A sparse occupancy model to quantify species interactions in time and space

Sadoune. Ait Kaci Azzou[a,b], Thierry. Aebischer[a,b], Liam. Singer[a,b], Beat. Wolf[c], Daniel. Wegmann[a,b,*]

[a] *Department of Biology, Université de Fribourg, Chemin du Musée 10, CH-1700 Fribourg, Switzerland*
[b] *Swiss Institute of Bioinformatics, 1700 Fribourg, Switzerland*
[c] *iCoSys, University of Applied Sciences Western Switzerland, Fribourg, Switzerland*

Summary

1. Camera traps are an essential tool to quantify the distribution, abundance and behavior of mobile species. As detection probabilities vary greatly among camera trap locations, they must be accounted for when analyzing such data, which is generally done using occupancy models.
2. We introduces a Bayesian Time-dependent Occupancy Model for Camera Trap data (`Tomcat`), suited to estimate relative event densities in space and time. `Tomcat` allows to learn about the environmental requirements and daily activity patterns of species while accounting for imperfect detection. It further implements a sparse model that deals well will a large number of potentially highly correlated environmental variables.
3. By integrating both spatial and temporal information, we extend the notation of overlap coefficient between species to time and space to study niche partitioning.
4. We illustrate the power of `Tomcat` through an application to camera trap data of eight sympatrically occurring duiker species in the the savanna - rainforest ecotone in the Central African Republic and show that most species pairs show little overlap. Exceptions are those for which one species is very rare, likely as a result of direct competition.

Keywords: Daily activity, Piece-wise constant function, sparsity, Camera trap data, spatio-temporal overlap coefficient

1. Introduction

Camera traps have become an essential part of many wildlife monitoring efforts that aim at quantifying the distribution, abundance and behavior of

*Corresponding author
Email address: `daniel.wegmann@unifr.ch` (Daniel. Wegmann)

mobile species. However, the inference of these biological characteristics is not trivial due to the confounding factor of detection, which may vary greatly among camera trapping locations. Hence, variation in the rates at which a species is recorded (the photographic rate) may indeed reflect differences in local abundance, but might just as well reflect differences in the probabilities with which individuals are detected, or more likely a combination of both (see Burton et al., 2015; Sollmann, 2018, for two excellent reviews).

Since local detection rates are generally not known, both processes have to be inferred jointly. The most often used methods are variants of so-called occupancy models that treat the detection probability explicitly (MacKenzie et al., 2002). The basic quantity of interest in these models is whether or not a particular site is occupied by the focal species. While the detection of a species implies that the species is present, the absence of a record does not necessarily imply it is absent. Since the probabilities of detection and occupation are confounded, they can not be inferred for each site individually. It is therefore common to use hierarchical models that express detection probabilities as a function of environmental variables (MacKenzie et al., 2002).

Here we introduce `Tomcat`, a Time-dependent Occupancy Model for CAmera Trap data, that extends currently used occupancy models in three important ways:

First, we explicitly account for the sparsity among environmental coefficients. This is relevant since many environmental variables are generally available and it is usually not known which ones explain the variation in abundance of a species. Enforcing sparsity on the vector of coefficients avoids the problem of over-fitting in case the number of camera trap locations is smaller or on the same order as the number of environmental coefficients.

Second, we propose to quantify a measure of relative species density, namely the rate at which animals pass through a specific location, rather than occupancy. Quantifying occupancy assumes there exists a well defined patch or site that is either occupied by a species or not. The notation of a discrete patch is, however, often difficult when analyzing camera trap data of mobile species, which complicates interpretation (Efford and Dawson, 2012; Steenweg et al., 2018). In addition, summarizing camera trap data by a simple presence-absence matrix ignores the information about differences in population densities at occupied sites. Occupancy is therefore not necessarily a good surrogate for abundance (Efford and Dawson, 2012; Steenweg et al., 2018; MacKenzie and Royle, 2005), even it has been advocated for birds (MacKenzie and Nichols, 2004).

By quantifying relative densities this limitation can be overcome as we do not need to make strict assumptions about the independence of camera trap locations. However, we note that relative densities do also not allow for an absolute quantification of density because it is not possible to distinguish mobility from abundance. However, they readily allow for the comparison of densities in space and hence to identify habitat important for a particular species. We further argue that it more useful than occupancy to monitor changes in species abundances over time as for many species, changes in population size will be

reflected in the rate at which a species is detected prior to local extinction.

Finally, we extend classic occupancy models by jointly estimating daily activity patterns. Several models have been proposed to estimate such patterns from camera trap data (Frey et al., 2017), including testing for non-random distributions of trap events in predefined time-bins (Bu et al., 2016) and circular kernel density functions (Oliveira-santos et al., 2013; Rowcliffe et al., 2014), with latter allowing for the quantification of activity overlap between species (Ridout and Linkie, 2009). Jointly inferring activity patterns with relative densities allows us not only account for imperfect detection, but also extent the idea of overlap to space, shedding additional light on species interactions.

In this article, we begin by describing the proposed model in great details. We then verify its performance using extensive simulations and finally illustrate this idea by inferring spatio-temporal overlap of six Duiker species within the forest-savanna ecotone of central Africa.

2. The method

We present a Bayesian Time-dependent Occupancy Model for CAmera Trap data (`Tomcat`), suited to estimate relative event densities in space and time. Let us denote by $\Lambda_j(t)$ the rate at which a camera trap at location $j = 1, \ldots, J$ takes pictures of a particular species (or guild) at the time of the day $t \in [0, T], T = 24h$. We assume that this rate is affected by three processes: 1) the average rate $\overline{\lambda_j}$ at which individuals pass through location j, 2) the daily activity patterns $\mathcal{T}(t)$, and 3) the probability p_j with which an individual passing through location j is detected by the camera trap:

$$\Lambda_j(t) = p_j \overline{\lambda_j} \mathcal{T}(t), \quad \int_0^T \mathcal{T}(t)dt = 1.$$

The number of pictures $W_j(d, t_1, t_2)$ taken by a camera trap at location j within the interval $[t_1, t_2)$ on day d is then given by the non-homogeneous Poisson process

$$W_j(d, t_1, t_2) \sim \text{Poisson}(\Lambda_j(t_1, t_2)),$$

with intensity function

$$\Lambda_j(t_1, t_2) = p_j \overline{\lambda_j} \int_{t_1}^{t_2} \mathcal{T}(t)dt. \tag{1}$$

A common problem to occupancy models is that the parameters related to species densities and detection probabilities, $\bar{\lambda}_j$ and p_j in our case, are confounded and can not be estimated individually for each location without extra information. However, it is possible to estimate relative differences between locations using a hierarchical model. Following others (e.g. Tobler et al., 2015), we assume that both parameters are functions of covariates (e.g. the environment), and hence only attempt to learn these hierarchical parameters. Here, we use

$$\log(\overline{\lambda_j}) = a + X_j A, \tag{2}$$

$$\text{logit}(p_j) = Y_j B, \tag{3}$$

where X_j and Y_j are known (environmental) covariates at location j and a, A and B are species specific coefficients. Note that to avoid non-identifiability issues, we did not include an intercept for p_j, and hence we set the average detection probability across locations $\frac{1}{J}\sum_j^J p_j = 0.5$ (assuming X and Y have mean zero). Also, X_j and Y_j should not contain strongly correlated covariates.

2.1. Non-independent events

Another issue specific to camera traps is that not every picture is necessarily reflective of an independent observation as the same individual might trigger multiple pictures while passing (or feeding) in front of a camera trap. It is often difficult and certainly laborious to identify such recurrent events. Here we account for non-independent events by dividing the day into n_o intervals of equal length $h_o = \frac{T}{n_o}$ (o for observation), and then only consider whether or not at least one picture was taken within each interval $[c_{m-1}, c_m), m = 1, \ldots, n_o$, where $c_0 = c_M = T$. Specifically, for an interval m,

$$\mathbb{P}(W_j(d, c_{m-1}, c_m) = w) = \begin{cases} e^{-\Lambda_j(c_{m-1}, c_m)} & \text{if } w = 0 \\ 1 - e^{-\Lambda_j(c_{m-1}, c_m)} & \text{if } w > 0 \end{cases}, \tag{4}$$

where $\Lambda_j(c_{m-1}, c_m)$ is given by (1).

2.2. Daily activity patterns

Here we assume that $\mathcal{T}(t)$ is a piece-wise constant function with n_a activity intervals of equal length $h_a = \frac{T}{n_a}$ (a for activity). While activity patterns are unlikely strictly piece-wise constant, we chose this function over a combination of periodic functions (e.g. Oliveira-santos et al., 2013) as they are fit to complicated, multi-peaked distributions with fewer parameters.

Since the best tiling of the day is unknown, we allow for a species-specific shift δ such that the first interval is $[\delta, h_a + \delta)$ and the last overlaps midnight and becomes $[T - h_a + \delta, \delta)$ (Figure 1). We therefore have

$$\mathcal{T}(t) = \sum_{i=1}^{n_{h_a}} k_i \mathbb{1}_{[(i-1)h_a, ih_a)}(t - \delta), \tag{5}$$

where the indicator function $\mathbb{1}_{[t_0,t_1)}(t)$ is 1 if $t \in [t_0, t_1)$ and zero otherwise and k_i reflects the relative activity of the focal species (or guild) in interval $i = 1, \ldots, n_h$ with $k_i = 0$ implying no activity and $k_i = 1$ implying average activity. Note that

$$\int_0^T \mathcal{T}(t)dt = h_a \sum_{l=1}^{n_a} k_l = T,$$

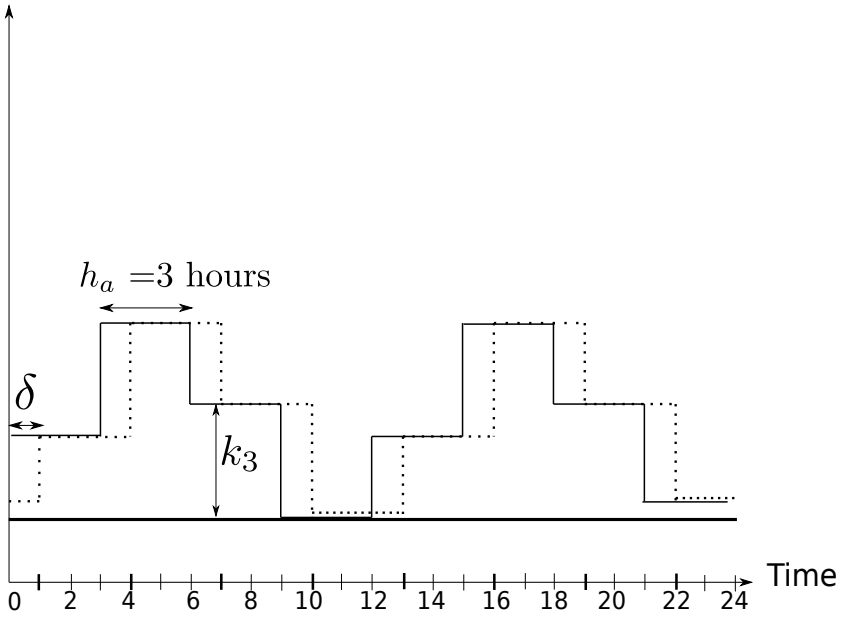

Figure 1: The solid line plot represents the piecewise-constant function $\mathcal{T}$ with $h_a = 3$ hours, and the dashed line plot represents $\mathcal{T}$ shifted with $\delta = 1$ hour.

and hence

$$\frac{1}{n_a}\sum_{l=1}^{n_a} k_l = 1.$$

2.3. Bayesian inference

We conduct Bayesian inference on the parameter vector $\boldsymbol{\theta} = \{a, A, B, \delta, \boldsymbol{k}\}$, where $\boldsymbol{k} = \{k_1, \ldots, k_{n_a}\}$, by numerically evaluating the posterior distribution $\mathbb{P}(\boldsymbol{\theta}|\boldsymbol{W}) \propto \mathbb{P}(\boldsymbol{W}|\boldsymbol{\theta})\mathbb{P}(\boldsymbol{\theta})$, where $\boldsymbol{W} = \{W_1, \ldots, W_J\}$ denotes the full data from all locations $j = 1, \ldots, J$.

The likelihood $\mathbb{P}(\boldsymbol{W}|\theta)$ is calculated as

$$\mathbb{P}(\boldsymbol{W}|\boldsymbol{\theta}) = \prod_{j=1}^{J}\prod_{d=D_{j1}}^{D_{j2}}\prod_{m=1}^{M}\mathbb{P}(W_j(d, c_{m-1}, c_m)|\boldsymbol{\theta}),$$

where $\mathbb{P}(W_j(d, c_{m-1}, c_m)|\boldsymbol{\theta})$ is given by equation (4) and the the product runs across all days $D_{j1}, \ldots, D_{j2}$ camera trap j was active.

Since it is usually not known which covariates X_j and Y_j are informative, nor at which spatial scale they should be evaluated, the potential number of covariates to be considered may be large. To render inference feasible, we enforce sparsity on the vectors of coefficients A and B. Specifically, we assume that $\mathbb{P}(A_i \neq 0) = \pi_\lambda$ and, correspondingly, $\mathbb{P}(B_i \neq 0) = \pi_p$.

We chose uniform priors on all other parameters, namely $\mathbb{P}(a) \propto 1$, $\mathbb{P}(\boldsymbol{k}) \propto 1$ for all vectors of $\boldsymbol{k}$ that satisfy $\sum_{l=1}^{n_h} k_l = n_h$, and $\mathbb{P}(\delta) \propto 1$ for all $0 \leq \delta < h_a$. For simplicity, we only consider cases in which h_a, the length of the activity intervals, is a multiple of h_o, the length of the observation intervals, and allow only for discrete $\delta \in \{0, \ldots, h_a/h_o\}$. Finally, we set $\pi_p = \pi_\lambda = 0.1$.

We use a reversible-jump MCMC algorithm (Green, 1995) to generate samples from the posterior distribution $\mathbb{P}(\boldsymbol{\theta}|\boldsymbol{W})$. The update $\boldsymbol{k} \to \boldsymbol{k}'$ is noteworthy. We begin by picking a random activity interval i and proposing a move $k_i \to k_i'$ according to a symmetric transition kernel. We then scale all other entries of $\boldsymbol{k}'$ to satisfy the constraint on the sum. Specifically, we set $k_j' = \alpha k_j$ for all $j \neq i$, where

$$\alpha = \frac{n - k_i'}{\sum_{j \neq i} k_j}.$$

2.4. Prediction

Using a set of S posterior samples $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_S \sim \mathbb{P}(\boldsymbol{\theta}|\boldsymbol{W})$, we project event densities to a not-surveyed location ι with covariates X_ι by calculating the mean $\hat{\overline{\lambda_\iota}}$ of the posterior $\mathbb{P}(\overline{\lambda_\iota}|X_\iota, \boldsymbol{W})$ as

$$\hat{\overline{\lambda_\iota}} \approx \frac{1}{S} \sum_{s=1}^{S} e^{a^{(s)} + X_\iota A^{(s)}}, \tag{6}$$

where $a^{(i)}$ and $A^{(i)}$ denote the i-th posterior sample of these parameters.

2.5. Species overlap in space and time

An important interest in ecology is to compare activity patterns among species and to see how overlapping patterns may relate to competition or predation (e.g. Ridout and Linkie, 2009; Rowcliffe et al., 2014).

We can quantify overlapping patterns of animal activity by estimating the coefficient of overlap Δ (Ridout and Linkie, 2009). This quantitative measure ranges from 0 (no overlap) to 1 (identical activity patterns) and is the area lying under two activity density curves (see Figure 4). For two known density functions $f(x)$ and $g(x)$, Δ is given by:

$$\Delta(f, g) = \int \min\{f(x), g(x)\} \tag{7}$$

The overlap measure $\Delta(f, g)$ can be related to the well known measure of distance between two densities L_1 as

$$\Delta(f, g) = 1 - \frac{1}{2} \int |f(x) - g(x)| dx, \tag{8}$$

which justifies the visualization of overlap coefficients between k species in a n-dimensional space using a Multidimensional Scaling (MDS) by considering $1 - \hat{\Delta}(f, g)$ as a measure of dissimilarity.

In practice, the true density functions $f(x)$ and $g(x)$ are usually not known. Here we obtain an estimate of Δ numerically from posterior samples. We distinguish three types of overlap coefficients, Δ_T for overlap in time, Δ_S for overlap in space and Δ_{ST} for overlap in time and space.

Overlap coefficient $\mathbf{\Delta_T}$. For a large number n_T of equally spaced time values $t_1, t_2, \ldots, t_{n_T} \in (0, T)$, we sample $\Delta_T^{(m)}$ from the posterior distribution $\mathbb{P}(\Delta_T|\boldsymbol{W}^{(1)}, \boldsymbol{W}^{(2)})$ where $\boldsymbol{W}^{(l)} = \{W_1^{(l)}, \cdots, W_J^{(l)}\}$ denotes the full data for a species $l = 1, 2$.

$$\widehat{\Delta}_T^{(m)} = \frac{1}{n_T} \sum_{i=1}^{n_T} \min\left\{\mathcal{T}_1^{(m)}(t_i), \mathcal{T}_2^{(m)}(t_i)\right\}, \tag{9}$$

where $\mathcal{T}_l^{(m)}$ is computed according to equation (5) with species specific parameters $\delta_l^{(m)}$ and $\boldsymbol{k}_l^{(m)} = (k_{1,l}^{(m)}, k_{2,l}^{(m)}, \cdots, k_{n_{h_a},l}^{(m)})$ sampled from $\mathbb{P}(\boldsymbol{\theta}_l|\boldsymbol{W}^{(l)})$.

Overlap coefficient $\mathbf{\Delta_S}$. For a given number n_S of sites reflecting the habitat in a region, we sample $\Delta_S^{(s)}$ from the posterior distribution $\mathbb{P}(\Delta_S|\boldsymbol{W}^{(1)}, \boldsymbol{W}^{(2)})$ as

$$\widehat{\Delta}_S^{(s)} = \frac{1}{n_S} \sum_{j=1}^{n_S} \min\left\{\overline{\lambda}_{1j}^{(s)}, \overline{\lambda}_{2j}^{(s)}\right\}, \tag{10}$$

where $\overline{\lambda}_{lj}^{(s)}, l = 1, 2$ is computed according to equation (2) and normalized such as $\sum_{j=1}^{n_S} \overline{\lambda}_{lj}^{(s)} = 1$ with species specific parameters $a_l^{(s)}$ and $A_l^{(s)}$ sampled from $\mathbb{P}(\boldsymbol{\theta}_l|\boldsymbol{W}^{(l)})$.

Overlap coefficient $\mathbf{\Delta_{ST}}$. For n_T time values and n_S number of sites, we sample $\Delta_{ST}^{(f)}$ from the posterior distribution $\mathbb{P}(\Delta_{ST}|\boldsymbol{W}^{(1)}, \boldsymbol{W}^{(2)})$ as

$$\widehat{\Delta}_{ST}^{(f)} = \frac{1}{n_T \times n_S} \sum_{i=1}^{n_T} \sum_{j=1}^{n_S} \min\left\{\overline{\lambda}_{1j}^{(f)} \mathcal{T}_1^{(f)}(t_i), \overline{\lambda}_{2j}^{(f)} \mathcal{T}_2^{(f)}(t_i)\right\}, \tag{11}$$

where for species $l = 1, 2$ we calculate $\mathcal{T}_l^{(f)}$ according to equation (5) and $\overline{\lambda}_{lj}^{(f)}$ according to equation (2) with species specific parameters $a_l^{(f)}$, $A_l^{(f)}$, and $\boldsymbol{k}_l^{(f)}$ sampled from $\mathbb{P}(\boldsymbol{\theta}_l|\boldsymbol{W}^{(l)})$, but normalized such that

$$\sum_{i=1}^{n_T} \sum_{j=1}^{n_S} \overline{\lambda}_{1j}^{(f)} \mathcal{T}_1^{(f)}(t_i) = 1.$$

Implementation

All methods were implemented in the `C++` program `Tomcat`, available through a git repository at `https://bitbucket.org/WegmannLab/tomcat/`.

3. Performance against simulations

We assessed the performance of our algorithm using 100 replicates for each combination of $J = 20$ or 100 camera trap locations and $D = 1, 2, 5, 10, 20, 50$ or 100 days at which data was collected. All simulations were conduced with one-dimensional $X_j \sim N(0, 1)$ and $Y_j \sim N(0, 1)$ and parameter choices such that on average one picture per species per location was expected, and hence the expected number of pictures per species was JD.

To evaluate the accuracy of our estimates, we then estimated the overlap between two species since errors in parameter estimates directly translate into biases in overlap coefficients. We thus simulated data for two species with little ($\Delta_T = 0.2$), moderate overlap ($\Delta_T = 0.5$) or large overlap in time ($\Delta_T = 0.8$) as described in Table 1, as well as for two species with varying overlap in space ($\Delta_S = 0.2$, 0.5, and 0.8) and for with varying overlap in space and time ($\Delta_{ST} = 0.2$, 0.5, and 0.8) as described in the Appendix.

As shown in Figure 2, the posterior means $\widehat{\Delta_T}$, $\widehat{\Delta_S}$ and $\widehat{\Delta_{ST}}$ were unbiased and highly accurate for all overlap coefficients if sufficient data is provided, i.e. if at least several hundred pictures were available ($J \times D \geq 500$). If less data was available, estimates were biased towards the prior expectations of $\Delta_T = 0.5$ and $\Delta_S = 1$.

4. Application to central African duikers

We applied `Tomcat` to camera trapping data obtained during the dry seasons from 2012 to 2018 from a region in the Eastern Central African Republic (CAR), a wilderness exceeding 100000 km^2 without permanent settlements, agriculture or commercial logging (Aebischer et al., 2017). The Eastern CAR consists of an ecotone of tropical moist closed canopy forests of the Northeastern Congolian lowland rain forest biome and a Sudanian-Guinean woodland savanna that is interspersed with small patches of edaphic grasslands on rocky ground or swampy areas (Boulvert, 1985; Olson and Dinerstein, 1998).

The available data was from 532 locations that cover the Chinko Nature Reserve (CNR), a protected area of about 20000 km^2 that was established in 2014 by the government of the CAR in former hunting zones. Here, we use `Tomcat` to study duikers (Cephalophinae), which are a diverse mammalian group common in the data set and observed often in sympatry, i.e. several species were captured by the same camera trap within a few hours. We detected a total of eight species in the data set (Table 2): *Cephalophus dorsalis castaneus* (Eastern Bay Duiker), *Cephalophus leucogaster arrhenii* (Uele White Bellied Duiker), *Cephalophus nigrifrons* (Black Fronted Duiker), *Cephalophus rufilatus* (Red Flanked Duiker), *Cephalophus silvicultor castaneus* (Western Yellow Backed Duiker), *Cephalophus weynsi* (Weyns Duiker), *Philantomba monticola aequatorialis* (Eastern Blue Duiker) and *Sylvicapra grimmia* (Bush Duiker).

To infer habitat preferences for these species, we benefited from an existing land cover classification at a 30m resolution that represents the five major habitat types of the Chinko region: Closed Canopy Forest (CCF), Open Savanna

Woodland (OSW), Dry Lakr Grassland (DLG), Wet Marshy Grassland (WMG) and Surface Water (SWA) (Aebischer et al., 2017). Around every camera trap location and 10,200 regular grid points spaced 2.5 km apart and spanning the entire CNR, we calculated the percentage of each of these habitats in 11 buffers of sizes 30, 65, 125, 180, 400, 565, 1260, 1785, 3,990, 5,640 and 17,840 meters. We complemented this information with the average value within every buffer for each of 15 additional environmental and bioclimatic variables from the WorldClim database version 2 (Table E.1 Fick and Hijmans, 2017) that we obtained at a resolution of 30 seconds, which translates into a spatial resolution of roughly 1km^2 per grid cell. To aid in the interpretation, we then processed our environmental data by 1) keeping only the additional effect of each variables after regressing out the habitat variables CCF and OSW at the same buffer, and by 2) keeping only the additional effect of every variable after regressing out the information contained in the same variable but at smaller buffers (see Appendix for details).

To avoid extrapolation, we restricted our analyses to 2,639 grid locations that exhibited similar environments to those at which camera traps were placed as measured by the Mahanalobis distance between each grid point and the average across all camera trap locations (see Appendix for details).

For each location we further used the binary classification of the four most common habitat types (CCF, OSW, MWG, DLG) and determined the presence or absence of six additional habitat characteristics: Animal path, road, salt lick, mud hole, riverine zone and bonanza.

The eight duiker species varied greatly in both their habitat preferences (Figures 3, S.2) and their daily activity patterns (Figures 4, S.1) as inferred by `Tomcat`). As shown in Figure 3, *C. dorsalis* and *C. weynsi* have both a strong preference for CCF over OSW habitat at the smallest buffers, in contrast to *S. grimmia* that shows a string preference of OSW. At higher buffers, the signal is less clear, probably owing to the heterogeneous nature of the habitat in which both CCF and OSW correlated negatively with WMG and DLG, habitats not well suited for all these species. Interestingly, two species (*P. monticola* and *C. silvicultor*) also seem to be true ecotome species preferring a mixture of the canonical habitats CCF and OSW (Figure S.2). Similarly, and as shown in Figure 4, some species appear to be almost exclusively nocturnal (*C. dorsalis* and *C. silvicultor*), some almost exclusively diurnal (*C. leucogaster*, *C. monticola*, *C. nigrifons*, *C. rufilatus*, *C. weynsi*) and one crepuscular (*S. grimmia*).

To better understand how these closely related duiker species of similar size and nutrition can occur sympatrically, we estimated pairwise overlap coefficients in space and time (Figure 4, Table E.2). Not surprisingly, most species pairs differed substantially either in their habitat preference of daily activity patterns. Of the two forest dwellers *C. dorsalis* and *C. weynsi* ($\widehat{\Delta_S} = 0.38$), for instance, one is almost exclusively nocturnal and the other almost exclusively diurnal ($\widehat{\Delta_T} = 0.22$), resulting in a small overlap in space and time ($\widehat{\Delta_{ST}} = 0.11$). Similarly, the nocturnal *C. dorsalis* and the crepuscular *S. grimmia* that share a lot of temporal overlap ($\widehat{\Delta_T} = 0.22$) use highly dissimilar habitats ($\widehat{\Delta_S} = 0.02$),

resulting in a very small overlap in time and space ($\widehat{\Delta_{ST}} = 0.01$).

A visualization using Multidimensional Scaling (MDS) of the pair-wise overlap coefficients of all six species with events from at least 50 independent camera trap locations is shown in Figure 5. For these species, 88.3% of variation in the temporal overlap can be explained by a single axis separating nocturnal from diurnal species. In contrast, only 45.6% of the variation in the spatial overlap is explained by the first axis distinguishing forest dwellers from savanna species.

When using both temporal and spatial information, it is striking that frequently observed and therefore evidently abundant species within a certain community tend to differ in their habitat preference and/or daily activity. In contrast, infrequently observed and therefore putative rare taxa seem to have large overlap with co-occurring species. The Uele white-bellied duiker (*C. leucogaster*), for instance, which is rather rare and was only observed at eleven distinct locations (Table 2), depends on similar food and is active at the same time as the Weyns duiker (*C. weynsi*), which is among the most common forest duikers within the CNR. In contrast, the Eastern Bay duikers (*C. dorsalis*), which is strictly nocturnal, seems to co-exist with the Weyns duikers at higher densities (Figure (4).

Conclusion

Despite a world-effort to assess biodiversity, there are still major areas for which almost no information is available on biodiversity (e.g. Hickisch et al., 2019). But several technological advances, and in particular camera traps and voice recorders, make it possible to obtain a first glimpse on the presence and distribution of larger or highly vocal animals such as mammals and birds in relatively short time with a reasonable budget. Thanks to technical advances, increased battery life and larger media to store data, such data sets can now be produced with comparatively little manpower, even under the demanding conditions in large and remote areas. In addition, the annotation of such data sets on the species level is now aided by machine learning algorithms that automatize the detection of at least common species and images without animals (e.g. **?**). Thanks to these developments, existing knowledge gaps may now be increasingly addressed, allowing for a re-evaluation of conservation strategies and optimization of conservation management.

Here we introduce `Tomcat`, a occupancy model that infers habitat preference and daily activities from such data sets. Unlike many previous methods that estimate the presence or absence of a species, `Tomcat` estimates relative species densities, from which overlap coefficients between species can be estimated in space and time, while accounting for variation in detection probabilities between locations. While estimates of overlap coefficients require larger data sets than the inference of pure occupancy, we believe they constitute a major step forward in understanding the complex species interactions in an area, which are particularly relevant for conservation planning in heterogeneous or fragmented habitat.

References

Aebischer, T., Siguindo, G., Rochat, E., 2017. First quantitative survey delineates the distribution of chimpanzees in the eastern central african republic. Biological Conservation 213, 84–94.

Boulvert, Y., 1985. Carte phytogographique de la rpublique centrafricaine. ORSTOM (Office de la recherche scientifique et technique Outre-Mer) .

Bu, H., Wang, F., Mcshea, W.J., Lu, Z., Wang, D., Li, S., 2016. Spatial Co-Occurrence and Activity Patterns of Mesocarnivores in the Temperate Forests of Southwest China. PLOS ONE , 1–15doi:10.1371/journal.pone.0164271.

Burton, A.C., Neilson, E., Moreira, D., Ladle, A., Steenweg, R., Fisher, J.T., Bayne, E., Boutin, S., 2015. Review: Wildlife camera trapping: a review and recommendations for linking surveys to ecological processes. Journal of Applied Ecology 52, 675–685. doi:10.1111/1365-2664.12432.

Efford, M.G., Dawson, D.K., 2012. Occupancy in continuous habitat. Ecosphere 3, art32. doi:10.1890/ES11-00308.1.

Fick, S.E., Hijmans, R.J., 2017. Worldclim 2: new 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology 37, 4302–4315. doi:10.1002/joc.5086.

Frey, S., Fisher, J.T., Burton, A.C., Volpe, J.P., 2017. Investigating animal activity patterns and temporal niche partitioning using camera-trap data: challenges and opportunities. Remote Sensing in Ecology and Conservation 3, 123–132. doi:10.1002/rse2.60.

Green, P., 1995. Reversible jump markov chain monte carlo computation and bayesian model determination. Biometrika 82, 711–732.

Hickisch, R., Hodgetts, T., Johnson, P.J., Sillero-Zubiri, C., Tockner, K., Macdonald, D.W., 2019. Effects of publication bias on conservation planning. Conservation Biology 33, 1151–1163. doi:10.1111/cobi.13326.

MacKenzie, D.I., Nichols, J.D., 2004. Occupancy as a surrogate for abundance estimation. Animal biodiversity and conservation 27, 461–467.

MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Andrew Royle, J., Langtimm, C.A., 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology 83, 2248–2255. doi:10.1890/0012-9658(2002)083[2248:ESORWD]2.0.CO;2.

MacKenzie, D.I., Royle, J.A., 2005. Designing occupancy studies: general advice and allocating survey effort. Journal of Applied Ecology 42, 1105–1114. doi:10.1111/j.1365-2664.2005.01098.x.

Oliveira-santos, L.G.R., Zucco, C.A., Agostinelli, C., 2013. Using conditional circular kernel density functions to test hypotheses on animal circadian activity. Animal Behaviour 85, 269–280. URL: http://dx.doi.org/10.1016/j.anbehav.2012.09.033, doi:10.1016/j.anbehav.2012.09.033.

Olson, D.M., Dinerstein, E., 1998. The global 200: A representation approach to conserving the earth's most biologically valuable ecoregions. Conserv. Biol. 12, 502–515.

Ridout, M.S., Linkie, M., 2009. Estimating overlap of daily activity patterns from camera trap data. Journal of Agricultural, Biological, and Environmental Statistics 14, 322–337. URL: https://doi.org/10.1198/jabes.2009.08038, doi:10.1198/jabes.2009.08038.

Rowcliffe, J.M., Kays, R., Kranstauber, B., Carbone, C., Jansen, P.A., 2014. Quantifying levels of animal activity using camera trap data. Methods in Ecology and Evolution 5, 1170–1179.

Sollmann, R., 2018. A gentle introduction to camera-trap data analysis. African Journal of Ecology 56, 740–749. doi:10.1111/aje.12557.

Steenweg, R., Hebblewhite, M., Whittington, J., Lukacs, P., McKelvey, K., 2018. Sampling scales define occupancy and underlying occupancyabundance relationships in animals. Ecology 99, 172–183. doi:10.1002/ecy.2054.

Tobler, M.W., Hartley, A., CarrilloPercastegui, S.E., Powell, G.V.N., 2015. Spatiotemporal hierarchical modelling of species richness and occupancy using camera trap data. Journal of Applied Ecology 43, 413–421.

Tables

Table 1: Values of $\boldsymbol{k}$ used for the simulation of the daily activity patterns $\mathcal{T}(t)$.

Δ_T	Species	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8
0.2	1	0.2	0.2	1.8	1.8	1.8	1.8	0.2	0.2
	2	1.8	1.8	0.2	0.2	0.2	0.2	1.8	1.8
0.5	1	0.5	0.5	1.5	1.5	1.5	1.5	0.5	0.5
	2	1.5	1.5	0.5	0.5	0.5	0.5	1.5	1.5
0.8	1	0.8	0.8	1.2	1.2	1.2	1.2	0.8	0.8
	2	1.2	1.2	0.8	0.8	0.8	0.8	1.2	1.2

Table 2: Available data on the eight detected species of duikers.

Species	Pictures	Locations
Cephalophus dorsalis castaneus	1,631	66
Cephalophus leucogaster arrhenii	432	11
Cephalophus nigrifrons	102	7
Cephalophus rufilatus	5,762	168
Cephalophus silvicultor castaneus	10,321	222
Cephalophus weynsi	10,037	146
Philantomba monticola aequatorialis	50,979	212
Sylvicapra grimmia	5,626	124

Figures

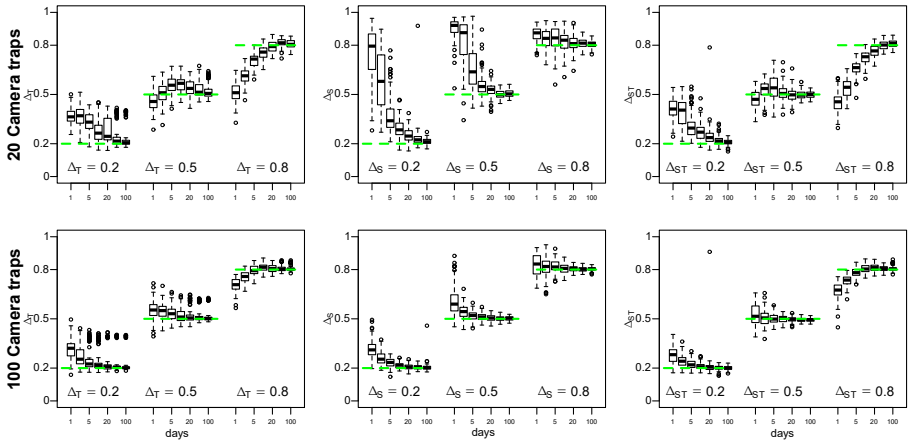

Figure 2: Distribution of bias in the estimated overlap coefficients $\hat{\Delta}_T$, $\hat{\Delta}_S$ and $\hat{\Delta}_{ST}$ for different sample sizes.

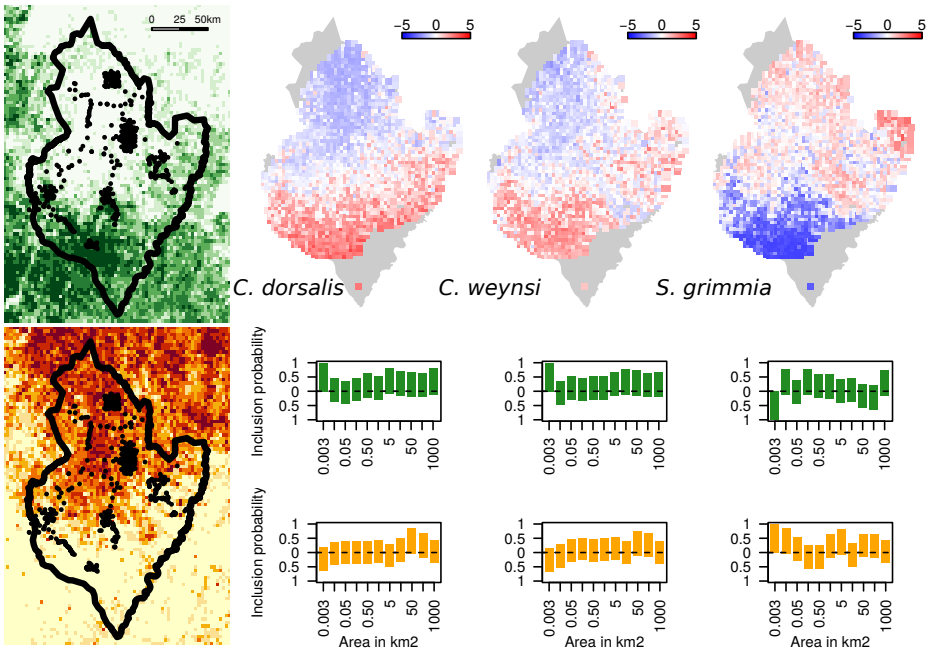

Figure 3: Habitat preference of the three duiker species *C. dorsalis*, *C. weynsi* and *S. grimmia*. a) distribution of close canopy forest (CCF, top, green) and open woodland savanna (OSW, bottom, yellow) across the study region with the CNR borders and camera trap locations (black dots). b) Relative densities d_{sj} of the three duikers predicted at 2,639 grid points. For each species the colors indicates $d_{sj} = \log_{10}\left(\bar{\lambda}_{sj}/\text{median}(\bar{\lambda}_s)\right)$, where $\text{median}(\bar{\lambda}_s)$ is the median value over all the grid points j. Red shades indicate $d_{sj} > 0$, blue shades $d_{sj} < 0$. c) Posterior inclusion probabilities for the CCF (green) and OSW (yellow) habitat variables for each buffer. Values above the dashed line indicate the posterior probability that the habitat correlates positively with the relative species density, values below the dashed line imply a negative correlation.

Appendix A. Simulating data with specific overlap coefficients

Simulating data with specific Δ_S

For the simulation of scenarios with a given Δ_S, we have simulated for each site j, an environmental variable X_j from $N(0,1)$. Δ_S will be estimated with $\hat{\Delta}_S$, where

$$\hat{\Delta}_S = \mathbb{E}\left(\min\left(\bar{\lambda}_1(X), \bar{\lambda}_2(X)\right)\right),$$

where $\bar{\lambda}_i = \{\bar{\lambda}_{i1}, \cdots, \bar{\lambda}_{ij}, \cdots, \bar{\lambda}_{in}\}$ in presence of n sites. For a species i and site j, $\bar{\lambda}_{ij}$ is given by equation (2), with the constraint :

$$\int \bar{\lambda}_{ij}(x_j) f_{X_j}(x_j) dx_j = 1. \tag{S.1}$$

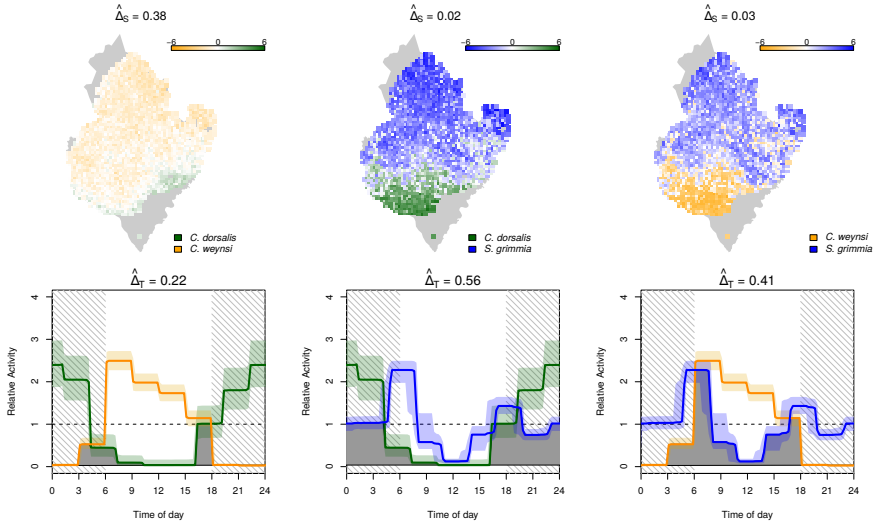

Figure 4: Interaction in space and time between the duiker species *C. dorsalis*, *C. weynsi*, and *S. grimma*. Top row: interactions in space quantified as $\log_{10}\left(\bar{\lambda}_{s_1 j}/\bar{\lambda}_{s_2 j}\right)$ between species 1 and 2. Bottom: posterior mean (solid line) and 90% credible intervals (shades) of temporal activity patterns. The area shaded in gray represents the overlap coefficient Δ_T

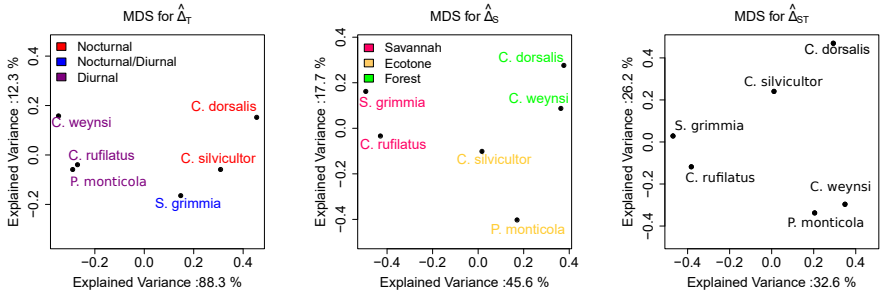

Figure 5: Illustration of the overlap coefficients in time and space between six duiker species visualized in two dimensions using the multidimensional scaling.

We have for species 1 :

$$\begin{aligned} (\text{S.1}) \quad & \Leftrightarrow \int_{-\infty}^{+\infty} \exp(\mu_1 + A_1 x_j) \times \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_j^2} dx_j = 1 \\ & \Leftrightarrow \exp(\mu_1 + \tfrac{1}{2}A_1^2) = 1 \\ & \Leftrightarrow \mu_1 = -\tfrac{1}{2}A_1^2. \end{aligned} \tag{S.2}$$

Equivalently, we have for species 2, $\mu_2 = -\frac{1}{2}A_2^2$. Therefore,

$$\bar{\lambda}_{i1} = \exp(-\frac{1}{2}A_i^2 + A_i x_j),\ i = 1, 2. \tag{S.3}$$

We have

$$\hat{\Delta}_S = \int \min\left(\bar{\lambda}_1(x_j), \bar{\lambda}_2(x_j)\right) f_{X_j}(x_j) dx_j. \tag{S.4}$$

We have

$$\begin{aligned} \bar{\lambda}_1(x_j) < \bar{\lambda}_2(x_j) &\Leftrightarrow \exp(-\tfrac{1}{2}A_1^2 + A_1 x_j) < \exp(-\tfrac{1}{2}A_2^2 + A_2 x_j) \\ &\Leftrightarrow x_j < \tfrac{A_1+A_2}{2} = B. \end{aligned}$$

The integral in equation (S.4) is given by:

$$\begin{aligned} \hat{\Delta}_S &= \int_{-\infty}^{B} \exp(-\tfrac{1}{2}A_1^2 + A_1 x_j) \times \tfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_j^2} dx_j + \int_{B}^{+\infty} \exp(-\tfrac{1}{2}A_2^2 + A_2 x_j) \times \tfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_j^2} dx_j \\ &= \Phi(\tfrac{1}{2}(A_2 - A_1)) + (1 - \Phi(\tfrac{1}{2}(A_1 - A_2))) \\ &= 2 - 2\Phi(\tfrac{1}{2}(A_1 - A_2)), \end{aligned}$$

where Φ represents the cumulative distribution function (CDF) of the standard normal distribution.

Finally, we have

$$\frac{2 - \hat{\Delta}_S}{2} = \Phi(\frac{1}{2}(A_1 - A_2)) \tag{S.5}$$

It is possible using equation (S.5) to simulate a model for which $\Delta_S = a$.

For example, for $\Delta_S = \delta_s$, and for $A_1 = a_1$, it is possible to get the value of A_2 from (S.5), which gives

$$\Phi^{-1}(\frac{2 - \delta_S}{2}) = \frac{1}{2}(a_1 - A_2) \Leftrightarrow A_2 = a_1 - 2\Phi^{-1}(\frac{2 - \delta_S}{2}). \tag{S.6}$$

After computing A_2, it is easy to deduce the value of μ_2 using equation (S.2).

Simulating data with specific Δ_{ST}

To simulate scenarios for a given Δ_{ST}, we have simulated for each site j an environmental variable X_j from $N(0,1)$, and $T \sim U_{[0,24]}$. Δ_{ST} is estimated using equation (11).

For a species i and site j, we have the constraint

$$\begin{aligned} \int_{-\infty}^{+\infty} \int_{0}^{24} \bar{\lambda}_{ij}(x_j) \mathcal{T}_i(t) f_T(t) f_{X_j}(x_j) dt\, dx_j &= \mathbb{E}\left(\mathcal{T}_i(T) \times \bar{\lambda}_{ij}(X_j)\right) = 1, \\ &= \mathbb{E}\left(\mathcal{T}_i(T)\right) \times \mathbb{E}\left(\bar{\lambda}_{ij}(X_j)\right), \end{aligned}$$

where $\mathcal{T}_i(t)$ is the daily activity for a species i, $i = 1, 2$ observed at time t.

We have

$$\begin{array}{rl} \mathbb{E}\left(\mathcal{T}_i(T)\right) & = \int_0^{24} \mathcal{T}_i(t) t_T(t) dt \\ & = \frac{1}{24} \int_0^{24} \mathcal{T}_i(t) dt \\ & = \frac{1}{24} \int_0^{24} \sum_{l=1}^{n_a} K_l \mathbb{1}_{[(l-1)h_a,\, l\times h_a]} dt \\ & = \frac{1}{24} h_a \sum_{l=1}^{n_a} K_l \\ & = \frac{1}{24} h_a \times n_a = 1 \end{array} ,$$

and

$$\begin{array}{rl} \mathbb{E}\left(\bar{\lambda}_{ij}(X_j)\right) & = \int_{-\infty}^{+\infty} \exp(\mu_i + A_i x_j) f_{X_j}(x_j) dx_j \\ & = \int_{-\infty}^{+\infty} \exp(\mu_i + A_i x_j) \times \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_j^2} dx_j \\ & \Leftrightarrow \exp(\mu_i + \frac{1}{2}A_i^2) \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_j - A_i)^2} dx_j \end{array} ,$$

which gives

$$\mathbb{E}\left(\mathcal{T}_i(T)\right) \times \mathbb{E}\left(\bar{\lambda}_{ij}(X_j)\right) = 1 \Leftrightarrow \mu_i = -\frac{1}{2}A_i^2. \tag{S.7}$$

We can therefore simulate the scenario $\Delta_{ST} = \delta_{ST}$ with known activity patterns $\mathcal{T}_1$, $\mathcal{T}_2$ respectively for species 1 and 2 as follows:

1. Simulate n_s $X_j \sim N(0,1)$, $j = 1, 2, \cdots n_s$.
2. Propose a value of A_1. (Should not be very large. Namely between -2 and 2).
3. Compute the value of μ_1 using equations (S.7).
4. Compute numerically the value of A_2 by solving the equation :

$$\delta_{ST} - \frac{1}{n_T \times n_S} \sum_{i=1}^{n_T} \sum_{j=1}^{n_S} \min\left\{\bar{\lambda}_{1j}^{(f)} \mathcal{T}_1^{(f)}(t_i), \bar{\lambda}_{2j}^{(f)} \mathcal{T}_2^{(f)}(t_i)\right\} = 0. \tag{S.8}$$

5. Using `Tomcat` for the two species, and by fixing the value A_i, $i = 1, 2$, we can sample from the posterior distribution of Δ_{ST} and compute the posterior mean of the values Δ_{ST} using equation (11).

Appendix B. Decorrelation environmental variables

While `Tomcat` readily handles correlated environmental variables, we chose to decorrelate specific variables to aid in interpretation. Specifically, we processed our environmental data as follow (two steps):

Step 1: A major interest in our application was to study the impact of the prevalence of close canopy forest (f) and savanna (s) habitat on species densities. For each scale (buffer) b, we therefore regress each environmental variable $V_{ib}, i \neq f, s$:

$$V_{ib} = \alpha_{fb} V_{fb} + \alpha_{sb} V_{sb} + \epsilon_{ib}, \tag{S.1}$$

where V_{fb} and V_{sb} represents, respectively, the forest and the Savannah habitat for a buffer $b = 1, \cdots, B$, $i = 1, \cdots, n_{env}$, and ϵ_{ib} is the error from the linear model described by equation (S.1), which captures the information of the variable V_{ib} independent of V_{fb} and V_{sb} at buffer b. We therefore replace the V_{ib} variables by $\widetilde{V}_{ib} = \epsilon_{ib}$ in our model, but kept $\widetilde{V}_{fb} = V_{fb}$ and $\widetilde{V}_{sb} = V_{sb}$.

Step 2: To evaluate relevant spatial scale of environmental variables, we also regressed out the larger buffers from the smaller one as:

$$\widetilde{V}_{ib} = \sum_{j=1}^{b-1} \beta_j \widetilde{V}_{i,j} + e_{ib}, \tag{S.2}$$

In the second step, and for a given environmental variable $\widetilde{V}_{ib}$ at the scale, we give priority to the smaller scales by keeping only the additional explanation by the studied variable to what we already know by replacing $\widetilde{V}_{ib}$ by $\widetilde{\widetilde{V}}_{ib} = e_{ib}$.

Appendix C. Restricting analysis to environmentally homogeneous regions

Let $\overrightarrow{X} = (x_1, x_2, \cdots, x_n)^T$ be a matrix where the rows represent the observed camera traps, and the columns the environmental variables, and let $\overrightarrow{Y} = (y_1, \cdots, y_m)^T$ a matrix containing the environmental variables for the locations for which we want to predict $\bar{\lambda}_s$, $s = 1, 2, \cdots, m$.

We define the Mahanalobis distance D_O of $\overrightarrow{X}$ given by

$$D_O = \sqrt{(\overrightarrow{X} - \overrightarrow{\mu})^T S (\overrightarrow{X} - \overrightarrow{\mu})},$$

where $\overrightarrow{\mu} = (\mu_1, \cdots, \mu_n)$, and S is the variance covariance matrix.
We define a second Mahalanobis distance D_P which measures the distance of $\overrightarrow{Y}$ from the the mean $\overrightarrow{\mu}$ of the camera traps environmental variables. D_P is given by

$$D_P = \sqrt{(\overrightarrow{Y} - \overrightarrow{\mu})^T S (\overrightarrow{Y} - \overrightarrow{\mu})}.$$

After computing D_O and D_P, we decided to remove the grid points for which we have $D_P > 1.2 \times \max(D_O)$.

Appendix D. Supplementary figures

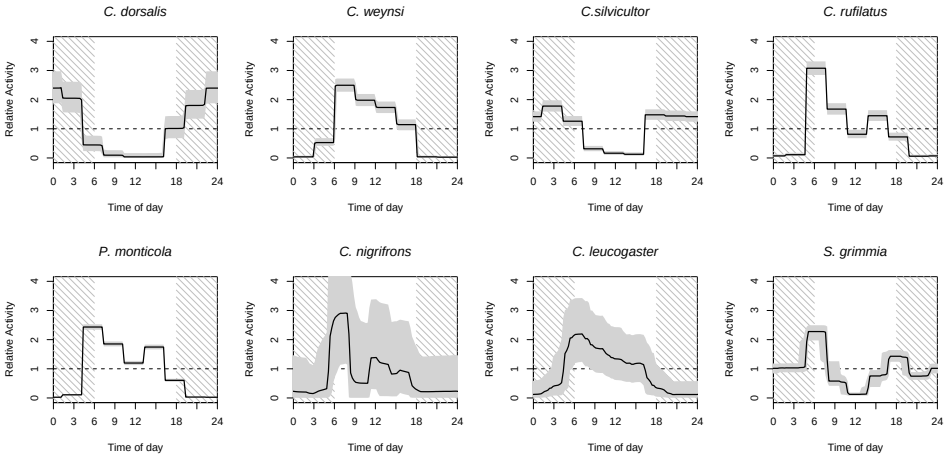

Figure S.1: Estimation of the daily activity of the 8 observed duickers. The solid line represents the posterior mean and the shades, 90% credible intervals of the temporal activity patterns.

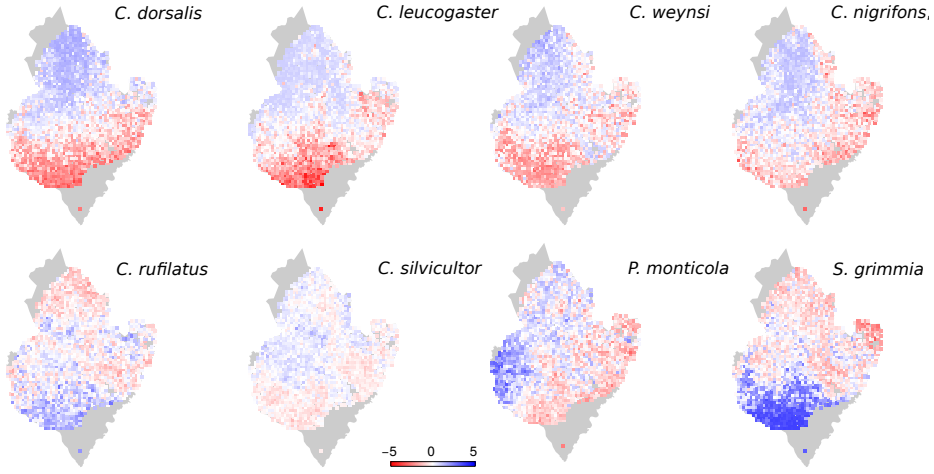

Figure S.2: Relative densities d_{sj} of the eight duikers predicted at 2,639 grid points. For each species the colors indicates $d_{sj} = \log_{10}\left(\bar{\lambda}_{sj}/\text{median}(\bar{\lambda}_s)\right)$, where $\text{median}(\bar{\lambda}_s)$ is the median value over all the grid points j. Red shades indicate $d_{sj} > 0$, blue shades $d_{sj} < 0$.

Appendix E. Supplementary tables

Table E.1: Environmental and bioclimatic variables obtained from the WorldClim database version 2.

Variable name	Description
dem	Digital elevation model
slope	Slope of a point
aspect	Direction of slope
TRI	Terrain Ruggedness Index
TWI	Topographic Wetness Index
wind	wind speed (m s^{-1})
bio2	Mean Diurnal Range (Mean of monthly (max temp - min temp))
bio5	Max Temperature of Warmest Month
bio6	Min Temperature of Coldest Month
bio12	Annual Precipitation
bio15	Precipitation Seasonality (Coefficient of Variation)
NDVI	Normalized Difference Vegetation Index
GC	General Curvature
MPI	Morphometric Protection Index
srad	Solar radiation (kJ m^{-2} day^{-1})

Table E.2: Summary of the overlap coefficients between the six duiker species.

	C. dorsalis	*C. leucogaster*	*C. nigrifeons*	*C. rufilatus*	*C. silvicultor*	*C. weynsi*	*P. monticola*	*S. grimmia*
				$\bar{\Delta}_T$: Overlap coefficient in time				
C. dorsalis	0	0.249	0.269	0.244	0.786	0.216	0.224	0.562
C. leucogaster	0.249	0	0.623	0.681	0.398	0.673	0.741	0.584
C. nigrifrons	0.269	0.623	0	0.527	0.403	0.515	0.576	0.551
C. rufilatus	0.244	0.681	0.527	0	0.431	0.789	0.871	0.535
C. silvicultor	0.786	0.398	0.403	0.431	0	0.300	0.421	0.753
C. weynsi	0.216	0.672	0.515	0.789	0.300	0	0.849	0.412
P. monticola	0.224	0.741	0.576	0.871	0.421	0.849	0	0.550
S. grimmia	0.562	0.584	0.551	0.536	0.753	0.412	0.550	0
				$\bar{\Delta}_S$: Overlap coefficient in space				
C. dorsalis	0	0.175	0.145	0.049	0.282	0.384	0.246	0.017
C. leucogaster	0.175	0	0.049	0.0144	0.104	0.167	0.090	0.010
C. nigrifrons	0.145	0.0486	0	0.076	0.166	0.176	0.194	0.050
C. rufilatus	0.049	0.0147	0.076	0	0.315	0.068	0.134	0.308
C. silvicultor	0.282	0.104	0.166	0.315	0	0.360	0.350	0.207
C. weynsi	0.384	0.167	0.176	0.068	0.360	0	0.294	0.030
P. monticola	0.246	0.090	0.194	0.134	0.350	0.294	0	0.100
S. grimmia	0.017	0.010	0.050	0.308	0.207	0.030	0.100	0
				$\bar{\Delta}_{ST}$: Overlap coefficient in space and time				
C. dorsalis	0	0.083	0.070	0.023	0.264	0.112	0.090	0.014
C. leucogaster	0.083	0	0.041	0.013	0.070	0.146	0.080	0.005
C. nigrifrons	0.070	0.041	0	0.059	0.106	0.129	0.151	0.039
C. rufilatus	0.023	0.013	0.059	0	0.178	0.066	0.132	0.223
C. silvicultor	0.264	0.070	0.106	0.178	0	0.161	0.185	0.194
C. weynsi	0.112	0.146	0.129	0.066	0.161	0	0.285	0.022
P. monticola	0.090	0.080	0.151	0.132	0.185	0.285	0	0.075
S. grimmia	0.014	0.005	0.039	0.223	0.194	0.022	0.075	0