

# SCIENTIFIC REPORTS



OPEN

## Defining habitat covariates in camera-trap based occupancy studies

Received: 05 January 2015

Accepted: 23 October 2015

Published: 24 November 2015

Jürgen Niedballa<sup>1</sup>, Rahel Sollmann<sup>1,2,†</sup>, Azlan bin Mohamed<sup>1</sup>, Johannes Bender<sup>1</sup> & Andreas Wilting<sup>1</sup>

In species-habitat association studies, both the type and spatial scale of habitat covariates need to match the ecology of the focal species. We assessed the potential of high-resolution satellite imagery for generating habitat covariates using camera-trapping data from Sabah, Malaysian Borneo, within an occupancy framework. We tested the predictive power of covariates generated from satellite imagery at different resolutions and extents (focal patch sizes, 10–500 m around sample points) on estimates of occupancy patterns of six small to medium sized mammal species/species groups. High-resolution land cover information had considerably more model support for small, patchily distributed habitat features, whereas it had no advantage for large, homogeneous habitat features. A comparison of different focal patch sizes including remote sensing data and an *in-situ* measure showed that patches with a 50-m radius had most support for the target species. Thus, high-resolution satellite imagery proved to be particularly useful in heterogeneous landscapes, and can be used as a surrogate for certain *in-situ* measures, reducing field effort in logistically challenging environments. Additionally, remote sensed data provide more flexibility in defining appropriate spatial scales, which we show to impact estimates of wildlife-habitat associations.

Understanding the distribution and habitat associations of wildlife species is a key topic in ecology, and important for their conservation<sup>1</sup>. Studying wildlife habitat associations requires appropriate definition of environmental covariates at spatial scales that are relevant to the species under study<sup>2</sup>. A variety of approaches and methods have been developed to generate potential explanatory variables for species distribution models. These include both information collected *in-situ*, such as measurements of vegetation, disturbances or terrain collected at and around the survey locations<sup>3</sup>, and information based on remote sensing or airborne land cover analyses<sup>4–6</sup>.

Data from *in-situ* habitat surveys are reliable, can provide information not readily available from remote sensing (e.g. ground cover, floristic or phenological information) and can serve for ground truthing remote sensing data. These surveys can, however, be logistically challenging, costly, time-consuming and physically demanding, depending on the terrain conditions, the habitat information of interest, and the spatial scale at which data are to be collected.

Advantages of remote sensing data include extensive data coverage over large regions, also allowing for extrapolation and mapping predicted distributions, a wide spectrum of available data sets<sup>7</sup> and user-friendly GIS software to extract information from data layers. Nevertheless, spatial resolution of remote sensing data is often a limiting factor for identification of smaller land cover features, with spatial resolutions ranging from 1-km resolution carbon stock data<sup>8</sup>, to 250-m resolution MODIS or land cover data<sup>9</sup>, 30-m Landsat imagery and derived data<sup>10</sup> to high-resolution (<1 m) satellite imagery.

<sup>1</sup>Leibniz Institute for Zoo and Wildlife Research, Alfred-Kowalke-Str. 17, 10315 Berlin, Germany. <sup>2</sup>North Carolina State University, Department of Forestry and Environmental Resources, Campus Box 8008, Raleigh, NC 27695-7646, USA. <sup>†</sup>Present address: USDA Forest Service, Pacific Southwest Research Station, 1731 Research Park Drive, Davis, CA 95618, USA. Correspondence and requests for materials should be addressed to J.N. (email: niedballa@izw-berlin.de)

Typical satellite based imagery is further restricted to the top vegetation layer, providing no information on three-dimensional vegetation structure or features below canopy cover. Studies using high-resolution airborne LiDAR data (down to < 1 m resolution and allowing for three-dimensional imaging) have overcome these problems and shown that fine-scale variations in habitat structure can influence species distributions<sup>11,12</sup>. These highly sophisticated data are, however, expensive to obtain and difficult to analyse, and thus unavailable to many wildlife studies.

Thus, both *in-situ* and remote sensing derived covariates have their advantages and disadvantages, but only few studies compared their usefulness in wildlife distribution and habitat modelling<sup>13</sup>. The choice of the suitable type of covariates used is mostly governed by knowledge of or hypotheses about the ecology and life histories of species of interest. If little or nothing is known, variables characterising the environment in general terms or proxy measures can be used.

Moreover, it is well known that ecological patterns and processes are scale-dependent<sup>14,15</sup>, and an adequate definition of spatial scale is important when modelling species-habitat associations<sup>16,17</sup>. In an ecological context, scale is the spatial (or temporal) dimension of an object or process, characterised by grain and extent<sup>16,18,19</sup>. Here, grain is the spatial resolution of remote sensing data (i.e., pixel size of a raster data set), and extent is characterised by focal patches of different sizes, i.e. circular areas of different radii surrounding the sampling points<sup>20</sup>.

Not determining the appropriate spatial scale (either grain or extent) may lead to failure to detect species habitat associations. Ideally, the definition of spatial scale is based on ecological reasoning<sup>21</sup>. If known, average species home range sizes or inference from related species can help in defining the appropriate spatial extent. If no information is available, various extents can be compared via model selection procedures<sup>22</sup>.

Here, we used camera-trapping data from Sabah, Malaysian Borneo, in an occupancy framework, one of the most common methods to study species-habitat association while accounting for imperfect species detection<sup>23–25</sup>, to 1) assess the sensitivity of occupancy models to the spatial resolution (grain size) of land cover data; and 2) investigate what focal patch size (extent) of remotely-sensed land cover information and *in-situ* habitat variables around camera-traps is most relevant to occupancy patterns of small to medium sized mammals. Our analysis aims to draw attention to scale sensitivity of model results, assess the usefulness of high-resolution land cover data, evaluate the need for *in-situ* habitat surveys, and thereby increase the efficiency and ecological relevance of future wildlife-habitat association studies.

## Methods

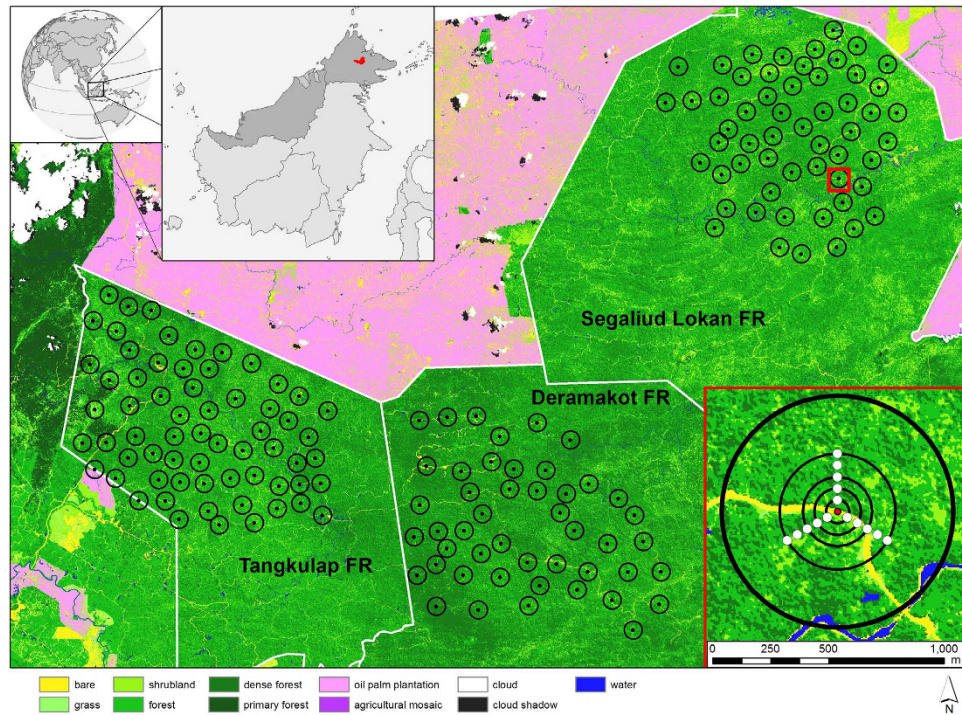
**Study sites.** This study was conducted in three commercial forest reserves in central Sabah on Malaysian Borneo: Deramakot Forest Reserve (DFR; 551 km<sup>2</sup>, 5°14′–28′N, 117°20′–38′E), Tangkulap-Pinangah Forest Reserve (TFR; 501 km<sup>2</sup>, 5°17′–31′N, 117°03′–20′E) and Segaliud Lokan Forest Reserve (SLFR; 5°20′–39′N, 573 km<sup>2</sup>, 117°25′–39′E; Fig. 1). The reserves are comprised of lowland rainforest (altitude between 50–250 m) and have all been selectively logged at least once. Because of more intensive and destructive logging in the past, TFR and SLFR show higher degrees of forest disturbance than DFR, where reduced impact logging was adopted in 1995 and certification by the Forest Stewardship Council followed in 1997<sup>3,26,27</sup>.

**Camera-trapping.** We set up 47, 64, and 55 camera-trap stations covering areas of 123 km<sup>2</sup>, 122 km<sup>2</sup>, and 114 km<sup>2</sup> in DFR, TFR, and SLFR, respectively (Fig. 1). Setups approximated a systematic array with random origin, adjusted to logistical circumstances, to achieve representative coverage of the study areas. DFR was sampled between September 2008 and January 2009, TFR between April and September 2009, and SLFR between January and April 2010. Camera stations were spaced approximately 1.4 km apart; each station consisted of 2 heat-in-motion sensor triggered camera-traps (models Expert and Capture; Cuddeback, De Pere, Wisconsin) facing each other (for details, see<sup>3</sup>).

**Occupancy modelling.** We used species detection information from camera-trapping in combination with occupancy modelling to investigate the effects of spatial resolution and extent of habitat covariates. Occupancy models use species detection/non-detection data from repeated visits to a collection of sampling sites to estimate the probability of species occurrence and its relationship with environmental covariates while accounting for imperfect species detection<sup>23–25</sup>. They consist of two components that explicitly model the ecological process (i.e. occupancy of sites) and the observation process<sup>28</sup>. The true occupancy state at site  $i$ ,  $z_i$  (1 if present, and 0 otherwise) is considered a Bernoulli trial with probability of occupancy  $\Psi_i$ :  $z_i \sim \text{Bernoulli}(\Psi_i)$ . Since non-detection of a species at a sampling site can either be caused by true absence or by failure of detection, repeated visits over  $k$  occasions to sampling sites are used to estimate detection probability  $p_{ik}$  conditional on occupancy. Observations  $y_{ik}$  are also assumed to be a Bernoulli trial with  $y_{ik}|z_i \sim \text{Bernoulli}(p_{ik} z_i)$ . Thus  $p_{ik} = 0$  where  $z_i = 0$ , i.e., the species is not present.

Both occupancy probability  $\Psi$  at a site  $i$  and detection probability  $p$  can be modelled as linear functions of covariates  $x_i$  using logit link functions, e.g.:

$$\text{logit}(\Psi_i) = \beta_0 + \beta_1 x_i \text{ and } \text{logit}(p_i) = \gamma_0 + \gamma_1 x_i,$$



**Figure 1.** Map of the study site in Sabah, Malaysian Borneo. The Borneo map highlights the three commercial forest reserves in red. The main map shows the RapidEye land cover classification. Black dots show camera-trap locations, each with their respective 500-m radius. The bottom right inset (indicated by the red frame in the main map) magnifies one camera-trap location (central red point) with its habitat survey points along the 250-m transect lines (white points). Land cover extraction radii (extent, focal patches) are overlaid (black circles). The bold outer circle has a 500-m radius, the others correspond to 250, 150, 100 and 50 m. The 10-m radius circle is equal in size to the red point. The map was generated using ArcGIS 10.1 (ESRI, Redlands, CA, USA).

where  $\beta_0$  and  $\gamma_0$  denote the intercepts and  $\beta_1$  and  $\gamma_1$  single regression coefficients<sup>25</sup>. To define sampling occasions, we divided the total sampling period for each study site into 6-day sampling intervals, resulting in 7 occasions in DFR and TFR and 8 in SLFR<sup>3</sup>. For each species, we constructed a site-by-occasion detection/non-detection matrix with values of 1 if the species was detected at a given site on a given occasion, 0 if not and NA if the cameras were not operational.

We implemented occupancy models<sup>25</sup> in R 3.1.1<sup>29</sup> using package “unmarked” version 0.10-3<sup>30</sup>. For every species, we first selected the most suitable model for detection probability  $p$  using the camera position (on/off road) and forest reserve (in all combinations) as detection probability covariates while holding occupancy probability constant across sites (i.e. we used no covariates to model occupancy probability, Table S1). These models will be termed ‘constant occupancy models’ for the sake of simplicity. Model selection was based on Akaike’s Information Criterion (AIC)<sup>31</sup>. Conditional on the best detection model we then evaluated the effect of different covariates at varying spatial resolutions and extents on species occupancy, as described below.

We generally assessed the effects of covariates on occupancy probabilities with one occupancy covariate per model. Therefore, model rankings and inferences were not affected by correlations between related covariates.

**Study species.** We built occupancy models for six relatively small mammal species covering different taxonomic clades and ecological groups: Banded Civet *Hemigalus derbyanus* ( $n = 35$  records), Long-tailed Macaque *Macaca fascicularis* ( $n = 76$ ), Malay Civet *Viverra zibetha* ( $n = 610$ ), Moonrat *Echinosorex gymnura* ( $n = 140$ ), Greater and Lesser Chevrotain *Tragulus napu* and *T. kanchil* ( $n = 561$ ), and Thick-spined Porcupine *Hystrix crassispinis* ( $n = 42$ ). As Greater and Lesser Chevrotain are difficult to distinguish with certainty on camera-trap photographs, we pooled both species and jointly analysed them.

Occupancy models assume spatial independence among sampling sites. Malay Civet<sup>32</sup>, Long-tailed Macaque<sup>33</sup> and Chevrotains<sup>34</sup> have average home range diameters smaller than our average camera-trap station spacing of 1.4 km; we assume that the same is true for the Banded Civet, Thick-spined Porcupine



and Moonrat, because of their smaller size compared to the Malay Civet, and because the latter two are not carnivorous.

**Habitat covariates.** We mapped land cover using multispectral classifications of RapidEye high-resolution (5 m) satellite imagery. We used seven images (Catalog-IDs: 10606784, 10606821, 9290487, 9290518, 6890479, 10129761, 6890524) acquired between 07/2011 and 09/2012 as data base for this analysis. The RapidEye data products were supplied by the RapidEye Science Archive program (Project-ID 654) and delivered in orthorectified L3A-format<sup>35</sup>.

To reduce scene-to-scene variability, radiometric corrections were applied as recommended for multi-temporal and multi-sensor data applications<sup>36,37</sup>. The image-based atmospheric corrections included 'dark object subtraction' and conversion to exoatmospheric (top-of-atmosphere) reflectance<sup>35,38,39</sup>. By applying pixel-based maximum-likelihood land cover classifications, nine different land cover types were identified (Fig. 1). Clouds and cloud shadows were eliminated consulting a Landsat-based classification<sup>40</sup>. All images used for land cover classification were processed with ERDAS Imagine 2013 (Hexagon Geospatial, Norcross, GA, USA). The overall accuracy of the classification as estimated from 211 ground control points was 82.4%.

Based on this land cover classification we calculated four habitat covariates: distance from every camera trap station to the nearest oil palm plantation (D.PLANT) and to the nearest water pixel (D.WATER), 'forest score' (FS) and land cover heterogeneity (HET). The first two covariates were used to assess the sensitivity of occupancy models to spatial resolution and the latter two to test the sensitivity of occupancy models to different focal patch sizes.

FS is the weighted mean of land cover percentages within extracted areas, the weights are integer numbers assigned to each land cover class ranking forest quality. Thus, FS is an index of the degree of forest cover and quality in the surroundings of camera-traps. Bare areas, grassland, oil palm plantations and water were assigned 0, shrub 1, forest 2, dense and primary forest 3, allowing FS to range from 0 to 3. Lower numbers indicate higher disturbance of the forest.

Heterogeneity was calculated using Pielou's evenness index, which is defined as the ratio between the actual and the highest possible Shannon diversity of members of an assemblage<sup>41–43</sup>. Values can range from 0 to 1, with 0 if a collection consists of only one class and 1 in case of perfect evenness between classes. In our context it can be interpreted as heterogeneity of land cover, because the more numerically similar the percentages of land cover classes in an area are, the more heterogeneous is land cover.

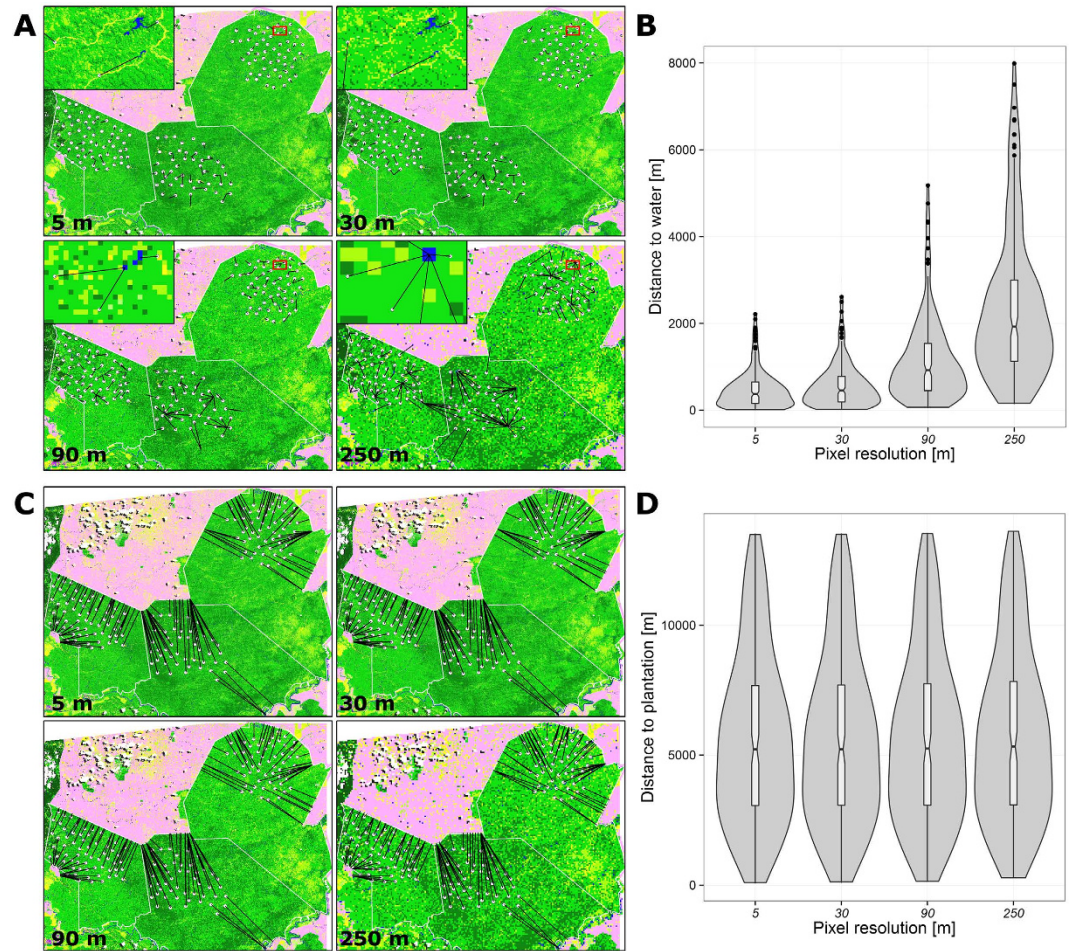
The ecological reasoning behind the choice of these covariates is that all animals depend on water to some degree and therefore access to water is a basic requirement<sup>44</sup>. Distance to oil palm plantations quantifies potential edge effects and can be interpreted as a proxy for human disturbance<sup>45</sup>. Forest score and heterogeneity are both metrics to characterise the habitat and describe the forest quality and disturbance.

In addition to habitat covariates derived from the high-resolution remote sensing data, we included one *in-situ* measured covariate into our analyses. At each camera-trap station, canopy closure (CC) was recorded every 50 m using a spherical densiometer along 3 line transects of 250 m in the direction of 0°, 120° and 240°, and the data were pooled by camera-trap station. Due to logistic constraints, not all transects could be carried out along the entire 250 m and mean effective transect length was 184 m ± 84 m. We computed CC covariates as the mean of CC values at distances of up to 50, 100 and 150 m from the camera-trap stations. 150 m was chosen as the maximum distance because 95% of all stations had at least one transect of at least that length. CC is related to forest disturbance: less disturbed forests are expected to have a more closed canopy, i.e. higher values of CC<sup>3</sup>.

**Goal 1: Sensitivity of occupancy models to spatial resolution of remotely sensed land cover information.** The 5 m land cover classification was resampled to lower resolutions commonly found in other remote sensing data (30-m Landsat; 90-m ASTER; 250-m MODIS) using the majority method (i.e. by assigning each new raster cell the most common pixel value within its extent) in ArcGIS 10.1 (ESRI, Redlands, CA, USA). For all 4 resolution levels we computed the distance from every camera-trap station to the nearest oil palm plantation (D.PLANT) and to the nearest water pixel (D.WATER) (D.PLANT<sub>5</sub>, D.PLANT<sub>30</sub> and so on, Fig. 2). The oil palm plantations represent a large continuous habitat feature, for which distances remained largely constant across spatial resolutions (Fig. 2C,D), whereas water resources were patchily distributed across our study areas and many water resources were smaller than the pixel sizes of the coarser resolutions. As a result, distance to water increased with the coarser resolutions (Fig. 2B); 30 m resolution resulted in the loss of very small ponds and streams while representing rivers well; at 90 m, most small ponds disappeared from the land cover map, medium rivers were represented in a discontinuous yet recognizable way and only large rivers were a continuous band of pixels, and at 250 m resolution even the largest river, Sungai Kinabatangan, was discontinuous, small and medium rivers and ponds mostly disappeared (Fig. 2A).

Conditional on the best constant occupancy models, we performed AIC-based model selection of occupancy covariates D.PLANT and D.WATER computed at 4 spatial resolutions for every species to assess the sensitivity of occupancy models to the spatial resolution of land cover information.

**Goal 2: Sensitivity of occupancy models to focal patch sizes around camera-traps.** We computed 'forest score' (FS) and land cover heterogeneity (HET) from the surroundings of the camera-trap stations



**Figure 2.** Maps and violin plots for distance to water (A,B) and distance to oil palm plantations (C,D) by pixel resolution (grain size) for three commercial forest reserves in Sabah, Malaysian Borneo. Red squares indicate position of magnified inset in A. The maps were generated using ArcGIS 10.1 (ESRI, Redlands, CA, USA).

using circles with radii of 10 m, 50 m, 100 m, 150 m, 250 m and 500 m (corresponding to focal patches of 0.03 ha, 0.8 ha, 3.1 ha, 7.1 ha, 16.9 ha, 78.5 ha). We chose 10 m as the minimum radius to achieve a sample size of at least 10 pixels per station, and 500 m as the maximum radius to avoid overlap between circles around neighbouring camera-trap stations. Further, we built occupancy models using the *in-situ* collected information on CC at 50, 100 and 150 m around each camera-trap.

We compared the six focal patch sizes (10 m, 50 m, 100 m, 150 m, 250 m and 500 m radii) of FS and HET land cover covariates and three focal patch sizes for *in-situ* CC measurements (50 m, 100 m, 150 m) to each other and their respective constant occupancy models using AIC-based model selection in order to find a radius at which habitat covariates had the highest predictive power for our set of species. We chose a consensus radius among those radii that were available for all covariates (50, 100, 150 m) using an ad hoc approach: We calculated the cumulative  $\Delta$ AIC for each radius over all six species. A lower cumulative  $\Delta$ AIC indicates that a given radius is, on average, closer to the top model than one with a higher cumulative  $\Delta$ AIC.

**Goodness of model fit.** Because AIC is only a relative measure of model quality, we conducted goodness of fit tests<sup>46</sup> for each species' global model, with covariates based on the consensus radius, using the R package AICcmodavg<sup>47</sup>. We found no evidence for lack of fit (bootstrapped p values > 0.1 and variance inflation factors < 1.5 in all global models, Supplementary Table S7) and therefore refrained from converting AIC to qAIC for model selection.

## Results

**Goal 1: Sensitivity of occupancy models to spatial resolution of land cover information.** For all species, occupancy models using D.PLANT as single covariates were not influenced by the spatial

Pixel size	AIC	$\Delta$ AIC	wAIC	$\beta_1^*$	SE	CV	p-value**
<i>Distance to oil palm plantation</i>							
90	331.48	0	0.250	-0.72	0.29	0.4	<b>0.012</b>
5	331.5	0.02	0.247	-0.72	0.29	0.4	<b>0.012</b>
30	331.5	0.02	0.247	-0.72	0.29	0.4	<b>0.012</b>
250	331.55	0.07	0.241	-0.72	0.29	0.4	<b>0.012</b>
—	337.05	5.57	0.015	—	—	—	—
<i>Distance to water</i>							
5	301.13	0	0.997	-3.59	0.92	0.26	< <b>0.001</b>
30	312.86	11.73	0.003	-2.32	0.64	0.28	< <b>0.001</b>
90	330.35	29.22	0	-0.96	0.39	0.41	<b>0.014</b>
250	335.41	34.28	0	-0.49	0.27	0.55	0.074
—	337.05	35.92	0	—	—	—	—

**Table 1.** Results of occupancy models for Long-tailed Macaque using distance to large continuous (distance to oil palm plantation) and small patchy (distance to water) remote sensed habitat features at different spatial resolutions as covariate, estimated from camera-trapping data collected between 2008 and 2010 in three commercial forest reserves in Sabah, Malaysian Borneo.  $\Delta$ AIC: difference in AIC to top model, wAIC = AIC model weights,  $\beta_1$  = regression coefficient, SE = regression coefficient standard error, CV = coefficient of variation of  $\beta_1$  ( $SE/|\beta_1|$ ), — denotes constant occupancy model. \*Positive regression coefficients indicate positive association with distance to features, i.e. negative association to features. Negative regression coefficients indicate negative association with distance to features, i.e. positive association to features. \*\*Bold font indicates significance at the 0.05 level.

resolution of land cover information. Regression coefficients and standard errors were very similar ( $\pm 0.02$ ) for spatial resolutions from 5 to 250 m; AIC values of individual models were virtually constant with  $\Delta$ AIC < 0.2, and AIC weights hardly differed across resolutions (see Table 1 for the Long-tailed Macaque and Supplementary Table S2 for other species).

In contrast, D.WATER models differed substantially by pixel resolution, particularly in species that exhibited strong association with this covariate. The effect was most pronounced in Long-tailed Macaque, which had a very strong ( $\beta_1 = -3.59 \pm 0.92$ ) and highly significant negative association with distance to water at high resolution, i.e., occupancy probability  $\Psi$  was higher closer to water resources. AIC increased drastically and significance of regression coefficients decreased gradually with lower resolutions (Table 1). The Thick-spined Porcupine showed a significant negative association with D.WATER at all resolutions, but the effect on occupancy was strongest at 90-m resolution (see Supplementary Table S3). Only the Chevrotain showed a positive association with D.WATER, but only the coarse (250 m) resolution had more support than the constant occupancy model.

**Goal 2: Sensitivity of occupancy models to focal patch sizes around camera-traps.** Generally, for FS and HET as well as CC, smaller focal patch sizes (i.e. smaller radii) had lower AIC values than larger radii in species whose occurrence was associated with the respective covariates (Table 2 for the Long-tailed Macaque and Supplementary Tables S4–S6 for other species). Particularly for the Long-tailed Macaque, the species with the strongest association with the covariates, the effect of focal patch size was pronounced, with smaller radii having more predictive power for occupancy than larger radii. Based on cumulative  $\Delta$ AIC, we chose 50 m as the consensus radius for all covariates (Table 3).

## Discussion

Species habitat associations and preferences are multi-factorial processes aimed at maximizing fitness<sup>48</sup> that integrate information and involve decisions made at various interacting spatial and temporal scales. To account for this complexity, analysis of habitat associations needs to be carried out at different spatial scales<sup>2,49</sup>. This is commonly achieved by analysing hierarchical selection of habitat use (e.g. at landscape, macrohabitat and microhabitat scale)<sup>2,50,51</sup> or by comparing multiple focal patch sizes<sup>20,17</sup>. Both approaches aim to find relevant spatial extents, but at the same time are often influenced by the available resolution (grain size) or, in case of *in-situ* measures, availability of habitat covariates.

Our analysis showed that the effect of grain size differed by the type of land cover feature. In the case of large continuous habitat features such as oil palm plantations a higher resolution did not improve the predictive power of the covariate, as distances to these large continuous features did not depend on resolution (Fig. 2). Analogous examples to oil palm plantations would be all kinds of large land cover features that have a well-defined linear border such as other large-scale agriculture or urban areas, but

Radius	AIC	$\Delta$ AIC	wAIC	$\beta_1^*$	SE	CV	p-value**
<i>Remote Sensing – Forest Score</i>							
10	332.48	0	0.5	−0.68	0.27	0.4	<b>0.013</b>
50	334.61	2.13	0.173	−0.55	0.27	0.49	<b>0.041</b>
100	335.61	3.13	0.105	−0.56	0.32	0.57	0.086
150	336.04	3.56	0.084	−0.56	0.36	0.64	0.116
250	336.8	4.32	0.058	−0.52	0.39	0.75	0.185
—	337.05	4.57	0.051	—	—	—	—
500	338.1	5.62	0.03	−0.27	0.28	1.04	0.339
<i>Remote Sensing - Heterogeneity</i>							
50	329.7	0	0.455	1.03	0.38	0.37	<b>0.007</b>
10	330.87	1.17	0.253	1.02	0.42	0.41	<b>0.014</b>
100	330.92	1.22	0.247	0.83	0.32	0.39	<b>0.009</b>
150	335.66	5.96	0.023	0.5	0.28	0.56	0.075
—	337.05	7.35	0.012	—	—	—	—
250	338.33	8.63	0.006	0.22	0.26	1.18	0.402
500	338.97	9.27	0.004	0.07	0.25	3.57	0.771
<i>In-situ - Canopy Closure</i>							
50	326.53	0	0.754	−1.1	0.37	0.34	<b>0.003</b>
100	329.18	2.65	0.201	−1.07	0.42	0.39	<b>0.01</b>
150	332.33	5.8	0.042	−0.91	0.42	0.46	<b>0.032</b>
—	337.05	10.52	0.004	—	—	—	—

**Table 2.** Results of occupancy models for Long-tailed Macaque using remote sensing information and *in-situ* canopy closure at different focal patch sizes as covariates on occupancy, estimated from camera-trapping data collected between 2008 and 2010 in three commercial forest reserves in Sabah, Malaysian Borneo.  $\Delta$ AIC: difference in AIC to top model, wAIC = AIC model weights,  $\beta_1$  = regression coefficient, SE = regression coefficient standard error, CV = coefficient of variation of  $\beta_1$  ( $SE/|\beta_1|$ ), — denotes constant occupancy model. \*Positive regression coefficients indicate positive association with features. Negative regression coefficients indicate negative association with features. \*\*Bold font indicates significance at the 0.05 level.

Extent	Forest Score	Heterogeneity	Canopy closure
10	17.02	22.99	—
50	13.33	7.17	7.15
100	15.32	18.68	9.60
150	18.63	23.82	11.19
250	21.31	25.69	—
500	17.67	33.02	—

**Table 3.** Cumulative  $\Delta$ AIC for occupancy models containing covariates at different focal patch sizes (extent) across six species/species groups, estimated from camera-trapping data collected between 2008 and 2010 in three commercial forest reserves in Sabah, Malaysian Borneo. ‘Extent’ refers to the radius around camera-trap stations from which covariate values were extracted. Cumulative  $\Delta$ AIC was calculated for each radius over all six species. A lower cumulative  $\Delta$ AIC indicates that a given radius is, on average, closer to the top model than one with a higher cumulative  $\Delta$ AIC.

also natural habitat edges or large lakes and coastlines. We expect that finer scale processes, such as movement or activity patterns, might be more sensitive to the resolution of habitat edges than the relatively coarse process of occupancy. In addition, higher resolution land cover data may have an advantage if the feature edge is less regular than in the present case.

In contrast to large land cover features, water resources were very localised, interspersed within a matrix of different land cover types and other fine-scale features in our study sites. Therefore, many of the small water resources were not visible in coarser resolution land cover data (Fig. 2B). In the set



of analysed species the Long-tailed Macaque is known to be closely associated with water resources, namely rivers<sup>52</sup>. Our results supported the strong association with water, but we further found that even small water bodies were important for occupancy of the species, as models with a higher resolution (i.e., accounting for small water sources) had much more support than low resolution models. Very little is known about the association of the other species with water, but similar to the Long-tailed Macaque, our results showed that the Thick-spined Porcupine had a significant negative association with distance to water at all resolutions, i.e. it occurred more frequently near water. In contrast to the Long-tailed Macaque, 90-m resolution had the highest predictive power, which might indicate that this species is associated with larger rather than small water bodies. The estimated positive association of Chevrotains with distance to water was only found in the model with the lowest resolution, whereas all other resolutions led to models with less support than the constant occupancy model. If Chevrotains really avoided water, however, we would expect this pattern to be found in the higher resolutions too, similar to the results for the Thick-spined Porcupine. Therefore, we think that this finding is either an artefact of the majority method used for the resampling or a spurious relationship caused by a correlation of distance to water at coarse resolution to a habitat feature not considered in our analysis.

Analogous to localised water resources, we expect a similar effect of grain size on the predictive power of habitat covariates for other small habitat features such as individual trees in savannahs, grass patches or clearings in forests, individual houses, small-scale agriculture, burnt areas, dump sites or small roads/skid trails, i.e. features that, although present in the landscape, are not visible at coarser resolutions (i.e. at larger grain sizes), because they are smaller than one pixel.

In addition to distance-based covariates, habitat association studies often use patch characteristics around the survey locations. As ecological patterns and processes are scale-dependent<sup>15</sup> adequate definition of the focal patch sizes is important<sup>16,17</sup>. Earlier studies already showed that home range size and other ecological parameters are highly important for defining the focal patch sizes and therefore, if possible, the definition of focal patch sizes should be species-specific<sup>2</sup>.

Often very little is known about the spatial ecology of the species of concern and therefore, species-specific focal patch sizes are difficult to define a priori. The possibility to test different focal patch sizes in order to define the adequate spatial extent for covariates and to adjust these to the spatial ecology of different species is a great advantage of remote-sensed covariates, highlighted by our study. This is particularly important in multiple-species data sets like those derived from camera-trap studies.

In this study we focused on species with small home-ranges, mainly to avoid spatial autocorrelation between camera-traps. For this set of species we found that smaller focal patches of land cover metrics and *in-situ* covariates (with a radius of 50 m) usually explained species occupancy patterns better or at least as good as larger patches. Smaller focal patches closely resemble the plots that are routinely used in vegetation assessments<sup>40</sup>, and it is extremely unlikely that *in-situ* covariates could be collected at focal patch sizes larger than 1 ha (corresponding approximately to the 50-m consensus radius identified in the present study) in challenging field conditions. Thus, our data indicate that for small species sampling squared plots around camera-traps would potentially have provided a more representative picture of the habitat conditions relevant for species occurrence compared to three 4-m wide 250-m long line transects and at the same time would have been easier to sample and more suitable for ground truthing remote sensing data.

Even for the species with small home ranges used in the present study, the focal patch size of 0.8 ha (corresponding to a 50-m radius around camera-trap locations) represents only a small fraction of their home-ranges<sup>32–34</sup>. We expect that for wider ranging species focal patch sizes smaller than the average home range would also have higher predictive power than home range based patches, especially in point survey based studies such as camera-trapping. We consider it unreasonable to assume that, for species with larger ranges, detections at a point can provide information about an area the size of an average home range.

The *in-situ* covariate canopy closure had high predictive power for two out of six species, but it was positively correlated with the remote sensing covariate forest score (Spearman's  $\rho = 0.58$  and  $p < 0.001$  between CC50 and FS50), indicating that remote sensing data can serve as a surrogate for canopy closure and potentially other *in-situ* variables. It should be noted, however, that these measures refer to different aspects of forest quality and may affect species occurrence via different mechanisms.

In addition, generating land cover classifications based on remote sensing data requires ground control points for ground truthing satellite imagery, i.e. vegetation plots that may include measurements of canopy cover or canopy closure. These plots can easily be placed around survey locations, if these represent the available land cover types. It is unnecessary, though, to perform ground truthing at all survey locations when using remote sensed covariates, consequently reducing overall field efforts.

Some habitat information, however, cannot be obtained without ground surveys. These can include human impacts like hunting (wire snares, camp sites, fire places or bullet cases) or harvesting of non-timber forest products. Proxies like distance to villages or roads derived from remote sensing data can potentially help to circumvent the need for ground based data in some circumstances, but relationships between anthropogenic impacts and remote sensed proxies need to be verified<sup>53</sup>. Apart from human impacts, some natural features such as salt-licks, soil types, fruiting trees or dead wood can also only be assessed on the ground. Whether *in-situ* information is needed thus depends on the research questions at hand.



Our results highlight the great potential high-resolution satellite imagery and derived landscape metrics offer for identifying species habitat associations with localised fine-scale land cover features in heterogeneous landscapes. Considering the predictive power of smaller focal patches and the advantages of high-resolution remote sensing information about certain habitat features in occupancy models, we expect that very high-resolution imagery of new satellites (grain sizes <1 m, e.g. WorldView, GeoEye or Quickbird) could further improve our ability to study species habitat associations. Even though three-dimensional and canopy structural LiDAR data<sup>6</sup> have great advantages over satellite imagery, we expect that due to the high costs of LiDAR data<sup>54</sup>, high-resolution remote sensing data will remain the main affordable data source for many wildlife studies and may be a compromise for studying fine-scale habitat variation with low costs and relative ease of use. Such satellite data also offer the opportunity to predict species occupancy and distribution to non-sampled areas and to evaluate changes in the distribution over time in wildlife monitoring.

In summary, we showed that both spatial resolution and spatial extent of habitat covariates influence camera-trap based occupancy models. Remote sensed land cover information and derived metrics provide more flexibility than *in-situ* data to tackle these issues, and can be a surrogate for, or at least complement, the labour-intensive on-the-ground habitat assessment. This is particularly beneficial in challenging environments such as tropical rainforests, ecosystems that are known for their rich biodiversity and number of endemic, threatened and little studied vertebrate species.

## References

- Guisan, A. *et al.* Predicting species distributions for conservation decisions. *Ecol. Lett.* **16**, 1424–1435 (2013).
- Mayor, S. J., Schneider, D. C., Schaefer, J. A. & Mahoney, S. P. Habitat selection at multiple scales. *Ecoscience* **16**, 238–247 (2009).
- Mohamed, A. *et al.* Density and habitat use of the leopard cat (*Prionailurus bengalensis*) in three commercial forest reserves in Sabah, Malaysian Borneo. *J. Mammal.* **94**, 82–89 (2013).
- Burton, A. C., Sam, M. K., Balangtaa, C. & Brashares, J. S. Hierarchical Multi-Species Modeling of Carnivore Responses to Hunting, Habitat and Prey in a West African Protected Area. *PLoS One* **7**, e38007 (2012).
- Gould, W. R. *et al.* Estimating Occupancy in Large Landscapes: Evaluation of Amphibian Monitoring in the Greater Yellowstone Ecosystem. *Wetlands* **32**, 379–389 (2012).
- Vierling, L. A., Vierling, K. T., Adam, P. & Hudak, A. T. Using satellite and airborne LiDAR to model woodpecker habitat occupancy at the landscape scale. *PLoS One* **8**, e80988 (2013).
- Kanagaraj, R., Wiegand, T., Mohamed, A. & Kramer-Schadt, S. Modelling species distributions to map the road towards carnivore conservation in the tropics. *Raffles Bull. Zool. Suppl. No.* **28**, 85–107 (2013).
- Saatchi, S. S. *et al.* Benchmark map of forest carbon stocks in tropical regions across three continents. *Proc. Natl. Acad. Sci. USA* **108**, 9899–9904 (2011).
- Miettinen, J., Shi, C., Tan, W. J. & Liew, S. C. 2010 land cover map of insular Southeast Asia in 250-m spatial resolution. *Remote Sens. Lett.* **3**, 11–20 (2012).
- Hansen, M. C. *et al.* High-resolution global maps of 21st-century forest cover change. *Science (80-)*. **342**, 850–853 (2013).
- Goetz, S. J. *et al.* Lidar remote sensing variables predict breeding habitat of a Neotropical migrant bird. *Ecology* **91**, 1569–1576 (2010).
- Palminteri, S., Powell, G. V. N., Asner, G. P. & Peres, C. a. LiDAR measurements of canopy structure predict spatial distribution of a tropical mature forest primate. *Remote Sens. Environ.* **127**, 98–105 (2012).
- Betts, M. G., Diamond, A. W., Forbes, G. J., Villard, M. A. & Gunn, J. S. The importance of spatial autocorrelation, extent and resolution in predicting forest bird occurrence. *Ecol. Modell.* **191**, 197–224 (2006).
- Levin, S. A. The problem of pattern and scale in ecology. *Ecology* **73**, 1943–1967 (1992).
- Rahbek, C. The role of spatial scale and the perception of large-scale species-richness patterns. *Ecol. Lett.* **8**, 224–239 (2005).
- Wiens, J. A. Spatial scaling in ecology. *Funct. Ecol.* **3**, 385–397 (1989).
- Holland, J. D., Bert, D. G. & Fahrig, L. Determining the Spatial Scale of Species' Response to Habitat. *Bioscience* **54**, 227–233 (2004).
- Turner, M. G., Dale, V. H. & Gardner, R. H. Predicting across scales: Theory development and testing. *Landsc. Ecol.* **3**, 245–252 (1989).
- Schneider, D. C. The Rise of the Concept of Scale in Ecology. *Bioscience* **51**, 545–554 (2001).
- Thornton, D. H., Branch, L. C. & Sunquist, M. E. The influence of landscape, patch, and within-patch factors on species presence and abundance: a review of focal patch studies. *Landsc. Ecol.* **26**, 7–18 (2010).
- Mazerolle, M. J. & Villard, M.-A. Patch characteristics and landscape context as predictors of species presence and abundance: A review. *Ecoscience* **6**, 117–124 (1999).
- Lechner, A. M., Langford, W. T., Jones, S. D., Bekessy, S. A. & Gordon, A. Investigating species-environment relationships at multiple scales: Differentiating between intrinsic scale and the modifiable areal unit problem. *Ecol. Complex.* **11**, 91–102 (2012).
- MacKenzie, D. I. *et al.* Estimating site occupancy rates when detection probabilities are less than one. *Ecology* **83**, 2248–2255 (2002).
- MacKenzie, D. I., Nichols, J. D., Hines, J. E., Knutson, M. G. & Frankling, A. B. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology* **84**, 2200–2207 (2003).
- MacKenzie, D. I. *et al.* *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence*. (Academic Press, 2006).
- Kitayama, K. *Co-benefits of Sustainable Forestry - Ecological Studies of a Certified Bornean Rain Forest*. (Springer, 2013).
- Lagan, P., Mannan, S. & Matsubayashi, H. Sustainable use of tropical forests by reduced-impact logging in Deramakot Forest Reserve, Sabah, Malaysia. *Ecol. Res.* **22**, 414–421 (2007).
- Royle, J. A. & Dorazio, R. M. *Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations and communities*. (Academic Press, 2008).
- R Core Team. R: A language and environment for statistical computing. (2014). at <<http://www.r-project.org/>>
- Fiske, I. J. & Chandler, R. B. unmarked: An R Package for Fitting Hierarchical Models of Wildlife Occurrence and Abundance. *J. Stat. Softw.* **43**, 1–23 (2011).
- Burnham, K. P. & Anderson, D. R. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. (Springer Verlag New York Inc., 2002).

32. Colón, C. P. Ranging behaviour and activity of the Malay civet (*Viverra zibetha*) in a logged and an unlogged forest in Danum Valley, East Malaysia. *J. Zool.* **257**, 473–485 (2002).
33. Wheatley, B. P. in *Macaques Stud. Ecol. Behav. Evol.* (Lindburg, D. G.) 215–46 (Van Nostrand Reinhold, 1980).
34. Matsubayashi, H., Bosi, E. & Kohshima, S. Activity and habitat use of lesser mouse-deer (*Tragulus javanicus*). *J. Mammal.* **84**, 234–242 (2003).
35. RapidEye A. G. *Satellite Imagery Product Specifications Version 3.2.* (2011).
36. Lu, D., Mausel, P., Brondizio, E. & Moran, E. Assessment of atmospheric correction methods for Landsat TM data applicable to Amazon basin LBA research. *Int. J. Remote Sens.* **23**, 2651–2671 (2002).
37. Chander, G., Markham, B. L. & Helder, D. L. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sens. Environ.* **113**, 893–903 (2009).
38. Chavez, P. S. Image-Based Atmospheric Corrections - Revisited and Improved. *Photogramm. Eng. Remote Sens.* **62**, 1025–1036 (1996).
39. Vanonckelen, S., Lhermitte, S. & Van Rompaey, A. The effect of atmospheric and topographic correction methods on land cover classification accuracy. *Int. J. Appl. Earth Obs. Geoinf.* **24**, 9–21 (2013).
40. Langner, A., Samejima, H., Ong, R. C., Titin, J. & Kitayama, K. Integration of carbon conservation into sustainable forest management using high resolution satellite imagery: A case study in Sabah, Malaysian Borneo. *Int. J. Appl. Earth Obs. Geoinf.* **18**, 305–312 (2012).
41. Pielou, E. C. The Measurement of Diversity in Different Types of Biological Collections. *J. Theor. Biol.* **13**, 131–144 (1966).
42. Mairota, P. *et al.* Using landscape structure to develop quantitative baselines for protected area monitoring. *Ecol. Indic.* **33**, 82–95 (2013).
43. Farina, A. Landscape structure and breeding bird distribution in a sub-Mediterranean agro-ecosystem. *Landsc. Ecol.* **12**, 365–378 (1997).
44. Rondinini, C. *et al.* Global habitat suitability models of terrestrial mammals. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **366**, 2633–2641 (2011).
45. Brodie, J. F., Giordano, A. J. & Ambu, L. Differential responses of large mammals to logging and edge effects. *Mamm. Biol. - Zeitschrift für Säugetierkd.* **80**, 7–13 (2014).
46. MacKenzie, D. I. & Bailey, L. L. Assessing the fit of site-occupancy models. *J. Agric. Biol. Environ. Stat.* **9**, 300–318 (2004).
47. Marc J. Mazerolle. AICcmodavg: Model selection and multimodel inference based on (Q)AIC(c). R package version 2.0-3. (2015). at <<http://cran.r-project.org/package=AICcmodavg>>
48. Chalfoun, A. D. & Martin, T. E. Assessments of habitat preferences and quality depend on spatial scale and metrics of fitness. *J. Appl. Ecol.* **44**, 983–992 (2007).
49. Orians, G. H. & Wittenberger, J. F. Spatial and Temporal Scales in Habitat Selection. *Am. Nat.* **137**, (1991).
50. Saab, V. Importance of Spatial Scale to Habitat Use by Breeding Birds in Riparian Forests: a Hierarchical Analysis. *Ecol. Appl.* **9**, 135–151 (1999).
51. Williams, S. E., Marsh, H. & Winter, J. Spatial Scale, Species Diversity, and Habitat Structure: Small Mammals in Australian Tropical Rain Forest. *Ecology* **83**, 1317–1329 (2002).
52. Rodman, P. S. Structural Differentiation of Microhabitats of Sympatric *Macaca fascicularis* and *M. nemestrina* in East Kalimantan, Indonesia. *Int. J. Primatol.* **12**, 357–375 (1991).
53. Shepard Jr., G. H., Levi, T., Neves, E. G., Peres, C. A. & Yu, D. W. Hunting in Ancient and Modern Amazonia: Rethinking Sustainability. *Am. Anthropol.* **114**, 652–667 (2012).
54. Hummel, S., Hudak, A. T., Uebler, E. H., Falkowski, M. J. & Megown, K. A. A Comparison of Accuracy and Cost of LiDAR versus Stand Exam Data for Landscape Management on the Malheur National Forest. *J. For.* **109**, 267–273 (2011).

## Acknowledgements

All authors thank the German Federal Ministry of Education and Research (BMBF FKZ: 01LN1301A) for funding this research. For supporting the field work we thank the Sabah Forestry Department, the Sabah Wildlife Department, and KTS Plantation Sdn. Bhd., and the Sabah Economic Planning Unit for issuing a research permit. Financial support was provided by the following institutions: Point Defiance Zoo and Aquarium, Clouded Leopard Project, WWF–Germany, WWF–Malaysia, Cleveland Metroparks, Minnesota Zoo, Houston Zoo, Nashville Zoo, Panthera Foundation, ZGAP, and the IZW. AM is grateful for a personal scholarship of WWF–Malaysia and for a Russel Fellowship of WWF United States. Underlying RapidEye data has been contributed on behalf of the German Aerospace Center through funding of the German Federal Ministry of Economy and Energy. We thank Aurelie Shapiro (WWF Germany) for her advice on the remote sensing techniques and Beth Gardner (NCSU) for insightful discussions about the modelling approaches.

## Author Contributions

J.N., R.S. and A.W. conceived the ideas, designed the analyses and wrote the manuscript. A.M. and A.W. collected the data. J.N. analysed the data. J.B. performed the remote sensing analyses. A.M. and J.B. contributed to writing of the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Niedballa, J. *et al.* Defining habitat covariates in camera-trap based occupancy studies. *Sci. Rep.* **5**, 17041; doi: 10.1038/srep17041 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>