predict observed richness by adopting the method applied by Macdonald *et al.* (2020): we computed the number of species predicted by our species richness layer at the camera trap stations used for validation, and performed GLM with Poisson distribution between the predicted and the empirical number of species, and then calculated the Nagelkerke-pseudo-R^2 (Nagelkerke, 1991; Fig. 2i). The Nagelkerke-pseudo-R^2 is an index ranging from 0 to 1 that can be used to compare the performances of different models. Additionally, we assessed the performances of the single-species GLMs by calculating their residual deviances, which indicate how well the response variables (i.e. species' detections) are predicted when the predictor variables are included in the models. The lower the residual deviance, the better the single-species GLM explains the species detection data.

We compared the multi-species predictive surface with a species richness surface produced by combining the IUCN geographic range layers of the sampled species (Fig. 2j) (IUCN, 2019) in ArcMap v10.6.1. To produce the IUCN surface, we used the polygons of the extent of occurrence in which the species were considered extant and resident. We evaluated the performance of the IUCN surface by calculating the Nagelkerke-pseudo-R^2 between the empirical number of species at the camera stations held out for validation and the number of species predicted by the IUCN surface (Fig. 2i).

Protected areas effectiveness

We assessed the effectiveness of the current system of protected areas network by calculating the ratio between the cumulative number of species predicted within protected areas as defined by the World Database on Protected Areas (WDPA; IUCN, UNEP-WCMC, 2017) relative to the cumulative number of species predicted in the entire islands (Grand *et al.*, 2004; Macdonald *et al.*, 2020), independently for Borneo and Sumatra. Hence, we evaluated the performances of each protected area in preserving predicted species richness (Fig. 2k).

Gap analysis

Additionally, we reclassified the species richness surface into low, medium and high priority areas for conservation, based on the percentiles distribution (>50th, >70th and >90th, respectively, Macdonald *et al.*, 2018*a*). Then, we identified the most important biodiversity hotspots not formally protected by calculating the ratio between the cumulative number of species predicted in the most biodiverse areas (>90th percentile) relative to the cumulative number of species predicted in the entire islands, independently for Borneo and Sumatra (Fig. 2l).

Indicator species effectiveness

To assess the importance of each species as an indicator of sampled biodiversity, we randomly sampled 1% of the pixels from the multi-species surface and from each single-species presence-absence surface and then performed GLMs independently for each species, in which the predictor variable was represented by the modelled number of species, while the response variables were represented by each species' presence-absence (Fig. 2m). The lower the residual deviance of the model, the more important that species was for explaining overall terrestrial vertebrate biodiversity (Macdonald *et al.*, 2020).

Results

Data collection

Camera traps were set in 15 sampling locations in Borneo (five in the Indonesian provinces of Kalimantan and 10 in the Malaysian state of Sabah) and in seven sampling locations in Sumatra, each covering an average area of 91.63 km^2 (±14.97 SE) and yielding a total of 1544 camera trap stations (801 in Borneo, of which 317 in Kalimantan and 484 in Sabah, and 743 in Sumatra). We achieved a combined sampling effort of 138 515 trap nights, with 77 687 trap nights in Borneo (mean = 96.99 ± 2.12 SE) and 60 828 in Sumatra (mean = 81.87 ± 0.72 SE; Table 1).

Species were recorded at all 22 sampling locations, including 747 camera trap stations in Borneo (93.26%) and 662 in Sumatra (89.10%). The 54 camera trap stations that failed to detect any species in Borneo were all in Kalimantan. The maximum number of species detected at a single camera trap station was 11 in Kalimantan, 33 in Sabah and 18 in Sumatra (Table 1). Overall, we sampled 58 species, including 47 mammals and 11 terrestrial birds. Additionally, we sampled 10 mammalian and two avian groups for which we retained data at broader taxonomic levels. Not all the species were detected in the two islands; total detections for sampled species are reported in Supporting Information Table S3. Moreover, we detected four types of anthropogenic disturbance: domestic cats, domestic cattle, domestic dogs and humans (Supporting Information Table S3). These anthropogenic elements were included only in the multivariate analyses as part of the response variable alongside the sampled species, to assess the most representative spatial scales, the covariates to be included in the final models and to produce the biplots. We included these anthropogenic elements in order to evaluate their contribution to the species-habitat relationships, as well as to assess their multivariate relationships with sampled species, but they were not used to produce the GLMs and the species richness surface, because we only wanted to predict wild species richness patterns.

Scale optimization

Ten covariates in Borneo and eight in Sumatra occurred at <10% of camera trap stations and were excluded due to lack of predictive power. Scale-optimization revealed that, on average, vertebrates were associated with medium (2–6 km) and broad (8–10 km) scale patterns. Four covariates in Borneo (8.16%) and three in Sumatra (5.88%) were identified as optimal at fine scales (≤1 km). Thirty-seven covariates in Borneo (75.51%) and 30 in Sumatra (58.82%) were selected at medium scales, while eight covariates in Borneo (16.33%) and 18 in Sumatra (35.29%) were selected at broad scales (Table 2).

We excluded 40 highly correlated covariates from Borneo and 39 from Sumatra. Following forward selection, we removed three covariates from Borneo and five from Sumatra, obtaining a final set of 15 and 16 covariates, respectively (Table 2).

The final sets of covariates explained 23.9 and 10.4% of absolute variance of the detection data in Borneo and Sumatra, respectively. Variance partitioning revealed that, in Borneo, species detection was most strongly predicted by camera effort (22% individual contribution of the single group of covariates and 67% joint contribution of the group of covariates with other groups of covariates), followed by environmental covariates (17% individually and 49% jointly). Variance explained by environmental covariates was in turn shared between landscape covariates (6% both individually and jointly) and anthropogenic covariates (11% both individually and jointly). Topographic and spatial covariates' independent contributions were comparatively lower (7 and 5%, respectively), but their shared contributions were more substantial (42 and 13%, respectively; Fig. 3a). In Sumatra, environmental covariates explained the greatest proportion of relative variance (23% individually and 70% jointly). Variance explained by environmental covariates was shared between landscape covariates (10% individually and 15% jointly) and anthropogenic covariates (8% individually and 13% jointly). Comparatively, topographic and spatial covariates explained less variance (8% individually and 55%

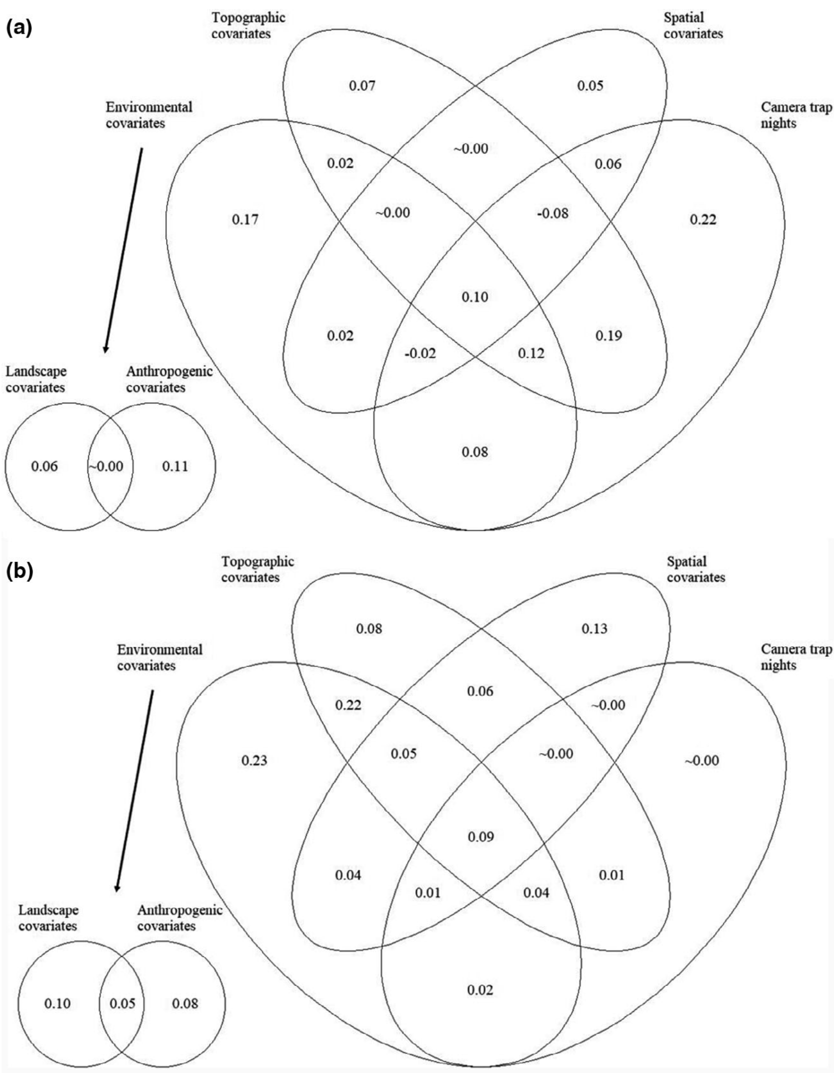

Figure 3 Variance partitioning analyses for Borneo (a) and for Sumatra (b), showing the proportion of variance in the detections of the sampled species explained in the Canonical Correspondence Analyses by different subsets of covariates. Shown are the relative values. Absolute variance explained for Borneo = 0.239. Absolute residual variance for Borneo = 0.761. Absolute variance explained for Sumatra = 0.104. Absolute residual variance for Sumatra = 0.896.

jointly, and 13% individually and 38% jointly, respectively; Fig. 3b).

Species richness and model performance

CCA performed between the number of detections of the sampled species and the final set of scale-optimized covariates confirmed that species-habitat relationships were primarily affected by environmental and topographic covariates. In Borneo, elevation, patch density of tree cover, slope position and lowland open areas, were the main factors positively affecting species detections. However, anthropogenic factors, like protected areas and forest loss, were also influential in affecting species detections (Fig. 4a; Supporting Information Table S4). Similarly, in Sumatra species detection patterns were most strongly associated with environmental factors like upper-montane forests, plantation and regrowth and lowland open areas, but also with topographic (elevation), anthropogenic (forest loss) and spatial (longitude) covariates (Fig. 4b; Supporting Information Table S5).

Analysis of the CCA biplots confirmed the positive association of many species with forests and elevation gradients. In Borneo, the majority of species appeared to be relative habitat generalists, as revealed by clustering at the centre of the biplot (Fig. 4a). Three carnivore species (Hose's civet *Diplogale hosei*, Bornean ferret badger *Melogale everetti* and masked palm civet *Paguma larvata*) showed strong affinity for high elevations and mountainsides (as indicated by slope position). Endangered flagship species were positively associated with anthropogenic factors: Asian elephant *Elephas*

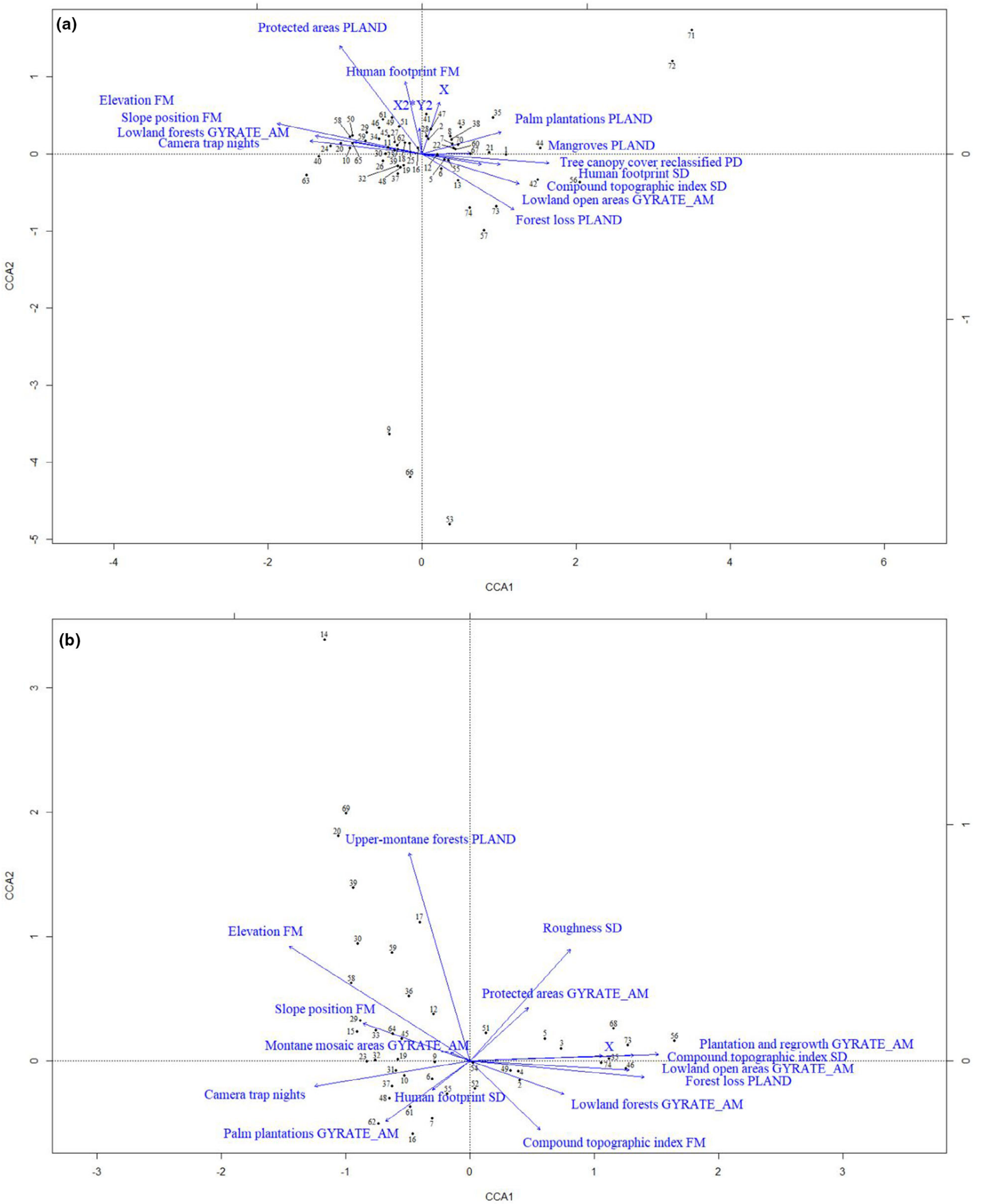

Figure 4 Biplots resulting from the Canonical Correspondence Analyses between the number of detections of the sampled species and the final sets of scale-optimized covariates, illustrating the relationships between species sampled in Borneo (a) and in Sumatra (b), and habitat covariates. See Table 2 for covariate description and Supporting Information Table S3 for species codes. FM, focal mean; PD, patch density; SD, standard deviation.

maximus was associated with palm plantations, while banteng *Bos javanicus* and Bornean orangutan *Pongo pygmaeus* occurred in strong association with forest loss.

In Sumatra, elevation was an important driver of many species including, from the most associated with high areas to the least, mainland serow *Capricornis sumatraensis*, Salvadori's pheasant *Lophura inornata*, masked palm civet *Paguma larvata*, yellow-throated marten *Martes flavigula*, civet species, Sunda clouded leopard *Neofelis diardi*, partridge and pheasant species, dhole *Cuon alpinus* and sambar *Rusa unicolor* Fig. 4b). Generalist species, including Sunda leopard cat *Prionailurus javanensis*, Malay crested fireback *Lophura rufa*, long-tailed porcupine *Trichys fasciculata* and common long-tailed macaque *Macaca fascicularis*, were positively associated with lowland open areas and forest loss. Additionally, the critically endangered Sunda pangolin *Manis javanica* was positively associated with palm plantations and human footprint. Overall, fewer species were clustered in the centre of the biplot, revealing a greater degree of specialization compared to Borneo.

Combining by addition the predicted occurrence surfaces produced for each sampled species based on the results of the single-species GLMs (Supporting Information Tables S1 and S2), we predicted species richness across Borneo and Sumatra (Fig. 5), compared it to the IUCN surfaces (Fig. 6) and evaluated the explanatory and predictive power of the models. For Borneo, analysis of the residual deviances of the single-species models revealed that they had strong explanatory power, even though the models for a few species revealed high residual deviances (mean residual deviance = 1365.0, median residual deviance = 767.8; Supporting Information Figure S3). Additionally, our species richness surface showed strong positive relationships with empirical species richness based on the independent holdout validation data (Nagelkerke-pseudo-$R^2 = 0.72$, $P < 0.0001$), while the IUCN surface performed poorly (Nagelkerke-pseudo-$R^2 = 0.06$, $P = 0.003$), and showed a contrasting geographic distribution of biodiversity (Fig. 6).

For Sumatra, the residual deviances revealed strong explanatory power of the single-species models (mean residual

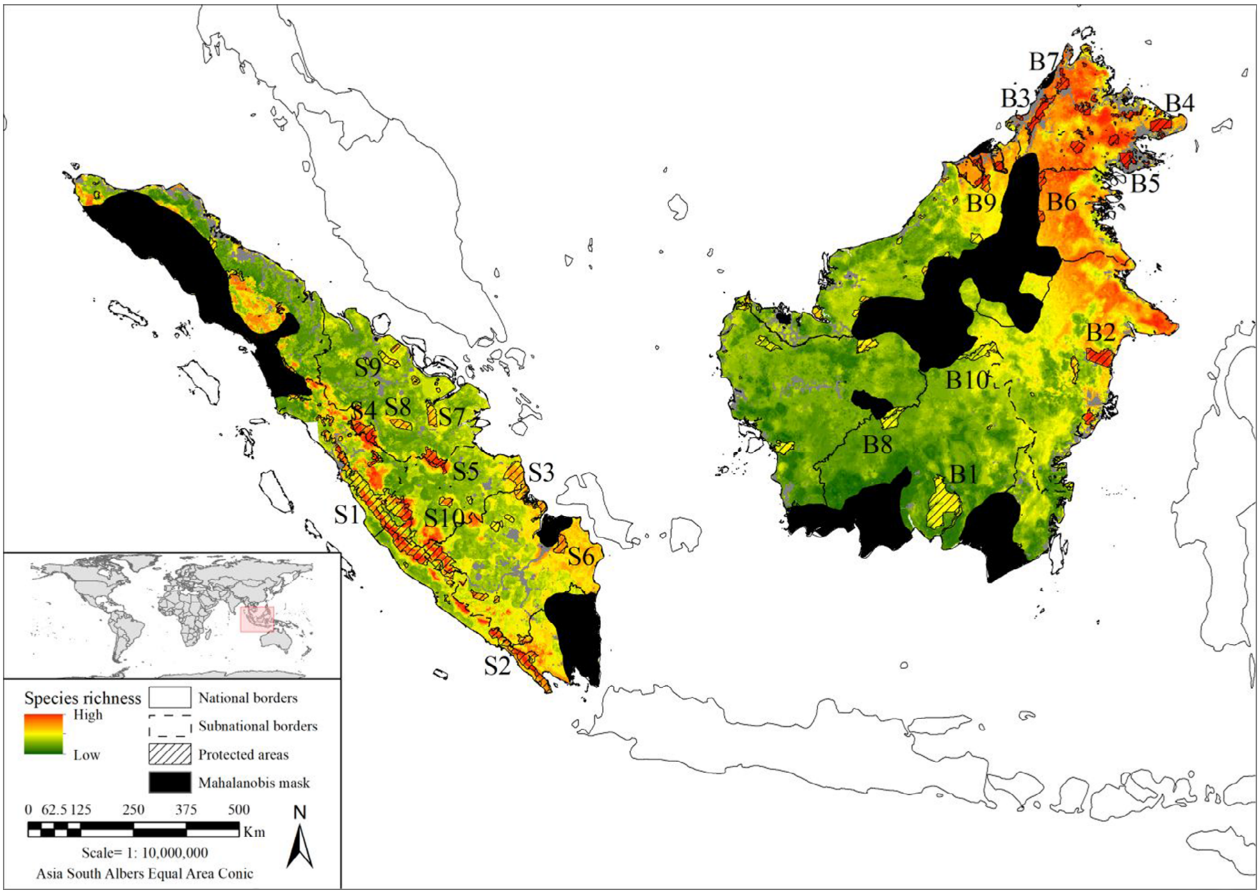

Figure 5 Predicted species richness in Borneo (right) and Sumatra (left), showing protected areas overlaid in black hash marks. Species richness predictions were based on single-species generalized linear models between the sampled species' number of detections and the scale-optimized covariates found important to describe overall sampled biodiversity geographic patterns in a Canonical Correspondence Analysis. Grey areas represent high human footprint. Black areas represent the most distinct regions from the sampled locations, as highlighted by a Mahalanobis distance surface, which were excluded from predictions. Numbers represent codes of the main systems of protected areas, as reported in Supporting Information Tables S6 and S7.

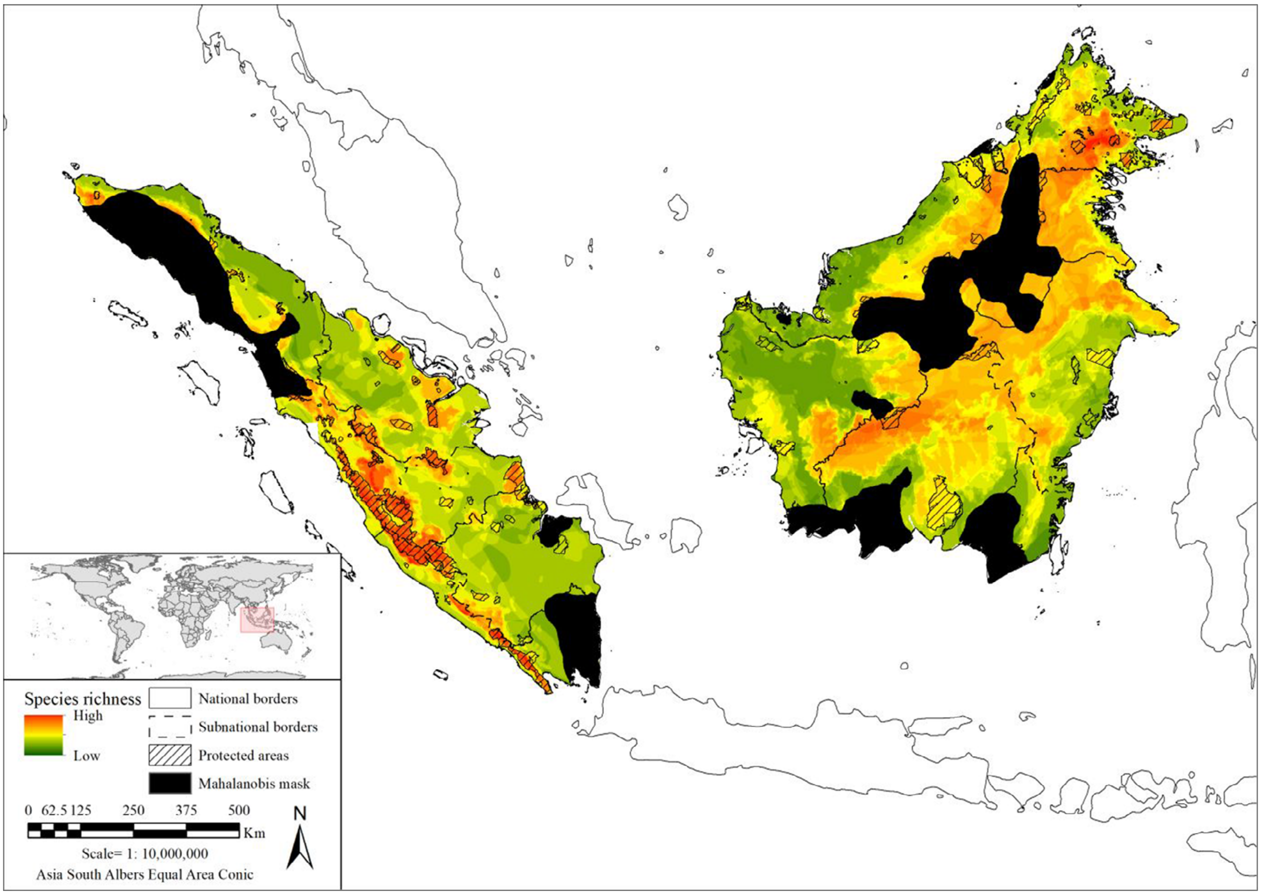

Figure 6 Predicted species richness in Borneo (right) and Sumatra (left) as obtained by summing the IUCN geographic range layers of the sampled species. Protected areas are shown in black hash marks. Black areas represent the most distinct regions from the sampled locations, as highlighted by a Mahalanobis distance surface, which were excluded from the predictions.

deviance = 564.6, median residual deviance = 251.1; Supporting Information Figure S4). However, our species richness surface and the IUCN prediction were much more similar with respect to the geographic distribution of biodiversity hotspots than in Borneo (Fig. 6). Importantly, in Sumatra neither our species richness surface nor the IUCN prediction were highly predictive of observed species richness based on the holdout validation data (Nagelkerke-pseudo-R^2 = 0.007, P = 0.34 for our species richness surface and Nagelkerke-pseudo-R^2 = 0.002, P = 0.59 for the IUCN prediction).

Protected areas effectiveness

The evaluation of the effectiveness of protected areas revealed that they preserved a relatively small proportion of biodiversity; only 9.2 and 18.2% of the cumulative modelled species richness occurred within protected areas in Borneo and Sumatra, respectively. In both islands, a small number of protected areas provided for most of the protected biodiversity. In Borneo, Sebangau, Kutai and Crocker Range National Parks represented the most important protected areas, harbouring 1.20, 0.82 and 0.63% of the overall biodiversity, and 12.94, 8.91 and 6.84% of the protected biodiversity, respectively (Supporting Information Table S6). The aforementioned most important protected areas for Bornean biodiversity occurred in Kalimantan, the Indonesian portion of Borneo and in Sabah, in the Malaysian side of the island. In Sumatra, Kerinci Seblat, Bukit Barisan Selatan and Berbak Sembilang National Parks protected 8.17, 1.99 and 1.44% of the overall biodiversity, and 44.98, 10.95 and 7.91% of the protected biodiversity, respectively (Supporting Information Table S7).

Gap analysis

The reclassification of the predictive surfaces into low, medium and high biodiverse areas allowed us to highlight the most critical areas for conservation (Fig. 7). In Borneo, the Malaysian state of Sabah and the Indonesian provinces of North and East Kalimantan were the regions where most of the biodiversity hotspots occurred. However, most of them fell outside the current system of protected areas (Fig. 7; Supporting Information Table S8). In Sabah, a critically important biodiverse area of more than 5500 km^2 fell

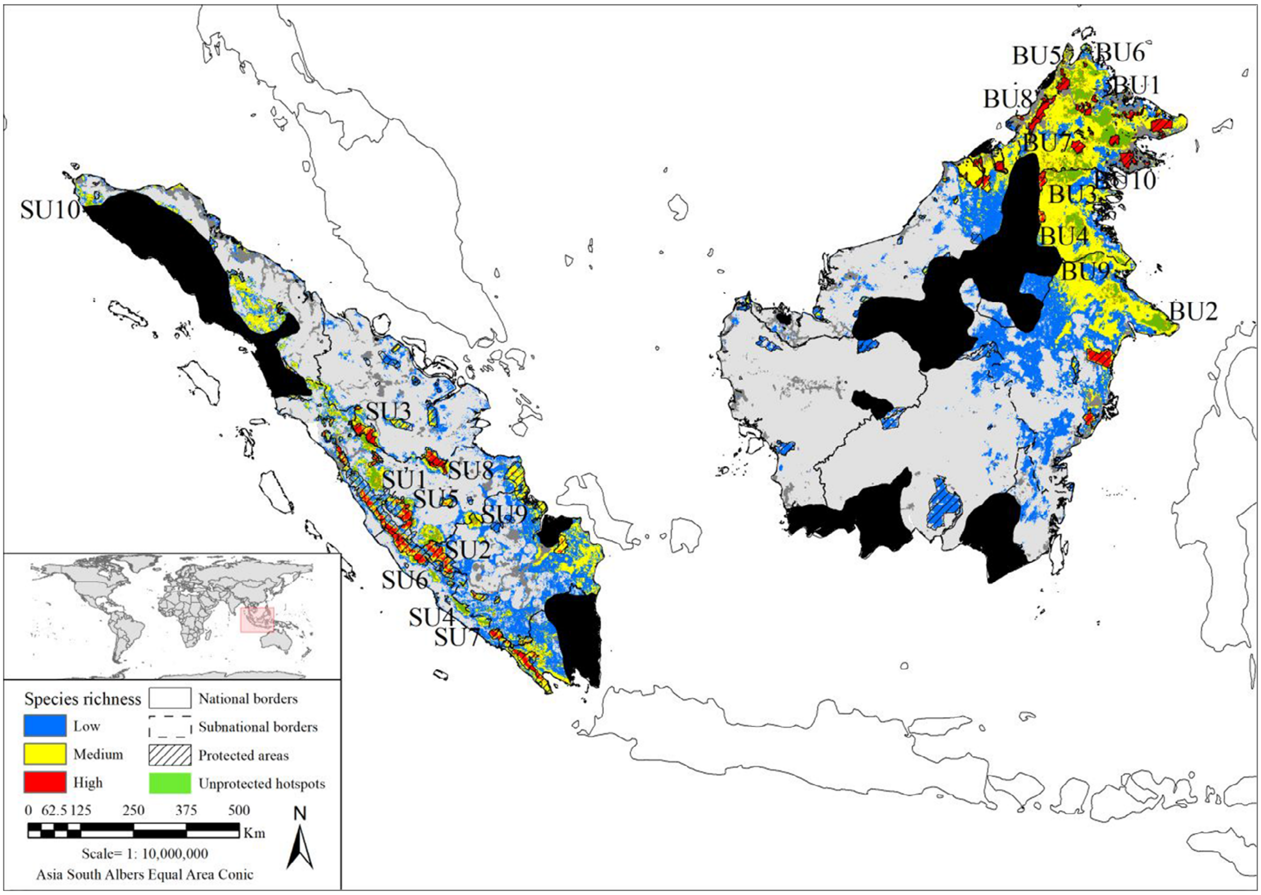

Figure 7 Predicted species richness in Borneo (right) and Sumatra (left), derived from the continuous species richness surface shown and described in Fig. 5, and reclassified into low, medium and high priority areas based on the percentiles distribution of the species richness values (>50th, >70th and >90th, respectively). Highlighted are the main biodiversity hotspots falling outside the current system of protected areas. Protected areas are overlaid in black hash marks. Black areas represent the most distinct regions from the sampled locations, as highlighted by a Mahalanobis distance surface, which were excluded from the predictions. Numbers represent codes of the main unprotected hotspots, as reported in Supporting Information Table S8.

entirely outside protected areas (Supporting Information Table S8). In North and East Kalimantan, at least five distinguished biodiversity hotspots have been highlighted, and none of them was protected (Fig. 7; Supporting Information Table S8).

In Sumatra, the most important predicted biodiversity hotspots coincided with protected areas in the western mountainous regions. Our results showed that in Sumatra there seems to have been a more severe pattern of habitat loss which resulted in biodiversity being more associated with, and constrained by, protected areas than it is in Borneo. However, we also highlighted important hotspots outside protected areas (Fig. 7; Supporting Information Table S8). Of these, the most important ones surrounded and connected existing protected areas, like the satellite hotspots connecting the Kerinci Seblat National Park and the Bukit Barisan Selatan National Park. Additionally, extensive areas of medium and low biodiversity occurred in the southern parts of the island, and these currently almost completely lack protection.

Indicator species effectiveness

Indicator species effectiveness analysis revealed that lesser oriental chevrotain *Tragulus kanchil* and Sunda leopard cat *Prionailurus javanensis* were the best indicators of biodiversity in Borneo and Sumatra, respectively. Surprisingly, rat species were the second best indicator in both islands. In Borneo, two civet species (banded civet *Hemigalus derbyanus* and Malay civet *Viverra tangalunga*) were among the top indicator species. Moonrat *Echinosorex gymnura*, Sunda stink-badger *Mydaus javanensis*, thick-spined porcupine *Hystrix crassispinis* and short-tailed mongoose *Herpestes brachyurus* were also strong indicators of biodiversity, along with Sabah partridge *Arborophila graydoni* and other partridge species (Supporting Information Table S9). In Sumatra, similar to mainland South-east Asia (Macdonald *et al*., 2020), various carnivores were among the top indicators, including members of the Felidae (Sunda leopard cat *Prionailurus javanensis* and Sunda clouded leopard *Neofelis diardi*), Viverridae (masked palm civet *Paguma larvata*) and

Canidae (dhole *Cuon alpinus*) families. Other important indicators of Sumatran biodiversity included bronze-tailed peacock-pheasant *Polyplectron chalcurum*, moonrat *Echinosorex gymnura*, bearded pig *Sus barbatus* and mainland serow *Capricornis sumatraensis* (Supporting Information Table S9).

Discussion

This study utilized the largest camera trap survey ever carried out in Borneo and Sumatra to (1) assess multivariate habitat relationships of terrestrial meso and mega vertebrate communities, (2) quantify the relative contributions of different factors driving community assemblages, (3) evaluate the level of biodiversity protection, (4) highlight areas where conservation actions should be prioritized, (5) compare rigorously modelled species richness patterns with more anecdotal knowledge about species distribution, and (6) investigate efficacy of different species as biodiversity indicators. The study has been conducted within a multivariate context, meaning that we accounted for the community-wide responses to habitat factors, while still considering the individual preferences of sampled species. Our results provide the most comprehensive assessment to date of the factors affecting terrestrial vertebrate biodiversity and its spatial patterns across Borneo and Sumatra.

Multi-scale selection of habitat factors

As hypothesized, we found that communities in Borneo and Sumatra respond to habitat factors at different spatial scales. Specifically, sampled species responded to habitat features primarily at medium (2–6 km) to broad (8–10 km) scales, confirming previous findings on mainland South-east Asian biodiversity (Macdonald *et al*., 2020). Our findings corroborate general patterns seen in mainland South-east Asia (Macdonald *et al*., 2020), demonstrating, in particular, the importance of extensive intact forest for maintaining high species richness. We also found that anthropogenic factors affect species at broad scales, as found by Hearn *et al*. (2018*b*) and Macdonald *et al*. (2020), confirming the broad negative effects that anthropogenic disturbance, particularly forest loss, has on biodiversity in South-east Asia and in the Sunda Islands. Elevation was the only covariate selected at the finest scale in both islands, consistent with the findings reported by Hearn *et al*. (2018*b*), likely reflecting the islands' highly heterogeneous topography. Interestingly, Macdonald *et al*. (2020) found elevation to be most related to biodiversity in mainland South-east Asian at very broad scales, likely reflecting the more extensive impact that altitudinal factors have on biodiversity in wider landscapes.

Environmental predictors of community assemblages

Variance in the multivariate detection counts of sampled species explained for Borneo was comparable with other studies using CCA with high species diversity (Palmer, 1993; ter Braak & Verdonschot, 1995; Ohmann & Spies, 1998; Macdonald *et al*., 2020), while the same variance explained for Sumatra, although lower, was in the 10–50% range of variance explained typically reported for CCA (Palmer, 1993).

As shown by the variance partitioning analysis, environmental factors were major drivers of the Sunda Islands' multivariate species assemblages, explaining the greatest proportion of total variance for Sumatra and the second greatest proportion of total variance for Borneo, reflecting that landscape composition and configuration, along with anthropogenic impacts on the landscape, strongly drive species assemblage patterns. Topographic and spatial covariates were overall less influential. However, CCA and relative biplots highlighted that topographic factors, and especially elevation, were nonetheless fundamental drivers of habitat selection for several species, both in Borneo and Sumatra.

In Borneo, carnivores were associated with the elevation gradient, supporting previous findings by Mathai *et al*. (2016) and Mathai *et al*. (2019), predicting that three endemic Bornean carnivores (Bornean ferret badger *Melogale everetti*, Hose's civet *Diplogale hosei* and Borneo bay cat *Catopuma badia*), along with masked palm civet *Paguma larvata*, tend to be associated with higher elevations. Similarly, Hearn *et al*. (2018*b*) found strong relationships with higher elevations and Sunda clouded leopard *Neofelis diardi* occurrence in Sabah. The negative impact of human footprint and forest loss was particularly evident for the critically endangered Bornean orangutan *Pongo pygmaeus* and the endangered banteng *Bos javanicus*. Bornean orangutan *Pongo pygmaeus* is known to mainly occur in primary forests and pristine forest habitats (Felton *et al*., 2003, Knop, Ward & Wich, 2004), while banteng *Bos javanicus* in Borneo prefers forest habitats as well as open grassland, mature timber plantations and abandoned cultivated lands (Melletti & Burton, 2014). However, we highlighted that in Borneo these species were associated with lowland habitats that have experienced extensive forest loss in recent decades, and their association with deforested areas is indicative of the fact that they persist only in areas with high degree of anthropogenic disturbance. This likely reflects their obligate association with lowland rainforest ecosystems which have had rapid and extensive impacts from deforestation and forest degradation. Recent studies have suggested that both these species might be more tolerant of deforestation than was previously thought. For example, Ancrenaz *et al*. (2010), Meijaard *et al*. (2010) and Davies *et al*. (2017), found that Bornean orangutan *Pongo pygmaeus* has a degree of tolerance towards disturbed forests and, as discussed above, it is known that banteng *Bos javanicus* can be found in disturbed habitats. However, the persistence of these species in areas of heavy human footprint might reflect an extinction debt of large recent changes which has yet to result in equilibrium levels of population decline and local extinction. Therefore, our results indicate that the ecological niches of the most endangered species are currently strongly limited to areas of their original ranges that have experienced the largest impacts of forest loss, and highlight their vulnerability to habitat loss in regions most vulnerable to deforestation. We highlighted also the striking association between Asian elephant *Elephas maximus* and palm plantations. Even though the species is commonly found in diverse habitats ranging from tropical evergreen and semi-evergreen forests to cultivated and secondary forests (Choudhury, 1999), this association also reflects that deforestation and land use changes are concentrated at lower elevations and flatter terrain, which are both critical habitat elements for the species (Cushman *et al.*, 2017). Therefore, current Asian elephant *Elephas maximus* occurrence patterns likely indicate niche displacement from optimal lowland forests, which are mostly converted to oil palm and industrial timber plantations.

In Sumatra, we observed a greater range of ecological specialization among sampled species, with higher elevation forests emerging as particularly important for biodiversity. A group of species including Sunda clouded leopard *Neofelis diardi* showed clear selection for high elevations and upper-montane forests, with predicted occurrence concentrated in the western mountainous region of Sumatra, in agreement with habitat suitability models for Sunda clouded leopard *Neofelis diardi* (Macdonald *et al.*, 2018*a*) and for overall biodiversity (Catullo *et al.*, 2008). More generalist species were associated with deforested lowland areas and this, coupled with the absence of specialists in these habitats, demonstrated the low suitability of disturbed areas for specialized species. Additionally, as highlighted in Borneo, in Sumatra we also found endangered flagship species to be associated with disturbed habitats. Specifically, we found Sunda pangolin *Manis javanica* to be associated with oil palm plantations and areas of higher human footprint, raising serious concerns about the conservation status of this critically endangered species, as this likely reflects the increased human impacts in the highly productive and most accessible lowland areas that provide critical habitat for this and other endangered species.

Our results suggest that anthropogenic impacts in Borneo and Sumatra have been so rapid and extensive that there is a disequilibrium between the distribution of many species and their habitat needs (e.g. a *lag effect*, Kaszta *et al.*, 2020). Hence, the association of several endangered flagship species, such as Bornean orangutan *Pongo pygmaeus* and Asian elephant *Elephas maximus*, with anthropogenic factors and disturbed habitat, reflects the degree of perturbation of the Sunda Islands, and must be a red flag about the conservation of their delicate communities.

Indicator species

In both islands, rat species were the second best indicator of biodiversity. However, the challenge of identifying individual rat species from camera trap images limited our analysis to the order level. We acknowledge this limitation and note that it is very unlikely that all the rat species have the same importance as indicators. In addition, we believe that this result is influenced by the large number of generalist species represented in this group. Future research is needed to clarify the best indicators among rodents.

In Borneo, Viverridae were crucial indicators, with two species (banded civet *Hemigalus derbyanus* and Malay civet *Viverra tangalunga*) occurring among the best indicator species, corroborating previous findings on the pivotal role of this family for Bornean (Ross, Hearn & Macdonald, 2017) and South-east Asian (Macdonald *et al.*, 2020) biodiversity.

In Sumatra, we found carnivores to be the strongest overall indicators of species richness. Members from the main carnivore families were among the best indicators, including Felidae (Sunda leopard cat *Prionailurus javanensis* and Sunda clouded leopard *Neofelis diardi*), Viverridae (masked palm civet *Paguma larvata*) and Canidae (dhole *Cuon alpinus*). Niche differentiation theory (Hardin, 1960) states that if two or more species compete, niche differentiation is a likely outcome to avoid extinction. Accordingly, and similarly to what found in mainland South-east Asia (Macdonald *et al.*, 2020), we found a single member from each family playing an important role as indicator; only the most common members appeared to be good indicators, while more specialized and rarer species emerged as weaker candidates.

Mammalian carnivores have often been proposed as umbrella species given their large habitat requirements and high mobility (Noss *et al.*, 1996; Carroll *et al.*, 2001; Dalerum *et al.*, 2008). Despite the appeal of this proposition, the assumption that ensuring habitat requirements for one species would automatically preserve sympatric species has been contested (e.g. Cushman *et al.*, 2010). However, Carroll *et al.* (2001), Roberge & Angelstam (2004), Cushman & Landguth (2012) and Penjor *et al.* (2021) demonstrated that conservation strategies based on combinations of multiple umbrella species cover broader ranges of habitat gradients, and can effectively ensure the protection of a higher fraction of sympatric species. Our analysis confirms the importance of carnivores as proxies of biodiversity and, in particular, of the more common and less specialized ones, such as Sunda stink-badger *Mydaus javanensis* in Borneo and Sunda leopard cat *Prionailurus javanensis* in Sumatra. Additionally, in a region as rich in endemism as the Sunda Islands, considering more than one carnivore within an umbrella grouping is critical to maximize protection of the regional biodiversity.

Importance of empirical predictions

Biodiversity richness patterns, protected area effectiveness and gap analysis are often assessed using species' inferred distribution layers or other proxies based on coarse-scale, non-empirical data, such as IUCN range maps. It is critical, therefore, to assess how they perform relative to optimized empirically based models. For Borneo, our species richness surface provided a much more reliable prediction of biodiversity hotspots than did the IUCN surface. The IUCN surface, in addition to having much lower Nagelkerke-pseudo-R^2, showed very different patterns of predicted species richness, with biodiversity spread more evenly over the central regions of Borneo. Worryingly, Hughes (2017) and Macdonald *et al.* (2020) also found the IUCN surface of the mainland South-east Asian biodiversity to drastically differ from the prediction produced with empirical data, with the latter also performing much better than the former. Since several studies use IUCN range layers to provide conservation recommendations, we urge a thorough investigation of factors leading to discrepancies between the expert-based maps and empirical data-based assessments of the sort we have explored.

For Sumatra, our predictive surface and the IUCN map showed similar distribution of biodiversity hotspots, and both had low Nagelkerke-pseudo-R^2, revealing weak predictive power, which was surprising given the strong explanatory power of the single-species models. We believe that the poor predictive power is due to the disequilibrium between species distributions and their habitat needs. In a perturbed ecological system where species are associated with habitat conditions because of anthropogenic pressure and not because of ecological associations, species-habitat relationships are difficult to define with precision, leading to weaker predictions (Guisan & Thuiller, 2005).

Additionally, a potential alternative, but not mutually exclusive, reason for the weak predictive power in Sumatra, is the spatial clustering of our camera trap surveys. In contrast to Borneo, where sampling locations occurred across a wider set of environmental gradients, in Sumatra camera traps have been deployed only in portions of the western mountainous region. In addition, both in Borneo and Sumatra, camera trap locations were mainly clustered within protected areas, providing a further potential source of spatial bias. However, by sampling within protected areas, we likely provided a more realistic view on species richness patters, reflecting the expected species-habitat relationships that would occur in undisturbed contexts, without accounting for adverse effects of human disturbance, poaching and other factors not associated with measured habitat factors. However, the camera trap surveys could have still been inadequate to represent the range of environmental conditions of the islands, which is known to affect the ability to correctly identify species-habitat relationships (Chiaverini *et al*., 2021). Therefore, we have minimized the potential errors arising from the spatially biased survey by masking the areas showing the strongest ecological differences from the sampling locations in the multivariate environment of the Mahalanobis distance surface. Future surveys focused on unsampled areas across the full range of ecological gradients are of vital importance to better assess the distribution of biodiversity in Borneo and Sumatra.

Conservation implications

Northern Borneo, specifically the Malaysian state of Sabah, the Indonesian provinces of North and East Kalimantan, and Brunei, are predicted to harbour the highest species richness in Borneo. Interestingly, Raes *et al*. (2009) found the same regions to be critical hotspots also for botanical species richness. These regions' high species richness is associated with moderate to higher elevations, rough topography and numerous remaining patches of primary forest. While only a few protected areas occur in this region, their critical importance is clear. In Sabah, where heavy logging followed by land conversion to oil palm plantations has altered the landscape, the few protected areas are among those with the highest values of protected biodiversity, such as Crocker Range National Park, Tabin Wildlife Reserve – Mount Hatton Protection Forest Reserve and Ulu Kalumpang Protection Forest Reserve – Tawau Hills Park. Also, Sabah harbours the highest levels of species richness outside of protected areas, with vast unprotected lands that we identified as a biodiversity hotspot. To assess the effectiveness of protected areas and perform gap analysis, however, we evaluated protected areas as those defined by the WDPA (IUCN, UNEP-WCMC, 2017) that, in Sabah, includes only areas assigned to Class 1 protection (Dudley *et al*., 2010). However, several other areas assigned from Class 2 to 5 occur in Sabah and, even though they are excluded from the WDPA, they cannot be deforested. In addition, even though we only evaluated protected areas, it has to be highlighted the important role that alternative conservation strategies such as sustainable forestry practices have on preserving regional biodiversity (Fuller, Jessup & Salim, 2004). Additionally, northern and eastern Kalimantan, regions almost completely lacking protection, harbour other critical biodiversity hotspots. Worryingly, this is the area where Indonesia's new capital will be developed (Teo *et al*., 2020). Hence, extending the current system of protected areas and the network of connectivity corridors in northern Borneo is critical for preserving the biodiversity of the island (Hearn *et al*., 2019; Kaszta *et al*., 2019). It is fundamental to highlight that, even though we did not assess the central highlands of Borneo due to their ecological differences from the areas that we sampled, which might have resulted in producing unreliable prediction, those areas have previously been found fundamental for Bornean biodiversity, particularly for carnivores (Mathai *et al*., 2016; Macdonald *et al*., 2018*b*), and future surveys at higher elevations are important for covering a broader altitudinal range.

In Sumatra, the western mountainous regions harbour the highest levels of species richness. However, we have to be cautious in asserting strong conclusions based on our results given the low predictive power of our species richness surface. A similar pattern characterized by large patches of suitable habitat along the western mountains, and smaller, fragmented patches occurring throughout the central and eastern Sumatran lowlands, was highlighted by Macdonald *et al*. (2018*a*) for Sunda clouded leopard *Neofelis diardi*. This pattern appears to be driven by anthropogenic factors, with habitat loss and deforestation predominating in eastern portions of the island and at lower elevations, and protected areas concentrated in the western highlands. Forest loss and subsequent conversion to oil palm plantations are the main factors negatively affecting species occurrence by driving habitat loss and fragmentation. Conversely, protected areas are a vital factor contributing positively to species richness; the main biodiversity hotspots in Sumatra occur within protected areas, such as Kerinci Seblat and Bukit Barisan Selatan National Parks. Together, these protected areas harbour more than 50% of predicted Sumatran biodiversity. However, in our analysis we excluded the northern portion of the island due to its ecological differences from the locations that we sampled. In northern Sumatra, Gunung Leuser National Park occurs, which is a critically important protected area and, even though in our work we did not assess its contribution in harbouring regional biodiversity, future studies at higher latitudes are strongly needed. Extending the current network of protected areas in Sumatra is critical, especially along the western mountains where we highlighted the most important unprotected hotspots. Establishing additional protected areas in the southern lowlands is of the essence where forest loss is a clear and imminent threat to biodiversity.

Scope and limitations

Even though the camera trap survey we used in the current work likely represents the most intense effort ever carried out in Borneo and Sumatra to collect camera trap data for meso and macro terrestrial fauna, a number of limitations to our study have to be acknowledged.

First, the camera trap survey did not sample the full gradients of all the environmental factors chosen to model species richness. Ideally, sampling strategies should be designed to identify the full range of gradients that are most influential to species' habitat selection processes, rather than being biased towards a fraction of the full spectrum of environmental conditions (Wessels *et al*., 1998). Failing to do so might result in models with erroneous parameter estimates and lower predictive performances (Chiaverini *et al*., 2021). However, in our case, it was impossible to implement a randomly stratified survey given the extent of the study area and the inaccessibility of some regions. Therefore, we produced a Mahalanobis distance surface in order to mask the regions of the study area that were more different from the sampled ones, reducing in this way the risk of producing unreliable predictions of species richness.

Additionally, another potential source of bias was represented by the imperfect and varying detection of sampled species (Sollmann *et al*., 2013). Most of the camera trap locations occurred within protected areas, where environmental conditions drastically differ from unprotected areas, where anthropic pressures might significantly reduce the detectability of several species (Lessa *et al*., 2017). In addition, deploying camera traps mainly along natural ridgelines and disused logging roads, might have determined an overestimation bias of species that are more likely to use these areas (e.g. felids), leading to underestimating smaller, rarer and more cryptic species (Kolowski & Forrester, 2017). However, having modelled the Mahalanobis distance surface accounting also for the influence of protected areas and human footprint, we minimized this geographic bias. Additionally, the same survey protocol was adopted for all the camera trap grids across Borneo and Sumatra, and this consistency minimized the variations in detection probabilities across sampling locations, reducing the potential bias in the species-habitat relationships' estimates.

Acknowledgments

The authors gratefully acknowledge the hard work of the field assistants who helped us in each study area, and the various authorities that approved our work. This work was largely funded by a programmatic grant to DWM from the Robertson Foundation. LC undertook this work while a doctoral student at WildCRU (Department of Zoology) Oxford, funded by the Horne Family Trust and by DWM's Robertson Foundation programme.

References

Ancrenaz, M., Ambu, L., Sunjoto, I., Ahmad, E., Manokaran, K., Meijaard, E. & Lackman, I. (2010). Recent surveys in the forests of Ulu Segama Malua, Sabah, Malaysia, show that orang-utans (*P. p. morio*) can be maintained in slightly logged forests. *PLoS One* **5**, e11510.

Austin, M.P., Nicholls, A.O., Doherty, M.D. & Meyers, J.A. (1994). Determining species response functions to an environmental gradient by means of a beta-function. *J. Veg. Sci.* **5**, 215–228.

Borcard, D., Legendre, P. & Drapeau, P. (1992). Partialling out the spatial component of ecological variation. *Ecology* **73**, 1045–1055.

Carroll, C., Noss, R.F. & Paquet, P.C. (2001). Carnivores as focal species for conservation planning in the rocky mountain region. *Ecol. Appl.* **11**, 961–980.

Catullo, G., Masi, M., Falcucci, A., Maiorano, L., Rondinini, C. & Boitani, L. (2008). A gap analysis of southeast Asian mammals based on habitat suitability models. *Biol. Conserv.* **141**, 2730–2744.

Chambers, C.L., Cushman, S.A., Medina-Fitoria, A., Martinez-Fonseca, J. & Chavez-Velasquez, M. (2016). Influences of scale on bat habitat relationships in a forested landscape in Nicaragua. *Landscape Ecol.* **31**, 1299–1318.

Chiaverini, L., Wan, H.Y., Hahn, B., Cilimburg, A., Wasserman, T.N. & Cushman, S.A. (2021). Effects of non-representative sampling design on multi-scale habitat models: flammulated owls in the Rocky Mountains. *Ecol. Model.* **450**, 109566.

Choudhury, A. (1999). Status and conservation of the Asian Elephant *Elephas* maximus in north-eastern India. *Mammal Rev.* **29**, 141–174.

CIESIN, CIAT. (2016) *Gridded Population of the World Version 4 (GPWV4): population density grids*. Palisades, NY: Socioeconomic Data and Applications Center (SEDAC), Columbia University. http://sedac.ciesin.columbia.edu/gpw

Coleman, B.T. & Hill, R.A. (2014). Living in a landscape of fear: the impact of predation, resource availability and habitat structure on primate range use. *Anim. Behav.* **88**, 165–173.

Cushman, S.A. & Landguth, E.L. (2012). Multi-taxa population connectivity in the northern Rocky Mountains. *Ecol. Model.* **231**, 101–112.

Cushman, S.A., Macdonald, E., Landguth, E., Malhi, Y. & Macdonald, D. (2017). Multiple-scale prediction of forest loss risk across Borneo. *Landscape Ecol.* **32**, 1581–1598.

Cushman, S.A. & McGarigal, K. (2002). Hierarchical, multi-scale decomposition of species-environment relationships. *Landscape Ecol.* **17**, 637–646.

Cushman, S.A. & McGarigal, K. (2004*a*). Hierarchical analysis of forest bird species-environment relationships in the Oregon coast range. *Ecol. Appl.* **14**, 1090–1105.

Cushman, S.A. & McGarigal, K. (2004*b*). Patterns in the species-environment relationship depend on both scale and choice of response variables. *Oikos* **105**, 117–124.

Cushman, S.A., McKelvey, K.S., Noon, B.R. & McGarigal, K. (2010). Use of abundance of one species as a surrogate for abundance of others. *Conserv. Biol.* **24**, 830–840.

Dalerum, F., Somers, M.J., Kunkel, K.E. & Cameron, E.Z. (2008). The potential for large carnivores to act as biodiversity surrogates in southern Africa. *Biodivers. Conserv.* **17**, 2939–2949.

Davies, A.B., Ancrenaz, M., Oram, F. & Asner, G.P. (2017). Canopy structure drives orangutan habitat selection in disturbed Bornean forests. *Proc. Natl. Acad. Sci. USA* **114**, 8307–8312.

Davis, C.L., Rich, L.N., Farris, Z.J., Kelly, M.J., Di Bitetti, M.S., Di Blanco, Y., Albanesi, S., Farhadinia, M.S., Gholikhani, N., Hamel, S., Harmsen, B.J., Wultsch, C., Kane, M.D., Martins, Q., Murphy, A.J., Steenweg, R., Sunarto, S., Taktehrani, A., Thapa, K., Tucker, J.M., Whittington, J., Widodo, F.A., Yoccoz, N.G. & Miller, D.A.W. (2018). Ecological correlates of the spatial co-occurrence of sympatric mammalian carnivores worldwide. *Ecol. Lett.* **21**, 1401–1412.

Dudley, N., Parrish, J.D., Redford, K.H. & Stolton, S. (2010). The revised IUCN protected area management categories: the debate and ways forward. *Oryx* **44**, 485–490.

ESRI. (2018) *Arcgis desktop: release 10.6.1*. Redlands, CA: Environmental system research Institute.

Evans, J. S., Oakleaf, J., Cushman, S. A. & Theobald, D. (2014). *An ArcGIS toolbox for surface gradient and geomorphometric modeling, version 2.0-0*. Available at http://evansmurphy.wix.com/evansspatial

Felton, A.M., Engstrom, L.M., Felton, A. & Knott, C.D. (2003). Orangutan population density, forest structure and fruit availability in hand-logged and unlogged peat swamp forests in west Kalimantan, Indonesia. *Biol. Conserv.* **114**, 91–101.

Fuller, D.O., Jessup, T.C. & Salim, A. (2004). Loss of forest cover in Kalimantan, Indonesia, since the 1997–1998 El Nino. *Conserv. Biol.* **18**, 249–254.

Grand, J., Buonaccorsi, J., Cushman, S.A., Griffin, C.R. & Neel, M.C. (2004). A multiscale landscape approach to predicting bird and moth rarity hotspots, in a threatened pitch pine-scrub oak community. *Conserv. Biol.* **18**, 1063.

Guisan, A. & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.* **8**, 993–1009.

Guisan, A. & Zimmermann, N.E. (2000). Predictive habitat distribution models in ecology. *Ecol. Model.* **135**, 147.

Haidir, I.A., Macdonald, D.W. & Linkie, M. (2018). Assessing the spatiotemporal interactions of mesopredators in Sumatra's tropical rainforest. *PLoS One* **13**, e0202876.

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., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O. & Townshend, J.R.G. (2013). High-resolution global maps of 21st-century forest cover change. *Science* **342**, 850–853.

Hardin, G. (1960). Competitive exclusion principle. *Science* **131**, 1292.

Hearn, A.J., Cushman, S.A., Goossens, B., Macdonald, E., Ross, J., Hunter, L.T.B., Abram, N.K. & Macdonald, D.W. (2018*a*). Evaluating scenarios of landscape change for Sunda clouded leopard connectivity in a human dominated landscape. *Biol. Conserv.* **222**, 232.

Hearn, A.J., Cushman, S.A., Goossens, B., Ross, J., Macdonald, E.A., Hunter, L.T.B. & Macdonald, D.W. (2019). Predicting connectivity, population size and genetic diversity of Sunda clouded leopards across Sabah, Borneo. *Landscape Ecol.* **34**, 275–290.

Hearn, A.J., Cushman, S.A., Ross, J., Goossens, B., Hunter, L.T.B. & Macdonald, D.W. (2018*b*). Spatio-temporal ecology of sympatric felids on Borneo. Evidence for resource partitioning? *PLoS One* **13**, e0200828.

Hearn, A.J., Ross, J., Bernard, H., Bakar, S.A., Goossens, B., Hunter, L.T.B. & Macdonald, D.W. (2017). Responses of Sunda clouded leopard Neofelis diardi population density to anthropogenic disturbance: refining estimates of its conservation status in Sabah. *Oryx* **1**, 643–653.

Hijmans, R. J., Phillips, S., Leathwick, J. & Elith, J. (2017). *Dismo: species distribution modelling. R package version 1.3-5*. Available at http://cran.R-project.org/package=dismo

Hughes, A.C. (2017). Mapping priorities for conservation in southeast Asia. *Biol. Conserv.* **209**, 395–405.

Hutchinson, G.E. (1957). Population studies – animal ecology and demography – concluding remarks. *Cold Spring Harb. Sym.* **22**, 415.

IUCN. (2019). *The IUCN red list of threatened species. Version 2019-5*. Available at https://www.iucnredlist.org

IUCN, UNEP-WCMC. (2017) *The World Database on Protected Areas (WDPA)*. Cambridge: UNEP-WCMC. www.protectedplanet.net

Jarvis, A., Reuter, H. I., Nelson, A. & Guevara, E. (2008). *Hole-filled seamless SRTM data v4*. International Centre for Tropical Agriculture (CIAT). Available at http://srtm.csi.cgiar.org

Kaszta, Z., Cushman, S.A., Hearn, A.J., Burnham, D., Macdonald, E.A., Goossens, B., Nathan, S.K.S.S. & Macdonald, D.W. (2019). Integrating sunda clouded leopard (*Neofelis diardi*) conservation into development and restoration planning in Sabah (Borneo). *Biol. Conserv.* **235**, 63–76.

Kaszta, Z., Cushman, S.A., Htun, S., Naing, H., Burnham, D. & Macdonald, D.W. (2020). Simulating the impact of Belt and Road initiative and other major developments in Myanmar on an ambassador felid, the clouded leopard, *Neofelis nebulosa*. *Landscape Ecol.* **35**, 727–746.

Knop, E., Ward, P.I. & Wich, S.A. (2004). A comparison of orang-utan density in a logged and unlogged forest on Sumatra. *Biol. Conserv.* **120**, 183–188.

Kolowski, J.M. & Forrester, T.D. (2017). Camera trap placement and the potential for bias due to trails and other features. *PLoS One* **12**, e0186679.

Lessa, I.C.M., Ferreguetti, A.C., Kajin, M., Dickman, C.R. & Bergallo, H.G. (2017). You can't run but you can hide: the negative influence of human presence on mid-sized mammals on an Atlantic Island. *J. Coast Conserv.* **21**, 829.

Macdonald, D.W., Bothwell, H.M., Hearn, A.J., Cheyne, S.M., Haidir, I., Hunter, L.T.B., Kaszta, A., Linkie, M., Macdonald, E.A., Ross, J. & Cushman, S.A. (2018*a*). Multi-scale habitat selection modeling identifies threats and conservation opportunities for the Sunda clouded leopard (*Neofelis diardi*). *Biol. Conserv.* **227**, 92–103.

Macdonald, D.W., Burnham, D., Hinks, A.E. & Wrangham, R. (2012). A problem shared is a problem reduced: seeking efficiency in the conservation of felids and primates. *Folia Primatol.* **83**, 171–215.

Macdonald, D.W., Chiaverini, L., Bothwell, H.M., Kaszta, Z., Ash, E., Bolongon, G., Can, O.E., Campos-Arceiz, A., Channa, P., Clements, G.R., Hearn, A.J., Hedges, L., Htun, S., Kamler, J.F., Macdonald, E.A., Moore, J., Naing, H., Onuma, M., Rasphone, A., Rayan, D.M., Ross, J., Singh, P., Tan, C.K.W., Wadey, J., Yadav, B.P. & Cushman, S.A. (2020). Predicting biodiversity richness in rapidly changing landscapes: climate, low human pressure or protection as salvation? *Biodivers. Conserv.* **29**, 4035–4057.

Macdonald, E.A., Cushman, S.A., Landguth, E.L., Hearn, A.J., Malhi, Y. & Macdonald, D.W. (2018*b*). Simulating impacts of rapid forest loss on population size, connectivity and genetic diversity of Sunda clouded leopards (*Neofelis diardi*) in Borneo. *PLoS One* **13**, e0196974.

Martin, T.G., Wintle, B.A., Rhodes, J.R., Kuhnert, P.M., Field, S.A., Low-Choy, S.J., Tyre, A.J. & Possingham, H.P. (2005). Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecol. Lett.* **8**, 1235–1246.

Mathai, J., Duckworth, J.W., Meijaard, E., Fredriksson, G., Hon, J., Sebastian, A., Ancrenaz, M., Hearn, A.J., Ross, J., Cheyne, S., Wilting, A. & Consortium, B.C. (2016). Carnivore conservation planning on Borneo: identifying key carnivore landscapes, research priorities and conservation interventions. *Raffles Bull. Zool.* **33**, 186–217.

Mathai, J., Niedballa, J., Radchuk, V., Sollmann, R., Heckmann, I., Brodie, J., Struebig, M., Hearn, A.J., Ross, J., Macdonald, D.W., Hon, J. & Wilting, A. (2019). Identifying refuges for Borneo's elusive hose's civet. *Glob. Ecol. Conserv.* **17**, e00531.

McGarigal, K., Cushman, S. & Stafford, S.G. (2000) *Multivariate statistics for wildlife and ecology research*. New York: Springer.

McGarigal, K., Cushman, S.A. & Ene, E. (2012) *Fragstats v4: spatial pattern analysis program for categorical and continuous maps*. Amherst, MA: University of Massachusetts. http://www.umass.edu/landeco/research/fragstats/fragstats.html

McGarigal, K., Wan, H.Y., Zeller, K.A., Timm, B.C. & Cushman, S.A. (2016). Multi-scale habitat selection modeling: a review and outlook. *Landscape Ecol.* **31**, 1161–1175.

Meijaard, E., Albar, G., Rayadin, Y., Ancrenaz, M. & Spehar, S. (2010). Unexpected ecological resilience in Bornean orangutans and implications for pulp and paper plantation management. *PLoS One* **5**, e12813.

Melletti, M. & Burton, J. (2014) *Ecology, evolution and behaviour of wild cattle implications for conservation introduction*. Cambridge: Cambridge University Press.

Miettinen, J., Shi, C.H., Tan, W.J. & Liew, S.C. (2012). 2010 land cover map of insular Southeast Asia in 250-m spatial resolution. *Remote Sens. Lett.* **3**, 11–20.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B. & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature* **403**, 853.

Nagelkerke, N.J.D. (1991). A note on a general definition of the coefficient of determination. *Biometrika* **78**, 691–692.

Noss, R.F., Quigley, H.B., Hornocker, M.G., Merrill, T. & Paquet, P.C. (1996). Conservation biology and carnivore conservation in the Rocky Mountains. *Conserv. Biol.* **10**, 949–963.

Ohmann, J.L. & Spies, T.A. (1998). Regional gradient analysis and spatial pattern of woody plant communities of oregon forests. *Ecol. Monogr.* **68**, 151.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Henry, M., Stevens, H., Szoecs, E. & Wagner, H. (2018). Vegan: community ecology package. R package version 2.4-6. Available at http://cran.r-project.org/package=vegan

Ovaskainen, O., Hottola, J. & Siitonen, J. (2010). Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. *Ecology* **91**, 2514–2521.

Palmer, M.W. (1993). Putting things in even better order – the advantages of canonical correspondence-analysis. *Ecology* **74**, 2215–2230.

Penjor, U., Kaszta, Z., Macdonald, D.W. & Cushman, S.A. (2021). Prioritizing areas for conservation outside the existing protected area network in Bhutan: The use of multi-species, multi-scale habitat suitability models. *Landscape Ecol.* **36**, 1281–1309.

PRISM Climate Group. (2016). *PRISM climate data*. Oregon State University. Available at http://prism.oregonstate.edu

R Core Team. (2018) *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. https://www.r-project.org/

Raes, N., Roos, M.C., Slik, J.W.F., van Loon, E.E. & ter Steege, H. (2009). Botanical richness and endemicity patterns of Borneo derived from species distribution models. *Ecography* **32**, 180.

Roberge, J.M. & Angelstam, P. (2004). Usefulness of the umbrella species concept as a conservation tool. *Conserv. Biol.* **18**, 76.

Ross, J., Hearn, A.J. & Macdonald, D.W. (2017) The Bornean carnivore community: lessons from a little-known guild. In *Biology and conservation of musteloids*: 326–339. Macdonald, D.W., Newman, C. & Harrington, L.A. (Eds). Oxford: Oxford University Press.

Simcharoen, A., Savini, T., Gale, G.A., Simcharoen, S., Duangchantrasiri, S., Pakpien, S. & Smith, J.L.D. (2014). Female tiger *Panthera tigris* home range size and prey abundance: important metrics for management. *Oryx* **48**, 370–377.

Sodhi, N.S., Koh, L.P., Brook, B.W. & Ng, P.K.L. (2004). Southeast Asian biodiversity: an impending disaster. *Trends Ecol. Evol.* **19**, 654–660.

Sollmann, R., Mohamed, A., Samejima, H. & Wilting, A. (2013). Risky business or simple solution – relative abundance indices from camera-trapping. *Biol. Conserv.* **159**, 405–412.

Teo, H.C., Lechner, A.M., Sagala, S. & Campos-Arceiz, A. (2020). Environmental impacts of planned capitals and lessons for Indonesia's new capital. *Land* **9**, 438.

ter Braak, C.J.F. (1986). Canonical correspondence-analysis – a new eigenvector technique for multivariate direct gradient analysis. *Ecology* **67**, 1167.

ter Braak, C.J.F. & Prentice, I.C. (1988). A theory of gradient analysis. *Adv. Ecol. Res.* **18**, 271.

ter Braak, C.J.F. & Verdonschot, P.F.M. (1995). Canonical correspondence-analysis and related multivariate methods in aquatic ecology. *Aquat. Sci.* **57**, 255–289.

Vincent, P.J. & Haworth, J.M. (1983). Poisson regression-models of species abundance. *J. Biogeogr.* **10**, 153.

Wang, Y.W., Allen, M.L. & Wilmers, C.C. (2015). Mesopredator spatial and temporal responses to large predators and human development in the Santa Cruz Mountains of California. *Biol. Conserv.* **190**, 23–33.

WCS, CIESIN. (2005). *Last of the wild data version 2 (LTW-2): global human footprint dataset (geographic)*. Available at http://sedac.ciesin.columbia.edu/wildareas/

Wessels, K.J., Van Jaarsveld, A.S., Grimbeek, J.D. & Van der Linde, M.J. (1998). An evaluation of the gradsect biological survey method. *Biodivers. Conserv.* **7**, 1093–1121.

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 Ecol. Evol.* **28**, 531–540.

Supporting information

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

Figure S1. Histograms showing the distribution of the dispersion values in relation to the number of single-species GLMs with Poisson distribution, for Borneo.

Figure S2. Histograms showing the distribution of the dispersion values in relation to the number of single-species GLMs with Poisson distribution, for Sumatra.

Figure S3. Histograms showing the distribution of the residual deviances of the single-species GLMs of species sampled in Borneo.

Figure S4. Histograms showing the distribution of the residual deviances of the single-species GLMs of species sampled in Sumatra.

Table S1. Covariates' coefficients of the single-species GLMs for Borneo.

Table S2. Covariates' coefficients of the single-species GLMs for Sumatra.

Table S3. Species sampled in Borneo and Sumatra with relative identification codes and detections.

Table S4. Borneo biplot covariates' scores.

Table S5. Sumatra biplot covariates' scores.

Table S6. Most efficient systems of protected areas in preserving predicted species richness in Borneo.

Table S7. Most efficient systems of protected areas in preserving predicted species richness in Sumatra.

Table S8. Most biodiverse areas falling outside the current system of protected areas in Borneo and Sumatra, and relative amount of cumulative species richness that they harbour.

Table S9. Indicator species for Bornean and Sumatran biodiversity.