Managing and Analysing Camera Trapping Data:

An Advanced Toolbox

Inaugural-Dissertation
to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry and Pharmacy
of Freie Universität Berlin

by

Jürgen Niedballa

from Dresden, Germany

Berlin, 2017

Diese Dissertation wurde am Leibniz-Institut für Zoo- und Wildtierforschung in Berlin im Zeitraum vom September 2013 bis September 2017 unter der Leitung von Dr. Andreas Wilting und Prof. Dr. Heribert Hofer angefertigt und am Institut für Biologie der Freien Universität Berlin eingereicht.

1. Gutachter: Prof. Dr. Heribert Hofer

2. Gutachter Prof. Dr. Florian Jeltsch

Disputation am 13.03.2018

This dissertation is based on the following manuscripts:

Jürgen Niedballa[1], Rahel Sollmann[2,3], Alexandre Courtiol[1], Andreas Wilting[1] (2016): camtrapR: an R package for efficient camera trap data management. Methods in Ecology and Evolution 7(12): 1457–1462 (https://doi.org/10.1111/2041-210X.12600)

Jürgen Niedballa[1], Rahel Sollmann[4], Azlan bin Mohamed[1], Johannes Bender[1], Andreas Wilting[1] (2015): Defining habitat covariates in camera-trap based occupancy studies. Scientific Reports 5: 17041 (https://doi.org/10.1038/srep17041)

Jürgen Niedballa[1], Andreas Wilting[1,] Rahel Sollmann[4], Heribert Hofer[1], Alexandre Courtiol[1]. Assessing spatiotemporal interactions between species from camera trapping data

[1] Leibniz Institute for Zoo and Wildlife Research, Alfred-Kowalke-Str. 17, 10315 Berlin, Germany

[2] North Carolina State University, Department of Forestry and Environmental Resources, Campus Box 8008, Raleigh, NC 27695-7646, USA

[3] Present address: USDA Forest Service, Pacific Southwest Research Station, 1731 Research Park Drive, Davis, CA 95618, USA.

[4] Department of Wildlife, Fish and Conservation Biology, 1088 Academic Surge, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA

Table of Contents

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

ACKNOWLEDGEMENTS

This thesis would not have been possible without the support of many, many people.

First of all my special thanks go to my supervisor Andreas Wilting for his guidance, motivation, his drive and support, and for giving me the freedom I needed to give my best throughout the writing of this thesis. I also wish to thank him for all the possibilities working in this project offered for growth on the professional and personal level. I sincerely appreciate it and will never forget it.

I further wish to thank Rahel Sollmann and Alexandre Courtiol for their guidance on the proper use of the methods I applied and the insights I gained during our discussions, and for extensive help with my manuscripts. Alexandre, I want to also thank you for introducing me to the art of creating R packages.

I wish to thank all colleagues from the ScreenForBio project, my fellow PhD students Azlan bin Mohamed, John Mathai, Andrew Tilker and Roshan Guharajan for the mutual support and for working together to make this project a success. I am particularly grateful to John and Azlan for their support during our time in Sabah and to John for being a great friend. I also thank Jan Axtner, Lisa Hörig, Johannes Bender, Tejas Bhagwat, Seth Timothy Wong for their help and support. Not least I thank our field assistants in Sabah, Azrie, Alex, Bendy, Jeff, Jereo and Len for making sure me and all other team members always returned safe and sound from the jungle.

I thank Prof. Heribert Hofer for inspiring and fruitful discussions and new ideas that arose from them. Furthermore I further thank Prof. Heribert and Prof. Florian Jeltsch for evaluating this thesis.

Stephanie Kramer-Schadt deserves a big thank you for first inviting me to IZW back in 2011 as an intern and for revealing to me the beauty and power of GIS work and scripting. I further wish to thank Pierre Gras enjoyable and rewarding discussions about R and GIS.

Many thanks go to all fellow PhD students at IZW, past and present, for the friendly and welcoming atmosphere you created at the institute, and I wish each and every one of you all the best for your future. In particular, I wish to thank Michaela Wetzel, Saskia Wutke, Juliane Maaß, Renata Martins, Riddhi Patel, Paula Ortega Peréz, Deniz Mengüllüoğlu, Niccolo Alfano, Linn Sophia Lehnert, Sara Troxell, Anne Seltmann, Susanne Pribbenow, Sónia Jesus Fontes and Milena Stillfried for making my

time at IZW really worthwhile. I am also grateful to Linn Sophia Lehnert and Milena Stillfried for proof-read parts of this thesis.

My gratitude extends to all users and testers of the camtrapR package, An The Truong Nguyen, Seth Timothy Wong, Ruben Portas, Jörg Melzheimer, and further to all users of the package near and far who contributed to the package with their constructive feedback, feature requests and bug reports. You are the ones for whom I wrote the package, and if it makes your life easier, I am happy.

I further wish to thank our institute admin team. Without your relentless work, nothing at all would be possible at the institute. In particular I wish to thank Katrin Hohnstädter and Daniel Spitzer for their help and patience when handling admin issues. Wolfgang Tauche, Paul Sobtzick and Luca Flauaus came to my rescue more than once with their IT support whenever it was most needed. Thank you!

I am grateful to Mike Meredith and Marc Kéry for statistical advice and for organising an unforgettable workshop on community modelling in Kuala Lumpur.

I thank the German Federal Ministry of Education and Reseach (BMBF) for funding our junior research group.

A deep heartfelt thank you goes to my entire family, particularly my parents Renate and Stefan. Thank you for always having been there for me with everything I needed throughout my life, for supporting me unconditionally and never questioning me, my ideas or my plans. I also thank Thomas and Sabine for their support throughout my life, and also Manuela, Georg, Sara, Max, Jannick and Konrad.

My deepest gratitude and love goes to Kelly for being the most loving, understanding, inspiring and uplifting partner I could ever have wished for. Love and light to you.

Last but surely not least I wish to thank my closest friends Franz, René, Hans, Christian and Mitsuki for their support and friendship.

Thank you very much to all of you!

Vielen herzlichen Dank euch allen!

CHAPTER 1

General introduction

Camera trapping for wildlife studies

Automatic camera traps revolutionised ecological inference about wildlife populations over the past decades due to their ability to continuously, simultaneously, and cost-efficiently survey animals in their natural environment over extensive periods of time and large areas with relatively low requirements of manpower (Kucera & Barrett 2011) . Even elusive, cryptic and rare species can be surveyed with their help, in some cases providing initial baseline data for species with virtually unknown ecology (Mathai *et al.* 2017).

The data camera traps collect are not only aesthetically appealing, but offer intriguing insights into the ecology and behaviour of wildlife, typically medium to large sized mammals. Consequently, the use of camera traps and publications about camera trapping surveys has seen a sharp increase over the past decades (see Figure 1.1 and McCallum 2012). Applications of camera trapping data are found in species management, conservation and ecological research. In the latter, inference ranges from the individual level (e.g. movement parameters; Royle *et al.* 2014) through population and metapopulation level (e.g. occupancy status, local abundance and population density, changes of occupancy over time; Karanth 1995; Nicholson & Van Manen 2009; Sollmann *et al.* 2012; Wilting *et al.* 2012; Mohamed *et al.* 2013) to community level inference as in species inventories, species interaction studies or community ecology (MacKenzie *et al.* 2006; Burton *et al.* 2012; Sunarto *et al.* 2013; Sollmann *et al.* 2017). Further applications of camera traps are in behavioural studies, e.g. for species activity patterns, the monitoring of denning behaviour or behavioural patterns that are difficult to infer from individual observations (Cutler & Swann 1999; Sunarto *et al.* 2013; Ancrenaz *et al.* 2014).

In the days when analog photography was the only available technology for camera trapping studies the amount of data was limited by technological constraints such as the maximum number of exposures available on film and consequently the manpower needed to keep cameras functional (Kays & Slauson 2008; Kucera & Barrett 2011). With the advent of digital photography in the 2000s and its application in camera trapping, the amount of data collected in camera trapping studies increased rapidly and began to pose a whole new challenge in terms of data management compared to film equipment used earlier, which is by now virtually non-existent in camera trapping. For illustration, in a typical camera trapping study involving dozens to hundreds of cameras that are deployed for weeks to months, hundreds of thousands or even millions of images can be collected (e.g. Swanson *et al.* 2015), making manual data management impractical and efficient data management imperative. Fortunately, standardised storage of metadata in the images captured by camera traps, most essentially date and time, allows for their automated extraction. Consequently, various software toolboxes have been developed to facilitate the management and automatic data extraction from large amounts of camera trapping data, each with their own standards for data storage (Harris *et al.* 2010; Fegraus *et al.* 2011; Sundaresan *et al.* 2011; Sanderson & Harris 2013; Krishnappa & Turner 2014; Tobler 2014; Ivan & Newkirk 2015; Zaragozí *et al.* 2015; Bubnicki *et al.* 2016; Hendry & Mann 2017).

Concomitantly to the rise of digital photography in camera trapping, the R language (R Core Team 2017) has become the de-facto standard environment for statistical analysis of ecological data, with some of the most highly developed tools for a multitude of ecological analyses of camera trapping data being native to R (e.g. Fiske & Chandler 2011; Laake 2013; Efford 2015; Meredith & Ridout 2016). The absence of a toolbox for camera trap data management in R was the incentive to develop an R package to fill this gap. The aim of its development was to seamlessly connect camera trap data acquisition and management with downstream analytical tool in the same environment. This R package is called camtrapR, was first released in July 2015 on CRAN (Niedballa *et al.* 2015a) and updated continuously afterwards. The development of camtrapR fell in a time in which there was a spate of development in camera trap data management software (see above), clearly illustrating the need for software that is capable of processing the large amounts of images that are typically collected in today's camera trapping studies. The R package camtrapR described in chapter 2 harnesses the power of the R language to efficiently manage and analyse camera trapping data (Niedballa *et al.* 2016).

Occupancy modelling using camera trapping data

One of the main applications of camera trapping data is their use for occupancy modelling. Occupancy models were first proposed by MacKenzie et al. (2002) and have been extended and steadily increased in popularity ever since. The main strength of occupancy models is that they explicitly account for imperfect detection, the possibility of not detecting a species that is in fact present. Imperfect detection is inevitable in wildlife studies in general and camera trapping studies in particular. Not accounting for imperfect detection can very severely affect model results and bias inferences (Gu & Swihart 2004; MacKenzie *et al.* 2005; Sollmann *et al.* 2013).

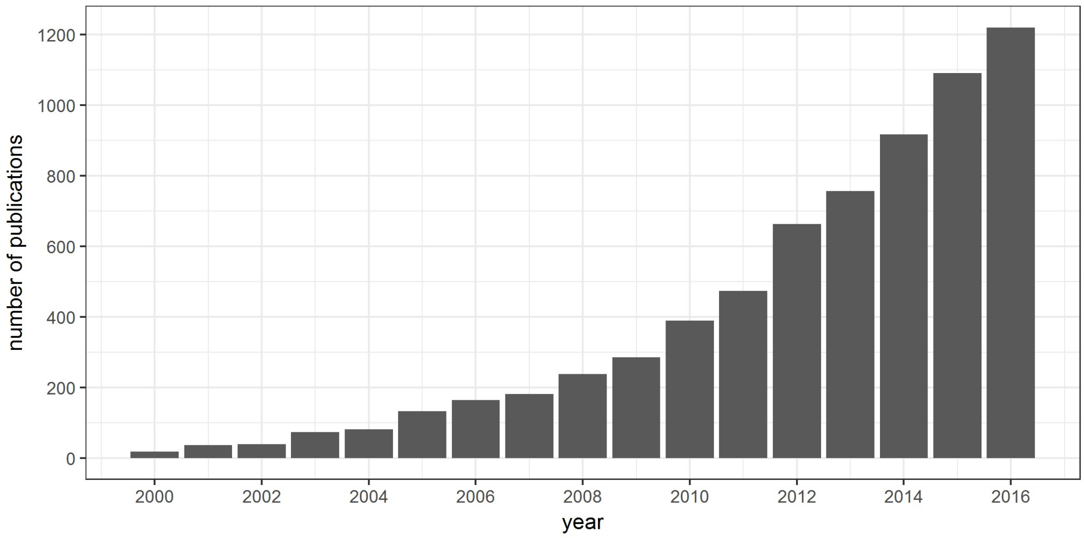


Figure 1.1: The increase of scientific publications related to camera trapping between 2000 and 2016 returned by Google Scholar for the search term: "Camera trap" OR "Trail camera" OR "Photo trap".

Occupancy models address imperfect detection explicitly by separating the ecological process from the imperfect detection process. The ecological process is the occupancy state of a site, i.e. the site being either occupied or unoccupied. Detections are conditional on this occupancy state and only possible if the site is occupied. An observation or detection model is applied to link observations to the unobserved (latent) state variable, which is occupancy (for details see Chapter 4). In that sense occupancy models classify as hierarchical models (Kéry & Royle 2015).

In order to estimate detection probabilities, occupancy models require repeated samples from the sampling sites. In the context of camera trapping, sampling sites are camera trap stations, and repeated samples are obtained by subdividing the survey period at each station into discrete time intervals (usually a few days) to obtain a pattern of detection / non-detection during these repeated samples over the course of the survey duration, from which detection probabilities can be inferred. Occupancy models can furthermore account for varying detection probability (e.g. due to habitat characteristics or temporally changing environmental conditions) and parameter estimates are unaffected by missing sample occasions (MacKenzie *et al.* 2006).

The simplest case of occupancy models are single-species, single season models. These can be expanded to multi-season models, two-species or community occupancy models (MacKenzie *et al.* 2006). Single-species single-season models can provide estimates of occupancy and detection probabilities of a single species while accounting for imperfect detection and factoring in the influence of environmental characteristic on occupancy and detection probability estimates

(MacKenzie *et al.* 2002). Single-species multi-season model allow the estimation of extinction and (re-)colonisation rates from repeated surveys during several seasons, e.g. in metapopulation studies (MacKenzie *et al.* 2003; Hamel *et al.* 2013). Two-species occupancy models are used to assess interactions between species, expressed as changes in occupancy probability of a species in the presence of another species (MacKenzie *et al.* 2004; Richmond *et al.* 2010; Lazenby & Dickman 2013; Rota *et al.* 2016). Community occupancy models allow for inferences about species richness, individual and community responses to habitat characteristics (Burton *et al.* 2012; Sollmann *et al.* 2017).

All of these different types of occupancy models can take into account covariates on detection and occupancy probabilities. These covariates are most commonly time-invariant habitat characteristics, but detection probabilities may also depend on time-variant covariates such as weather conditions. Habitat can be characterised either via ground surveys, which are often labour-intensive and limited in their spatial coverage and consequently don't allow extrapolation to larger areas or continuous monitoring through time. In addition, ground surveys inevitably lead to disturbance and interference with natural processes on the ground, possibly biasing results. Thus, an increasingly popular alternative to ground surveys which overcomes many of their limitations is the use of remote sensing technology for habitat characterisation and the generation of covariates for occupancy models (Turner 2014; Bush *et al.* 2017; Steenweg *et al.* 2017). Remote sensing covariates hold potential for extrapolation of species distribution models to unsampled areas and monitoring of changes in habitat structure over time without physical interference at these sites.

The application of remote sensing data in occupancy modelling

Since the first satellite-based remote sensing surveys of the earth's surface in the early 1970s, the use of remote sensing technologies revolutionised earth observation and the way ecologists gather and make use of spatial data in a similar way to the revolution which the use of automatic camera traps meant for the observation of cryptic and elusive wildlife species (Leimgruber *et al.* 2005; Williams *et al.* 2006). Remote sensing means measuring properties of objects without making physical contact, and here refers to spaceborne and airborne earth observation. While airborne remote sensing offers more detail due to greater proximity and can be adapted more specifically to the needs of an individual study, satellite data are more readily available, cheaper (often available for free), and provide a wider and more consistent spatial and temporal coverage of the study area.

Earth observation satellites are equipped with a plethora or different instruments for a multitude of scientific purposes. Amongst these, optical sensors are the most accessible, widely available and suitable data source for land cover mapping, which is highly relevant for the analysis of species distribution data. While older satellites such as Landsat had a moderate spatial resolution of 30 metres per pixel, more recent generations of sensor offer higher resolution and thus make much more detailed habitat mapping possible (available sensors range from 5-m RapidEye data to < 1 m in commercial satellites such as WorldView-4 (RapidEye AG 2012)). Higher resolution, however, is not invariably favourable because it comes with a number of drawbacks, namely much larger data volumes, smaller swath width (smaller spatial extent of images), excessive image detail which complicates image analysis and classification, and frequently high costs associated with image acquisition. Consequently, there is a trade-off to be made between using lower-resolution data and running the risk of missing crucial habitat features, and high-resolution data which are more difficult to work with, in many cases expensive, and may not cover the entire study area.

The potential applications of remote sensing in the realm of wildlife ecology are manifold, with land cover mapping and ecosystem monitoring, land cover change detection or terrain description being prominent examples (Karanth *et al.* 2009; Pettorelli *et al.* 2011; Sunarto *et al.* 2012; Carter *et al.* 2013; García-Rangel & Pettorelli 2013). They offer repeated temporal and extensive spatial coverage of the earth’s surface, allowing continuous monitoring and spatial extrapolation of point processes observed on the ground (e.g. at camera trap stations; Bush *et al.* 2017).

Yet, remote sensing data require validation by means of in-situ (ground) surveys. These in-situ surveys are crucial to link remotely sensed data to processes and patterns on the ground and guide interpretation of remote sensing data (Asner *et al.* 2010; Langner *et al.* 2012). Such in-situ surveys also allow for measurements that are difficult or impossible to obtain using remote sensing techniques, e.g. tree diameters or floristic information. But at the same time they can be logistically challenging, are expensive, and limited in spatial scope. Hence, both in situ and remote sensing surveys have their benefits and drawbacks and support one another in terms of thorough habitat characterisation.

The concept of scale in ecology

In all of these applications, both of remote sensing and in situ habitat surveys, observed landscape patterns are scale-dependent, drawing attention to the widely-known and much-debated concept of scale in ecology, and how it may affect inference from occupancy models that utilise remote sensing covariates (Wiens & Milne 1987; Wiens 1989; Wu 2004). The impact of scale on ecological inferences is a long-known phenomenon and has received much coverage in scientific literature. In the words of Wiens (1989), “[...] different patterns emerge at different scales of investigation of virtually any aspect of any ecological system”. As a consequence, not accounting for scaling relationships may lead to biased or even wrong conclusions about the system under investigation.

Ecological scale consists of two constituents, extent and grain (Turner *et al.* 1989). Generally speaking, extent is the overall area studied while grain is the size of an individual unit of observation. It is impossible to detect any element smaller than the grain size, and generalisation beyond the study extent is only possible if scale-independence of processes and patterns is assumed, which is uncommon (Wiens 1989). In the context of remote sensing, extent is the area surveyed and grain is the spatial resolution of the data, meaning the pixel size in satellite imagery and derived data. In in-situ surveys, extent equally refers to the area surveyed, and grain could, for example, be the size of subplots within a vegetation plot for which data are aggregated.

The question of scale is related to the question of how animals perceive their environment. In the context of connecting remote sensing to species occurrence data, the questions raised are 1) how the immediate and wider surroundings of a location (i.e., different extents/radii around that location) influence species occupancy patterns, and 2) what spatial resolution (grain size, pixel size) is adequate to describe habitat characteristics that influence species occurrence, i.e., do habitat covariates derived from high-resolution remote sensing data yield higher explanatory value than those from lower-resolution data? To investigate how this concept of scale affects inference from occupancy models based on camera trapping data, in chapter 3 we compared the effect of varying extents and grain sizes in habitat covariates derived from remote sensing data and in situ surveys in terms of their explanatory power (Niedballa *et al.* 2015b).

Detecting species interactions in camera trapping data

The potential applications of camera trapping data are by far not limited to occupancy modelling. While the inference of spatial interactions between species can be achieved with two-species occupancy models as mentioned above, these models are only sensitive to observable patterns in spatial distributions, thus assuming that interactions between species impact the spatial occurrence of species, e.g. via exclusion of a species from sites where a competitor or predator is present (in the case of avoidance). Nevertheless, an alternative, more subtle form of avoidance between species may involve temporal avoidance rather than spatial avoidance, with species partitioning time instead of space to avoid encounters and interactions while still co-occurring at the same localities. These temporal interactions between species cannot be detected with two-species occupancy models because the occupancy state (the state variable of occupancy models) is not affected by and also not sensitive to temporal avoidance.

In the case of avoidance between species (due to predation risk or competition), temporal avoidance may take different forms. The first is the temporary avoidance of a site after a dominant species or individual was present (e.g. a predator or stronger competitor). Alternatively, species may shift their activity periods relative to one another in an effort to segregate temporally and thus reduce the chances of encounters and potentially detrimental interactions. Both of these interactions were observed in wildlife data, e.g. between sympatric intraguild competitors or in predator-prey relationships (e.g. Harmsen *et al.* 2009; Monterroso *et al.* 2014; Parsons *et al.* 2016). Both of these types of temporal interactions are avoidance strategies and ultimately behavioural adaptations aimed at increasing fitness by reducing predation risk or the risk of disadvantageous interactions (Schuette *et al.* 2013).

Because of its more subtle nature and the randomness in species detections in camera trapping, the inference of temporal interactions between species is a challenging task, yet rewarding in terms of ecological insights gained (e.g. Hayward & Slotow 2009; Tambling *et al.* 2015; Wang *et al.* 2015). Notwithstanding the ecological interest and despite a number of methods that were suggested and applied (Harmsen *et al.* 2009; Ridout & Linkie 2009; Parsons *et al.* 2016), the detection of temporal interactions from camera trapping data has received little methodological scrutiny compared to two-species occupancy models. The main obstacle for a comparative assessment of different methods for detecting temporal interactions between species is the unknown and unobserved true state of the system. The only way to overcome this obstacle and arrive at an unbiased assessment of the different methods is to simulate data and apply the methods to these simulated data, which was done in chapter 4.

Structure of the dissertation

The overarching aim of this work was to facilitate the handling of and improve inferences from camera trapping studies by firstly providing a new toolbox for camera trap data management, secondly exploring the utility of high-resolution remote sensing data in camera trap-based species occupancy models, and thirdly assessing and comparing methods for detecting temporal interactions between species from camera trapping data. The results are presented in the form of three manuscripts in **chapters 2 to 4**:

1) Modern camera trapping is a very data-intensive discipline and therefore requires thorough data management. **Chapter 2** (“camtrapR: an R package for efficient camera trap data management”) describes the R package camtrapR which provides a camera trap data management workflow within the widely-used R environment. It seamlessly connects camera trap data acquisition and management with ecological analyses provided by a multitude of other R packages and software tools.

2) **Chapter 3** (“Defining habitat covariates in camera-trap based occupancy studies”) deals with the incorporation of high-resolution remote sensing data into camera trap based occupancy models and the arising questions of ecological scale. Using six sample species from Sabah, Malaysian Borneo, the manuscript explores the influence that spatial resolution (pixel size of land cover data) and the extent of patches around camera trap localities have on inferences of habitat associations from single-species occupancy models. It shows that high-resolution land cover data can have considerably more model support than lower resolution land cover data, suggesting their application in occupancy models is justified and may lead to improved inferences. Remote sensing data for habitat characterisation can further reduce field effort by serving as a surrogate for labour-intensive in-situ measures. The manuscript thus demonstrates the utility of high-resolution land cover data in species occupancy models and confirms the decades-old yet frequently ignored notion that ecological scale still matters when using state-of-the-art methods.

3) **Chapter 4** (“Assessing spatiotemporal interactions between species from camera trapping data”) explores the use of camera trapping data for inferring temporal interactions between species. More specifically, it investigates the properties of different statistical tests for detecting two types of temporal interactions between species: 1) spatiotemporal avoidance, i.e. temporary avoidance of a site by a subordinate species after a dominant species was present, and 2) temporal segregation, i.e. shifts in activity patterns between species which lead to reduced activity overlaps. Both of these strategies can serve to avoid encounters or interference between species, thus facilitating co-existence between species. In a simulation study, we compared four methods for detecting spatiotemporal avoidance: linear models, U-test, a permutation test and a test based on randomly created records; and assessed a permutation test for detecting temporal segregation. All tests were generally robust and suitable for detecting temporal interactions. Linear models had highest power and greatest flexibility for detecting spatiotemporal avoidance. The results can guide practitioners in their choice of methods and furthermore provide a flexible and extensible framework for simulation and exploration of temporal species interactions in camera trapping data.

Chapter 5 is a general discussion which summarises the results of the dissertation and discusses them in the wider context of camera trapping and quantitative wildlife ecology.

References

Ancrenaz, M., Sollmann, R., Meijaard, E., Hearn, A.J., Ross, J., Samejima, H., Loken, B., Cheyne, S.M., Stark, D.J., Gardner, P.C., Goossens, B., Mohamed, A., Bohm, T., Matsuda, I., Nakabayasi, M., Lee, S.K., Bernard, H., Brodie, J., Wich, S., Fredriksson, G., Hanya, G., Harrison, M.E., Kanamori, T., Kretzschmar, P., Macdonald, D.W., Riger, P., Spehar, S., Ambu, L.N. & Wilting, A. (2014). Coming down from the trees: Is terrestrial activity in Bornean orangutans natural or disturbance driven? *Scientific Reports*, **4**, 1–6.

Asner, G.P., Powell, G.V.N., Mascaro, J., Knapp, D.E., Clark, J.K., Jacobson, J., Kennedy-Bowdoin, T., Balaji, A., Paez-Acosta, G., Victoria, E., Secada, L., Valqui, M. & Hughes, R.F. (2010). High-resolution forest carbon stocks and emissions in the Amazon. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 16738–42.

Bubnicki, J.W., Churski, M. & Kuijper, D.P.J. (2016). TRAPPER: an open source web-based application to manage camera trapping projects. *Methods in Ecology and Evolution*, **7**, 1209–1216.

Burton, A.C., Sam, M.K., Balangtaa, C. & Brashares, J.S. (2012). Hierarchical Multi-Species Modeling of Carnivore Responses to Hunting, Habitat and Prey in a West African Protected Area. *PloS one*, **7**, e38007.

Bush, A., Sollmann, R., Wilting, A., Bohmann, K., Cole, B., Balzter, H., Martius, C., Zlinszky, A., Calvignac-Spencer, S., Cobbold, C.A., Dawson, T.P., Emerson, B.C., Ferrier, S., Gilbert, M.T.P., Herold, M., Jones, L., Leendertz, F.H., Matthews, L., Millington, J.D.A., Olson, J.R., Ovaskainen, O., Raffaelli, D., Reeve, R., Rödel, M.-O., Rodgers, T.W., Snape, S., Visseren-Hamakers, I., Vogler, A.P., White, P.C.L., Wooster, M.J. & Yu, D.W. (2017). Connecting Earth observation to high-throughput biodiversity data. *Nature Ecology & Evolution*, **1**, 176.

Carter, N.H., Gurung, B., Vina, A., III, H.C., Karki, J.B. & Liu, J. (2013). Assessing spatiotemporal changes in tiger habitat across different land management regimes. *Ecosphere*, **4**, 1–19.

Cutler, T.L. & Swann, D.E. (1999). Using Remote Photography in Wildlife Ecology: A Review. *Wildlife Society Bulletin*, **27**, 571–581.

Efford, M.G. (2015). secr: Spatially explicit capture-recapture models.

Fegraus, E.H., Lin, K., Ahumada, J.A., Baru, C., Chandra, S. & Youn, C. (2011). Data acquisition and 7management software for camera trap data: A case study from the TEAM Network. *Ecological Informatics*, **6**, 345–353.

Fiske, I.J. & Chandler, R.B. (2011). unmarked : An R Package for Fitting Hierarchical Models of Wildlife Occurrence and Abundance. *Journal of Statistical Software*, **43**, 1–23.

García-Rangel, S. & Pettorelli, N. (2013). Thinking spatially: The importance of geospatial techniques for carnivore conservation. *Ecological Informatics*, **14**, 84–89.

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

Hamel, S., Killengreen, S.T., Henden, J.A., Yoccoz, N.G. & Ims, R.A. (2013). Disentangling the importance of interspecific competition, Food availability, And habitat in species occupancy: Recolonization of the endangered Fennoscandian arctic fox. *Biological Conservation*, **160**, 114–120.

Harmsen, B.J., Foster, R.J., Silver, S.C., Ostro, L.E.T. & Doncaster, C.P. (2009). Spatial and Temporal Interactions of Sympatric Jaguars (Panthera onca) and Pumas (Puma concolor) in a Neotropical Forest. *Journal of Mammalogy*, **90**, 612–620.

Harris, G., Thompson, R., Childs, J.L. & Sanderson, J.G. (2010). Automatic Storage and Analysis of Camera Trap Data. *Bulletin of the Ecological Society of America*, **91**, 352–360.

Hayward, M.W. & Slotow, R. (2009). Temporal Partitioning of Activity in Large African Carnivores: Tests of Multiple Hypotheses. *South African Journal of Wildlife Research*, **39**, 109–125.

Hendry, H. & Mann, C. (2017). Camelot (Occupancy models.

Ivan, J.S. & Newkirk, E.S. (2015). CPW Photo Warehouse: a custom database to facilitate archiving, identifying, summarizing and managing photo data collected from camera traps. *Methods in Ecology and Evolution*, **7**, 499–504.

Karanth, K.U. (1995). Estimating tiger Panthera tigris populations from camera-trap data using capture-recapture models. *Biological Conservation*, **71**, 333–338.

Karanth, K.K., Nichols, J.D., Hines, J.E., Karanth, K.U. & Christensen, N.L. (2009). Patterns and determinants of mammal species occurrence in India. *Journal of Applied Ecology*, **46**, 1189–1200.

Kays, R.W. & Slauson, K.M. (2008). Remote cameras. *Noninvasive Survey Methods for Carnivores* (eds R.A. Long, P. MacKay, J. Ray & W. Zielinski), pp. 110–140. Island Press, Washington, DC.

Kéry, M. & Royle, J.A. (2015). *Applied Hierarchical Modeling in Ecology. Analysis of distribution,* *abundance and species richness in R and BUGS: Volume 1: Prelude and Static Models*. Academic Press.

Krishnappa, Y.S. & Turner, W.C. (2014). Software for minimalistic data management in large camera trap studies. *Ecological Informatics*, **24**, 11–16.

Kucera, T.E. & Barrett, R.H. (2011). A History of Camera Trapping. *Camera Traps in Animal Ecology*, pp. 9–26. Springer Japan.

Laake, J.L. (2013). *RMark: An R Interface for Analysis of Capture-Recapture Data with MARK*. Seattle, WA.

Langner, A., Samejima, H., Ong, R.C., Titin, J. & Kitayama, K. (2012). Integration of carbon conservation into sustainable forest management using high resolution satellite imagery: A case study in Sabah, Malaysian Borneo. *International Journal of Applied Earth Observation and Geoinformation*, **18**, 305–312.

Lazenby, B.T. & Dickman, C.R. (2013). Patterns of Detection and Capture Are Associated with Cohabiting Predators and Prey. *PLoS ONE*, **8**, e59846.

Leimgruber, P., Christen, C.A. & Laborderie, A. (2005). The impact of landsat satellite monitoring on conservation biology. *Environmental Monitoring and Assessment*, **106**, 81–101.

MacKenzie, D.I., Bailey, L.L. & Nichols, J.D. (2004). Investigating species co-occurrence patterns when species are detected imperfectly. *Journal of Animal Ecology*, **73**, 546–555.

MacKenzie, D.I., Nichols, J.D., Hines, J.E., Knutson, M.G. & Frankling, A.B. (2003). Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology*, **84**, 2200–2207.

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

MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L. & Hines, J.E. (2006). *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurence*. Academic Press, London, United Kingdom.

MacKenzie, D.I., Nichols, J.D., Sutton, N., Kawanishi, K. & Bailey, L.L. (2005). Improving inferences in population studies of rare species that are detected imperfectly. *Ecology*, **86**, 1101–1113.

Mathai, J., Sollmann, R., Michael E. Meredith, J.L.B., Niedballa, J., Buckingham, L., Wong, S.T., Asad, S. & Wilting, A. (2017). Fine-scale distributions of carnivores in a logging concession in Sarawak, Malaysian Borneo. *Mammalian Biology - Zeitschrift für Säugetierkunde*, **86**, 56–65.

McCallum, J. (2012). Changing use of camera traps in mammalian field research: Habitats, taxa and study types. *Mammal Review*, **43**, 196–206.

Meredith, M. & Ridout, M. (2016). overlap: Estimates of coefficient of overlapping for animal activity patterns. R package version 0.2.6.

Mohamed, A., Sollmann, R., Bernard, H., Ambu, L.N., Lagan, P., Mannan, S., Hofer, H. & Wilting, A. (2013). Density and habitat use of the leopard cat (Prionailurus bengalensis) in three commercial forest reserves in Sabah, Malaysian Borneo. *Journal of Mammalogy*, **94**, 82–89.

Monterroso, P., Alves, P.C. & Ferreras, P. (2014). Plasticity in circadian activity patterns of mesocarnivores in Southwestern Europe: Implications for species coexistence. *Behavioral Ecology and Sociobiology*, **68**, 1403–1417.

Nicholson, J.M. & Van Manen, F.T. (2009). Using occupancy models to determine mammalian responses to landscape changes. *Integrative Zoology*, **4**, 232–239.

Niedballa, J., Courtiol, A., Sollmann, R. & Wilting, A. (2015a). camtrapR: Camera Trap Data Management and Preparation of Occupancy and Spatial Capture-Recapture Analyses.

Niedballa, J., Sollmann, R., Courtiol, A. & Wilting, A. (2016). camtrapR: An R package for efficient camera trap data management. *Methods in Ecology and Evolution*, **7**, 1457–1462.

Niedballa, J., Sollmann, R., Mohamed, A. bin, Bender, J. & Wilting, A. (2015b). Defining habitat covariates in camera-trap based occupancy studies. *Scientific Reports*, **5**, 17041.

Parsons, A.W., Bland, C., Forrester, T., Baker, M.C., Schuttler, S.G., McShea, W.J., Costello, R. & Kays, R. (2016). The ecological impact of dogs on wildlife in protected areas in eastern North America. *Biological Conservation*, **203**, 75–88.

Pettorelli, N., Ryan, S., Mueller, T., Bunnefeld, N., Jedrzejewska, B., Lima, M. & Kausrud, K. (2011). The Normalized Difference Vegetation Index (NDVI): unforeseen successes in animal ecology. *Climate Research*, **46**, 15–27.

R Core Team. (2017). R: A language and environment for statistical computing.

RapidEye AG. (2012). Satellite Imagery Product Specifications Version 4.1. URL https://earth.esa.int/documents/10174/896711/RE_Product_Specifications_ENG.pdf [accessed 4 July 2013]

Richmond, O.M.W., Hines, J.E. & Beissinger, S.R. (2010). Two-species occupancy models: a new paramaterization applied to co-occurence of secretive rails. *Ecological Applications*, **20**, 2036–2046.

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.

Rota, C.T., Wikle, C.K., Kays, R.W., Forrester, T.D., McShea, W.J., Parsons, A.W. & Millspaugh, J.J. (2016). A two-species occupancy model accommodating simultaneous spatial and interspecific dependence. *Ecology*, **97**, 48–53.

Royle, J.A., Chandler, R.B., Sollmann, R. & Gardner, B. (2014). *Spatial Capture-Recapture*. Academic Press.

Sanderson, J. & Harris, G. (2013). Automatic data organization, storage, and analysis of camera trap pictures. *Journal of Indonesian Natural History*, **1**, 6–14.

Schuette, P., Wagner, A.P., Wagner, M.E. & Creel, S. (2013). Occupancy patterns and niche partitioning within a diverse carnivore community exposed to anthropogenic pressures. *Biological Conservation*, **158**, 301–312.

Sollmann, R., Furtado, M.M., Hofer, H., Jácomo, A.T. a., Tôrres, N.M. & Silveira, L. (2012). Using occupancy models to investigate space partitioning between two sympatric large predators, the jaguar and puma in central Brazil. *Mammalian Biology - Zeitschrift für Säugetierkunde*, **77**, 41–46.

Sollmann, R., Mohamed, A. bin, Niedballa, J., Bender, J., Ambu, L., Lagan, P., Mannan, S., Ong, R.C., Langner, A., Gardner, B. & Wilting, A. (2017). Quantifying mammal biodiversity co-benefits in certified tropical forests. *Diversity and Distributions*, 317–328.

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

Steenweg, R., Hebblewhite, M., Kays, R., Ahumada, J., Fisher, J.T., Burton, C., Townsend, S.E., Carbone, C., Rowcliffe, J.M., Whittington, J., Brodie, J., Royle, J.A., Switalski, A., Clevenger, A.P., Heim, N. & Rich, L.N. (2017). Scaling up camera traps: monitoring the planet's biodiversity with networks of remote sensors. *Frontiers in Ecology and the Environment*, **15**, 26–34.

Sunarto, Kelly, M.J., Parakkasi, K., Klenzendorf, S., Septayuda, E. & Kurniawan, H. (2012). Tigers need cover: multi-scale occupancy study of the big cat in Sumatran forest and plantation landscapes. *PloS one*, **7**, e30859.

Sunarto, Sollmann, R., Mohamed, A. & Kelly, M.J. (2013). Camera trapping for the study and conservation of tropical carnivores. *Raffles Bulletin of Zoology*, **28**, 21–42.

Sundaresan, S.R., Riginos, C. & Abelson, E.S. (2011). Management and Analysis of Camera Trap Data: Alternative Approaches (Response to Harris et al. 2010). *Bulletin of the Ecological Society of America*, **92**, 188–195.

Swanson, A., Kosmala, M., Lintott, C., Simpson, R., Smith, A. & Packer, C. (2015). Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. *Scientific Data*, **2**, 150026.

Tambling, C.J., Minnie, L., Meyer, J., Freeman, E.W., Santymire, R.M., Adendorff, J. & Kerley, G.I.H. (2015). Temporal shifts in activity of prey following large predator reintroductions. *Behavioral Ecology and Sociobiology*, **69**, 1153–1161.

Tobler, M.W. (2014). Camera base version 1.6.1.

Turner, W. (2014). Sensing biodiversity. *Science*, **346**, 301–302.

Turner, M.G., Dale, V.H. & Gardner, R.H. (1989). Predicting across scales: Theory development and testing. *Landscape Ecology*, **3**, 245–252.

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

Wiens, J.A. (1989). Spatial scaling in ecology. *Functional Ecology*, **3**, 385–397.

Wiens, J.A. & Milne, B.T. (1987). Scaling of 'landscapes' in landscape ecology, or, landscape ecology from a beetle's perspective. *Landscape Ecology*, **3**, 87–96.

Williams, D.L., Goward, S.N. & Arvidson, T. (2006). Landsat: Yesterday, today, and tomorrow. *Photogrammetric Engineering and Remote Sensing*, **72**, 1171–1178.

Wilting, A., Mohamed, A., Ambu, L.N., Lagan, P., Mannan, S., Hofer, H. & Sollmann, R. (2012). Density of the Vulnerable Sunda clouded leopard Neofelis diardi in two commercial forest reserves in Sabah, Malaysian Borneo. *Oryx*, **46**, 423–426.

Wu, J. (2004). Effects of changing scale on landscape pattern analysis: scaling relations. *Landscape Ecology*, **19**, 125–138.

Zaragozí, B., Belda, A., Giménez, P., Navarro, J.T. & Bonet, A. (2015). Advances in camera trap data management tools: Towards collaborative development and integration with GIS. *Ecological Informatics*, **30**, 6–11.

CHAPTER 2

`camtrapR`: an R package for efficient camera trap data management

JÜRGEN NIEDBALLA[1*], RAHEL SOLLMANN[2], ALEXANDRE COURTIOL[1], ANDREAS WILTING[1]

Methods in Ecology and Evolution 7.12 (2016): 1457-1462 (https://doi.org/10.1111/2041-210X.12600)

1) Leibniz Institute for Zoo and Wildlife Research, Alfred-Kowalke-Str. 17, 10315 Berlin, GERMANY

2) Department of Wildlife, Fish and Conservation Biology, University of California Davis, Davis, CA, USA

** corresponding author: niedballa@izw-berlin.de*

Author's contribution statement

JN, RS and AW conceived the ideas for the R package and the manuscript. JN wrote the R code of the package. AC advised on R package development and package structure. RS and AW advised on the methods implemented in the package. RS, AW and AC extensively commented on the manuscript.

Niedballa, J., Sollmann, R., Courtiol, A., Wilting, A. (2016), camtrapR: an R package for efficient camera trap data management. Methods Ecol Evol, 7: 1457-1462

camtrapR: an R package for efficient camera trap data management

JÜRGEN NIEDBALLA[1*], RAHEL SOLLMANN[2], ALEXANDRE COURTIOL[1], ANDREAS WILTING[1]

1) Leibniz Institute for Zoo and Wildlife Research, Alfred-Kowalke-Str. 17, 10315 Berlin, GERMANY

2) Department of Wildlife, Fish and Conservation Biology, University of California Davis, Davis, CA, USA

Summary

1. Camera trapping is a widely applied method to study mammalian biodiversity and is still gaining popularity. It can quickly generate large amounts of data which need to be managed in an efficient and transparent way that links data acquisition with analytical tools.
2. We describe the free and open-source R package `camtrapR`, a new toolbox for flexible and efficient management of data generated in camera trap-based wildlife studies. The package implements a complete workflow for processing camera trapping data. It assists in image organization, species and individual identification, data extraction from images, tabulation and visualization of results and export of data for subsequent analyses. There is no limitation to the number of images stored in this data management system; the system is portable and compatible across operating systems.
3. The functions provide extensive automation to minimize data entry mistakes and, apart from species and individual identification, require minimal manual user input. Species and individual identification are performed outside the R environment, either via tags assigned in dedicated image management software or by moving images into species directories.
4. Input for occupancy and (spatial) capture–recapture analyses for density and abundance estimation, for example in the R packages `unmarked` or `secr`, is computed in a flexible and reproducible manner. In addition, survey summary reports can be generated, spatial distributions of records can be plotted and exported to GIS software, and single- and two-species activity patterns can be visualized.
5. `camtrapR` allows for streamlined and flexible camera trap data management and should be most useful to researchers and practitioners who regularly handle large amounts of camera trapping data.

Key-words

biodiversity surveys and monitoring, occupancy models, spatial capture-recapture models, detection history, data management, photo-trapping, camera-trapping, wildlife studies

Introduction

Camera trapping is a powerful and widely used method for the rapid assessment of mammalian biodiversity, particularly in challenging environments (Tobler *et al*. 2008; Sunarto *et al*. 2013; Burton *et al*. 2015). A multitude of ecological analyses utilize camera trap data, including estimation of occupancy probabilities (MacKenzie *et al*. 2002) or abundance, density and demographic rates with capture–recapture (Karanth 1995; Silver *et al*. 2004) and spatial capture–recapture models (Efford 2004; Royle *et al*. 2009; Gardner *et al*. 2010). These methods are implemented in R packages [e.g. unmarked (Fiske & Chandler 2011), secr (Efford 2015) or RMark (Laake 2013)] and stand-alone computer programs [e.g. program mark (White & Burnham 1999) or presence (Hines 2006)].

Efficient use of these analytical tools requires efficient and systematic management of the large numbers of images that can be generated in short periods of time. A variety of approaches using different software have been developed for that purpose (Harris *et al*. 2010; Fegraus *et al*. 2011; Sundaresan, Riginos & Abelson 2011; Sanderson & Harris 2013; Krishnappa & Turner 2014; Tobler 2014; Zaragozí *et al*. 2015; McShea *et al*. 2016; Ivan & Newkirk 2016; see the latter and Table S2.1 for a comparison of approaches). These software approaches have different foci and offer different sets of features. In developing `camtrapR`, we aimed at incorporating and expanding upon these capabilities within a unified camera trap data management tool. In addition to functionalities already available (e.g. automatic import of images, generation of reports and input files for subsequent analyses), `camtrapR` (i) uses the increasingly popular R language, (ii) is free and fully open-source, (iii) is fully compatible with Windows, MacOS and Linux, (iv) reads and allows the user to create arbitrary image metadata tags, (v) supports different methods for identifying species and individuals and (vi) has mapping and GIS export capabilities.

Here, we describe `camtrapR`, the first toolbox for the management of camera trap data available for the R community. Our R package provides a flexible and coherent workflow for efficient camera trap data organization, exploration and processing in the R statistical language, which seamlessly connects data acquisition with downstream analytical tools. We outline the `camtrapR` workflow for organizing camera trap images as well as extracting, exploring and visualizing the resulting data and illustrate its use with a sample data set from a camera trapping study conducted in Sabah, Malaysian Borneo (Mohamed *et al*. 2013). Detailed vignettes, help files, sample data and analyses are available in the `camtrapR` package available on CRAN (https://cran.r-project.org/web/packages/camtrapR/).

Functionality

The `camtrapR` standard workflow can be divided into five main functionalities, listed here and described in sequence below.

1. *Image organization and management*: Setting up a directory structure for storing raw camera trap images and optionally renaming images by station identity (station ID), date and time.
2. *Species/individual identification*: Species and individual identification by metadata tagging in image management software or drag and drop of images into directories. Functions for checking species lists with taxonomic databases, verifying identification and appending species names to files are provided.
3. *Image data extraction*: Tabulation of species records and extraction of image metadata.
4. *Data exploration*: Visualization of spatial species occurrence patterns (including export to gis software), single- and two-species activity patterns.
5. *Data export*: Preparation of input files for subsequent analyses in occupancy and (spatial) capture–recapture frameworks. Generation of survey summary reports.

Table 2.1 provides a list and a short description of all functions in the `camtrapR` package. The functions of the package are described below and in more detail in the package help files and vignettes.

Package description

Overview

The `camtrapR` package, now in version 0.99.1, is written in the R language (R Core Team 2015) and was first released on CRAN in July 2015. It can be used under R version 3.1 (R Core Team 2015) and higher on Windows, MacOS and Linux. The key functions of the package make use of the free command line software ExifTool (Harvey 2015) via system calls to extract metadata from camera trap images in JPEG format. `camtrapR` provides extensive automation of processes, performs rigorous consistency checks on input data and has no inherent limitation in terms of the image number held in the data management system.

`camtrapR` was designed for studies utilizing arrays of camera trap stations, each consisting of one or more (often two) camera trap units (termed cameras for sake of simplicity). Cameras within a station are set in relative proximity to each other compared to between-station distances.

Table 2.1: List of `camtrapR` functions**.**

Functionality	**Function**	**Description**
Image organisation and management	*createStationFolders*	Create directories for storing raw camera trap images
	timeShiftImages	Apply time shifts to JPEG images
	imageRename	Copy and rename images based on station ID and image creation date
	appendSpeciesNames	Add or remove species names from image filenames
Species / individual identification	*checkSpeciesNames*	Check species names against the ITIS taxonomic database
	createSpeciesFolders	Create directories for species identification
	checkSpeciesIdentification	Consistency check on species identification
	getSpeciesImages	Gather all images of a species in a new directory
Image data extraction	*recordTable*	Create a species record table from camera trap images
	recordTableIndividual	Create a single-species record table from camera trap images with individual IDs
	exifTagNames	Return metadata tags and tag names from JPEG images (for use in *recordTable* functions)
	exiftoolPath	Add the directory containing exiftool.exe to PATH temporarily
Data exploration and visualisation	*detectionMaps*	Generate maps of observed species richness and species detections by station
	activityHistogram	Plot histograms of single-species activity
	activityDensity	Plot kernel density estimations of single-species activity
	activityRadial	Radial plots of single-species activity
	activityOverlap	Plot two-species diurnal activity overlap and compute activity overlap
Data export	*cameraOperation*	Create a camera operation matrix
	detectionHistory	Species detection histories for occupancy analyses
	spatialDetectionHistory	Detection histories of individuals for spatial capture-recapture analyses
	surveyReport	Summarise a camera trapping survey

Image Organization and Management

Image organization begins with saving raw images into camera trap station directories (e.g. myStudy/rawImages/stationA). Station directories may contain camera subdirectories (e.g. myStudy/rawImages/stationA/camera1) if more than one camera was used at a station. The function *createStationFolders* can create these directories.

Date and time of images can be changed using the function *timeShiftImages*, for example if internal camera date and/or time values were set incorrectly, reset accidentally, or if users wish to synchronize camera pairs. It uses the date/time shift module of ExifTool.

If desired, all images can be renamed automatically with station ID, camera ID, date and time with the function *imageRename*.

Image Metadata and Metadata Tagging

Digital images contain metadata in standardized Exif format, for example date and time, geotags, camera settings, ambient data, trigger event number and many more. In addition, users can assign information to images via custom metadata tags in image management software, for example species or individual identification, sex, behaviour, group size counts or group membership of individuals. These metadata tags become part of the images and are portable without depending on a relational database structure. Both types of metadata can be extracted, tabulated and used subsequently, for example for data filtering prior to analyses. The package vignettes contain a performance estimate for metadata extraction using ExifTool.

We recommend the free and open-source software DigiKam (www.digikam.org) for tagging because it provides a customizable, hierarchical tag structure and has powerful filtering, querying and batch-tagging capabilities. Adobe Lightroom and Adobe Bridge are also suitable.

Species Identification

Species identification is a laborious but most crucial step in the workflow because all analyses rely on correct species identification and many models are sensitive to false positives (Miller *et al*. 2011). It is also the only task that cannot be automated readily (both in this and other software packages), as automatic identification tools are currently still too unreliable and need reference data for all species potentially present in the study area (Yu *et al*. 2013; but see McShea *et al*. 2016). `camtrapR` supports two different ways of identifying species: (i) by assigning species tags to images in image management software, and (ii) by moving images into species directories [drag and drop, an approach also used by Harris *et al*. (2010) and Sanderson & Harris (2013)].

Users are free to use any species names (or abbreviations or codes) they wish. If scientific or common species names are used, the function *checkSpeciesNames* can check them against the ITIS taxonomic database (www.itis.gov) and returns their matching counterparts (utilizing the R package taxize (Chamberlain & Szöcs 2013) internally), making sure species names and spelling are standardized and taxonomically sound, and thus making it easier to combine data sets from different studies.

To improve reliability of species identification, multiple observers can replicate species assignment (if metadata tags are used for species identification). In order to reconcile their species assignments, and because of the scope for incorrect species assignment even by one observer, the function *checkSpeciesIdentification* finds conflicting species assignments from multiple observers and assesses temporal proximity between images assigned to different species within the same station.

After species identification, the function *appendSpeciesNames* optionally appends species names to file names. The function *getSpeciesImages* can create a species image report by copying all images of a focal species into a separate species directory (e.g. myStudy/speciesImages/Malay Civet), thus facilitating checks on species identification or gathering images for expert identification. If species identification changes at a later point (e.g. after expert identification), these images can easily be copied back into the image directory structure and functions can be rerun.

Individual Identification

Individual identification is a prerequisite for spatial (as well as traditional, non-spatial) capture–recapture analyses. After identifying images to species level and collecting images of the focal species, individual identification is performed in the same way as species identification described above, using either metadata tags or directories for individual identification.

Table 2.2: Example record table. *Station* is the camera trap station ID, *Species* are Leopard Cat *Prionailurus bengalensis* (PBE) and Malay Civet *Viverra tangalunga* (VTA). *"delta.time…"* denotes lag between a record and the last record of the same species at the same station (in seconds, minutes, hours and days). Columns *Directory* and *FileName* were omitted.

Station	Species	DateTimeOriginal	Date	Time	delta.time.mins	delta.time.hours	delta.time.days
StationA	PBE	2009-04-21 00:40:00	2009-04-21	00:40:00	0	0·0	0·0
StationA	PBE	2009-04-22 20:19:00	2009-04-22	20:19:00	2619	43·6	1·8
StationA	PBE	2009-04-23 00:07:00	2009-04-23	00:07:00	226	3·8	0·2
StationA	PBE	2009-05-07 17:11:00	2009-05-07	17:11:00	21182	353·0	14·7
StationA	VTA	2009-04-10 05:07:00	2009-04-10	05:07:00	0	0·0	0·0
StationA	VTA	2009-05-06 19:06:00	2009-05-06	19:06:00	38279	638·0	26·6

Image Data Extraction

After species identification, the function *recordTable* organizes species records in a table containing (at the minimum) station IDs, species names, date and time of records (see Table 2.2). The function *recordTableIndividual* offers analogous capabilities for individually identified animals. In order to use the capabilities of `camtrapR` on record tables from prior work (created manually or with other software), these data sets can easily be converted into the simple data format provided by the *recordTable* functions.

Both functions can extract custom and manufacturer-specific metadata tags from the images. Because metadata tag names are generally unknown, the function *exifTagNames* returns metadata tags and tag names, thereby helping users to identify the relevant tags they wish to include in the tables.

A filter for temporal independence between images of the same species at the same station is implemented (argument minDeltaTime, in minutes). If set to 0, the *recordTable* functions return all records. Any higher number will only return records that were taken at least minDeltaTime minutes after the last record of the same species/individual at the same station or, alternatively, minDeltaTime minutes after the last independent record of the same species/individual. All functions for downstream analyses depend on the results of *recordTable/recordTableIndividual* and thus on the argument minDeltaTime.

Camera Trap Station Information

A simple data frame is used to store information about camera trap stations and, if applicable, individual cameras (see Table 2.3). It contains station/camera IDs, geographic coordinates, setup and retrieval dates, and possibly station-level covariates. It can be created in standard spreadsheet software and imported into R. Periods in which cameras malfunctioned (once or repeatedly) can be defined. Both format and names of date and coordinate columns can be specified by the user.

Based on setup, retrieval and malfunctioning dates, the function *cameraOperation* computes a day-by-station camera operation matrix, coding whether stations were operational, partly operational, not operational (malfunctioning) or not set up. The camera operation matrix reflects the daily trapping effort per station, that is the number of active cameras per station and day. Depending on their placement, multiple cameras within a sampling point can increase the probability of detecting an animal. If cameras are set up directly opposite each other, they may be considered one operational unit. If they are set up further apart, it may be desirable to count them as two units

Table 2.3: Example camera trap station table. *Station* is the camera trap station ID, *utm_y* and *utm_x* are station coordinates. *setup_date* and *retrieval_date* are the dates the stations were set up and retrieved. *Problem1_from* and *Problem1_to* define malfunctioning dates.

Station	**utm_y**	**utm_x**	**detup_date**	**retrieval_data**	**Problem1_from**	**Problem2_from**
StationA	604000	526000	02/04/2009	14/05/2009		
StationB	606000	523000	03/04/2009	16/05/2009		
StationC	607050	525000	04/04/2009	17/05/2009	12/05/2009	17/05/2009

accumulating effort independently. Therefore, the camera operation matrix can return either the number of operational cameras (if effort is accumulated independently) or an indicator for station operability (if effort is not accumulated independently). The camera operation matrix is used in creating detection histories for occupancy and spatial capture–recapture analyses (see description of the functions *detectionHistory* and *spatialDetectionHistory* below).

Data Exploration and Visualization

camtrapR can plot maps of species records (number of observed species by station and number of independent detections by species; see Figure 2.1) with the function detectionMaps. The function allows shapefile export for use in gis software. Single-species activity patterns can be visualized in three different ways: as histograms of hourly activity, activity kernel density estimations and radial plots (functions activityHistogram, activityDensity and activityRadial). Two-species activity overlaps (Ridout & Linkie 2009) are estimated and plotted with the function activityOverlap. These functions use code from the packages overlap and plotrix (Meredith & Ridout 2014; Lemon et al. 2015).

Data Export for Occupancy Analyses

Occupancy models are used to gain insight into species habitat associations while accounting for imperfect detection. The function *detectionHistory* computes species detection/non-detection matrices for use in occupancy models, for example in package unmarked (Fiske & Chandler 2011) or program presence (Hines 2006) by combining the record table created with the function *recordTable* and the camera operation matrix created with the function *cameraOperation*. In the detection/non-detection matrices, rows represent stations and

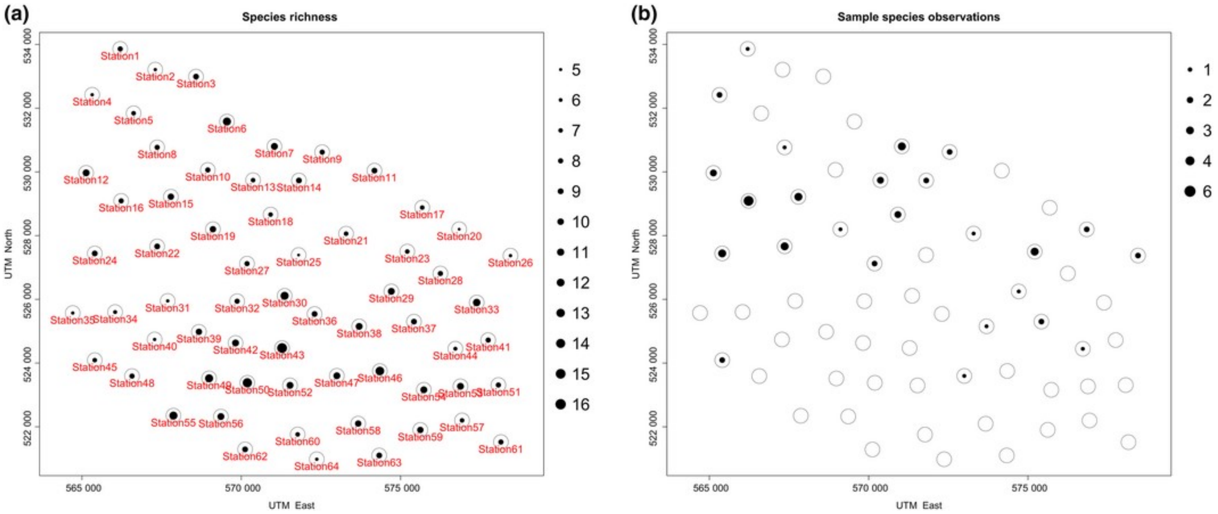


Figure 2.1: Example maps created with the function *detectionMaps*. A) Number of observed species, B) Number of independent observation of an example species.

columns survey occasions. Survey occasions consist of one or more days. The matrix cell becomes 1 if a species was detected at a station during an occasion, 0 in case of non-detection, and NA if the station was not operational. Users have complete freedom over occasion start dates and time, occasion length (in days) and the length of the trapping period per station. Occasions can begin on a fixed date, the day the first station was set up or each station's individual setup date (optionally with a buffer between the setup date and the beginning of the first occasion).

Trapping effort by station and occasion can be returned alongside species detection histories for use as a covariate/offset on detection probability. Incomplete occasions (occasions in which cameras were only partly operational) may contain records in the detection/non-detection matrix if effort is returned. Otherwise, any incomplete occasion will cause corresponding detection matrix cells to be NA.

Data Export for (Spatial) Capture–Recapture Analyses

Spatial capture–recapture methods use repeated detections of marked individuals of a species at an array of sampling locations (camera trap stations) to estimate species density while accounting for imperfect detection and movement of individuals about their home ranges (Efford 2004; Royle & Young 2008; Royle *et al*. 2014). In order to prepare species data for spatial capture–recapture analyses, the function *spatialDetectionHistory* can build capthist objects as defined in the secr package (Efford 2015), containing information about where (station) and when (occasion) individuals were detected. The camera trap station table, the camera operation matrix and the record table are combined for that purpose. The record table needs to contain individual IDs (see sections 'Individual Identification' and 'Image Data Extraction') and may contain individual covariates (from metadata tags). The stations' geographic coordinates and station-level covariates are read from the camera trap station table. The camera operation matrix provides information about station operation dates and trapping effort. In creating the capthist objects, we provide the same flexibility regarding occasion length and starting time as in the function *detectionHistory*. Trapping effort (trap usage) can also be returned in the capthist object. For non-spatial capture–recapture analyses, the function can also return an RMark data frame, containing only individual-by-occasion information without the spatial component.

Creating a Survey Report

The function *surveyReport* summarizes camera trapping surveys. It returns station operation and image date ranges, the number of trap days (total and by station), observed numbers of species and the number of independent observations by species and station. A zip file containing essential data and tables, detection maps and activity plots can be generated. It also contains an example script for reproducing all of these and for creating the input for occupancy analyses. The summary report and zip file can further be used for data sharing and archiving, for example in online repositories such as the Knowledge Network for Biocomplexity (KNB; https://knb.ecoinformatics.org/).

Conclusion

`camtrapR` is the first R package to bridge the gap between camera trap data acquisition and the well-developed downstream analytical tools by providing a workflow for camera trap data management, exploration and preparation of subsequent analyses. Its main advantages are flexibility, ease of use, extensive automation of many of the otherwise labour-intensive tasks, and compatibility with software for further analyses of camera trapping data.

`camtrapR` offers a standardized camera trap data management, and we expect it to be most useful to researchers and practitioners who regularly handle large numbers of camera trap images and wish to generate input for activity, occupancy and/or (spatial) capture–recapture analyses with minimal manual effort. We will keep improving and extending `camtrapR` functionalities and welcome both feedback and collaborations to further increase the usefulness to its users.

Acknowledgements

All authors thank the German Federal Ministry of Education and Research (BMBF FKZ: 01LN1301A) and Point Defiance Zoo & Aquarium for funding this research. We further thank Peter Lagan, Johny Kissing, Albert Radin and Datuk Sam Mannan of the Sabah Forestry Department and Laurentius Ambu of Sabah Wildlife Department for supporting the fieldwork that generated the sample data set. We are grateful to John Mathai, Seth Timothy Wong, An The Truong Nguyen, Andrew Tilker, Nicolas

Deere, Patrick Lorch, Brent Pease and Azlan bin Mohamed for extensive testing of the package and valuable feature requests during development. In addition, we thank three anonymous reviewers whose comments helped us to improve earlier versions of the manuscript and the package.

References

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

Chamberlain, S. A, & Szöcs, E. (2013). taxize: taxonomic search and retrieval in R. F1000Research, 2, 191.

Efford, M. (2004). Density estimation in live-trapping studies. *Oikos*, **106**, 598–610.

Efford, M.G. (2015). secr: Spatially explicit capture-recapture models. R package version 2.9.5. http://cran.r-project.org/package=secr

Fegraus, E.H., Lin, K., Ahumada, J.A., Baru, C., Chandra, S. & Youn, C. (2011). Data acquisition and management software for camera trap data: A case study from the TEAM Network. *Ecological Informatics*, **6**, 345–353.

Fiske, I.J. & Chandler, R.B. (2011). unmarked : An R Package for Fitting Hierarchical Models of Wildlife Occurrence and Abundance. *Journal of Statistical Software*, **43**, 1–23.

Gardner, B., Reppucci, J., Lucherini, M. & Royle, J.A. (2010). Spatially explicit inference for open populations: Estimating demographic parameters from camera-trap studies. *Ecology*, **91**, 3376–3383.

Harris, G., Thompson, R., Childs, J.L. & Sanderson, J.G. (2010). Automatic Storage and Analysis of Camera Trap Data. *Bulletin of the Ecological Society of America*, **91**, 352–360.

Harvey, P. (2015). ExifTool: Read, write and edit meta information. Version 10.05. http://www.sno.phy.queensu.ca/~phil/exiftool/

Hines, J.E. (2006). PRESENCE2 - Software to estimate patch occupancy and related parameters.

Ivan, J. S., & Newkirk, E. S. (2016). CPW Photo Warehouse: a custom database to facilitate archiving, identifying, summarizing and managing photo data collected from camera traps. Methods in Ecology and Evolution, **7**, 499-504

Karanth, K.U. (1995). Estimating tiger Panthera tigris populations from camera-trap data using capture-recapture models. *Biological Conservation*, **71**, 333–338.

Krishnappa, Y.S. & Turner, W.C. (2014). Software for minimalistic data management in large camera trap studies. *Ecological Informatics*, **24**, 11–16.

Laake, J.L. (2013). *RMark: An R Interface for Analysis of Capture-Recapture Data with MARK*. Seattle, WA.

Lemon, J., Bolker, B., Oom, S., Klein, E., Rowlingson, B., Wickham, H., Tyagi, A., Eterradossi, O., Gabor Grothendieck, M.T., Kane, J., Turner, R., Witthoft, C., Stander, J., Petzoldt, T., Duursma, R., Biancotto, E., Levy, O., Dutang, C., Solymos, P., Robby Engelmann, M.H., Steinbeck, F., Borchers, H., Singmann, H., Toal, T. & Ogle, D. (2015). plotrix: Various Plotting Functions. R package version 3.5-12. http://cran.r-project.org/package=plotrix

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

McShea, W.J., Forrester, T., Costello, R., He, Z. & Kays, R. (2016). Volunteer-run cameras as distributed sensors for macrosystem mammal research. Landscape Ecology, **31**, 55–66.

Meredith, M. & Ridout, M. (2014). overlap: Estimates of coefficient of overlapping for animal activity patterns. R package version 0.2.4. http://cran.r-project.org/package=overlap

Miller, D.A., Nichols, J.D., McClintock, B.T., Grant, E.H.C., Bailey, L.L. & Weir, L.A. (2011). Improving occupancy estimation when two types of observational error occur: non-detection and species misidentification. *Ecology*, **92**, 1422–1428.

Mohamed, A., Sollmann, R., Bernard, H., Ambu, L.N., Lagan, P., Mannan, S., Hofer, H. & Wilting, A. (2013). Density and habitat use of the leopard cat (Prionailurus bengalensis) in three commercial forest reserves in Sabah, Malaysian Borneo. *Journal of Mammalogy*, **94**, 82–89.

R Core Team. (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org

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.

Royle, J.A., Chandler, R.B., Sollmann, R. & Gardner, B. (2014). *Spatial Capture-Recapture*. Academic Press, Waltham, Massachusetts, USA.

Royle, J.A., Karanth, K.U., Gopalaswamy, A.M. & Kumar, N.S. (2009). Bayesian inference in camera trapping studies for a class of spatial capture-recapture models. *Ecology*, **90**, 3233–3244.

Royle, J.A. & Young, K. V. (2008). A hierarchical model for spatial capture-recapture data. *Ecology*, **89**, 2281–2289.

Sanderson, J. & Harris, G. (2013). Automatic data organization, storage, and analysis of camera trap pictures. *Journal of Indonesian Natural History*, **1**, 6–14.

Silver, S.C., Ostro, L.E.T., Marsh, L.K., Maffei, L., Noss, A.J., Kelly, M.J., Wallace, R.B., Gómez, H. & Ayala, G. (2004). The use of camera traps for estimating jaguar Panthera onca abundance and density using capture/recapture analysis. *Oryx*, **38**, 148–154.

Sunarto, Sollmann, R., Mohamed, A. & Kelly, M.J. (2013). Camera trapping for the study and conservation of tropical carnivores. *Raffles Bulletin of Zoology*, 21–42.

Sundaresan, S.R., Riginos, C. & Abelson, E.S. (2011). Management and Analysis of Camera Trap Data: Alternative Approaches (Response to Harris et al. 2010). *Bulletin of the Ecological Society of America*, **92**, 188–195.

Tobler, M.W. (2014). Camera base version 1.6.1. Botanical Research Institute of Texas. http://www.atrium-biodiversity.org/tools/camerabase/

Tobler, M.W., Carrillo-Percastegui, S.E., Leite Pitman, R., Mares, R. & Powell, G. (2008). An evaluation of camera traps for inventorying large- and medium-sized terrestrial rainforest mammals. *Animal Conservation*, **11**, 169–178.

White, G.C. & Burnham, K.P. (1999). Program MARK: survival estimation from populations of marked animals. *Bird Study*, **46**, S120–S139.

Yu, X., Wang, J., Kays, R., Jansen, P. a, Wang, T. & Huang, T. (2013). Automated identification of animal species in camera trap images. *EURASIP Journal on Image and Video Processing*, **52**, 1–10.

Zaragozí, B., Belda, A., Giménez, P., Navarro, J.T. & Bonet, A. (2015). Advances in camera trap data management tools: Towards collaborative development and integration with GIS. *Ecological Informatics*, **30**, 6–11.

Supporting Information

Supplementary Table S2.1: Comparison of software for camera trap data management. Only full workflow solutions that create input files for subsequent analyses are shown. Adapted from Ivan & Newkirk (2016)

Feature	**Renamer + CamTrap**	**CameraBase**	**eMammal**	**CPW Photo Warehouse**	**camtrapR**
Website	*http://www.snapfiles.com/downloads/denrenamer/dldenrenamer.html; http://esapubs.org/archive/bulletin/B091/002/suppl-1.htm*	*http://www.atrium-biodiversity.org/tools/camerabase/*	*http://emammal.si.edu/*	*http://cpw.state.co.us/learn/Pages/ResearchMammalsSoftware.aspx*	*https://cran.r-project.org/web/packages/camtrapR/*
Operating system	Windows	Windows	Windows, MacOS	Windows	Windows, MacOS, Linux
Single relational platform	No	**Yes**	No	**Yes**	No
Data storage	local	local	cloud	local	local
Storage capacity	unlimited	c. 2,000,000	unlimited	c. 800,000 - 2,000,000	unlimited
Automatic import[1]	No	**Yes**	**Yes**	**Yes**	**Yes**
Taxonomic checks on species names	No	No	**Yes**	No	**Yes**
Image recognition[2]	No	No	**Yes**	No	No
Assign multiple species	No[3]	No[4]	**Yes**	**Yes**	**Yes**
Double observer ID	No	No	**Yes**	**Yes**	**Yes**

Feature	**Renamer + CamTrap**	**CameraBase**	**eMammal**	**CPW Photo Warehouse**	**camtrapR**
Batch ID	No	**Yes**	**Yes**	**Yes**	**Yes**
Crowd source ID[5]	No	No	No	**Yes**	**Yes**
Assign individual IDs	No	**Yes**	**Yes**	**Yes**	**Yes**
Assign and tabulate custom image tags	No	Sex only	No	Comments only	**Yes**[6]
Extract and tabulate Exif metadata tags[7]	No	Date/Time	Date/Time	Date/Time	**Any**
Record active days[8]	**Yes**	**Yes**	No	**Yes**	**Yes**
Filter, query data	No	**Yes**	**Yes**	**Yes**	**Yes**
Auto-generate input files	**Yes**[9]	**Yes**[10]	**Yes**[11]	**Yes**[12]	**Yes**[13]
Auto-generate reports	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Mapping species records	No	**Yes**	No	No	**Yes**
Direct GIS export	No	No	No	No	**Yes**[14]
Free software	**Yes**	**Yes**	No	**Yes**	**Yes**
Open source software	**Yes**	No[15]	No	No[15]	**Yes**
Usable without coding skills	No	**Yes**	**Yes**	**Yes**	No

[1] Photos and associated metadata can be imported to database structure automatically without the need to manually enter or manipulate data

[2] automatic species identification

[3] Multiple species or number of individuals can be assigned to each photo if the user copies photos to multiple folders

[4] Multiple species can be assigned to each photo if the user imports photos multiple times, once for each species present

[5] Identification of sub-datasets by different users

[6] metadata tags assigned in image management software, e.g. digiKam or Adobe Bridge can be extracted and tabulated

[7] e.g. ambient temperature or air pressure (depending on camera model), or custom metadata tags

[8] Allows users to record and/or manage the days over which each camera was active and operating properly.

[9] Software produces input files for use in Program PRESENCE; limited to a single occasion length (10 days).

[10] Software produces input files for Programs MARK (closed capture), CAPTURE, PRESENCE, R 'RMark' (occupancy), DENSITY, and EstimateS.

[11] Software produces input files for Program PRESENCE and R 'unmarked' and produces output graphs from R 'overlap', and R 'vegan'.

[12] Software produces input files for Programs MARK, PRESENCE, DENSITY, R 'secr', and R 'overlap'.

[13] Software produces input files for R 'unmarked', R 'secr', R 'RMark', R 'overlap', programs MARK, PRESENCE, DENSITY.

[14] Software can save a point shapefile of camera trap station locations with the number independent species observations as attributes.

[15] Microsoft Access® database

CHAPTER 3

Defining habitat covariates in camera-trap based occupancy studies

JÜRGEN NIEDBALLA[1*], RAHEL SOLLMANN[1,2], AZLAN BIN MOHAMED[1], JOHANNES BENDER[1], ANDREAS WILTING[1]

Scientific Reports 5 (2015): 17041 (https://doi.org/10.1038/srep17041)

1) Leibniz Institute for Zoo and Wildlife Research, Alfred-Kowalke-Str. 17, 10315 Berlin, Germany

2) North Carolina State University, Department of Forestry and Environmental Resources, Campus Box 8008, Raleigh, NC 27695-7646, USA

Current address: US Forest Service, Pacific Southwest Research Station, 1731 Research Park Drive, Davis, CA 95618, USA

** corresponding author: niedballa@izw-berlin.de*

Author's contribution statement

JN, RS and AW conceived the ideas for this manuscript. AM and AW collected the field data in Sabah, Borneo. JB conducted the land cover classification. JN conducted the statistical analyses and wrote the manuscript. RS and AW commented the manuscript extensively. All authors read and agreed on the manuscript.

Defining habitat covariates in camera-trap based occupancy studies

JÜRGEN NIEDBALLA[1*], RAHEL SOLLMANN[1,2], AZLAN BIN MOHAMED[1], JOHANNES BENDER[1], ANDREAS WILTING[1]

1) Leibniz Institute for Zoo and Wildlife Research, Alfred-Kowalke-Str. 17, 10315 Berlin, Germany

2) North Carolina State University, Department of Forestry and Environmental Resources, Campus Box 8008, Raleigh, NC 27695-7646, USA

Current address: US Forest Service, Pacific Southwest Research Station, 1731 Research Park Drive, Davis, CA 95618, USA

Abstract

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.

Introduction

Understanding the distribution and habitat associations and of wildlife species is a key topic in ecology, and important for their conservation (Guisan *et al.* 2013). Studying wildlife habitat associations requires appropriate definition of environmental covariates at spatial scales that are relevant to the species under study (Mayor *et al.* 2009). 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 (Mohamed *et al.* 2013), and information based on remote sensing or airborne land cover analyses (Burton *et al.* 2012; Gould *et al.* 2012; Vierling *et al.* 2013).

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 (Kanagaraj *et al.* 2013) 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 (Saatchi *et al.* 2011), to 250 m resolution MODIS or land cover data (Miettinen *et al.* 2012), 30-m Landsat imagery and derived data (Hansen *et al.* 2013) to high-resolution (<1m) satellite imagery.

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 <1m resolution and allowing for three-dimensional imaging) have overcome these problems and shown that fine-scale variations in habitat structure can influence species distributions (Goetz *et al.* 2010; Palminteri *et al.* 2012). 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 (Betts *et al.* 2006). 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 (Levin 1992; Rahbek 2005), and an adequate definition of spatial scale is important when modelling species-habitat associations (Wiens 1989; Holland *et al.* 2004).

In an ecological context, scale is the spatial (or temporal) dimension of an object or process, characterized by grain and extent (Turner *et al.* 1989; Wiens 1989; Schneider 2001). Here, grain is the spatial resolution of remote sensing data (i.e., pixel size of a raster data set), and extent is characterized by focal patches of different sizes, i.e. circular areas of different radii surrounding the sampling points (Thornton *et al.* 2010).

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 (Mazerolle & Villard 1999). 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 (Lechner *et al.* 2012).

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 (MacKenzie *et al.* 2002, 2003, 2006), 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 km², 5°14’-28’N, 117°20’-38’E), Tangkulap-Pinangah Forest Reserve (TFR; 501 km², 5°17’-31’N, 117°03’-20’E) and Segaliud Lokan Forest Reserve (SLFR; 5°20’-39’N, 573 km², 117°25’-39’E; Figure 3.1). 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 (Lagan *et al.* 2007; Kitayama 2013; Mohamed *et al.* 2013).

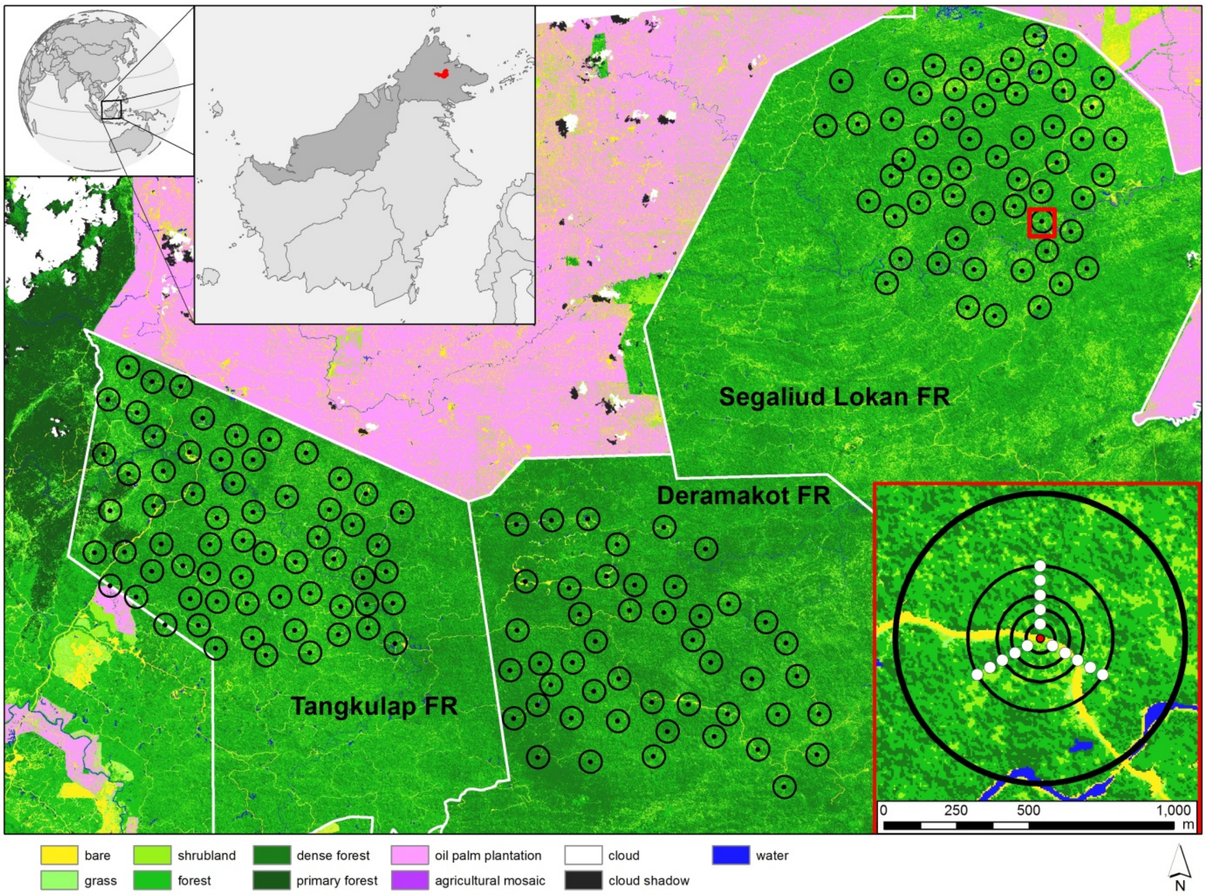


Figure 3.1: Map of the study site in Sabah, Malaysian Borneo.

Camera-trapping

We set up 47, 64, and 55 camera-trap stations covering areas of 123 km², 122 km², and 114 km² in DFR, TFR, and SLFR, respectively (Figure 3.1). 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 Mohamed *et al.* 2013).

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 (MacKenzie *et al.* 2002, 2003, 2006). They consist of two components that explicitly model the ecological process (i.e. occupancy of sites) and the observation process (Royle & Dorazio 2008). The true occupancy state at site *i*, z_i (1 if present, and 0 otherwise) is considered a Bernoulli trial with probability of occupancy Ψ_i : $z_i \sim$ Bernoulli(Ψ_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 $y_{ik}|z_i \sim$ Bernoulli($p_{ik}\, z_i$). Thus $p_{ik} = 0$ where $z_i = 0$, i.e., the species is not present.

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

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

where β_0 and γ_0 denote the intercepts andβ_1 and γ_1 single regression coefficients (MacKenzie *et al.* 2006). 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 (Mohamed *et al.* 2013). 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 (MacKenzie *et al.* 2006) in R 3.1.1(R Core Team 2014) using package “unmarked” version 0.10-3 (Fiske & Chandler 2011). 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 S3.1). These models will be termed ‘constant occupancy models’ for the sake of simplicity. Model selection was based on Akaike’s Information Criterion (AIC, Burnham & Anderson 2002). 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 *Hemigalusderbyanus* (*n* = 35 records), Long-tailed Macaque *Macacafascicularis* (*n* = 76), Malay Civet *Viverratangalunga* (*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 (Colón 2002), Long-tailed Macaque (Wheatley 1980) and Chevrotains (Matsubayashi *et al.* 2003) 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 (RapidEye AG 2012).

To reduce scene-to-scene variability, radiometric corrections were applied as recommended for multi-temporal and multi-sensor data applications (Lu *et al.* 2002; Chander *et al.* 2009). The image-based atmospheric corrections included 'dark object subtraction' and conversion to exoatmospheric (top-of-atmosphere) reflectance (Chavez 1996; RapidEye AG 2012; Vanonckelen *et al.* 2013). By applying pixel-based maximum-likelihood land cover classifications, nine different land cover types were identified (Figure 3.1). Clouds and cloud shadows were eliminated consulting a Landsat-based classification (Langner *et al.* 2012). 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 (Pielou 1966; Farina 1997; Mairota *et al.* 2013). 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 (Rondinini *et al.* 2011). Distance to oil palm plantations quantifies a potential edge effects can be interpreted as a proxy for human disturbance (Brodie *et al.* 2014). Forest score and heterogeneity are both metrics to characterize 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 150m 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. CCis related to forest disturbance: less disturbed forests are expected to have a more closed canopy, i.e. higher values of CC (Mohamed *et al.* 2013).

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) ($D.PLANT_5$, $D.PLANT_{30}$ and so on, Figure 3.2). The oil palm plantations represent a large continuous habitat feature, for which distances remained largely constant across spatial resolutions (Figure 3.2C, 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 (Figure 3.2B); 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 (Figure 3.2A).

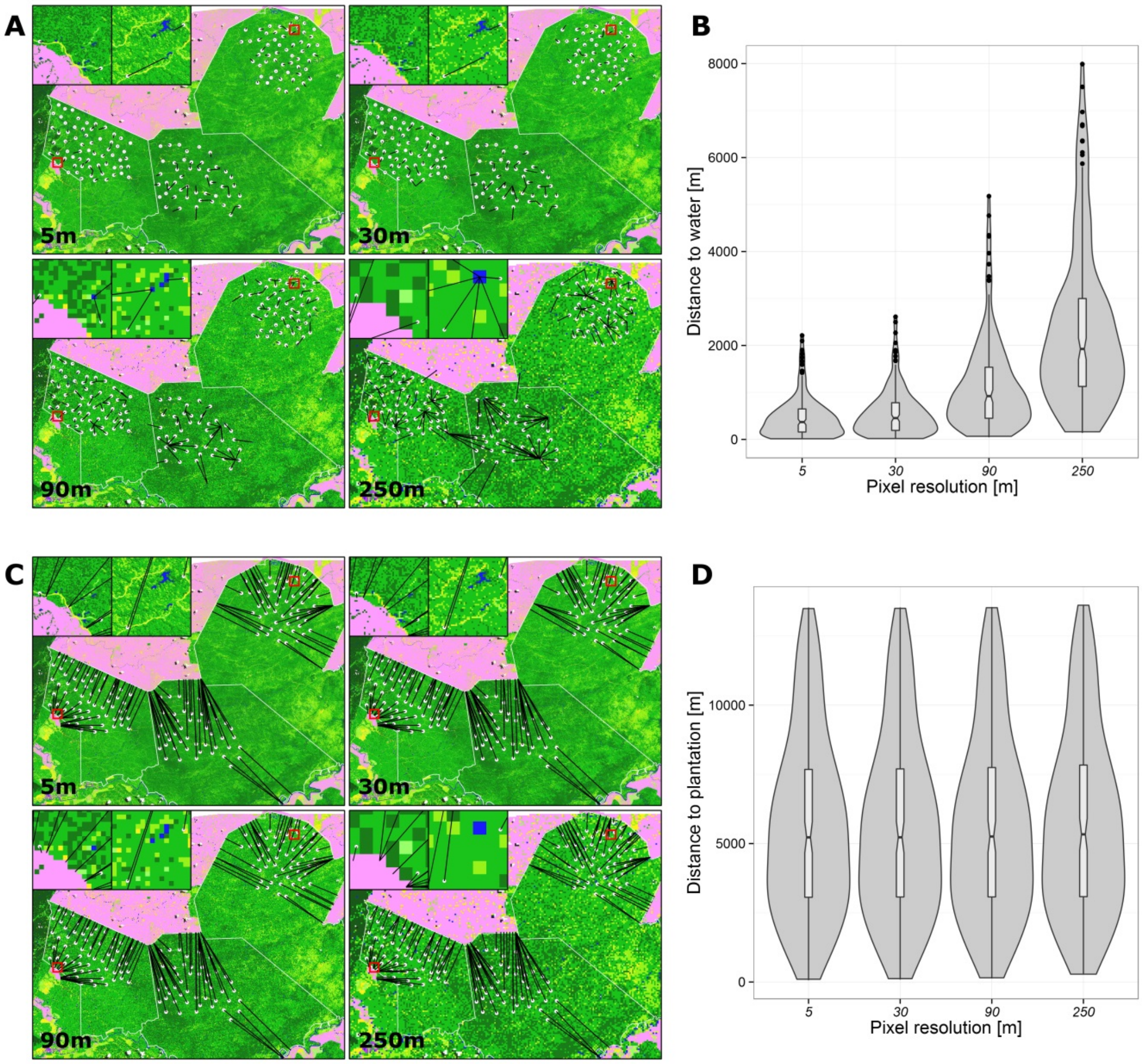


Figure 3.2: Maps and violin plots for distance to water (A,B) and distance to oil palm plantations (C,D) by pixel resolution (grain size) for three commercial forest reserves in Sabah, Malaysian Borneo

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 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 ΔAIC for each radius over all six species. A lower cumulative ΔAIC indicates that a given radius is, on average, closer to the top model than one with a higher cumulative ΔAIC.

Goodness of model fit

Because AIC is only a relative measure of model quality, we conducted goodness of fit tests (MacKenzie & Bailey 2004) for each species' global model, with covariates based on the consensus radius, using the R package AICcmodavg (Marc J. Mazerolle 2015). 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 S3.7) 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 resolution of land cover information. Regression coefficients and standard errors were very similar (±0.02) for spatial resolutions from 5 to 250 m, AIC values of individual models were virtually constant with ΔAIC < 0.2, and AIC weights hardly differed across resolutions (see Table 3.1 for the Long-tailed Macaque and Supplementary Table S3.2 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 (β_1 = -3.59 ± 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 3.1). 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 S3.3). 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.

Table 3.1: 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.

Pixel size	AIC	ΔAIC	wAIC	β_1*	SE	CV	p-value**
Distance to oil palm plantation							
90	331.48	0	0.250	-0.72	0.29	0.4	**0.012**
5	331.5	0.02	0.247	-0.72	0.29	0.4	**0.012**
30	331.5	0.02	0.247	-0.72	0.29	0.4	**0.012**
250	331.55	0.07	0.241	-0.72	0.29	0.4	**0.012**
-	337.05	5.57	0.015	-	-	-	-
Distance to water							
5	301.13	0	0.997	-3.59	0.92	0.26	**<0.001**
30	312.86	11.73	0.003	-2.32	0.64	0.28	**<0.001**
90	330.35	29.22	0	-0.96	0.39	0.41	**0.014**
250	335.41	34.28	0	-0.49	0.27	0.55	0.074
-	337.05	35.92	0	-	-	-	-

ΔAIC: relative difference in AIC to top model, wAIC = AIC model weights, β1 = regression coefficient, SE = regression coefficient standard error, CV = coefficient of variation of β1 (SE / |β1|), – 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*

Table 3.2: 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.

Radius	AIC	ΔAIC	wAIC	β_1*	SE	CV	p-value**
Remote Sensing – Forest Score							
10	332.48	0	0.5	-0.68	0.27	0.4	**0.013**
50	334.61	2.13	0.173	-0.55	0.27	0.49	**0.041**
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
Remote Sensing - Heterogeneity							
50	329.7	0	0.455	1.03	0.38	0.37	**0.007**
10	330.87	1.17	0.253	1.02	0.42	0.41	**0.014**
100	330.92	1.22	0.247	0.83	0.32	0.39	**0.009**
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
In-situ - Canopy Closure							
50	326.53	0	0.754	-1.1	0.37	0.34	**0.003**
100	329.18	2.65	0.201	-1.07	0.42	0.39	**0.01**
150	332.33	5.8	0.042	-0.91	0.42	0.46	**0.032**
-	337.05	10.52	0.004	-	-		-

ΔAIC: relative difference in AIC to top model, wAIC = AIC model weights, β1 = regression coefficient, SE = regression coefficient standard error, CV = coefficient of variation of β1 (SE / |β1|), – 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*

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 3.2 for the Long-tailed Macaque and Supplementary Tables S3.4-S3.6 for other species). Particularly for the Long-tailed Macaque, the species with the strongest associations 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 ΔAIC, we chose 50 m as the consensus radius for all covariates (Table 3.3).

Discussion

Species habitat associations and preferences are multi-factorial processes aimed at maximizing fitness (Chalfoun & Martin 2007) 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 (Orians & Wittenberger 1991; Mayor *et al.* 2009). This is commonly achieved by analysing hierarchical selection of habitat use (e.g. at landscape, macrohabitat and microhabitat scale; Saab 1999; Williams *et al.* 2002; Mayor *et al.* 2009) or by comparing multiple focal patch sizes (Holland *et al.* 2004; Thornton *et al.* 2010). 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 (Figure 3.2). 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 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 (Figure 3.2B). In the set of analysed species the Long-tailed Macaque is known to be closely associated with water resources, namely rivers (Rodman 1991). 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 significantly negative association with distance to water at all resolutions, i.e. occurred more frequently near water. In contrast to the Long-tailed Macaque, a 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 (Rahbek 2005) adequate definition of the focal patch sizes is important (Wiens 1989; Holland *et al.* 2004). 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 (Mayor *et al.* 2009).

Table 3.3: Cumulative ΔAIC 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 (m)	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	-

'Extent' refers to the radius around camera-trap stations from which covariate values were extracted. Cumulative ΔAIC was calculated for each radius over all six species. A lower cumulative ΔAIC indicates that a given radius is, on average, closer to the top model than one with a higher cumulative ΔAIC.

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 agreat 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 (Langner *et al.* 2012), 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 (Wheatley 1980; Colón 2002; Matsubayashi *et al.* 2003). 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 (Spearmans $\rho = 0.58$ 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 (Shepard Jr. *et al.* 2012). 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 < 1m, e.g. WorldView, GeoEye or Quickbird) could further improve our ability to study species habitat association. Even though three-dimensional and canopy structural LiDAR data (Vierling *et al.* 2013) have great advantages over satellite imagery, we expect that due to the high costs of LiDAR data (Hummel *et al.* 2011), 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.

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.

References

Betts, M.G., Diamond, A.W., Forbes, G.J., Villard, M.A. & Gunn, J.S. (2006). The importance of spatial autocorrelation, extent and resolution in predicting forest bird occurrence. *Ecological Modelling*, **191**, 197–224.

Brodie, J.F., Giordano, A.J. & Ambu, L. (2014). Differential responses of large mammals to logging and edge effects. *Mammalian Biology - Zeitschrift für Säugetierkunde*, **80**, 7–13.

Burnham, K.P. & Anderson, D.R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd edition. Springer Verlag New York Inc.

Burton, A.C., Sam, M.K., Balangtaa, C. & Brashares, J.S. (2012). Hierarchical Multi-Species Modeling of Carnivore Responses to Hunting, Habitat and Prey in a West African Protected Area. *PloS one*, **7**, e38007.

Chalfoun, A.D. & Martin, T.E. (2007). Assessments of habitat preferences and quality depend on spatial scale and metrics of fitness. *Journal of Applied Ecology*, **44**, 983–992.

Chander, G., Markham, B.L. & Helder, D.L. (2009). Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sensing of Environment*, **113**, 893–903.

Chavez, P.S. (1996). Image-Based Atmospheric Corrections - Revisited and Improved. *Photogrammetric Engineering & Remote Sensing*, **62**, 1025–1036.

Colón, C.P. (2002). Ranging behaviour and activity of the Malay civet (Viverra tangalunga) in a logged and an unlogged forest in Danum Valley, East Malaysia. *Journal of Zoology*, **257**, 473–485.

Farina, A. (1997). Landscape structure and breeding bird distribution in a sub-Mediterranean agro-ecosystem. *Landscape Ecology*, **12**, 365–378.

Fiske, I.J. & Chandler, R.B. (2011). unmarked : An R Package for Fitting Hierarchical Models of Wildlife Occurrence and Abundance. *Journal of Statistical Software*, **43**, 1–23.

Goetz, S.J., Steinberg, D., Betts, M.G., Holmes, R.T., Doran, P.J., DUbayah, R. & Hofton, M. (2010). Lidar remote sensing variables predict breeding habitat of a Neotropical migrant bird. *Ecology*, **91**, 1569–1576.

Gould, W.R., Patla, D. a., Daley, R., Corn, P.S., Hossack, B.R., Bennetts, R. & Peterson, C.R. (2012). Estimating Occupancy in Large Landscapes: Evaluation of Amphibian Monitoring in the Greater

Yellowstone Ecosystem. *Wetlands*, **32**, 379–389.

Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I.T., Regan, T.J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T.G., Rhodes, J.R., Maggini, R., Setterfield, S. a, Elith, J., Schwartz, M.W., Wintle, B. a, Broennimann, O., Austin, M., Ferrier, S., Kearney, M.R., Possingham, H.P. & Buckley, Y.M. (2013). Predicting species distributions for conservation decisions. *Ecology Letters*, **16**, 1424–1435.

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

Holland, J.D., Bert, D.G. & Fahrig, L. (2004). Determining the Spatial Scale of Species' Response to Habitat. *BioScience*, **54**, 227–233.

Hummel, S., Hudak, A.T., Uebler, E.H., Falkowski, M.J. & Megown, K.A. (2011). A Comparison of Accuracy and Cost of LiDAR versus Stand Exam Data for Landscape Management on the Malheur National Forest. *Journal of Forestry*, **109**, 267–273.

Kanagaraj, R., Wiegand, T., Mohamed, A. & Kramer-Schadt, S. (2013). Modelling species distributions to map the road towards carnivore conservation in the tropics. *The Raffles Bulletin of Zoology Supplement No.*, **28**, 85–107.

Kitayama, K. (2013). *Co-benefits of Sustainable Forestry - Ecological Studies of a Certified Bornean Rain Forest* (K. Kitayama, Ed.). Springer.

Lagan, P., Mannan, S. & Matsubayashi, H. (2007). Sustainable use of tropical forests by reduced-impact logging in Deramakot Forest Reserve, Sabah, Malaysia. *Ecological Research*, **22**, 414–421.

Langner, A., Samejima, H., Ong, R.C., Titin, J. & Kitayama, K. (2012). Integration of carbon conservation into sustainable forest management using high resolution satellite imagery: A case study in Sabah, Malaysian Borneo. *International Journal of Applied Earth Observation and Geoinformation*, **18**, 305–312.

Lechner, A.M., Langford, W.T., Jones, S.D., Bekessy, S.A. & Gordon, A. (2012). Investigating species-environment relationships at multiple scales: Differentiating between intrinsic scale and the modifiable areal unit problem. *Ecological Complexity*, **11**, 91–102.

Levin, S.A. (1992). The problem of pattern and scale in ecology. *Ecology*, **73**, 1943–1967.

Lu, D., Mausel, P., Brondizio, E. & Moran, E. (2002). Assessment of atmospheric correction methods for Landsat TM data applicable to Amazon basin LBA research. *International Journal of Remote Sensing*, **23**, 2651–2671.

MacKenzie, D.I. & Bailey, L.L. (2004). Assessing the fit of site-occupancy models. *Journal of Agricultural, Biological, and Environmental Statistics*, **9**, 300–318.

MacKenzie, D.I., Nichols, J.D., Hines, J.E., Knutson, M.G. & Frankling, A.B. (2003). Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology*, **84**, 2200–2207.

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

MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L. & Hines, J.E. (2006). *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurence*. Academic Press, London, United Kingdom.

Mairota, P., Cafarelli, B., Boccaccio, L., Leronni, V., Labadessa, R., Kosmidou, V. & Nagendra, H. (2013). Using landscape structure to develop quantitative baselines for protected area monitoring. *Ecological Indicators*, **33**, 82–95.

Marc J. Mazerolle. (2015). AICcmodavg: Model selection and multimodel inference based on (Q)AIC(c). R package version 2.0-3.

Matsubayashi, H., Bosi, E. & Kohshima, S. (2003). Activity and habitat use of lesser mouse-deer (Tragulus javanicus). *Journal of Mammalogy*, **84**, 234–242.

Mayor, S.J., Schneider, D.C., Schaefer, J.A. & Mahoney, S.P. (2009). Habitat selection at multiple scales. *Ecoscience*, **16**, 238–247.

Mazerolle, M.J. & Villard, M.-A. (1999). Patch characteristics and landscape context as predictors of species presence and abundance: A review. *Ecoscience*, **6**, 117–124.

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

Mohamed, A., Sollmann, R., Bernard, H., Ambu, L.N., Lagan, P., Mannan, S., Hofer, H. & Wilting, A.

(2013). Density and habitat use of the leopard cat (Prionailurus bengalensis) in three commercial forest reserves in Sabah, Malaysian Borneo. *Journal of Mammalogy*, **94**, 82–89.

Orians, G.H. & Wittenberger, J.F. (1991). Spatial and Temporal Scales in Habitat Selection. *The American Naturalist*, **137**, S29–S49.

Palminteri, S., Powell, G.V.N., Asner, G.P. & Peres, C. a. (2012). LiDAR measurements of canopy structure predict spatial distribution of a tropical mature forest primate. *Remote Sensing of Environment*, **127**, 98–105.

Pielou, E.C. (1966). The Measurement of Diversity in Different Types of Biological Collections. *Journal of Theoretical Biology*, **13**, 131–144.

R Core Team. (2014). R: A language and environment for statistical computing.

Rahbek, C. (2005). The role of spatial scale and the perception of large-scale species-richness patterns. *Ecology Letters*, **8**, 224–239.

RapidEye AG. (2012). Satellite Imagery Product Specifications Version 4.1. URL https://earth.esa.int/documents/10174/896711/RE_Product_Specifications_ENG.pdf [accessed 4 July 2013]

Rodman, P.S. (1991). Structural Differentiation of Microhabitats of Sympatric Macaca fascicularis and M. nemestrina in East Kalimantan, Indonesia. *International Journal of Primatology*, **12**, 357–375.

Rondinini, C., Di Marco, M., Chiozza, F., Santulli, G., Baisero, D., Visconti, P., Hoffmann, M., Schipper, J., Stuart, S.N., Tognelli, M.F., Amori, G., Falcucci, A., Maiorano, L. & Boitani, L. (2011). Global habitat suitability models of terrestrial mammals. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **366**, 2633–2641.

Royle, J.A. & Dorazio, R.M. (2008). *Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations and communities*. Academic Press.

Saab, V. (1999). Importance of Spatial Scale to Habitat Use by Breeding Birds in Riparian Forests: a Hierarchical Analysis. *Ecological Applications*, **9**, 135–151.

Saatchi, S.S., Harris, N.L., Brown, S., Lefsky, M., Mitchard, E.T. a, Salas, W., Zutta, B.R., Buermann, W., Lewis, S.L., Hagen, S., Petrova, S., White, L., Silman, M. & Morel, A. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 9899–904.

Schneider, D.C. (2001). The Rise of the Concept of Scale in Ecology. *BioScience*, **51**, 545–554.

Shepard Jr., G.H., Levi, T., Neves, E.G., Peres, C.A. & Yu, D.W. (2012). Hunting in Ancient and Modern Amazonia: Rethinking Sustainability. *American Anthropologist*, **114**, 652–667.

Thornton, D.H., Branch, L.C. & Sunquist, M.E. (2010). The influence of landscape, patch, and within-patch factors on species presence and abundance: a review of focal patch studies. *Landscape Ecology*, **26**, 7–18.

Turner, M.G., Dale, V.H. & Gardner, R.H. (1989). Predicting across scales: Theory development and testing. *Landscape Ecology*, **3**, 245–252.

Vanonckelen, S., Lhermitte, S. & Van Rompaey, A. (2013). The effect of atmospheric and topographic correction methods on land cover classification accuracy. *International Journal of Applied Earth Observation and Geoinformation*, **24**, 9–21.

Vierling, L.A., Vierling, K.T., Adam, P. & Hudak, A.T. (2013). Using satellite and airborne LiDAR to model woodpecker habitat occupancy at the landscape scale. *PloS one*, **8**, e80988.

Wheatley, B.P. (1980). Feeding and ranging of east Bornean Macaca fascicularis. *The Macaques: Studies in Ecology, Behavior, and Evolution* (ed D.G. Lindburg), pp. 215–46. Van Nostrand Reinhold, New York.

Wiens, J.A. (1989). Spatial scaling in ecology. *Functional Ecology*, **3**, 385–397.

Williams, S.E., Marsh, H. & Winter, J. (2002). Spatial Scale, Species Diversity, and Habitat Structure: Small Mammals in Autralian Tropical Rain Forest. *Ecology*, **83**, 1317–1329.

Supporting Information

S 3.1: Best constant occupancy models with covariates on the detection component while holding occupancy constant (based on AIC) for six mammal species/species groups, estimated from camera-trapping data collected between 2008 and 2010 in three commercial forest reserves in Sabah, Malaysian Borneo. γ_1(SE) are estimates of regression coefficients for detection probability covariates with their standard errors on the logit scale; 2.5% and 97.5% CI are confidence interval bounds. Occupancy probability ψ is held constant within each model.

Species	Parameter*	γ_1(SE)	2.5% CI	97.5% CI	p-value**
Banded Civet	Roads	-1.06 (0.521)	-2.09	-0.04	**0.04**
	TFR/SLFR	-3.22 (0.648)	-4.49	-1.95	**<0.001**
Long-tailed Macaque	Roads	1.18 (0.532)	0.14	2.22	**0.02**
	SLFR	-1.98 (0.631)	-3.22	-0.74	**0.002**
Malay Civet	Roads	0.822 (0.158)	0.51	1.13	**<0.001**
	TFR	-0.797 (0.169)	-1.13	-0.47	**<0.001**
Moonrat	TFR/SLFR	-0.66 (0.288)	-1.22	-0.10	**0.02**
Chevrotain	TFR/SLFR	-0.732 (0.158)	-1.04	-0.42	**<0.001**
Thick-spined Porcupine	Roads	-1.046 (0.557)	-2.14	0.05	**0.06**
	TFR/SLFR	-1.444 (0.668)	-2.75	-0.14	**0.03**

**Roads = effect of camera-trap position on/off-road on detection probability p (positive coefficient signifies higher p on roads, negative coefficient higher p off roads);*

TFR: Tangkulap (TFR) has different detection probability than both Segaliud Lokan (SLFR) and Deramakot (DFR)(negative coefficient signifies lower p in TFR compared to DFR and SLFR)

SLFR: SLFR has different detection probability than both TFR and DFR (negative coefficient signifies lower p in SLFR compared to TFR and DFR)

TFR/SLFR: TFR and SLFR both show a different detection probability than DFR (negative coefficient signifies lower p in TFR and SLFR as compared to DFR)

***Bold font indicates significance at the 0.05 level*

S 3.2: Results of occupancy models using distance to oil palm plantation at different spatial resolutions as covariate; estimated from camera-trapping data on six mammal species/species groups, collected between 2008 and 2010 in three commercial forest reserves in Sabah, Malaysian Borneo. ΔAIC: relative difference in AIC to top model, wAIC= AIC model weights, β_1 = regression coefficient, SE = regression coefficient standard error, CV = coefficient of variation of β_1 (SE / $|\beta_1|$), – denotes constant occupancy model.

Species	Pixel size	AIC	ΔAIC	*w*AIC	β_1*	SE	CV	p-value**
Banded Civet	250	195.80	0	0.248	1.38	0.65	0.47	**0.034**
	90	195.81	0.01	0.246	1.38	0.65	0.47	**0.035**
	5	195.83	0.03	0.243	1.37	0.65	0.47	**0.035**
	30	195.83	0.03	0.243	1.37	0.65	0.47	**0.035**
	-	200.83	5.03	0.020	-	-	-	-
Long-tailed Macaque	90	331.48	0	0.250	-0.72	0.29	0.4	**0.012**
	5	331.50	0.02	0.247	-0.72	0.29	0.4	**0.012**
	30	331.50	0.02	0.247	-0.72	0.29	0.4	**0.012**
	250	331.55	0.07	0.241	-0.72	0.29	0.4	**0.012**
	-	337.05	5.57	0.015	-	-	-	-
Malay Civet	-	1344.11	0	0.383	-	-	-	-
	250	1345.91	1.8	0.155	0.13	0.30	2.31	0.66
	5	1345.93	1.82	0.154	0.13	0.30	2.31	0.673
	30	1345.93	1.82	0.154	0.13	0.30	2.31	0.671
	90	1345.94	1.83	0.153	0.12	0.30	2.5	0.681
Moonrat	-	629.62	0	0.403	-	-	-	-
	5	631.61	1.99	0.149	0.02	0.18	9	0.913
	30	631.61	1.99	0.149	0.02	0.18	9	0.912
	90	631.61	1.99	0.149	0.02	0.18	9	0.916
	250	631.61	1.99	0.149	0.02	0.18	9	0.932
Chevrotain (Greater & Lesser)	5	1304.20	0	0.227	0.38	0.21	0.55	0.066
	90	1304.20	0	0.227	0.38	0.21	0.55	0.066
	30	1304.21	0.01	0.226	0.38	0.21	0.55	0.067
	250	1304.27	0.07	0.220	0.38	0.21	0.55	0.069
	-	1305.84	1.64	0.100	-	-	-	-
Thick-spined Porcupine	30	250.70	0	0.231	-0.53	0.30	0.57	0.079
	5	250.71	0.01	0.231	-0.53	0.30	0.57	0.079
	90	250.80	0.1	0.220	-0.52	0.30	0.58	0.083
	250	250.89	0.19	0.211	-0.51	0.30	0.59	0.086
	-	252.22	1.52	0.108	-	-	-	-

** Positive regression coefficients indicate positive association with distance to features, i.e. negative association with features. Negative regression coefficients indicate negative association with distance to features, i.e. positive association with features.*

***Bold font indicates significance at the 0.05 level*

S 3.3: Results of occupancy model using distance to water at different spatial resolutions as covariate; estimated from camera-trapping data on six mammal species/species groups, collected between 2008 and 2010 in three commercial forest reserves in Sabah, Malaysian Borneo. ΔAIC: relative difference in AIC to top model, wAIC= AIC model weights, β_1 = regression coefficient, SE = regression coefficient standard error, CV = coefficient of variation of β_1 (SE / $|\beta_1|$), – denotes constant occupancy model.

Species	Pixel size	AIC	ΔAIC	wAIC	β_1*	SE	CV	p-
Banded Civet	90	198.49	0	0.532	0.71	0.38	0.54	0.062
	-	200.83	2.34	0.165	-	-	-	-
	30	201.09	2.6	0.145	0.44	0.36	0.82	0.22
	5	201.87	3.38	0.098	0.33	0.36	1.09	0.354
	250	202.83	4.34	0.061	-0.03	0.48	16	0.948
Long-tailed Macaque	5	301.13	0	0.997	-3.59	0.92	0.26	**<0.001**
	30	312.86	11.73	0.003	-2.32	0.64	0.28	**<0.001**
	90	330.35	29.22	0	-0.96	0.39	0.41	**0.014**
	250	335.41	34.28	0	-0.49	0.27	0.55	0.074
	-	337.05	35.92	0	-	-	-	-
Malay Civet	-	1344.11	0	0.331	-	-	-	-
	30	1344.85	0.74	0.229	0.32	0.32	1	0.309
	5	1345.47	1.36	0.167	0.23	0.31	1.35	0.46
	250	1345.83	1.72	0.140	0.15	0.29	1.93	0.607
	90	1345.93	1.82	0.133	0.13	0.31	2.38	0.686
Moonrat	-	629.62	0	0.374	-	-	-	-
	250	630.84	1.22	0.203	-0.16	0.18	1.12	0.384
	90	631.54	1.92	0.143	-0.05	0.18	3.6	0.778
	30	631.58	1.96	0.141	0.04	0.17	4.25	0.83
	5	631.62	2	0.138	0.01	0.17	17	0.976
Chevrotain (Greater & Lesser)	250	1300.83	0	0.844	0.59	0.26	0.44	**0.02**
	-	1305.84	5.01	0.069	-	-	-	-
	90	1307.22	6.39	0.035	0.16	0.21	1.31	0.449
	30	1307.73	6.9	0.027	0.06	0.19	3.17	0.745
	5	1307.81	6.98	0.026	0.03	0.19	6.33	0.882
Thick-spined Porcupine	90	243.04	0	0.604	-1.37	0.54	0.39	**0.011**
	30	245.75	2.71	0.156	-1.08	0.46	0.43	**0.019**
	250	245.88	2.84	0.147	-0.96	0.4	0.42	**0.016**
	5	246.93	3.89	0.086	-0.99	0.46	0.46	**0.03**
	-	252.22	9.18	0.006	-	-	-	-

** Positive regression coefficients indicate positive association with distance to features, i.e. negative association with features. Negative regression coefficients indicate negative association with distance to features, i.e. positive association with features.*

***Bold font indicates significance at the 0.05 level*

S 3.4: Results of occupancy models using forest score (index of forest quality, ranging from 0 = bare land/oil palm plantation to 3 = dense forest) extracted for different focal patch sizes (radius) around camera traps, as covariate; estimated from camera-trapping data on six mammal species/species groups, collected between 2008 and 2010 in three commercial forest reserves in Sabah, Malaysian Borneo. ΔAIC: relative difference in AIC to top model, wAIC= AIC model weights, β_1 = regression coefficient, SE = regression coefficient standard error, CV = coefficient of variation of β_1 (SE / $|\beta_1|$), – denotes constant occupancy model.

Species	radius	AIC	ΔAIC	wAIC	β_1*	SE	CV	p-value**
Banded Civet	500	197.16	0	0.55	-1.76	0.92	0.52	0.055
	250	200.12	2.96	0.125	-1.34	0.95	0.71	0.161
	10	200.63	3.47	0.097	0.62	0.43	0.69	0.152
	-	200.83	3.67	0.088	-	-	-	-
	50	201.23	4.07	0.072	0.57	0.46	0.81	0.215
	150	202.63	5.47	0.036	-0.5	1.64	3.28	0.759
	100	202.81	5.65	0.033	-0.1	0.75	7.5	0.892
Long-tailed Macaque	10	332.48	0	0.5	-0.68	0.27	0.40	**0.013**
	50	334.61	2.13	0.173	-0.55	0.27	0.49	**0.041**
	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	336.80	4.32	0.058	-0.52	0.39	0.75	0.185
	-	337.05	4.57	0.051	-	-	-	-
	500	338.10	5.62	0.03	-0.27	0.28	1.04	0.339
Malay Civet	10	1339.60	0	0.552	-0.72	0.32	0.44	**0.024**
	50	1341.91	2.31	0.173	-0.57	0.30	0.53	0.054
	100	1343.04	3.44	0.099	-0.5	0.30	0.60	0.091
	-	1344.11	4.51	0.058	-	-	-	-
	150	1344.13	4.53	0.057	-0.41	0.29	0.71	0.168
	250	1344.90	5.3	0.039	-0.33	0.29	0.88	0.269
	500	1346.03	6.43	0.022	-0.07	0.26	3.71	0.773
Moonrat	100	626.41	0	0.322	0.44	0.20	0.45	**0.029**
	50	626.55	0.14	0.301	0.43	0.20	0.47	**0.03**
	150	628.22	1.81	0.13	0.35	0.20	0.57	0.076
	10	628.59	2.18	0.108	0.32	0.19	0.59	0.087
	-	629.62	3.21	0.065	-	-	-	-
	250	630.59	4.18	0.04	0.19	0.19	1	0.318
	500	630.95	4.54	0.033	0.15	0.18	1.20	0.414
Chevrotain (Greater & Lesser)	100	1297.52	0	0.309	0.63	0.21	0.33	**0.003**
	150	1298.04	0.52	0.239	0.63	0.22	0.35	**0.005**
	500	1298.60	1.08	0.18	0.62	0.22	0.35	**0.005**
	50	1299.11	1.59	0.139	0.57	0.20	0.35	**0.005**
	250	1299.36	1.84	0.123	0.58	0.22	0.38	**0.008**
	-	1305.84	8.32	0.005	-	-	-	-
	10	1306.09	8.57	0.004	0.25	0.19	0.76	0.189
Thick-spined Porcupine	500	248.96	0	0.42	-0.77	0.42	0.55	0.067
	250	251.67	2.71	0.109	-0.48	0.38	0.79	0.204
	150	251.70	2.74	0.107	-0.45	0.31	0.69	0.155
	10	251.76	2.8	0.104	-0.42	0.28	0.67	0.125
	50	252.05	3.09	0.09	-0.4	0.27	0.68	0.146
	100	252.06	3.1	0.089	-0.4	0.28	0.70	0.159
	-	252.22	3.26	0.082	-	-	-	-

** Positive regression coefficients indicate positive association with features. Negative regression coefficients indicate negative association features.*

*** Bold font indicates significance at the 0.05 level*

S 3.5: Results for occupancy models using heterogeneity (index of habitat heterogeneity), extracted for different focal patch sizes (radius) around camera traps, as covariate; estimated from camera-trapping data on six mammal species/species groups, collected between 2008 and 2010 in three commercial forest reserves in Sabah, Malaysian Borneo. ΔAIC: relative difference in AIC to top model, wAIC= AIC model weights, β_1 = regression coefficient, SE = regression coefficient standard error, CV = coefficient of variation of β_1 (SE / $|\beta_1|$), – denotes constant occupancy model.

Species	radius	AIC	ΔAIC	wAIC	β_1*	SE	CV	p-value**
Banded Civet	50	198.99	0	0.393	-0.97	0.61	0.63	0.112
	500	200.70	1.71	0.167	0.52	0.39	0.75	0.188
	-	200.83	1.84	0.156	-	-	-	-
	250	201.69	2.7	0.102	0.57	0.78	1.37	0.465
	100	202.65	3.66	0.063	-0.23	0.55	2.39	0.676
	10	202.73	3.74	0.061	0.17	0.53	3.12	0.745
	150	202.82	3.83	0.058	-0.05	0.49	9.8	0.921
Long-tailed Macaque	50	329.70	0	0.455	1.03	0.38	0.37	0.007
	10	330.87	1.17	0.253	1.02	0.42	0.41	0.014
	100	330.92	1.22	0.247	0.83	0.32	0.39	0.009
	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
Malay Civet	50	1337.90	0	0.866	0.74	0.29	0.39	0.011
	-	1344.11	6.21	0.039	-	-	-	-
	10	1344.46	6.56	0.033	0.3	0.23	0.77	0.179
	150	1345.80	7.9	0.017	0.16	0.28	1.75	0.586
	500	1345.90	8	0.016	0.13	0.27	2.08	0.642
	250	1346.01	8.11	0.015	0.09	0.29	3.22	0.752
	100	1346.08	8.18	0.015	0.05	0.27	5.4	0.855
Moonrat	100	624.73	0	0.495	-0.49	0.19	0.39	0.012
	150	626.05	1.32	0.256	-0.44	0.19	0.43	0.023
	250	627.91	3.18	0.101	-0.36	0.19	0.53	0.062
	50	628.87	4.14	0.063	-0.3	0.18	0.6	0.102
	-	629.62	4.89	0.043	-	-	-	-
	10	630.56	5.83	0.027	0.2	0.19	0.95	0.312
	500	631.61	6.88	0.016	0.02	0.18	9	0.922
Chevrotain (Greater & Lesser)	250	1304.42	0	0.344	-0.36	0.20	0.56	0.068
	-	1305.84	1.42	0.169	-	-	-	-
	100	1306.11	1.69	0.148	-0.26	0.20	0.77	0.197
	150	1306.76	2.34	0.107	-0.2	0.20	1	0.301
	10	1307.12	2.7	0.089	-0.17	0.21	1.24	0.417
	50	1307.45	3.03	0.075	-0.12	0.19	1.58	0.539
	500	1307.67	3.25	0.068	-0.08	0.19	2.38	0.679
Thick-spined Porcupine	50	250.28	0	0.418	-0.48	0.24	0.5	0.048
	-	252.22	1.94	0.158	-	-	-	-
	150	252.75	2.47	0.122	-0.32	0.27	0.84	0.232
	10	253.27	2.99	0.094	-0.24	0.24	1	0.32
	250	253.35	3.07	0.09	-0.24	0.26	1.08	0.356
	500	254.19	3.91	0.059	-0.05	0.25	5	0.851
	100	254.21	3.93	0.059	-0.03	0.26	8.67	0.907

** Positive regression coefficients indicate positive association with features. Negative regression coefficients indicate negative association features.*

**Bold font indicates significance at the 0.05 level*

S 3.6: Results for occupancy models using canopy closure, collected in-situ along transects around camera traps and averaged across transects of different length (radius), as covariate on occupancy; estimated from camera-trapping data on six mammal species/species groups, collected between 2008 and 2010 in three commercial forest reserves in Sabah, Malaysian Borneo. ΔAIC: relative difference in AIC to top model, wAIC= AIC model weights, β_1 = regression coefficient, SE = regression coefficient standard error, CV = coefficient of variation of β_1 (SE / $|\beta_1|$), – denotes constant occupancy model.

Species	radius	AIC	ΔAIC	wAIC	β_1*	SE	CV	p-
Banded Civet	-	200.83	0	0.311	-	-	-	-
	50	200.95	0.12	0.293	0.79	0.57	0.72	0.165
	150	201.51	0.68	0.221	0.7	0.59	0.84	0.236
	100	201.98	1.15	0.175	0.54	0.56	1.04	0.335
Long-tailed Macaque	50	326.53	0	0.754	-1.1	0.37	0.34	**0.003**
	100	329.18	2.65	0.201	-1.07	0.42	0.39	**0.01**
	150	332.33	5.8	0.042	-0.91	0.42	0.46	**0.032**
	-	337.05	10.52	0.004	-	-	-	-
Malay Civet	-	1344.11	0	0.372	-	-	-	-
	50	1344.82	0.71	0.261	-0.34	0.32	0.94	0.29
	100	1345.44	1.33	0.192	-0.25	0.32	1.28	0.438
	150	1345.61	1.5	0.176	-0.21	0.31	1.48	0.502
Moonrat	-	629.62	0	0.467	-	-	-	-
	100	631.51	1.89	0.182	0.06	0.18	3	0.735
	150	631.58	1.96	0.176	0.04	0.18	4.5	0.834
	50	631.59	1.97	0.175	-0.03	0.18	6	0.856
Chevrotain (Greater & Lesser)	150	1288.66	0	0.6	0.83	0.20	0.24	**<0.001**
	100	1290.16	1.5	0.283	0.80	0.20	0.25	**<0.001**
	50	1291.93	3.27	0.117	0.77	0.20	0.26	**<0.001**
	-	1305.84	17.18	0	-	-	-	-
Thick-spined Porcupine	-	252.22	0	0.37	-	-	-	-
	50	253.30	1.08	0.216	-0.28	0.30	1.07	0.344
	100	253.30	1.08	0.216	-0.28	0.29	1.04	0.342
	150	253.47	1.25	0.199	-0.25	0.28	1.12	0.387

** Positive regression coefficients indicate positive association with features. Negative regression coefficients indicate negative association features.*

***Bold font indicates significance at the 0.05 level*

S 3.7:Results of the Goodness-of-fit tests for each species' global model as computed by the mb.gof.test function from the R package AICcmodavg with the consensus radius of 50 m for Forest Score, Heterogeneity and canopy closure and distance to water and distance to oil palm plantation computed at 5-m resolution.

Species	p-value*	Variance inflation factor
Banded Civet	0.182	1.3
Long-tailed Macaque	0.108	1.44
Malay Civet	0.112	1.13
Moonrat	0.104	1.36
Chevrotain (Greater & Lesser)	0.117	1.15
Thick-spined Porcupine	0.241	1.05

** p-values < 0.05 indicate lack of fit*

CHAPTER 4

Assessing spatiotemporal interactions between species from camera trapping data

JÜRGEN NIEDBALLA[1*], ANDREAS WILTING[1], RAHEL SOLLMANN[2], HERIBERT HOFER[1], ALEXANDRE COURTIOL[1]

1) Leibniz Institute for Zoo and Wildlife Research, Alfred-Kowalke-Str. 17, 10315 Berlin, Germany

2) Department of Wildlife, Fish and Conservation Biology, University of California Davis, Davis, CA, USA

**** corresponding author: niedballa@izw-berlin.de***

Author's contribution statement

JN, RS and AW and AC conceived the ideas for this manuscript. HH gave conceptual guidance. AC gave extensive statistical advice. JN conducted the statistical analyses and wrote the manuscript. All authors read, commented and agreed on the manuscript.

Assessing spatiotemporal interactions between species from camera trapping data

JÜRGEN NIEDBALLA[1*], ANDREAS WILTING[1], RAHEL SOLLMANN[2], HERIBERT HOFER[1], ALEXANDRE COURTIOL[1]

1) Leibniz Institute for Zoo and Wildlife Research, Alfred-Kowalke-Str. 17, 10315 Berlin, Germany

2) Department of Wildlife, Fish and Conservation Biology, University of California Davis, Davis, CA, USA

Abstract

Understanding spatiotemporal patterns of species interactions is of fundamental interest to behavioural and community ecology. Observer-independent, unbiased methods such as camera trapping facilitate the study of interactions in challenging habitats, but analyses are hampered by a lack of suitable approaches with well-defined properties. We present a flexible and expandable framework to simulate and explore spatiotemporal species interactions which can improve methods to detect spatiotemporal interactions. We simulated a two-species system with two types of (spatio)temporal interactions, spatiotemporal avoidance (of a site by a species after the presence of another species) and temporal segregation (shifts in daily activity patterns), across a range of activity patterns and interaction strengths. For spatiotemporal avoidance, we compared different time intervals between species records using linear models, Mann-Whitney U-test, a permutation test and a test based on randomly generated records. For temporal segregation, we applied a permutation test. We assessed the statistical power (the ability to detect an existing effect) and robustness (the ability to detect no effect when none is present) of all tests. Power for detecting spatiotemporal avoidance between species was strongly affected by interaction strength and highest for linear models, and reliable above 50 records per species. Reliably detecting strong temporal segregation required fewer records but depended heavily on the underlying activity pattern. All tests were robust even at low sample sizes, above a minimum of 10 records per species. Linear models were most suitable to analyse spatiotemporal avoidance and can easily correct for confounding effects of other sources of variation in interactions.

Introduction

Investigating biotic interactions is a central topic in ecology (Kissling *et al.* 2012; Wisz *et al.* 2013). Species interactions can take various forms, ranging from mutually detrimental, to antagonistic (detrimental to one species and beneficial to the other) to mutualistic (beneficial to both partners). Detrimental and antagonist interactions may vary considerably in their intensity and include exploitation competition, interference competition, predation, harassment and kleptoparasitism, amongst others (Palomares & Caro 1999; Caro & Stoner 2003; Arim & Marquet 2004; St-Pierre *et al.* 2006; Vanak *et al.* 2013). Possible adaptations to avoid the negative consequences of detrimental interactions include the suite of morphological, physiological and behavioural traits responsible for niche partitioning, which can minimise the impact of detrimental interactions and promote co-existence between species (Di Bitetti *et al.* 2010). Behavioural partitioning of niches can occur in space, e.g. by avoidance of sites that are occupied by a competitor, a predator or pathogen, or by habitat partitioning along environmental gradients (e.g. terrain, proximity to resources, habitat structure). Niches can also be partitioned temporally or spatiotemporally, e.g. by the adjustment of activity patterns to avoid interaction with other species, or by temporarily avoiding sites after the presence of a predator or competitor (Kronfeld-Schor & Dayan 2003; Schuette *et al.* 2013; Karanth *et al.* 2017).

It is often difficult to measure species interactions directly, particularly for cryptic or rare species and in dense habitats such as tropical rainforests where direct observations are challenging. Amongst observer-independent and hence unbiased methods, telemetry studies can provide information on spatiotemporal species interactions (Downs *et al.* 2015; Long 2015), but are not always feasible and often suffer from low sample sizes. Camera trapping is a widely used and cost-effective alternative in many habitats difficult to access, but only offers the detection of species or individuals at point localities without detailed movement information. Detecting species interactions from camera trapping data therefore requires different methods than telemetry studies.

Spatial species interactions can be investigated from camera trapping data using two-species occupancy models, in which the detection or occupancy probability of a species depends on the detection of or occupancy by another species (MacKenzie *et al.* 2004; Richmond *et al.* 2010). While two-species occupancy models use spatial information, the precise time stamps recorded in camera trap images are not used to directly assess (spatio)temporal interactions in these models.

For directly assessing (spatio)temporal interactions several methods and measures have been developed and applied. One group of methods estimates to what extent a subordinate species temporarily avoids a site that was recently visited by a dominant species (e.g. Harmsen *et al.* 2009; Parsons *et al.* 2016; Karanth *et al.* 2017). Such avoidance behaviour can be mediated by olfactory cues (Apfelbach *et al.* 2005). These methods measure spatiotemporal interactions, and we will refer to these interactions as spatiotemporal avoidance here. The second, more commonly used method assesses temporal species interactions. Here the temporal overlap in activity between two species is estimated to assess if the activity patterns may be shifted to avoid encounters (e.g. Ridout & Linkie 2009; Linkie & Ridout 2011; Foster *et al.* 2013; Lynam *et al.* 2013; Ross *et al.* 2013; Farris *et al.* 2015;

Sunarto *et al.* 2015). Often, multiple camera trap stations are pooled in this method, thus omitting the spatial component and consequently assessing temporal interactions only. We will refer to these interactions as temporal segregation here. Such a shift in activity patterns could be the result of avoiding the dominant species by a subordinate species, a consequence of competition between species or it may reflect other aspects of the species' ecology, such as differences in physiological adaptation to diurnal rhythms of ambient conditions (Haim & Fourie 1980; Fuller *et al.* 2010).

Despite the common use of camera trap based methods to assess these species interactions, there is no comparative assessment available on how suitable these methods are to detect such interactions. We therefore simulated camera trap data to assess the ability of these methods to correctly identify species interactions for a wide range of activity patterns. Simulations allowed us to overcome the main obstacle in field data, the unknown (latent) true state of the study system that generates the observations by explicitly specifying the system properties (Peck 2004).

The aim of this study was to assess different measures and statistical tests for detecting the two types of species interactions: spatiotemporal avoidance, expressed as differences in the time intervals between detections of the interacting species at camera trap stations, and temporal segregation, expressed as shifts in activity patterns between two species. We compared the tests in terms of their statistical power and robustness under different interaction (avoidance) strengths and for the most commonly found daily activity patterns, cathemeral, diurnal, nocturnal and crepuscular. Specifically, we assessed whether and under which circumstances spatiotemporal avoidance of a dominant species by a subordinate species and temporal segregation can be detected, how many species records are needed for a reliable detection and which method is the most powerful and robust.

Methods

Outline of the study system

We simulated the interactions of a dominant species A and a subordinate species B. We assumed the dominant species A is unaffected by the presence of subordinate species B, whereas subordinate species B has two possibilities to avoid an interaction with species A. It may either avoid a site after species A was present, expressed as a reduced probability of recording the subordinate species B after the dominant species A was recorded. In this case, we assumed a subsequent linear recovery of the chance that species B visits the site (spatiotemporal avoidance). Alternatively, species B may shift its diurnal activity peaks relative to species A to reduce activity overlap and thus reduce the chance to encounter species A (temporal segregation).

The activity patterns of both species was chosen to be uniform (flat), unimodal (with one activity peak per day), or bimodal (two peaks per day). Uniform activity patterns are representative of cathemeral species (absence of a fixed activity rhythm). Unimodal and bimodal activity patterns are more common, the former being typical of diurnal or nocturnal species, the latter of crepuscular species (Ridout & Linkie 2009; Levy *et al.* 2012; for examples see e.g. Foster *et al.* 2013; Lynam *et al.* 2013; Monterroso *et al.* 2013; Ross *et al.* 2013; Farris *et al.* 2015; Ikeda *et al.* 2016).

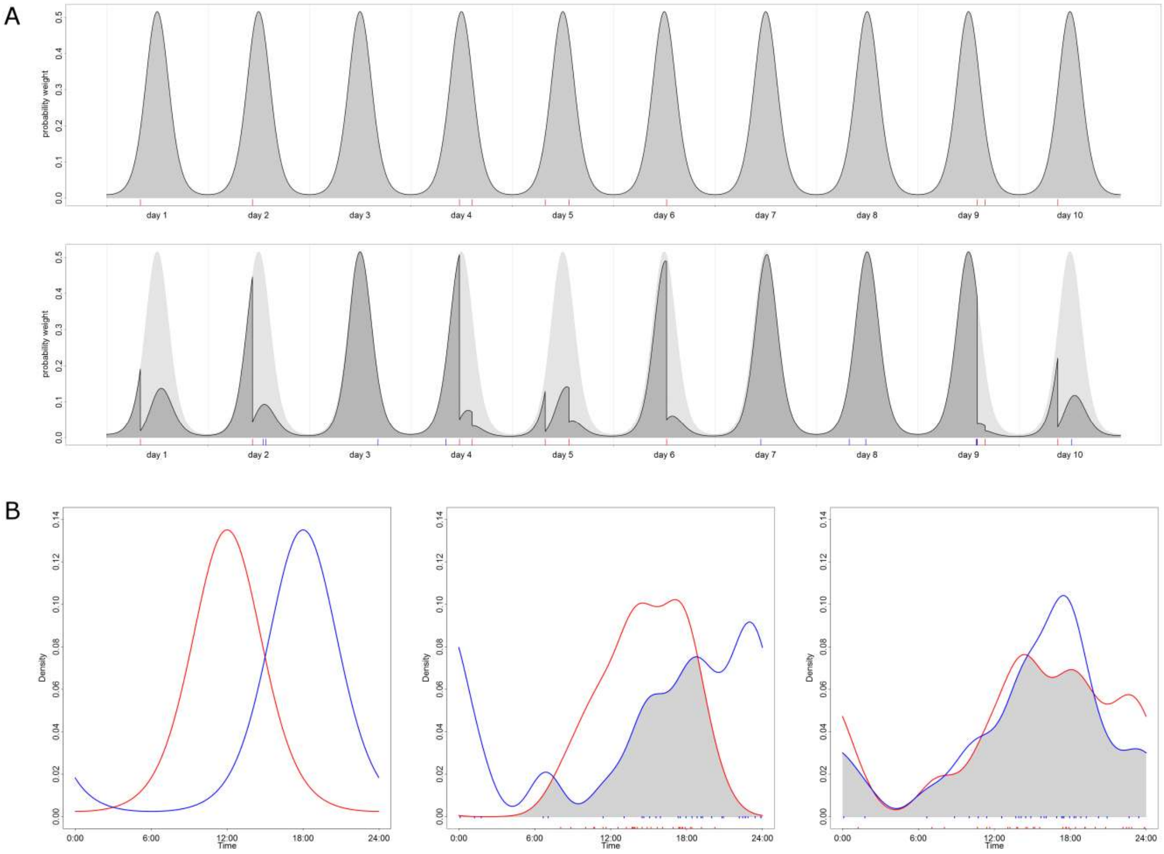


Figure 4.1: Types of (spatio)temporal interactions examined. A) Spatiotemporal avoidance: the top row shows the probability weights for realising records of a dominant species during an example of 10 simulated survey days, with realised records shown as red ticks. The bottom row shows a subordinate species (which avoids the dominant species), the probability weights of which (darker grey) are reduced after records of a dominant species compared to their original value (light grey). Realised records of the subordinate species are shown in blue. B) Temporal segregation. The left plot shows the daily activity density curves of two species with a time shift of 6 hours between activity peaks. These activity density curves are used as probability weights for realising records in the simulation. The central plot shows the activity overlap (in grey) of those realised records (observed data). The right plot shows the increased activity overlap after one possible species label permutation

Simulation of species records

We wrote an R function that simulates species records for two temporally interacting species under different scenarios, mimicking data from camera trapping surveys. The scenarios are defined by the parameters of this function which describe the survey duration in days, number of records of both species, the diurnal activity pattern for both species, and the interaction, expressed as intensity and duration of avoidance of the dominant species A by subordinate species B or its shift in activity period (see Table 4.1 for function parameters and supplementary R code of the function simulating the records). All computations were performed in R 3.3.3 (R Core Team 2017).

Uni- and bimodal patterns were approximated using von Mises distributions. These are continuous circular probability distributions used here to represent the density for a species activity by time of day (expressed in radian ranging from 0 to 2π). We modelled the bimodal activity patterns using a mixture of two von Mises distributions. The density functions of those distributions were calculated using the R package CircStats version 0.2-4 (Lund & Agostinelli 2012). Their shape is determined by two parameters, the mean μ and concentration parameter κ (kappa).For the unimodal distributions, we set the mean to $\mu = \pi$, corresponding to maximum activity at 12pm (noon, a diurnal species). For the bimodal distributions, we set the two means to $\mu_1 = \frac{1}{2}\pi$ and $\mu_2 = \frac{3}{2}\pi$, corresponding to activity peaks at 6am and 6pm, a crepuscular species. In the assessment of spatiotemporal avoidance, the mean was identical for both species to isolate spatiotemporal avoidance and avoid confounding effects of temporal segregation. In the assessment of temporal segregation, the means (i.e., the activity peaks) were allowed to vary between species (see "temporal segregation" below in the methods section). κ determines how strongly records are concentrated around the mean. We chose one value of κ in our assessment of spatiotemporal avoidance (because we expected κ to have little influence) and three different values in our tests for temporal segregation (because here we expect a strong influence of κ). See Table 4.1 for more details and Figure 4.1 for a plot of activity patterns with different κ).

The function creates records of dominant species A by randomly sampling the time axis along the whole study duration (in 1-minute intervals) using the activity density of species A in each of these 1-minute intervals as probability weights for the random draws from the time axis. The binary outcome for each 1-minute interval corresponds to observations or non-observations. We created records of the subordinate species B in a similar manner using the activity density of species B as probability weights for the random draws along the time axis. However, as species B is also affected by prior presence of species A, these probability weights can be reduced to a fraction of their original values after a record of species A. This reduction in probability weights is defined so as to obtain constant odds ratios:

$$Odds\ Ratio = \frac{\frac{p\ (B\mid no\ A)}{1-p\ (B\mid no\ A)}}{\frac{p\ (B\mid A)}{1-\ p\ (B\mid A)}}$$

where $p(B|A)$ is the probability of detecting species B directly after species A was detected and $p\ (B\mid noA)$ is the probability of detecting species B in the absence of species A. The reduction of probability weights only depends on the odds ratio and is independent of the absolute values of activity probability density, and thus superimposed on the activity patterns (see Figure 4.1 for an example). After a detection of species A, the detection rate of species B recovers linearly to the original value for a user-defined amount of time. In the assessment of spatiotemporal interactions, the odds ratio was varied between 1 and 100 (1 corresponding to no avoidance and 100 to very strong avoidance). In the assessment of temporal segregation, it was fixed at 1. Because the simulated records of both species are analogous to data generated in camera trapping studies we term them *observed data* here.

Spatiotemporal avoidance

For each parameter combination as shown in Table 4.1, the number of records and the activity patterns were identical for both species; we only varied the strength and duration of the spatiotemporal avoidance. Analogous to the common practice in camera trapping data, we removed records of a species if they fell within 60-minutes from the last record of the same species at the same camera trap. We calculated various time intervals between both species in the *observed data* to assess which is most suitable for detecting spatiotemporal avoidance: AB, BA, AA, BB, ABA, BAB, and the ratios AB/BA and BAB/BB. AB and BA were used by Harmsen *et al.*(2009) and Karanth *et al.*(2017). The two ratios correspond to the attraction-avoidance-ratios (AARs) T2/T1 and T4/T3 in Parsons et al. (2016) and compare the time intervals between a dominant species and a subordinate species to the converse situation (T2/T1) or the time intervals between records of subordinate species with/without the passage of a dominant species in between (T4/T3). If there was a sequence of records of the other species before a given record, we calculate the time intervals since the last record of the other species in that sequence.

Linear models

We first directly compared the time intervals AB and BA (dominant-subordinate and subordinate-dominant) by fitting two linear models using either log-transformed (linear model 1) or

Table 4.1: Parameters of the R function used for simulating species records

Function input	**Details**	**Temporal segregation**	**Spatiotemporal avoidance**
Number of days	Number of simulated days	100, 300	100, 300
Number of records A	Number of records of species A	5, 6, 7, 8, 9, 10, 20, 30, 40, 50	10, 20, ..., 90, 100
Number of records B	Number of records of species B	5, 6, 7, 8, 9, 10, 20, 30, 40, 50	10, 20, ..., 90, 100
activity pattern $^{A/B}$	Uniform, unimodal (von Mises) or bimodal (von Mises mixture)	unimodal, bimodal	uniform, unimodal, bimodal
κ (kappa) $^{A/B}$	κ (kappa), concentration parameter of the von Mises distributions used for uni- and bimodal activity patterns	1,2,3	2
Spatiotemporal avoidance strength ^{B}	The odds ratio between the odds for detecting B in the absence of A relative to the odds for detecting B after A was present	1 (no effect)	1, 2, 10, 100
Spatiotemporal avoidance duration ^{B}	Duration for which the effect of A on B persists until full recovery (in days)	- (no spatiotemporal avoidance)	1, 3
Activity peak difference ^{B}	Time difference between the activity peaks of A and B (in hours)	unimodal: 1,2,3,…,12 bimodal: 1,2,3,…,6	- (synchronous activity patterns)

^{A} *argument impacts on the probability distribution of the dominant species A*

^{B} *argument impacts on the probability distribution of the subordinate species B* untransformed (linear model 2) time intervals as the response variable. For both linear models, the only predictor considered was the type of the time interval (AB or BA). The interval AB was used as the reference level in the model. The linear models thus estimate the difference in BA (log) time intervals relative to AB (log) time intervals. The `summary.lm()` function returns the significance level for this difference. In this simple situation, the p-value of the parameter estimate from the linear model fit corresponds to the p-value one would obtain preforming a traditional t-test on the data without fitting a linear model (`t.test(…, var.equal = TRUE)` in R).

We further checked whether the main assumptions about the errors in linear models were met, namely homoscedasticity (using the Breusch-Pagan test), independence (i.e. absence of serial autocorrelation assessed using the Durbin-Watson test), and normality of the model residuals (using the Shapiro-Wilk test).

Mann-Whitney U-test

We also compared the time intervals AB and BA using the non-parametric Mann-Whitney U-test. This test is usually considered as the non-parametric counterpart of the t-test: it is used to test for difference in central locations between two independent groups but it does not require observations to be normally distributed.

Permutation test

Using the *observed data* of both species, we performed a permutation test to generate data under the null hypothesis of no spatiotemporal avoidance between species. We first calculated the median of the observed time intervals and ratios mentioned above (AB, BA, AA, BB, ABA, BAB, AB/BA and BAB/BB. We then generated the distributions of these time intervals/ratios under the null hypothesis by randomly permuting species labels from the original data at a camera trap station 1000 times, each time filtering for temporal independence (60 minutes, see above) and recalculating all time intervals. We then compared the median observed time intervals/ratios to the median of these 1000 permutation time intervals/ratios and performed a two-sided significance test on these (because the permutation values may be higher or lower than in *observed data*). The p-value for this test is directly deduced from the distribution of the test statistic:

$$p = min(q, 1 - q) \times 2,$$

where q is the quantile of the observed values within the distribution of the randomised values. Thus, p is equivalent to the percentage of permutation tests whose values are more extreme than the observed values.

Random records

To assess an alternative method to simulate the null hypothesis of no spatiotemporal avoidance between species we created random records of the subordinate species B based on its overall observed activity pattern (because it is how activity patterns are estimated from field data) without influence of presence/absence of species A. To that end we derived a kernel density estimate of the diurnal activity pattern from the *observed data* of subordinate species B using the R package overlap (Meredith & Ridout 2016). Using this kernel density estimate as a weight for a random draw from 1-minute intervals covering one day, we created random times of day. By combining these with randomly selected days from the study period we obtained random records of subordinate species B. We created a number of random records for B that was identical to the one in original data. This procedure was also used by Karanth *et al.* (2017), but without explicitly taking the activity patterns into account. The newly generated records of B were again filtered for temporal independence (60 minutes). Records of A were not manipulated.

We then calculated the median of all time intervals mentioned above for 1000 independent sets of random records generated in this manner and compared the median observed values to the distribution of values from the random records. As in the permutation test above, we performed a two-sided significance test comparing the observed test statistic and obtained the p-value as:

$$p = min(q, 1 - q) \times 2,$$

where q is the quantile of the observed values within the distribution of the values derived from the random data.

Temporal segregation

We created records of both species with temporal segregation between both species (time shifts of the activity peaks) ranging from 0 - 12 hours for unimodal activity patterns and 0 - 6 hours for bimodal activity patterns (in 1-hour steps). The number of records was equal between both species and varied between 5 and 50. The temporal segregation between species A and B consisted of a shift in the activity peak of species B relative to species A. Both species had the same type of activity pattern (unimodal or bimodal) and no spatiotemporal avoidance. See Table 4.1 for the function parameters that were varied. We calculated activity overlap coefficient $\hat{\Delta}_1$(Ridout & Linkie 2009) of these *observed data* using function `overlapEst` from the R package overlap (Meredith & Ridout 2016). $\hat{\Delta}_1$ is the integral of (= the area under) the probability density functions of the estimated daily activity density curves of both species (denoted $\hat{f}(t)$ and $\hat{g}(t)$):

$$\hat{\Delta}_1 = \int_0^1 \min\{\hat{f}(t), \hat{g}(t)\}\, dt.$$

Analogous to the permutation test above, we randomised the species labels 1000 times (or the maximum number of possible permutations for 5 and 6 records) and calculated $\hat{\Delta}_1$ for each of these randomised data sets to obtain the distribution of $\hat{\Delta}_1$ under the null hypothesis (no temporal segregation between species).

If the time of activity peaks differs between both species, we expect the $\hat{\Delta}_1$ values from the randomized datasets to be higher than the observed $\hat{\Delta}_1$, i.e. we expect a higher activity overlap coefficient in data with randomised species labels than the *observed data* (see Figure 4.1). The statistic-value for this comparison corresponds to the quantile of the observed $\hat{\Delta}_1$ compared to the distribution of 1000 randomised $\hat{\Delta}_1$ values. It was calculated as

$$p = \frac{\sum_1^{\mathrm{n}} \hat{\Delta}_{1_{obs}} > \hat{\Delta}_{1_{H0}}}{n+1}$$

with $\hat{\Delta}_{1_{obs}}$ being the observed $\hat{\Delta}_1$, $\hat{\Delta}_{1_{H0}}$ being the permutation $\hat{\Delta}_1$ values, and n being the number of permutations. It is a one-sided test, and its p-value corresponds to the null hypothesis of no time shift in activity peaks between both species.

Power analysis

Statistical power is the probability that a test correctly rejects the null hypothesis when it is false. Thus, it quantifies the ability of a test to detect an effect that actually exists. To assess the power of the tests we conducted, each of the above tests was performed on 1000 sets of independently generated records of both species (*observed data*) for each combination of function parameters detailed in Table 4.1. Power was calculated as the percentage of significant tests at α = 0.05 out of these 1000 independent tests. We considered a test as reliable if an effect was present and power was > 0.8.

Robustness assessment

Robustness is the property of a statistical test to have a type I error rate (i.e, the probability of erroneously rejecting the null hypothesis) that corresponds to the significance level under the null hypothesis. A test that is not robust will reject the null hypothesis more or less frequently than suggested by the significance level, leading to biased conclusions.

We assessed the robustness of all methods using the empirical cumulative density functions (ECDFs) of the p-values from the 1000 independent test with data generated under the null hypothesis of no temporal interaction (no spatiotemporal avoidance or identical activity peaks, respectively). If a test is robust, its p-values are expected to follow a uniform distribution under the null hypothesis. This expectation was assessed in three complementary ways, namely 1) by visually assessing the ECDFs of p-values, which should follow a straight line, indicating a uniform distribution, 2) by comparing the distribution of p-values obtained to a uniform distribution using a Kolmogorov-Smirnov test (expecting that there is no evidence for significant deviations from a uniform distribution), and 3) by computing the value of the ECDFs at a significance level of $\alpha = 0.05$, expecting an ECDF value of about 0.05 at $\alpha = 0.05$ if the distribution of p-values is uniform. Major deviations from an expected value of 0.05 would indicate the test has a higher or lower false positive rate than suggested by the nominal significance level, i.e. it is not robust.

Results

Spatiotemporal Avoidance

Power

All four methods we tested (linear models, Mann-Whitney U test, permutation test and the test based on randomly generated records) were able to detect spatiotemporal avoidance of a subordinate species relative to a dominant species (given sufficient records) and were not affected by the type of activity pattern considered. For all methods, higher numbers of records, stronger and longer avoidance resulted in higher power (Figure 4.2).

Overall, the highest power was achieved with a linear model comparing log-transformed time intervals AB and BA, followed by the U-test, linear model with untransformed data and the comparison of the time interval AB to those from randomly generated records or the permutation test. Between the linear models and the Mann-Whitney U-test (both of which compared the intervals AB and BA), power was generally similar in range, but highest when using a linear model on log-transformed time intervals, intermediate for U-tests, and lowest in linear models considering untransformed interval values as the response variable. The power of these three tests was generally higher than for the permutation test or for the test based on randomly generated records (Figure 4.2).

In both the permutation procedure and test based on randomly generated records, using the time interval AB consistently resulted in the highest power of all time intervals calculated. The avoidance-attraction ratio AB/BA had second highest power followed by time interval ABA. The power of the remaining time intervals was consistently lower across all tests (Figure 4.2). For the time interval AB, the power was slightly higher when using random records compared to the species label permutation. The same was observed for the time interval ABA (but more pronounced), whereas in the ratio AB/BA, the test using species label permutation achieved slightly higher power (Figure 4.2).

Generally, even with the time intervals that performed best (AB), high power (>0.8) was only achievable for high numbers of records and strong avoidance effects (> 50 records per species, odds ratio ≥ 10). Below 50 records per species, power dropped sharply and even very strong avoidance effects could not be detected reliably, with power up to 50% for 40 records and about 25% for 30 records (Figure 4.2). The number of necessary records to reliably detect spatiotemporal avoidance was lower when the avoidance effect lasted longer relative to the survey duration. If avoidance was more subtle (odds ratio = 2), even 100 records per species were insufficient to reliably detect spatiotemporal avoidance.

Robustness

All tests were robust for all types of activity patterns and all factor combinations tested. The ECDFs of p-values under the null hypothesis were linear and there was no evidence for systematic deviations from a uniform distribution (Supplementary Figures S4.1 and S4.2).

In the linear models, p-values were distributed uniformly under the null hypothesis despite frequent significant deviations from the test assumptions of normally distributed, homoscedastic residuals. The assumption of homoscedasticity was commonly violated significantly, but the distributions of test statistics of the Breusch-Pagan test indicated only subtle heteroscedasticity under the null hypothesis. However, these test statistics deviated substantially from expectation when the avoidance effect was strong (odds ratio = 10 or 100), and particularly when numbers of records was high. This problem was generally more pronounced when using log-transformation, and most severe in unimodal activity patterns. The assumption of normality of residuals was violated very frequently, particularly when the number of data points (records) was high. Nevertheless, the test statistics of the Sharpiro-Wilk tests were consistently close to 1, indicating the violations were not severe, albeit statistically significant. Log-transformation resulted in test statistics closer to 1, indicating a less severe non-normality of regression errors than in untransformed data. There was no evidence for systematic autocorrelation.

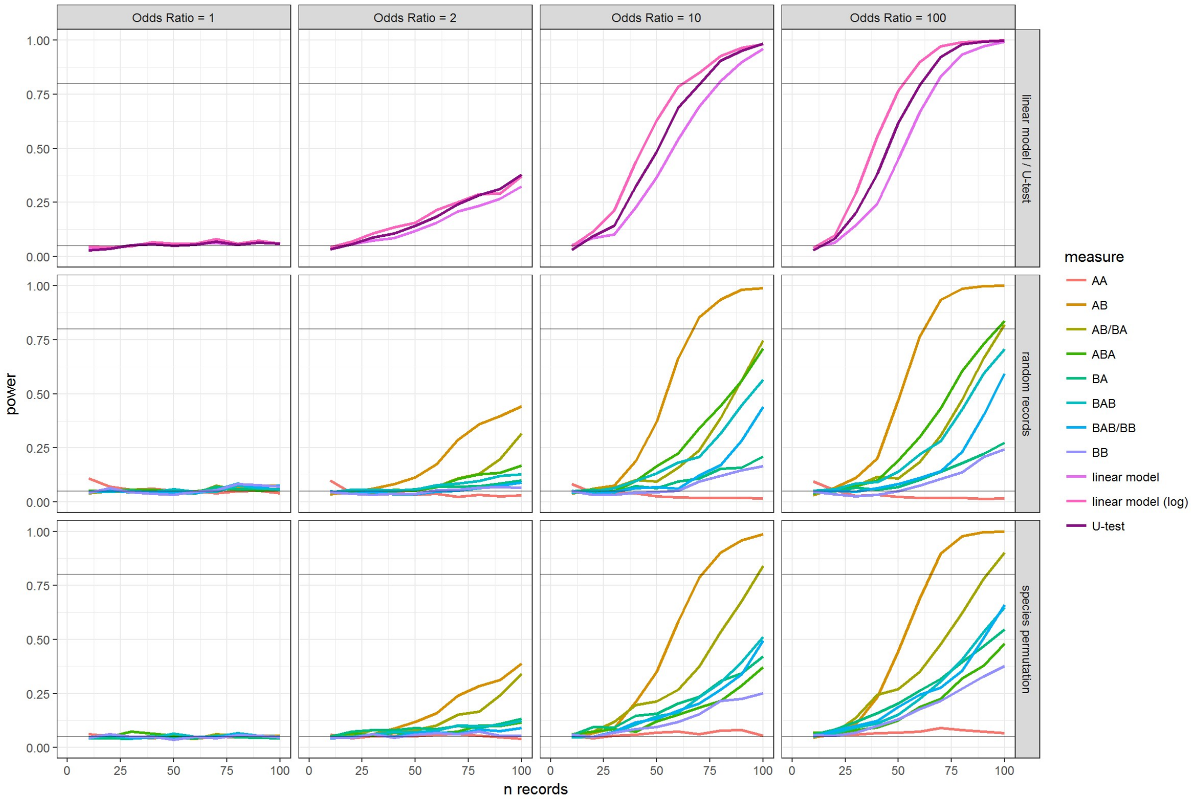


Figure 4.2: Power of four methods for detecting spatiotemporal avoidance between a dominant and a subordinate species in camera trapping data (a species label permutation test, randomly created records, linear models and Mann-Whitney U-test). Each point is based on 1000 independent tests. Data shown are for 100 simulated survey days, and the detection rate of the subordinate species takes 1 day to recover to its original level after the presence of the dominant species. This plot shows data from unimodal activity patterns, but it is essentially the same for bimodal and uniform activity. The grey horizontal lines represents power = 0.05 and 0.8. The four columns in which plots are arranged show the interaction (avoidance) strengths expressed as odds ratios between the odds for detecting the subordinate species in the absence of the dominant species relative to the odds for detecting the subordinate species after the presence of the dominant species (100 = very strong avoidance, 1 = no avoidance).

Temporal segregation

Power

The power of a species label permutation test to find temporal segregation between species varied considerably and was influenced by multiple factors. It increased with the number of records and the magnitude of the time shift between activity peaks, and when activity peaks became narrower. For a given number of records, power was higher if species had unimodal activity patterns compared to bimodal activity patterns (Figure 4.3).

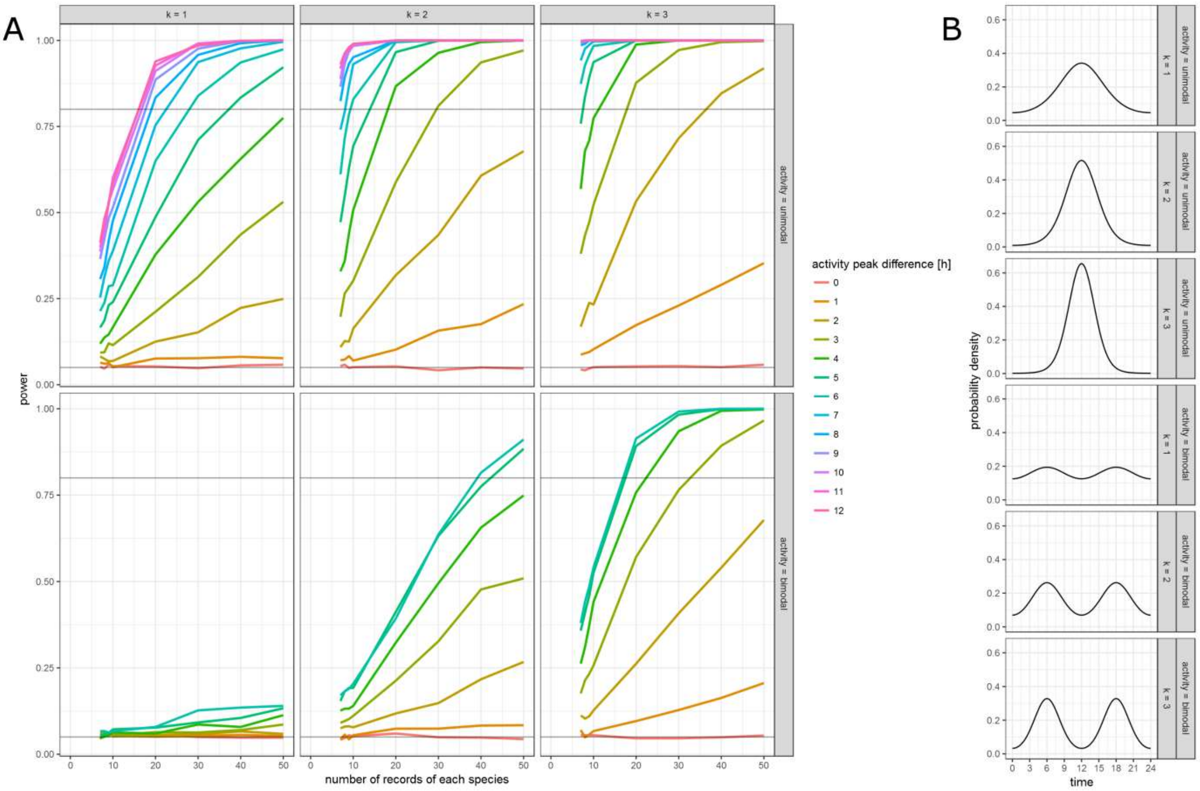


Figure 4.3: A) Power of species label permutation test for detecting temporal segregation for uni- and bimodal activity patterns with different concentration parameters κ and activity peak differences (0… 12 hours for unimodal activity and 0…6 hours for bimodal activity). Each point is based on 1000 independent tests. The grey horizontal lines corresponds to a significance level of α = 0.05. Power of test for 5 and 6 records of each species are not shown due to lacking robustness. B) The underlying diurnal probability distributions for species detections. In unimodal distributions, 95% of activity density lies approximately between μ±10 h (2am – 10pm) for κ = 1, between μ±7 h (5am - 7 pm) for κ = 2, and between μ±5 h (7am – 5 pm) for κ = 3.

The conditions under which power was high (i.e. > 0.8) varied widely depending on these factors. In narrow unimodal activity patterns, even smaller time difference between activity peaks (2 hours) could be detected reliably given sufficient records (40 or more). Under these conditions, differences in activity peaks of 5 hours and above were detectable reliably with less than 10 records. On the opposite extreme, it was impossible to reliably detect even considerable temporal segregation in large numbers of records with less pronounced bimodal activity (κ = 1). Similarly to unimodal activity patterns, power approached 1 for more pronounced bimodal activity patterns (κ = 3) with higher numbers of records, and activity peak shifts of 3 hours and above.

Robustness

Robustness of the species label permutation test for differences in species activity peaks was satisfying for all parameter combinations with 7 or more records of each species. Below 7 records per species, the observed false positive rate at α= 0.05 was 20-30% (see Supplementary Figure S4.3), demonstrating that the tests are not robust in this situation irrespectively of the other simulation parameters.

Discussion

We demonstrated the use of different methods for detecting two different types of spatiotemporal species interactions in camera trapping data, namely four methods for spatiotemporal avoidance and a permutation test for temporal segregation (shifts in activity patterns), using a flexible simulation approach and assessed their statistical power and robustness. The methods were generally robust and capable of detecting both of these types of interactions irrespective of the species' activity patterns, but they varied in statistical power.

The most powerful methods for detecting spatiotemporal avoidance behaviour were linear models (corresponding to traditional t-tests for a single station) on the time intervals between the dominant and subordinate species and vice versa (AB and BA). Log-transforming time intervals slightly increased the power. Harmsen et al. (2009) used a similar linear modelling approach and found evidence of spatiotemporal avoidance between jaguars and pumas in a neotropical forest. Even though linear model assumptions were violated, power or robustness were not affected negatively and we do not consider it a serious problem in this application. However, users can check linear model assumptions to ensure power and robustness are acceptable without having to perform a simulation study. In case of concern, non-parametric alternative methods for robust detection of spatiotemporal avoidance behaviour described here (e.g. U-test) can be applied with slightly lower power (Adams & Anthony 1996).

The preferred time interval for detecting spatiotemporal avoidance behaviour with the permutation test and the test based on randomly generated records is the interval between the dominant and the subordinate species (AB). There was no indication that Avoidance-Attraction-Ratios AB/BA and BAB/BB provided an advantage in detecting spatiotemporal avoidance between species compared to the other time intervals in terms of power or robustness. Karanth et al. (2017) used a similar test based on randomly generated records compared times intervals between co-occurring species (times-to-encounters, akin to AB here) from observed data and random records and found indications for potential spatiotemporal interactions between dholes, leopards and tigers in Indian wildlife reserves.

While the presented tests are suitable to reliably (power < 0.8) detect spatiotemporal avoidance and temporal segregation, failure to detect these may be due to low power while spatiotemporal avoidance or temporal segregation between species is present. Spatiotemporal avoidance can be detected if avoidance is sufficiently strong (odds ratio ≥ 10) and long-lasting (≥ 1 day). However, if avoidance was weak (i.e., a mild decrease in detection probability of the subordinate species after the presence of a dominant species, odds ratio = 2) or very short-term (e.g. in the range of hours), avoidance cannot be detected reliably even with high numbers of records. Analogously, for temporal segregation, subtle shifts in activity peaks of 1-2 hours cannot be detected reliably even with large numbers of records unless activity peaks are very narrow (narrower than simulated in the present study). Hence, failure to detect temporal segregation or spatiotemporal avoidance in data may be due to low power while there is an actual interaction between species.

Tests based on low numbers of records generally have low power, thus impeding the detection of spatiotemporal avoidance or temporal segregation. In addition, when calculating activity densities (for the creation of random records or activity overlaps), low numbers of records give disproportionate weight to individual records and can prevent the inference of actual activity patterns. As a consequence, the activity densities estimated from low numbers of records do not reflect the underlying distributions adequately. Current studies were generally aware of this limitation and used from 10 to well over 1000 records for estimating activity densities (e.g. Lynam *et al.* 2013; Ross *et al.* 2013; Farris *et al.* 2015). The use of hourly accumulation curves or a comparison of activity density estimates with non-negative trigonometric sums were suggested to help deciding whether sufficient records are available to adequately describe the species' activity patterns (Ridout & Linkie 2009; Tambling *et al.* 2015).

For the reliable detection of differences in activity peaks between species with unimodal activity patterns, 20 records per species were usually sufficient to detect a difference of a few hours. The necessary number of records is larger for smaller differences (2-3 hours) or bimodal activity patterns and more likely in the range of 50 records per species. As stated above, power strongly depends on the shape of the underlying detection probability distribution and the actual activity shift, with chances being that subtle changes in activity peaks (e.g. 1-2 hour) would go unnoticed.

While low numbers of records may result in low power for detecting temporal segregation, power may be sufficient in high numbers of records for the detection of small shifts in activity patterns that may be ecologically irrelevant. Additionally, differences in the underlying activity density functions between the species (e.g., one having a unimodal and the other a bimodal activity pattern) can give significant results. Even if sufficient numbers of records are available and activity patterns are similar between species, the main constraints on the interpretation of such an analysis is that (in contrast to spatiotemporal avoidance results) a significant finding in the presented activity overlap permutation test does not necessarily indicate interaction, niche partitioning, or avoidance between these species, but may simply reflect individual species autecology.

This work is theoretical in nature and does not incorporate the manifold sources of uncertainty found in real world data. Examples of such sources of uncertainty are imperfect detection, confounding effects of multiple species from the community, or attraction between species. Imperfect detection is inevitable in camera trapping and likely reduces the power of the tests for spatiotemporal avoidance, as it adds noise to the signal and alters the distribution of time intervals between records, thereby likely increasing the number of records needed to reliably detect spatiotemporal interactions. However, imperfect detection should not affect the test for differences in activity patterns, given sufficient records and under the assumption that detection probabilities of both species are constant throughout the day. To assess the impact of the detection process, a potential extension of the presented simulation framework would be to (randomly or non-randomly) remove records of both species.

Potential other competitors or predators besides the two focal species complicate the system and may introduce further unobserved interaction patterns, which would negatively impact the power of detecting the species interactions of interest. Removing time intervals between focal species from the analyses when a third species was detected in between (as in Parsons *et al.* 2016) may partly alleviate this problem at the cost of losing data and consequently reduced power. Removing those time intervals, however, would not be enough if occupancy by some third species alone changes activity patterns of one of the study species (e.g. Ross *et al.* 2013),. In this situation, the third species may introduce spurious relationships between species, e.g. an apparent interaction effect between two species that in reality is caused by the additional species. Including a third species in the analyses presented here would be possible in theory, but sharply increases the number of necessary comparisons, and thus the data requirements to reach a given power, and complicates interpretation of findings (there would be three activity overlaps instead of one and six time intervals equivalent to AB and BA instead of two). A further alternative may be to take a functional, trait-based approach and treat records of different species from the same guild as one species group at the cost of reduced taxonomic resolution.

Spatiotemporal attraction of one species by another should be detectable with the methods we present for spatiotemporal avoidance. Therefore, an ecologically meaningful extension of the presented simulating framework could be to include attraction between species (e.g. a predator following prey).

The methods we present may also be applied to intra-specific relationships, e.g. between individuals, animals of different age classes or sexes, between seasons, different areas, or between stations with specific properties. For example, a significant shift in activity peaks of a species in the absence or presence of a potential competitor or predator could give an indication of behavioural adaptation in terms of the activity periods (Ross *et al.* 2013). Nevertheless, even strong shifts in activity patterns do not prevent interactions between species, as shown for dholes in Laos that preyed on species they had little activity overlap with (Kamler *et al.* 2012).

In order to make the results more easily interpretable, we simulated data from a simple system consisting of a single camera trap station only. However, all methods we present here can easily be expanded to multiple camera trap stations, as will be found in almost all field camera trapping studies.

For detecting spatiotemporal avoidance, our results showed that the most sensitive method to study this interaction is a linear model comparing the (log-transformed) time intervals AB and BA. By considering the camera trap station as an additional random or fixed effect, one could capture geographic differences in interval times. Which of these two formalisms should be used (fixed or random) depends on the number of stations and on the number of observations per station. Considering stations as a random effect implies the additional assumption that the geographic differences are normally distributed between stations (other distributions can also be considered in some R packages such as spaMM). If this assumption is legitimate, then fitting the model as a mixed-effects model considering the type of interval as a fixed effect and the camera trap station as a random effect should lead to higher power than the alternative full fixed effect parameterisation. If the number of stations is low (e.g. < 6), the mixed model approach is not recommended because the estimation of the variance of the random effect will be poor. In such a case, one should consider the camera trap station as an additional fixed effect and introducing an interaction term to account for the statistical interaction between the spatiotemporal interaction and the station. This latter approach requires many observations per station to ensure sufficient power. The approach based on linear models (whether mixed or not) has the additional advantage that other factors potentially affecting the interaction between species (such as habitat characteristics) could also be included. If the violations of linear model assumptions are severe and non-parametric U-tests are performed, the p-values of these independent U-tests from individual stations can be combined using Fisher's method (Fisher 1932) at the cost of reduced power.

Similarly, in both the random record method and the permutation test for spatiotemporal avoidance, the p-values from tests at each station can be aggregated using Fisher's method to test the global null hypothesis of no interaction. The species label permutation test can only be applied if sufficient numbers of records are available at each station to ensure a sufficiently high number of possible permutations and avoid problems with robustness shown above. Therefore, a high number of records at each station will be required and it is likely that this method can only be used for two commonly recorded species. When using randomly generated records to detect spatiotemporal avoidance, pooling records from different stations can also give a more accurate reflection of species' general activity patterns, thus providing more realistic random records at each station (assuming activity is constant between stations).

For the temporal segregation permutation test, records from all stations or stations at which both species co-occur can be pooled to give sufficient numbers of records, assuming activity patterns are constant between stations. All stations are then jointly analysed in one permutation test. If activity patterns within species are assumed to differ between stations, independent permutation tests can be run on different (sets of) stations. Pooling data from many stations or over extended periods of time will however induce an increase observed overlap, which is an artefact of pooling (Nouvellet *et al.* 2012). Fisher's method can also be applied here if multiple permutation tests are performed (e.g. on different subsets of camera trap stations).

The simulation approach we presented provides a flexible, extensible framework for the development and testing of statistical methods for detecting species interactions in camera trapping data under well-defined conditions. We recommend field researchers to adapt our approach to simulate data closer to their biological reality to assess their methods prior to analyses or to assess the sampling design prior to the data collection. Our results provide guidance to field researcher as to when their data will be sufficient to test spatiotemporal avoidance and temporal segregation and how such tests could be implemented. Irrespective of the type of analysis, our simulation study showed that in order to detect subtle spatiotemporal interactions, sufficient numbers of records (often in the range of 100 per species and above) are needed. As understanding species interactions is a key topic in ecology, our results support calls to standardise data collection schemes and combine camera trap datasets from different studies in joint analyses (Forrester *et al.* 2016; Steenweg *et al.* 2017).

Acknowledgements

We thank the German Federal Ministry of Education and Research (BMBF FKZ: 01LN1301A) for funding this research.

References

Adams, D.C. & Anthony, C.D. (1996). Using randomization techniques to analyze behavioural data. *Animal Behaviour*, **51**, 733–738.

Apfelbach, R., Blanchard, C.D., Blanchard, R.J., Hayes, R.A. & McGregor, I.S. (2005). The effects of predator odors in mammalian prey species: A review of field and laboratory studies. *Neuroscience and Biobehavioral Reviews*, **29**, 1123–1144.

Arim, M. & Marquet, P.A. (2004). Intraguild predation: A widespread interaction related to species biology. *Ecology Letters*, **7**, 557–564.

Di Bitetti, M.S., De Angelo, C.D., Di Blanco, Y.E. & Paviolo, A. (2010). Niche partitioning and species coexistence in a Neotropical felid assemblage. *Acta Oecologica*, **36**, 403–412.

Caro, T.M. & Stoner, C.J. (2003). The potential for interspecific competition among African carnivores. *Biological Conservation*, **110**, 67–75.

Downs, J.A., Lamb, D., Hyzer, G., Loraamm, R., Smith, Z.J. & O'Neal, B.M. (2015). Quantifying spatio-temporal interactions of animals using probabilistic space-time prisms. *Applied Geography*, **55**, 1–8.

Farris, Z.J., Gerber, B.D., Karpanty, S., Murphy, A., Andrianjakarivelo, V., Ratelolahy, F. & Kelly, M.J. (2015). When carnivores roam: Temporal patterns and overlap among Madagascar's native and exotic carnivores. *Journal of Zoology*, **296**, 45–57.

Fisher, R.A. (1932). *Statistical Methods for Research Workers*, 4th Edition. Oliver and Boyd, Edinburgh.

Forrester, T.D., O'Brien, T., Fegraus, E., Jansen, P.A., Palmer, J., Stern, B., Kays, R., Ahumada, J. & McShea, W. (2016). An Open Standard for Camera-Trap Data. *Biodiversity Data Journal*, **4**, e10197.

Foster, V.C., Sarmento, P., Sollmann, R., Tôrres, N., Jácomo, A.T.A., Negroes, N., Fonseca, C. & Silveira, L. (2013). Jaguar and Puma Activity Patterns and Predator-Prey Interactions in Four Brazilian Biomes. *Biotropica*, **45**, 373–379.

Fuller, A., Dawson, T., Helmuth, B., Hetem, R.S., Mitchell, D. & Maloney, S.K. (2010). Physiological mechanisms in coping with climate change. *Physiological and Biochemical Zoology*, **83**, 713–720.

Haim, A. & Fourie, F. le R. (1980). Heat production in nocturnal (Praomys natalensis) and diurnal (Rhabdomys pumilio) South African murids. *South African Journal of Zoology*, **15**, 91–94.

Harmsen, B.J., Foster, R.J., Silver, S.C., Ostro, L.E.T. & Doncaster, C.P. (2009). Spatial and Temporal Interactions of Sympatric Jaguars (Panthera onca) and Pumas (Puma concolor) in a Neotropical Forest. *Journal of Mammalogy*, **90**, 612–620.

Ikeda, T., Uchida, K., Matsuura, Y., Takahashi, H., Yoshida, T., Kaji, K. & Koizumi, I. (2016). Seasonal and diel activity patterns of eight sympatric mammals in northern Japan revealed by an intensive camera-trap survey. *PLoS ONE*, **11**, 1–16.

Kamler, J.F., Johnson, A., Vongkhamheng, C. & Bousa, A. (2012). The diet, prey selection, and activity of dholes (Cuon alpinus) in northern Laos. *Journal of Mammalogy*, **93**, 627–633.

Karanth, K.U., Srivathsa, A., Vasudev, D., Puri, M., Parameshwaran, R. & Kumar, N.S. (2017). Spatio-temporal interactions facilitate large carnivore sympatry across a resource gradient. *Proceedings of the Royal Society of London B: Biological Sciences*, **284**.

Kissling, W.D., Dormann, C.F., Groeneveld, J., Hickler, T., Kühn, I., Mcinerny, G.J., Montoya, J.M., Römermann, C., Schiffers, K., Schurr, F.M., Singer, A., Svenning, J.C., Zimmermann, N.E. & O'Hara, R.B. (2012). Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. *Journal of Biogeography*, **39**, 2163–2178.

Kronfeld-Schor, N. & Dayan, T. (2003). Partitioning of time as an ecological resource. *Annual Review of Ecology, Evolution, and Systematics*, **34**, 153–181.

Levy, O., Dayan, T., Kronfeld-Schor, N. & Porter, W.P. (2012). Biophysical Modeling of the Temporal Niche: From First Principles to the Evolution of Activity Patterns. *The American Naturalist*, **179**, 794–804.

Linkie, M. & Ridout, M.S. (2011). Assessing tiger-prey interactions in Sumatran rainforests. *Journal of Zoology*, **284**, 224–229.

Long, J.A. (2015). Quantifying spatial-temporal interactions from wildlife tracking data: issues of space, time, and statistical significance. *Procedia Environmental Sciences*, **26**, 3–10.

Lund, U. & Agostinelli, C. (2012). Circstats: circular statistics, from 'topics in circular statistics'(2001).

Lynam, A.J., Jenks, K.E., Tantipisanuh, N., Chutipong, W., Ngoprasert, D., Gale, G.A., Steinmetz, R., Sukmasuang, R., Bhumpakphan, N., Grassman, L.I., Cutter, P., Kitamura, S., Reed, D.H., Baker, M.C., Mcshea, W., Songsasen, N. & Leimgruber, P. (2013). Terrestrial activity patterns of wild cats from camera-trapping. *Raffles Bulletin of Zoology*, **61**, 407–415.

MacKenzie, D.I., Bailey, L.L. & Nichols, J.D. (2004). Investigating species co-occurrence patterns when species are detected imperfectly. *Journal of Animal Ecology*, **73**, 546–555.

Meredith, M. & Ridout, M. (2016). overlap: Estimates of coefficient of overlapping for animal activity patterns. R package version 0.2.6.

Monterroso, P., Alves, P.C. & Ferreras, P. (2013). Catch me if you can: Diel activity patterns of mammalian prey and predators. *Ethology*, **119**, 1044–1056.

Nouvellet, P., Rasmussen, G.S.A., MacDonald, D.W. & Courchamp, F. (2012). Noisy clocks and silent sunrises: Measurement methods of daily activity pattern. *Journal of Zoology*, **286**, 179–184.

Palomares, F. & Caro, T.M. (1999). Interspecific Killing among Mammalian Carnivores. *The American Naturalist*, **153**, 492–508.

Parsons, A.W., Bland, C., Forrester, T., Baker, M.C., Schuttler, S.G., McShea, W.J., Costello, R. & Kays, R. (2016). The ecological impact of dogs on wildlife in protected areas in eastern North America. *Biological Conservation*, **203**, 75–88.

Peck, S.L. (2004). Simulation as experiment: A philosophical reassessment for biological modeling. *Trends in Ecology and Evolution*, **19**, 530–534.

R Core Team. (2017). R: A language and environment for statistical computing.

Richmond, O.M.W., Hines, J.E. & Beissinger, S.R. (2010). Two-species occupancy models: a new paramaterization applied to co-occurence of secretive rails. *Ecological Applications*, **20**, 2036–2046.

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.

Ross, J., Hearn, A.J., Johnson, P.J. & Macdonald, D.W. (2013). Activity patterns and temporal avoidance by prey in response to Sunda clouded leopard predation risk. *Journal of Zoology*, **290**, 96–106.

Schuette, P., Wagner, A.P., Wagner, M.E. & Creel, S. (2013). Occupancy patterns and niche partitioning within a diverse carnivore community exposed to anthropogenic pressures. *Biological Conservation*, **158**, 301–312.

St-Pierre, C., Ouellet, J.P. & Crête, M. (2006). Do competitive intraguild interactions affect space and habitat use by small carnivores in a forested landscape? *Ecography*, **29**, 487–496.

Steenweg, R., Hebblewhite, M., Kays, R., Ahumada, J., Fisher, J.T., Burton, C., Townsend, S.E., Carbone, C., Rowcliffe, J.M., Whittington, J., Brodie, J., Royle, J.A., Switalski, A., Clevenger, A.P., Heim, N. & Rich, L.N. (2017). Scaling up camera traps: monitoring the planet's biodiversity with networks of remote sensors. *Frontiers in Ecology and the Environment*, **15**, 26–34.

Sunarto, S., Kelly, M.J., Parakkasi, K. & Hutajulu, M.B. (2015). Cat coexistence in central Sumatra: Ecological characteristics, spatial and temporal overlap, and implications for management. *Journal of Zoology*, **296**, 104–115.

Tambling, C.J., Minnie, L., Meyer, J., Freeman, E.W., Santymire, R.M., Adendorff, J. & Kerley, G.I.H. (2015). Temporal shifts in activity of prey following large predator reintroductions. *Behavioral Ecology and Sociobiology*, **69**, 1153–1161.

Vanak, A.T., Fortin, D., Thaker, M., Ogden, M., Owen, C., Greatwood, S. & Slotow, R. (2013). Moving to stay in place: Behavioral mechanisms for coexistence of African large carnivores. *Ecology*, **94**, 2619–2631.

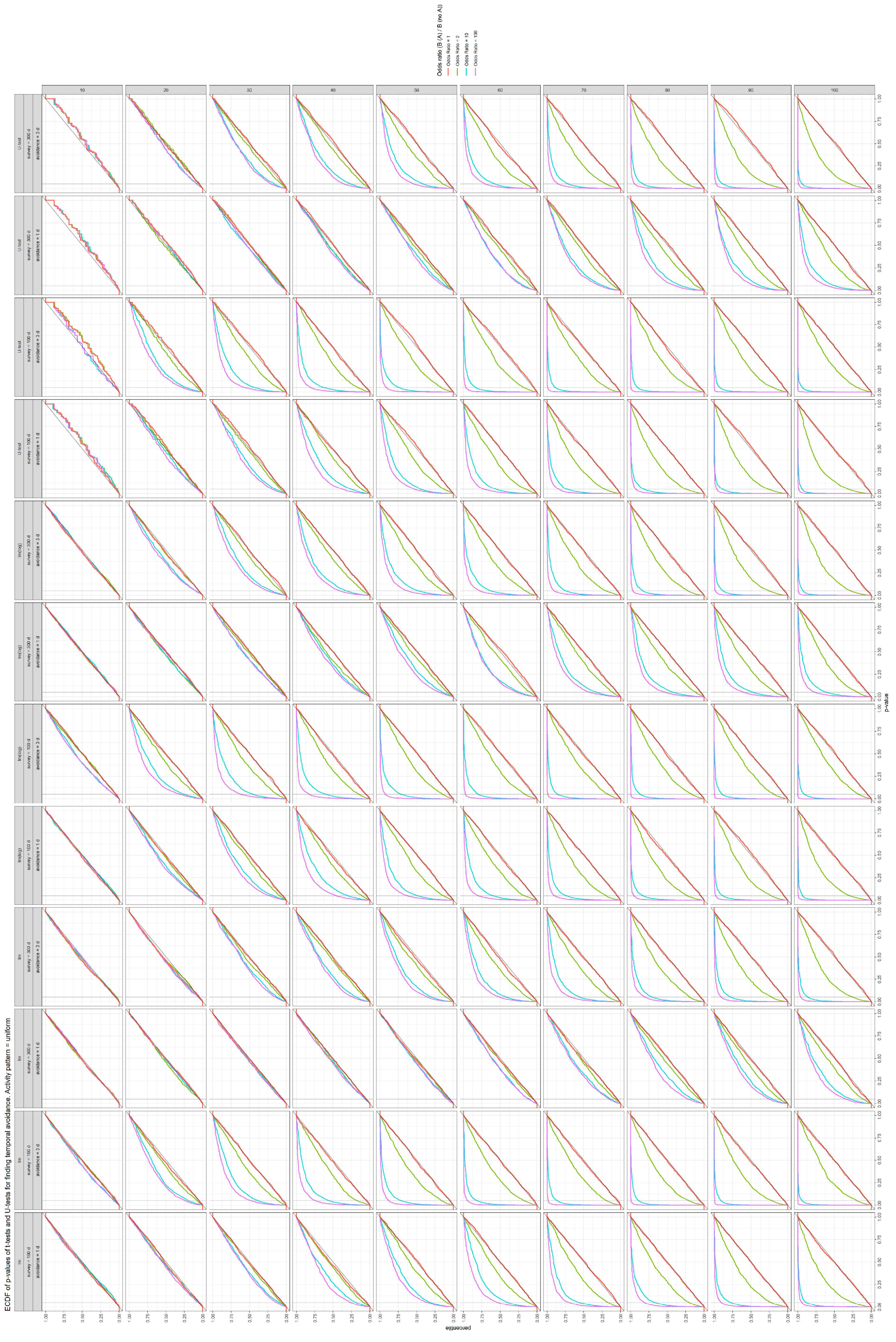
Wisz, M.S., Pottier, J., Kissling, W.D., Pellissier, L., Lenoir, J., Damgaard, C.F., Dormann, C.F., Forchhammer, M.C., Grytnes, J.A., Guisan, A., Heikkinen, R.K., Høye, T.T., Kühn, I., Luoto, M., Maiorano, L., Nilsson, M.C., Normand, S., Öckinger, E., Schmidt, N.M., Termansen, M., Timmermann, A., Wardle, D.A., Aastrup, P. & Svenning, J.C. (2013). The role of biotic interactions in shaping distributions and realised assemblages of species: Implications for species distribution modelling. *Biological Reviews*, **88**, 15–30.

Supplementary Figures

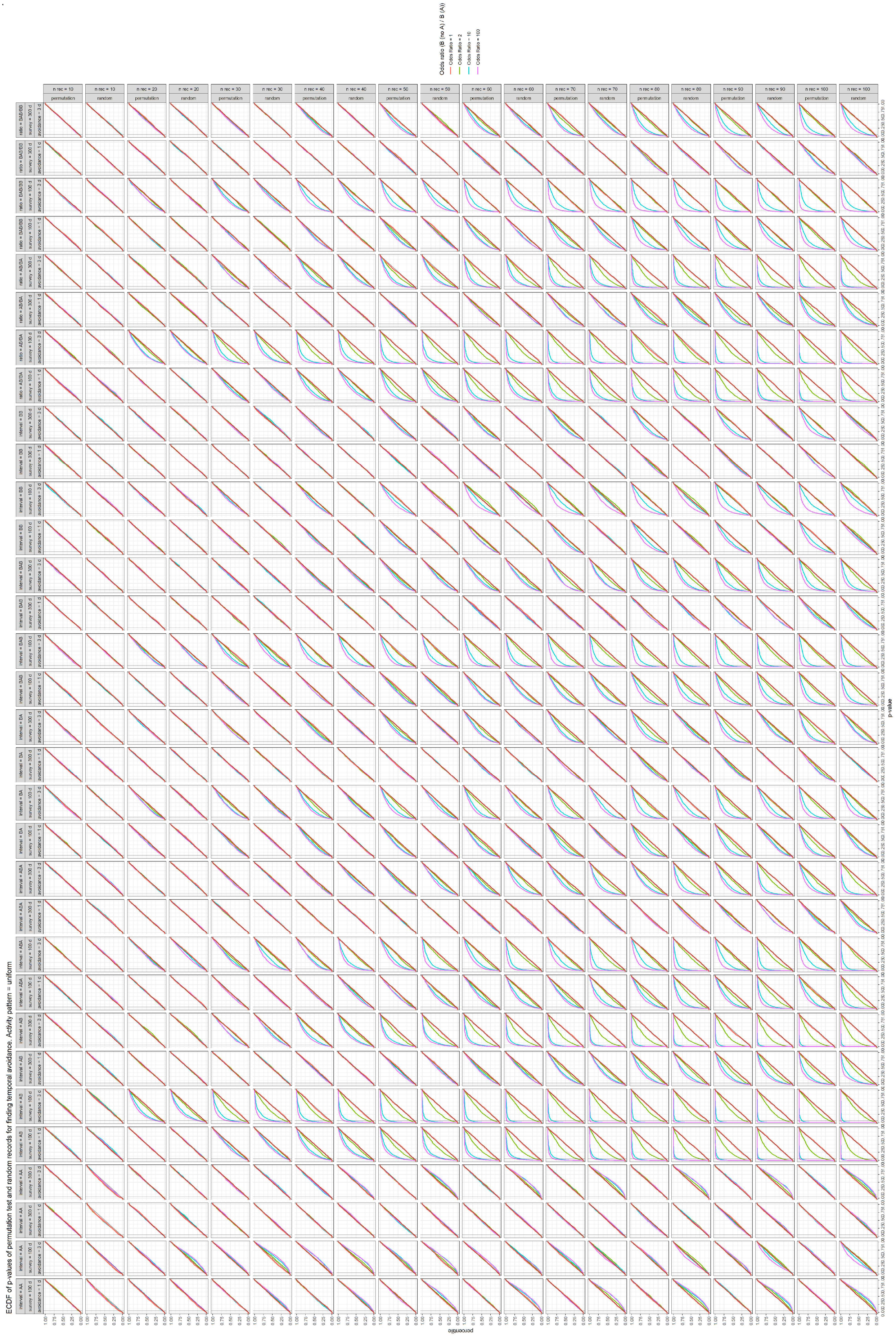
S4.1. Empirical cumulative density functions (ECDFs) of p-values of linear models and U-test for detecting spatiotemporal avoidance between two species from 1000 iterations for each factor combination. The margins indicate the factor combinations. In the columns: the method applied (lm = linear model, lm(log) = linear model with log-transformed data, U-test = Mann-Whitney U-test), survey duration (100/300 days), time until recovery of detection rate of the subordinate species after presence of the dominant species (1/3 days). The rows show the numbers of simulated records. Data shown are for uniform activity patterns, but almost identical to unimodal and bimodal activity patterns. The distribution of p-values under the null hypothesis corresponds to a uniform distribution if it follows the grey diagonal line (page 89).

S4.2. ECDFs of the random record method for detecting spatiotemporal avoidance between two species from 1000 iterations for each factor combination. The margins indicate the factor combinations. In the columns: the time interval assessed, survey duration (100/300 days), time until recovery of detection rate of the subordinate species after presence of the dominant species (1/3 days). The rows show the numbers of simulated records and both methods, the permutation test and randomly generated records. Data shown are for uniform activity patterns, but almost identical to unimodal and bimodal activity patterns. The distribution of p-values under the null hypothesis corresponds to a uniform distribution if it follows the grey diagonal line (page 90).

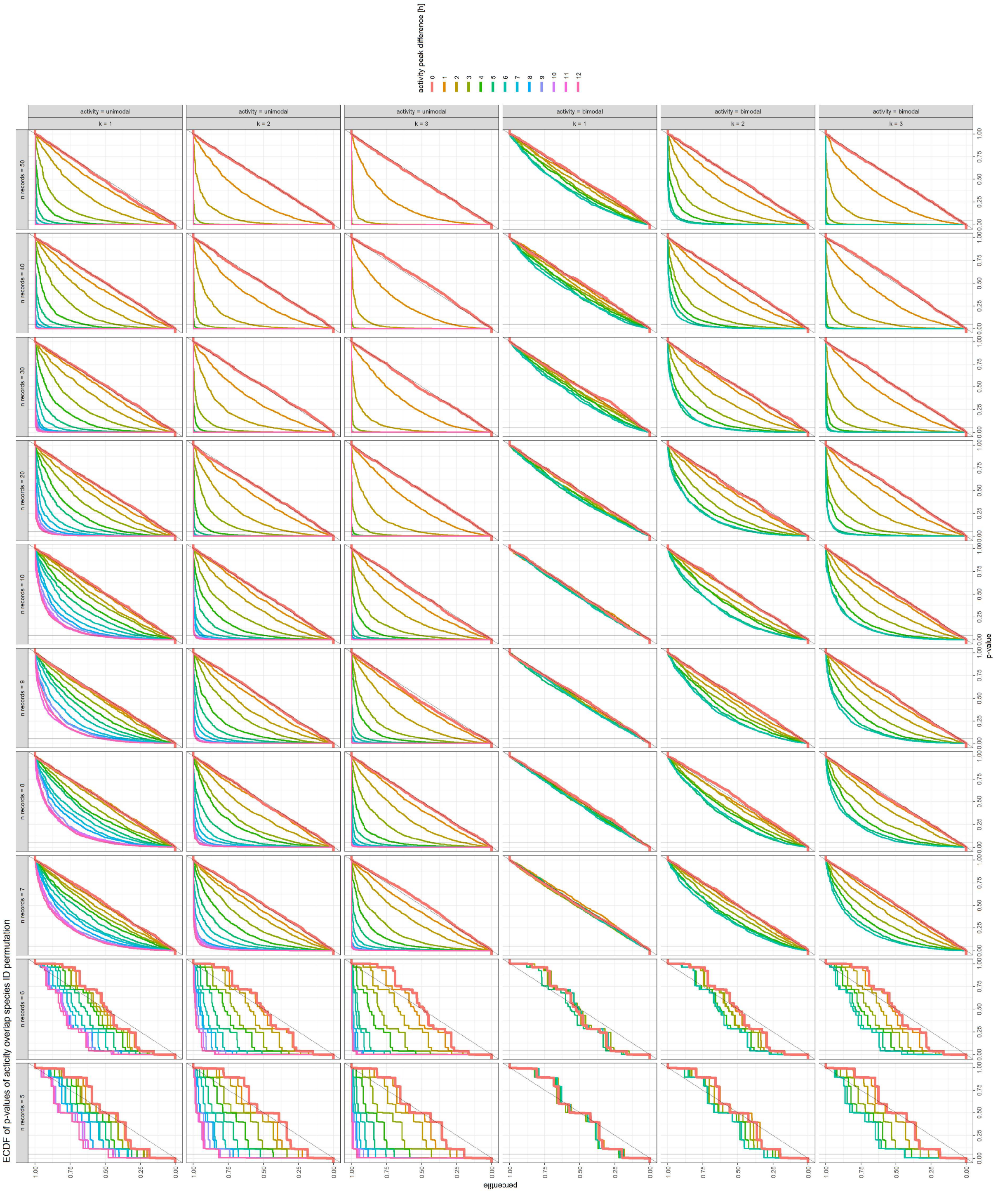
S4.3. ECDFs of permutation tests for finding temporal segregation between two species from 1000 iterations for each factor combination. The columns show the number of simulated records of each species. The rows show the activity pattern (uni- or bimodal) and its respective concentration parameter k. An activity peak difference of 0 hours means equal activity peaks between two species, hence it corresponds to the null hypothesis (thick orange line. The distribution of p-values under the null hypothesis corresponds to a uniform distribution if it follows the grey diagonal line (page 91).



Supplementary Figure S4.1



Supplementary Figure S4.2



Supplementary Figure S4.3

Supplement

Supplement 4.1: R function for simulating camera trappingrecords of interacting species

```
library(CircStats)


### function arguments


#    n_records_A                          # number of records of the dominant species
#    n_records_B                          # number of records of the subordinate species
#    n_days                               # simulated survey duration in days
#    effect_duration_days       # how many days until full recovery of detection probability of subordinate species after a detection of the dominant species
#    effect_reduction_factor    # the odds ratio between the odds of detecting the subordinate species after the dominant species was present relative to the odds of detecting the subordinate species in the absence of the dominant species
#    linearRecoveryOfB          # after records of dominant species, does detection probability of subordinate species recover linearly?
#    do_plots = FALSE                     # make a plot?
#     family = c("uniform", "von Mises", "Cauchy", "von Mises Mixture")    # the type of probability density distribution describing daily activity patterns.
#    species_offset_hours = 0    # if set, species B activity peaks are shifted relative to peaks of A (in hours)
#    densityFunctionParameters           # a list with parameters of von Mises / Cauchy distributions



simulateInteractionRecordsActivity_simple <- function(n_records_A,
                                  n_records_B,
                                  n_days,
                                  effect_duration_days,
                                  oddsRatio,
                                  linearRecoveryOfB,
                                  do_plots = FALSE,
                                  family = c("uniform", "von Mises", "Cauchy", "von Mises Mixture"),
                                  species_offset_hours = 0,
                                  densityFunctionParameters
)
{

  if(oddsRatio < 0) stop("oddsRatio must be be > 0")
```

```
if(oddsRatio < 1) warning("oddsRatio < 1 suggests attraction. This is not fully supported in this function.")

stopifnot(length(family) == 1)

family <- match.arg(family, choices =  c("uniform", "von Mises", "Cauchy", "von Mises Mixture"))



# check function input according to distribution function family
if(family == "uniform")  {
  densityFunctionParameters <- list()
  mu    <- 0
  kappa <- 0
} else { stopifnot(hasArg(densityFunctionParameters)); stopifnot(is.list(densityFunctionParameters))}

if(family == "von Mises") {
  stopifnot(all(c("mu", "kappa") %in% names(densityFunctionParameters)))
  mu    <- densityFunctionParameters$mu
  kappa <- densityFunctionParameters$kappa
  if(mu < 0 | mu > 2 * pi) stop("mu must be between 0 and 2*pi")
}

if(family == "Cauchy"){
  stopifnot(all(c("mu", "rho") %in% names(densityFunctionParameters)))
  mu  <- densityFunctionParameters$mu
  rho <- densityFunctionParameters$rho
  if(rho < 0 | rho > 1) stop("rho must be between 0 and 1")
}

if(family == "von Mises Mixture") {
  stopifnot(all(c("mu1", "mu2", "kappa1", "kappa2") %in% names(densityFunctionParameters)))
  mu1     <- densityFunctionParameters$mu1
  mu2     <- densityFunctionParameters$mu2
  kappa1 <- densityFunctionParameters$kappa1
  kappa2 <- densityFunctionParameters$kappa2
  if(mu1 < 0 | mu1 > 2 * pi) stop("mu1 must be between 0 and 2*pi")
  if(mu2 < 0 | mu2 > 2 * pi) stop("mu2 must be between 0 and 2*pi")
}
```

```
  if(species_offset_hours > 24) stop("species_offset_hours can't be greater than 24")

  det_prob_baseline <- 1
  tz <- "UTC"

  # generate a time sequece with 1-minute intervals for 1 day (00:00:00 - 23:59:00) in radial and clock time
  seq0 <- seq(0, (2*pi), length.out = 1441)
  seq0 <- seq0[-length(seq0)]

  date0 <- base::as.Date(rep(seq(0, n_days - 1, by = 1), each = length(seq0)), origin = "1970-01-01",  tz = tz)
# 0... n_days -1 because origin = day 0!!
  n_events <- length(date0)

  det_prob_0 <- rep(det_prob_baseline, times = n_events)

  species_offset_rad <- species_offset_hours/24*2*pi

  if(family %in% c("uniform", "von Mises")){  # density of von Mises distribution (it is uniform if kappa = 0)

    density_distribution_A <- dvm(theta = seq0,
                                  mu    = mu,
                                  kappa = kappa)

    if(species_offset_hours == 0){
      density_distribution_B <- density_distribution_A
    } else {
      if(mu + species_offset_rad > 2*pi) stop("mu + species_offset_hours is > 2 * pi")
      density_distribution_B <- dvm(theta = seq0,
mu    = mu + species_offset_rad,
kappa = kappa)
    }
  }
  if(family == "Cauchy"){        # Wrapped Cauchy Density Function

    density_distribution_A <- dwrpcauchy(theta = seq0,
                                         mu    = mu,
                                         rho   = rho)
```

```
    if(species_offset_hours == 0){
      density_distribution_B <- density_distribution_A
    } else {
      if(mu + species_offset_rad > 2*pi) stop("mu + species_offset_hours is > 2 * pi")
      density_distribution_B <- dwrpcauchy(theta = seq0,
mu    = mu + species_offset_rad,
rho   = rho)
    }
  }

  if(family == "von Mises Mixture"){  # Mixture of 2 von Mises Distributions
    density_distribution_A <- dmixedvm(theta  = seq0,
mu1    = mu1,
                                        mu2    = mu2,
                                        kappa1 = kappa1,
kappa2 = kappa2,
                                        p      = 0.5)

    if(species_offset_hours == 0){
      density_distribution_B <- density_distribution_A
    } else {
      if(mu1 + species_offset_rad > 2*pi) stop("mu1 + species_offset_hours is > 2 * pi")
      if(mu2 + species_offset_rad > 2*pi) stop("mu2 + species_offset_hours is > 2 * pi")

      density_distribution_B <- dmixedvm(theta  = seq0,
mu1    = mu1 + species_offset_rad,
                                          mu2    = mu2 + species_offset_rad,
                                          kappa1 = kappa1,
                                          kappa2 = kappa2,
                                          p      = 0.5)
}
  }

  #  observation probabilites (relative, not absolute) for each minute in study period

  det_prob_A <- det_prob_0 * density_distribution_A
  if(species_offset_hours == 0){
    det_prob_B_0 <- det_prob_A
```

```
 } else {
  det_prob_B_0 <- det_prob_0 * density_distribution_B
 }

 # generate records of species A conditional on detection probability of A

 time_random_A_tmp <- sort(sample(x = seq(1, n_events), size = n_records_A, prob = det_prob_A, replace
= FALSE))

 # create probability density for species B dependent on records of species A and the strength of the effect

 if(!isTRUE(linearRecoveryOfB)){          # constant effect of strength "oddsRatio" for length of
"effect_duration_days"

  det_prob_B <- det_prob_B_0

  which_have_reduced_p <- unique(as.vector(sapply(time_random_A_tmp, FUN = function(X){seq(from =
X, to = X + (effect_duration_days * 1440))})))
  if(any(which_have_reduced_p > n_events)){
   which_have_reduced_p <- which_have_reduced_p[-which(which_have_reduced_p > n_events)]
  }

  det_prob_B[which_have_reduced_p] <- det_prob_B_0[which_have_reduced_p] / oddsRatio

 } else {   #linear recovery of det_prob_B

  det_prob_B <- det_prob_B_0

  which_have_reduced_p <- unique(as.vector(sapply(time_random_A_tmp, FUN = function(X){seq(from =
X, to = X + (effect_duration_days * 1440))})))
  # which_have_reduced_p <- sapply(lapply(time_random_A_tmp, FUN = function(X){seq(from = X, to = X
+ (effect_duration_days * 1440))}),
  #                             FUN  = as.vector)

  if(max(which_have_reduced_p) > n_events){
   which_have_reduced_p <- which_have_reduced_p[-which(which_have_reduced_p > n_events)]
  }
  # for each time point affected by presence of A, calculate distance from last record of A
```

```
    test <- sapply(time_random_A_tmp, FUN = function(x) {which_have_reduced_p - x})
    test[test < 0] <- NA
    distance_from_event <- apply(test, MARGIN = 1, FUN = min, na.rm = TRUE)   # = time steps since last detection of A

    n_steps_until_recovery <- effect_duration_days * 1440                       # number of steps after which detection probability of B is back to normal

    # calculate detection probability taking into account linear recovery
    det_prob_B[which_have_reduced_p] <-  det_prob_B[which_have_reduced_p] / oddsRatio +
      (distance_from_event / n_steps_until_recovery) * (1 - (1 / oddsRatio)) * det_prob_B[which_have_reduced_p]
  }

  # random sample from these (realised observations)
 time_random_weighted_B_tmp <- sort(sample(x = seq(1, n_events), size = n_records_B, replace = FALSE, prob = det_prob_B))

  # make data frame with results for non-independent detections
  simu2 <- data.frame(Station         = "station 1",
                      Species         = rep(c("Species A", "Species B"), times = c(n_records_A, n_records_B)),
                      DateTimeOriginal  = strptime(c(time_random_A_tmp, time_random_weighted_B_tmp) * 60, format = "%s", tz = tz))

  simu2 <- cbind(simu2,
                 Date     = base::as.Date(simu2$DateTimeOriginal, tz = tz),
                 Time     = format(simu2$DateTimeOriginal, format = "%H:%M:%S"),
                 Time_sec = as.numeric(format(simu2$DateTimeOriginal, format = "%s")),
                 TimeRad  = ClocktimeToRadialTime(simu2$DateTimeOriginal))

  if(isTRUE(do_plots)){

    mfrow0 <- par()$mfrow
    on.exit(par(mfrow = mfrow0))

    # generate tile and subheadings for plot
    main_title <- paste("family = ", family,
                        ";  odds ratio = ", oddsRatio,
```

```
                  "; linear recovery of B = ", linearRecoveryOfB, sep = "")
  subtitle <- paste("n_records A = ", n_records_A, "; total length = ", n_days, " days", "; effect_duration =
", effect_duration_days, " days", sep = "")
  subtitle2 <- paste("n_records B = ", n_records_B, "; species_offset_hours = ", species_offset_hours, sep =
"")

  if(family == "von Mises") subtitle <- paste(subtitle, "; mu = ", round(mu, 2), " (= ", round(mu * 24/(2*pi)),
" o'clock); kappa = ", kappa,
                                    "; odds ratio (B(no A) / B(A)) =", oddsRatio , sep = "")
  if(family == "Cauchy")    subtitle <- paste(subtitle, "; mu = ", round(mu, 2), " (= ", round(mu * 24/(2*pi)), "
o'clock); rho = ", rho,
                                    "; odds ratio (B(no A) / B(A)) =", oddsRatio, sep = "")
  if(family == "von Mises Mixture")  subtitle <- paste(subtitle, "; mu1 = ", round(mu1, 2), " (= ", round(mu1
* 24/(2*pi)), " o'clock)",
"; mu2 = ", round(mu2, 2), " (= ", round(mu2 * 24/(2*pi)), " o'clock)",
"; kappa1 = ", kappa1, "; kappa2 = ", kappa2,
                                         "; odds ratio (B(no A) / B(A)) =", oddsRatio, sep = "")

  # set graphical parameters
  col_abline <- rgb(0, 0, 0, 0.2)
  col_abline2 <- rgb(0, 0, 0, 0.1)
  col_rug_A <- "red"
  col_rug_B <- "blue"
  lwd_rug <- 3



  # create plot

  # top plot: dominant species A
  par(mfrow = c(2,1))
  plot(det_prob_A, type = "l", axes = F, ylim = c(0, max(det_prob_A)),
     main = main_title, sub = subtitle, xlab = "", ylab = "probability weight")
  at_tmp_label <- seq(720, length(det_prob_A), by = 1440)          # location for x axis label (noon)
  at_tmp_vline <- seq(1, length(det_prob_A), by = 1440)            # location for abline at midnight each day
  axis(1, at = at_tmp_vline, labels = FALSE, tick = TRUE)    # make ticks
  axis(1, at = at_tmp_label, labels = paste("day", seq(1, (length(at_tmp_label)))), tick = FALSE)
  axis(2)
  abline(v = at_tmp_vline, col = col_abline)
```

```
  rug(time_random_A_tmp, lwd = lwd_rug, col = col_rug_A); box()
  polygonToPlot_A <- data.frame(x = c(1, 1:length(det_prob_A), length(det_prob_A), 1),
                                y = c(0, det_prob_A, 0, 0))
  polygon(x = polygonToPlot_A$x, y = polygonToPlot_A$y, border = NA, col = col_abline)


  # bottom plot: subordinate species B
  plot(det_prob_B, type = "l", axes = F, ylim = c(0, max(det_prob_B)), xlab = "", ylab = "probability weight",
sub = subtitle2)
  axis(1, at = at_tmp_vline, labels = FALSE, tick = TRUE)    # make ticks
  axis(1, at = at_tmp_label, labels = paste("day", seq(1, (length(at_tmp_label)))), tick = FALSE)
  axis(2)
  abline(v = at_tmp_vline, col = col_abline)
  rug(time_random_A_tmp, lwd = lwd_rug, col = col_rug_A)
  rug(time_random_weighted_B_tmp, lwd = lwd_rug, col = col_rug_B); box()

  polygonToPlot_B <- data.frame(x = c(1, 1:length(det_prob_B), length(det_prob_B), 1),
                                y = c(0, det_prob_B, 0, 0))
  polygon(x = polygonToPlot_A$x, y = polygonToPlot_A$y, border = col_abline, col = col_abline2)     # plot
where it would be
  polygon(x = polygonToPlot_B$x, y = polygonToPlot_B$y, border = NA, col = col_abline)

 }
 return(simu2)
}
```

```
# Helper function to  convert clock time to radial time (0, ..., 2 pi)


ClocktimeToRadialTime <- function(Clocktime,
                              timeformat = "%Y-%m-%d %H:%M:%S"
){
  DateTime2 <- strptime(as.character(Clocktime), format = timeformat, tz = "UTC")
  Time2     <- format(DateTime2, format = "%H:%M:%S", usetz = FALSE)
  Time.rad  <- (as.numeric(as.POSIXct(strptime(Time2, format = "%H:%M:%S", tz = "UTC"))) -
                as.numeric(as.POSIXct(strptime("0", format = "%S", tz = "UTC")))) / 3600 * (pi/12)
  return(Time.rad)
}




# a sample call to the function


###  function parameters explained
# 20 records of both species over 50 days
# after record of A, probability of encounter of B is reduced to 10% of its original value and recovers linearly within 1 day
# unimodal activity pattern for both species, activity peaks at noon for both species

records_simulated <- simulateInteractionRecordsActivity_simple (n_records_A = 20,
                                                    n_records_B = 20,
                                                    n_days = 50,
                                                    effect_duration_days = 1,
                                                    oddsRatio = 10,
                                                    linearRecoveryOfB = TRUE,
                                                    do_plots = TRUE,
                                                    family = "von Mises",
                                                    species_offset_hours = 0,
                                                    densityFunctionParameters = list(mu = pi, kappa = 2))
```

CHAPTER 5

General discussion

The purpose of this study was to develop new and advance already available methods for the management and analyses of wildlife data from camera trapping surveys. The study focused on streamlining camera trap data management workflows by providing a new toolbox in the widely used R language for that purpose. The second focus of this study was on the advancement of analytical tools for camera trapping data in spatial and temporal contexts, the two fundamental axes shaping ecological systems (Wolkovich *et al.* 2014), by assessing the usefulness of high-resolution remote sensing data in camera trap based occupancy models and scrutinising methods for detecting spatiotemporal interactions between species in camera trapping data. While this study is largely conceptual, it is broad in scope and its findings should be applicable to most camera trapping studies.

The importance of wildlife monitoring in a changing world

Globally, numerous species of mammals are threatened (Schipper *et al.* 2008), and while the threat status of species and the individual causes of threats differ between regions and species, the underlying causes are almost invariably anthropogenic, with habitat loss due to land conversion and degradation (Newbold *et al.* 2015; Struebig *et al.* 2015a), overexploitation due to hunting and poaching (Fa & Brown 2009; Harrison *et al.* 2016), and the dangers of climate change (Struebig *et al.* 2015b) being the most prominent ones. Large, ground-dwelling mammals are particularly threatened, and at the same time of high importance for the functioning of ecosystems (Ripple *et al.* 2014; Wolf & Ripple 2016). Large carnivores exert trophic control over lower trophic levels (by means of direct predation and induced avoidance behaviour), which can have cascading effects and shape whole food webs (Terborgh *et al.* 1999; Suraci *et al.* 2016). Large ungulates can change vegetation structure and succession and play an important roles as seed dispersers, and thus have an ecological landscaping effect (McShea & Rappole 1992; Redford 1992; Sinclair 2003). For the same reasons, large ungulates affect ecosystem carbon storage and thus may indirectly influence mechanisms associated with climate change (Bello *et al.* 2015; Osuri *et al.* 2016). Hence, many large mammal species are considered keystone species with particular value for ecosystem functioning and severe consequences ecosystem health should they be lost, making them a conservation priority (Simberloff 1998; Camargo-Sanabria *et al.* 2014).

Larger species have higher extinction risk than smaller species because, to a higher degree than in smaller species (< 3kg), their extinction risk is not only determined by environmental factors, but by a combination of environmental factors and intrinsic traits (Cardillo *et al.* 2005). This may contribute to unexpectedly rapid future losses of large mammal biodiversity under increasing anthropogenic pressure by growing human populations, e.g. due to increased hunting and poaching (Corlett 2007; Brodie *et al.* 2014), and it may be further exacerbated by potential lags in species extinction from an extinction debt accumulated from historical land cover conversion (Rosa *et al.* 2016).

Camera trapping surveys typically targeted these large ground-dwelling mammals, which are of tremendous ecological relevance, yet particularly vulnerable. Analyses of camera trapping data thus have high conservation and management relevance and can be used for monitoring and robust assessment of population status and trends of wildlife and serve as sensitive indicators for ecosystem-wide effects of anthropogenic pressures (Cheyne *et al.* 2016). Computational advances and collaborative efforts between practitioners are further expected to enable near real-time monitoring and analyses of mammalian biodiversity from camera trapping data at local, regional and global scales through networks of camera traps and other sensors in the near future (Steenweg *et al.* 2017). Such studies are of particular importance because current assessments of species vulnerability tend to be restricted to local and regional scales and rarely take place at continental or global scales (Pacifici *et al.* 2015). To enable such collaborative efforts, sound data management enabling efficient and reproducible analyses is imperative.

Camera trapping for wildlife monitoring

Given the multitude of survey methods for wildlife monitoring that were developed and applied in the past, the question arises why the potential to conduct such comprehensive analyses is seen in camera trapping in particular?

Wildlife surveys are conducted to answer a variety of ecological questions, ranging from individual behavioural observations to population assessments or trends to community level inference, and the choice of methods depends on the question of interest (Gese 2001). Most traditional methods are rather specific in their aims or too labour-intensive to allow broader inferences. Telemetry studies provide more detailed animal movement data than any other survey method, but they require live capture, which is often difficult to achieve and regularly result in low sample sizes, and are complicated by statistical challenges (Aarts *et al.* 2008). Therefore, telemetry studies are more suitable for specific ecological questions, less so for population-level inferences, let alone global ecological inferences (Hebblewhite & Haydon 2010). Direct counts of animals, e.g. along transects, is labour-intensive and was found to be of little use as a stand-alone technique due to low numbers of detection in random transects and biases introduced in non-random transects (Mathai *et al.* 2013). Track plots are similarly labour-intensive due to the need for frequent checks (preferably daily) and susceptible to being negatively impacted by adverse weather conditions.

Camera trapping, on the other hand, was often found to be superior in terms of data collected and more cost-efficient than other survey methods (Silveira *et al.* 2003; Weckel *et al.* 2006; Lyra-Jorge *et al.* 2008) Camera trap networks can be operated synchronously in for months without requiring human presence or further habitat disturbance and provide information about the mammal community instead of targeting individual animals only, allowing broader analyses. Methodological advances in analyses of camera trapping now allow for more sophisticated and robust population and community assessments far beyond species inventories, tapping on the wealth of ecological data collected by networks of camera traps (MacKenzie *et al.* 2006; Royle *et al.* 2014).

In addition, innovative uses of camera traps kept emerging in recent years, suggesting that camera traps still hold unexplored potentials for ecological analyses. Examples include arboreal camera trapping for studying the ecology and behaviour of arboreal species that are not readily detected using typical terrestrial camera traps (but see Ancrenaz *et al.* 2014; Gregory *et al.* 2014), analyses of temporal interactions between species (Ridout & Linkie 2009), applications in biodiversity monitoring within the framework of forest certification schemes (Sollmann *et al.* 2017), assessments of frugivory and seed dispersal by ruminants (Prasad *et al.* 2010), use in law enforcement to link confiscated skins of poached animals to their origin (Hiby *et al.* 2009), or underwater cameras for marine monitoring (Williams *et al.* 2014). Furthermore, as the saying goes, a picture is worth a thousand words, and the images that are collected by camera traps are highly suitable for raising public awareness and appreciation of wildlife and can thus increase public engagement in science and conservation alike (Toomey & Domroese 2013; Swanson *et al.* 2015; McShea *et al.* 2016).

Novel methods such as sampling invertebrate-derived or environmental DNA (iDNA / eDNA) hold potential for rapid and detailed mammalian wildlife assessments (Schnell *et al.* 2012; Bohmann *et al.* 2014). But at the same time, the increased complexity of these molecular methods (in DNA extraction, sequencing, species assignment, and the availability of reference sequences), issues of spatiotemporal uncertainty, and potential seasonal unavailability of invertebrate species for iDNA sampling severely complicate analyses (Schnell *et al.* 2015). Hence, much development and technological refinement will be needed in addition to rigorous comparison of results with existing methods before these novel molecular methods can achieve comparable reliability to camera trapping.

In summary, camera trapping can be considered the most versatile method for surveying wildlife available at the moment and combines numerous advantages of other methods. It is widely used and has proven its potential in diverse ecological applications. In addition, it is often more cost-efficient and less labour-intensive than other methods in terms of the amount of data collected and the analytical possibilities offered. Therefore, camera trapping is legitimately considered the method of choice for global mammalian biodiversity assessments and monitoring.

Advances and challenges in camera trap data management

Over the past decades, camera trapping equipment has become more sophisticated, reliable, user-friendly and affordable. Enabled by the development of new analytical approaches, the use of camera traps has been extended to a multitude of ecological analyses, granting insights and allowing analyses that were unthinkable just two decades ago. This increase in popularity is reflected in the publication of hundreds of papers pertaining to camera trapping every year (see Chapter 1 and McCallum 2012). Irrespective of the aim of these studies, be it occupancy, behavioural analyses or population assessments, the basic workflow for data management and preparation of analyses is similar and consists of assigning images to locations, identifying species (and/or individuals) and preparing data for subsequent analyses.

A number of approaches were available to facilitate and streamline camera trap data management, usually with slightly different foci and workflows, but all with the intention to automate repetitive tasks, reduce manual labour and thus increase efficiency and reliability of data generated in camera trapping. The popularity of the R language for ecological analyses of camera trapping data and its efficiency when working on large amounts of data induced the development of camtrapR, the first R package for camera trap data management. Since its publication in 2015 (Niedballa *et al.* 2015), it has quickly gained popularity and is now being used around the world by academic researchers and conservationists to manage camera trapping projects. The use of the package was encouraged in a number of training courses that were held since the package release, particularly in Southeast Asia, and the user base is supported by a Google group[1] with 75 members (as of September 2017) which the author founded and moderates. The package is being actively maintained, expanded and updated regularly to further increase its usefulness to the camera trapping community.

The multitude of different software toolboxes for camera trap data management available today has a number of advantages and disadvantages. Different projects have different aims and study designs, complicating the development of a single toolbox that can accommodate all possible study designs.

---

[1] https://groups.google.com/forum/#!forum/camtrapr

Consequently, more available toolboxes increase the chances for finding suitable software for specific needs of particular studies. But, different software adhering to different data storage schemes and paradigms, this diversity may on the other hand hamper data exchange and complicate collaborative efforts. Future directions for the development for software for camera trap data management should therefore aim at standardising data storage to improve compatibility between data sets and allow scalability of analyses (Steenweg *et al.* 2017). To ensure long-term data storage and availability, local data storage should be linked to online repositories (Global Biodiversity Information Facility 2014). Modern camera traps can usually records videos, but dependable video support in camera trap data management software is currently not possible because of inconsistent metadata storage in video files, which may easily lead to loss of essential information like date and time. Species and individual identification is generally done manually and can be very time-consuming. Alternative approaches, such as machine learning algorithms for automated species and individual identification or citizen science approaches hold potential to significantly reduce the time effort needed to manually identify species and individuals from camera trapping data (Pimm *et al.* 2015; Swanson *et al.* 2015).

Combining advances in camera trapping and remote sensing

The growing number of camera trapping studies and calls to combine data sets regionally and even globally not only requires standardisation in terms of camera trap data management, but also in terms of the data used to explain the distribution and population parameters of animal populations and communities. In parallel to the developments in camera trapping hardware and software, remote sensing technology and analytical possibilities thereof have advanced rapidly in recent years. While for many years 30-m Landsat data were the best available data for use in conservation biology and still provide highly valuable information on the global scale (Leimgruber *et al.* 2005; Hansen *et al.* 2013), more recent generations of satellites provide higher resolution and more detailed earth observations. Remote sensing is the method of choice to deliver earth observation data for use in global biodiversity assessments due to its standardised, continuous, near real-time coverage of the earth's surface, capturing patterns and processes in unprecedented detail and spatial extent (Pettorelli *et al.* 2014; Proença *et al.* 2016). However, the effective use of remote sensing in biodiversity conservation and research requires unrestricted and free access to data and must ensure data continuity (Turner *et al.* 2015).

This study was the first to assess the applicability of 5-m RapidEye satellite imagery in camera trap based occupancy models and found strong evidence for higher explanatory value of covariates derived from these high-resolution data, highlighting their potential for explaining species-habitat associations at local and regional scales. While the amount of detail provided by these data was shown to be of value, further studies will be needed to confirm their usefulness at larger scales, e.g. at the national, continental or global level, as applicability at those scales may be impeded by excessive detail in high resolution data and soaring requirement in terms of data storage and computational power.

Whatever the data basis of global biodiversity assessments will be, the question of scale is universal in ecology and affects remote sensing data, in-situ field surveys and biodiversity assessments alike. Consequently, all inferences drawn from these models are likewise affected by their respective scale and should only be interpreted at that scale. Analyses at multiple spatial scales can help better understand mechanisms underlying observed patterns of species distribution and co-occurrence (Gillingham *et al.* 2012; Lindström *et al.* 2013; Harms & Dinsmore 2016).

Species interactions in predictive distribution modelling

While the development of predictive models for species distributions and abundance under climate change is central to future biodiversity conservation, most of these predictive models exclude important biological mechanisms (Urban *et al.* 2016). Six biological mechanisms were identified that are often neglected in current predictive models due to lack of data, but which could improve prediction accuracy and robustness of these models to inform conservation efforts under climate change considerably if they were included.

Species interactions are one of these key biological processes and can limit species distributions across large spatial and temporal scales (competition in particular; Pigot & Tobias 2013). The high-priority parameters about species interactions that are needed include specialist interactions, top-down food web interactions and timing mismatches among interacting species (Urban *et al.* 2016). Incorporating these species interaction parameters may enable the development of more realistic, mechanistic species distribution models (Urban *et al.* 2013).

Camera traps can deliver information about spatiotemporal species interactions at point localities and may thus provide the species interaction information needed for mechanistic predictions of species and community responses to climate change. A number of approaches were developed to assess spatiotemporal interactions in camera trapping data, but these methods remain poorly studied and have not been assessed in terms of their statistical properties, mainly because the true state of system is unobserved in field data. Data simulation, mimicking basic properties of real-world systems at reduced complexity, makes it possible to experiment with these systems, which would often be too difficult or costly to do in real world scenarios (Peck 2004) and was applied here to assess different methods for detecting spatiotemporal interactions in camera trapping data.

The assessment of different methods for detecting spatiotemporal interactions in camera trapping data presented in chapter 4 and the recommendations about which method to choose in which situation can help derive more robust inferences about species interactions from camera trapping data. Besides answering basic ecological questions, such information can also be used in mechanistic and thus more realistic predictive frameworks of future biodiversity scenarios under climate change. Furthermore, given the potential of global change for altering species interaction patterns (Tylianakis *et al.* 2008), spatiotemporal interactions may be a more sensitive indicator for ecosystem changes than occupancy patterns.

Conclusion

Camera trapping is a prime method for mammalian biodiversity monitoring and for studying rare and elusive species, or those occurring at low densities, where direct observations are too challenging. It is particularly suitable in closed habitats (e.g. forests) where direct observations are even more challenging than in open habitats such as savannahs. Because of its suite of favourable properties, camera traps are expected to stay a main tool in the hands of ecologists and conservation biologists for the rapid and robust assessment of mammalian biodiversity in the years to come. This study has contributed to these efforts by facilitating camera trap data management, assessing the applicability of high-resolution remote sensing data in the analyses of camera trapping data, and providing a theoretical framework and recommendations for the inference of spatiotemporal species interactions from camera trapping data.

References

Aarts, G., MacKenzie, M., McConnell, B., Fedak, M. & Matthiopoulos, J. (2008). Estimating space-use and habitat preference from wildlife telemetry data. *Ecography*, **31**, 140–160.

Ancrenaz, M., Sollmann, R., Meijaard, E., Hearn, A.J., Ross, J., Samejima, H., Loken, B., Cheyne, S.M., Stark, D.J., Gardner, P.C., Goossens, B., Mohamed, A., Bohm, T., Matsuda, I., Nakabayasi, M., Lee, S.K., Bernard, H., Brodie, J., Wich, S., Fredriksson, G., Hanya, G., Harrison, M.E., Kanamori, T., Kretzschmar, P., Macdonald, D.W., Riger, P., Spehar, S., Ambu, L.N. & Wilting, A. (2014). Coming down from the trees: Is terrestrial activity in Bornean orangutans natural or disturbance driven? *Scientific Reports*, **4**, 1–6.

Bello, C., Galetti, M., Pizo, M.A., Magnago, L.F.S., Rocha, M.F., Lima, R.A.F., Peres, C.A., Ovaskainen, O. & Jordano, P. (2015). Defaunation affects carbon storage in tropical forests. *Science Advances*, **1**, 1–11.

Bohmann, K., Evans, A., Gilbert, M.T.P., Carvalho, G.R., Creer, S., Knapp, M., Yu, D.W. & de Bruyn, M. (2014). Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology and Evolution*, **29**, 358–367.

Brodie, J.F., Giordano, A.J., Zipkin, E.F., Bernard, H., Mohd-Azlan, J. & Ambu, L. (2014). Correlation and Persistence of Hunting and Logging Impacts on Tropical Rainforest Mammals. *Conservation Biology*, **29**, 110–121.

Camargo-Sanabria, A.A., Mendoza, E., Guevara, R., Martinez-Ramos, M. & Dirzo, R. (2014). Experimental defaunation of terrestrial mammalian herbivores alters tropical rainforest understorey diversity. *Proceedings of the Royal Society B: Biological Sciences*, **282**, 20142580–20142580.

Cardillo, M., Mace, G.M., Jones, K.E., Bielby, J., Bininda-Emonds, O.R.P., Sechrest, W., Orme, C.D.L. & Purvis, A. (2005). Multiple causes of high extinction risk in large mammal species. *Science*, **309**, 1239–41.

Cheyne, S.M., Sastramidjaja, W.J., Muhalir, Rayadin, Y. & Macdonald, D.W. (2016). Mammalian communities as indicators of disturbance across Indonesian Borneo. *Global Ecology and Conservation*, **7**, 157–173.

Corlett, R.T. (2007). The impact of hunting on the mammalian fauna of tropical Asian forests. *Biotropica*, **39**, 292–303.

Fa, J.E. & Brown, D. (2009). Impacts of hunting on mammals in African tropical moist forests: A review and synthesis. *Mammal Review*, **39**, 231–264.

Gese, E.M. (2001). *Monitoring of terrestrial carnivore populations*.

Gillingham, P.K., Palmer, S.C.F., Huntley, B., Kunin, W.E., Chipperfield, J.D. & Thomas, C.D. (2012). The relative importance of climate and habitat in determining the distributions of species at different spatial scales: A case study with ground beetles in Great Britain. *Ecography*, **35**, 831–838.

Global Biodiversity Information Facility. (2014). *Publishing Camera Trap Data, a Best Practice Guide* (M. Cadman & A. González-Talaván, Eds.).

Gregory, T., Carrasco Rueda, F., Deichmann, J., Kolowski, J. & Alonso, A. (2014). Arboreal camera trapping: Taking a proven method to new heights. *Methods in Ecology and Evolution*, **5**, 443–451.

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

Harms, T.M. & Dinsmore, S.J. (2016). Spatial scale matters when modeling avian co-occurrence. *Ecosphere*, **7**, 1–13.

Harrison, R.D., Sreekar, R., Brodie, J.F., Brook, S., Luskin, M., O'Kelly, H., Rao, M., Scheffers, B. & Velho, N. (2016). Impacts of hunting on tropical forests in Southeast Asia. *Conservation Biology*, **30**, 972–981.

Hebblewhite, M. & Haydon, D.T. (2010). Distinguishing technology from biology: a critical review of the use of GPS telemetry data in ecology. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, **365**, 2303–2312.

Hiby, L., Lovell, P., Patil, N., Kumar, N.S., Gopalaswamy, A.M. & Karanth, K.U. (2009). A tiger cannot change its stripes: using a three-dimensional model to match images of living tigers and tiger skins. *Biology letters*, **5**, 383–386.

Leimgruber, P., Christen, C.A. & Laborderie, A. (2005). The impact of landsat satellite monitoring on conservation biology. *Environmental Monitoring and Assessment*, **106**, 81–101.

Lindström, Å., Green, M., Paulson, G., Smith, H.G. & Devictor, V. (2013). Rapid changes in bird community composition at multiple temporal and spatial scales in response to recent climate change. *Ecography*, **36**, 313–322.

Lyra-Jorge, M.C., Ciocheti, G., Pivello, V.R. & Meirelles, S.T. (2008). Comparing methods for sampling large- and medium-sized mammals: camera traps and track plots. *European Journal of Wildlife Research*, **54**, 739–744.

MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L. & Hines, J.E. (2006). *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurence*. Academic Press, London, United Kingdom.

Mathai, J., Jathanna, D. & Duckworth, J.W. (2013). How useful are transect surveys for studying carnivores in the tropical rainforests of Borneo? *Raffles Bulletin of Zoology*, 9–20.

McCallum, J. (2012). Changing use of camera traps in mammalian field research: Habitats, taxa and study types. *Mammal Review*, **43**, 196–206.

McShea, W.J., Forrester, T., Costello, R., He, Z. & Kays, R. (2016). Volunteer-run cameras as distributed sensors for macrosystem mammal research. *Landscape Ecology*, **31**, 55–66.

McShea, W. & Rappole, J. (1992). White-Tailed Deer as Keystone Species within Forest Habitats of Virginia. *Virginia Journal of Science*, **43**, 177–186.

Newbold, T., Hudson, L.N., Hill, S.L.L., Contu, S., Lysenko, I., Senior, R. a, Börger, L., Bennett, D., Choimes, A., Collen, B., Day, J., De Palma, A., Dıáz, S., Echeverria-Londoño, S., Edgar, M.J., Feldman, A., Garon, M., Harrison, M.L.K., Alhusseini, T., Ingram, D.J., Itescu, Y., Kattge, J., Kemp, V., Kirkpatrick, L., Kleyer, M., Laginha Pinto Correia, D., Martin, C.D., Meiri, S., Novosolov, M., Pan, Y., Phillips, H.R.P., Purves, D.W., Robinson, A., Simpson, J., Tuck, S.L., Weiher, E., White, H.J., Ewers, R.M., Mace, G.M., Scharlemann, J.P. & Purvis, A. (2015). Global effects of land use on local terrestrial biodiversity. *Nature*, **520**, 45–50.

Niedballa, J., Courtiol, A., Sollmann, R. & Wilting, A. (2015). camtrapR: Camera Trap Data Management and Preparation of Occupancy and Spatial Capture-Recapture Analyses.

Osuri, A.M., Ratnam, J., Varma, V., Alvarez-Loayza, P., Astaiza, J.H., Bradford, M., Fletcher, C., Ndoundou-Hockemba, M., Jansen, P.A., Kenfack, D., Marshall, A.R., Ramesh, B.R., Rovero, F. & Sankaran, M. (2016). Contrasting effects of defaunation on aboveground carbon storage across the global tropics. *Nature Communications*, **7**, 11351.

Pacifici, M., Foden, W.B., Visconti, P., Watson, J.E.M., Butchart, S.H.M., Kovacs, K.M., Scheffers, B.R., Hole, D.G., Martin, T.G., Akçakaya, H.R., Corlett, R.T., Huntley, B., Bickford, D., Carr, J. a., Hoffmann, A. a., Midgley, G.F., P., P.-K., Pearson, R.G., Williams, S.E., Willis, S.G., Young, B. & Rondinini, C. (2015). Assessing species vulnerability to climate change. *Nature Climate Change*, **5**, 215–225.

Peck, S.L. (2004). Simulation as experiment: A philosophical reassessment for biological modeling. *Trends in Ecology and Evolution*, **19**, 530–534.

Pettorelli, N., Laurance, W.F., O'Brien, T.G., Wegmann, M., Nagendra, H. & Turner, W. (2014). Satellite remote sensing for applied ecologists: Opportunities and challenges. *Journal of Applied Ecology*, **51**, 839–848.

Pigot, A.L. & Tobias, J.A. (2013). Species interactions constrain geographic range expansion over evolutionary time. *Ecology Letters*, **16**, 330–338.

Pimm, S.L., Alibhai, S., Bergl, R., Dehgan, A., Giri, C., Jewell, Z., Joppa, L., Kays, R. & Loarie, S. (2015). Emerging Technologies to Conserve Biodiversity. *Trends in ecology & evolution*, **30**, 685–696.

Prasad, S., Pittet, A. & Sukumar, R. (2010). Who really ate the fruit? A novel approach to camera trapping for quantifying frugivory by ruminants. *Ecological Research*, **25**, 225–231.

Proença, V., Martin, L.J., Pereira, H.M., Fernandez, M., McRae, L., Belnap, J., Böhm, M., Brummitt, N., García-Moreno, J., Gregory, R.D., Honrado, J.P., Jürgens, N., Opige, M., Schmeller, D.S., Tiago, P. & van Swaay, C.A.M. (2016). Global biodiversity monitoring: From data sources to Essential Biodiversity Variables. *Biological Conservation*.

Redford, K.H. (1992). The Empty Forest. *BioScience*, **42**, 412–422.

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.

Ripple, W.J., Estes, J. a, Beschta, R.L., Wilmers, C.C., Ritchie, E.G., Hebblewhite, M., Berger, J., Elmhagen, B., Letnic, M., Nelson, M.P., Schmitz, O.J., Smith, D.W., Wallach, A.D. & Wirsing, A.J. (2014). Status and ecological effects of the world's largest carnivores. *Science*, **343**, 1241484.

Rosa, I.M.D., Smith, M.J., Wearn, O.R., Purves, D. & Ewers, R.M. (2016). The Environmental Legacy of Modern Tropical Deforestation. *Current Biology*, **26**, 1–6.

Royle, J.A., Chandler, R.B., Sollmann, R. & Gardner, B. (2014). *Spatial Capture-Recapture*. Academic

Press.

Schipper, J., Chanson, J.S., Chiozza, F., Cox, N. a, Hoffmann, M., Katariya, V., Lamoreux, J., Rodrigues, A.S.L., Stuart, S.N., Temple, H.J., Baillie, J., Boitani, L., Lacher, T.E., Mittermeier, R. a, Smith, A.T., Absolon, D., Aguiar, J.M., Amori, G., Bakkour, N., Baldi, R., Berridge, R.J., Bielby, J., Black, P.A., Blanc, J.J., Brooks, T.M., Burton, J. a, Butynski, T.M., Catullo, G., Chapman, R., Cokeliss, Z., Collen, B., Conroy, J., Cooke, J.G., da Fonseca, G. a B., Derocher, A.E., Dublin, H.T., Duckworth, J.W., Emmons, L., Emslie, R.H., Festa-Bianchet, M., Foster, M., Foster, S., Garshelis, D.L., Gates, C., Gimenez-Dixon, M., Gonzalez, S., Gonzalez-Maya, J.F., Good, T.C., Hammerson, G., Hammond, P.S., Happold, D., Happold, M., Hare, J., Harris, R.B., Hawkins, C.E., Haywood, M., Heaney, L.R., Hedges, S., Helgen, K.M., Hilton-Taylor, C., Hussain, S.A., Ishii, N., Jefferson, T. a, Jenkins, R.K.B., Johnston, C.H., Keith, M., Kingdon, J., Knox, D.H., Kovacs, K.M., Langhammer, P., Leus, K., Lewison, R., Lichtenstein, G., Lowry, L.F., Macavoy, Z., Mace, G.M., Mallon, D.P., Masi, M., McKnight, M.W., Medellín, R. a, Medici, P., Mills, G., Moehlman, P.D., Molur, S., Mora, A., Nowell, K., Oates, J.F., Olech, W., Oliver, W.R.L., Oprea, M., Patterson, B.D., Perrin, W.F., Polidoro, B. a, Pollock, C., Powel, A., Protas, Y., Racey, P., Ragle, J., Ramani, P., Rathbun, G., Reeves, R.R., Reilly, S.B., Reynolds, J.E., Rondinini, C., Rosell-Ambal, R.G., Rulli, M., Rylands, A.B., Savini, S., Schank, C.J., Sechrest, W., Self-Sullivan, C., Shoemaker, A., Sillero-Zubiri, C., De Silva, N., Smith, D.E., Srinivasulu, C., Stephenson, P.J., van Strien, N., Talukdar, B.K., Taylor, B.L., Timmins, R., Tirira, D.G., Tognelli, M.F., Tsytsulina, K., Veiga, L.M., Vié, J.-C., Williamson, E. a, Wyatt, S. a, Xie, Y. & Young, B.E. (2008). The status of the world's land and marine mammals: diversity, threat, and knowledge. *Science*, **322**, 225–30.

Schnell, I.B., Sollmann, R., Calvignac-Spencer, S., Siddall, M.E., Yu, D.W., Wilting, A. & Gilbert, M.T.P. (2015). iDNA from terrestrial haematophagous leeches as a wildlife surveying and monitoring tool - prospects, pitfalls and avenues to be developed. *Frontiers in zoology*, **12**, 24.

Schnell, I.B., Thomsen, P.F., Wilkinson, N., Rasmussen, M., Jensen, L.R.D., Willerslev, E., Bertelsen, M.F. & Gilbert, M.T.P. (2012). Screening mammal biodiversity using DNA from leeches. *Current Biology*, **22**, R262-3.

Silveira, L., Jácomo, A.T. a. & Diniz-Filho, J.A.F. (2003). Camera trap, line transect census and track surveys: a comparative evaluation. *Biological Conservation*, **114**, 351–355.

Simberloff, D. (1998). Flagships, umbrellas, and keystones: Is single-species management passe in the landscape era? *Biological Conservation*, **83**, 247–257.

Sinclair, A.R.E. (2003). Mammal population regulation, keystone processes and ecosystem dynamics.

Philosophical transactions of the Royal Society of London. Series B, Biological sciences, **358**, 1729–1740.

Sollmann, R., Mohamed, A. bin, Niedballa, J., Bender, J., Ambu, L., Lagan, P., Mannan, S., Ong, R.C., Langner, A., Gardner, B. & Wilting, A. (2017). Quantifying mammal biodiversity co-benefits in certified tropical forests. *Diversity and Distributions*, 317–328.

Steenweg, R., Hebblewhite, M., Kays, R., Ahumada, J., Fisher, J.T., Burton, C., Townsend, S.E., Carbone, C., Rowcliffe, J.M., Whittington, J., Brodie, J., Royle, J.A., Switalski, A., Clevenger, A.P., Heim, N. & Rich, L.N. (2017). Scaling up camera traps: monitoring the planet's biodiversity with networks of remote sensors. *Frontiers in Ecology and the Environment*, **15**, 26–34.

Struebig, M.J., Fischer, M., Gaveau, D.L.A., Meijaard, E., Wich, S.A., Gonner, C., Sykes, R., Wilting, A. & Kramer-Schadt, S. (2015a). Anticipated climate and land-cover changes reveal refuge areas for Borneo's orang-utans. *Global Change Biology*, **21**, 2891–2904.

Struebig, M.J., Wilting, A., Gaveau, D.L.A., Meijaard, E., Smith, R.J., Fischer, M., Metcalfe, K., Kramer-Schadt, S., Abdullah, T., Abram, N., Alfred, R., Ancrenaz, M., Augeri, D.M., Belant, J.L., Bernard, H., Bezuijen, M., Boonman, A., Boonratana, R., Boorsma, T., Breitenmoser-Würsten, C., Brodie, J., Cheyne, S.M., Devens, C., Duckworth, J.W., Duplaix, N., Eaton, J., Francis, C., Fredriksson, G., Giordano, A.J., Gonner, C., Hall, J., Harrison, M.E., Hearn, A.J., Heckmann, I., Heydon, M., Hofer, H., Hon, J., Husson, S., Khan, F.A.A., Kingston, T., Kreb, D., Lammertink, M., Lane, D., Lasmana, F., Liat, L.B., Lim, N.T.L., Lindenborn, J., Loken, B., Macdonald, D.W., Marshall, A.J., Maryanto, I., Mathai, J., McShea, W.J., Mohamed, A., Nakabayashi, M., Nakashima, Y., Niedballa, J., Noerfahmy, S., Persey, S., Peter, A., Pieterse, S., Pilgrim, J.D., Pollard, E., Purnama, S., Rafiastanto, A., Reinfelder, V., Reusch, C., Robson, C., Ross, J., Rustam, R., Sadikin, L., Samejima, H., Santosa, E., Sapari, I., Sasaki, H., Scharf, A.K., Semiadi, G., Shepherd, C.R., Sykes, R., Van Berkel, T., Wells, K., Wielstra, B. & Wong, A. (2015b). Targeted conservation to safeguard a biodiversity hotspot from climate and land-cover change. *Current Biology*, **25**, 372–378.

Suraci, J.P., Clinchy, M., Dill, L.M., Roberts, D. & Zanette, L.Y. (2016). Fear of large carnivores causes a trophic cascade. *Nature Communications*, **7**, 10698.

Swanson, A., Kosmala, M., Lintott, C., Simpson, R., Smith, A. & Packer, C. (2015). Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. *Scientific Data*, **2**, 150026.

Terborgh, J., Estes, J., Paquet, P., Rails, K., Boyd-Heger, D., Miller, B. & Noss, R. (1999). The role of top carnivores in regulating terrestrial ecosystems. *Continental Conservation* (eds M.E. Soule & J. Terborgh), pp. 35–64. Island Press, Washington, DC.

Toomey, A.H. & Domroese, M.C. (2013). Can citizen science lead to positive conservation attitudes and behaviors? *Human Ecology Review*, **20**, 50–62.

Turner, W., Rondinini, C., Pettorelli, N., Mora, B., Leidner, A.K., Szantoi, Z., Buchanan, G., Dech, S., Dwyer, J., Herold, M., Koh, L.P., Leimgruber, P., Taubenboeck, H., Wegmann, M., Wikelski, M. & Woodcock, C. (2015). Free and open-access satellite data are key to biodiversity conservation. *Biological Conservation*, **182**, 173–176.

Tylianakis, J.M., Didham, R.K., Bascompte, J. & Wardle, D.A. (2008). Global change and species interactions in terrestrial ecosystems. *Ecology Letters*, **11**, 1351–1363.

Urban, M.C., Bocedi, G., Hendry, A.P., Mihoub, J.-B., Peer, G., Singer, A., Bridle, J.R., Crozier, L.G., De Meester, L., Godsoe, W., Gonzalez, A., Hellmann, J.J., Holt, R.D., Huth, A., Johst, K., Krug, C.B., Leadley, P.W., Palmer, S.C.F., Pantel, J.H., Schmitz, A., Zollner, P.A. & Travis, J.M.J. (2016). Improving the forecast for biodiversity under climate change. *Science*, **353**, aad8466-aad8466.

Urban, M.C., Zarnetske, P.L. & Skelly, D.K. (2013). Moving forward: Dispersal and species interactions determine biotic responses to climate change. *Annals of the New York Academy of Sciences*, **1297**, 44–60.

Weckel, M., Giuliano, W. & Silver, S. (2006). Jaguar (Panthera onca) feeding ecology: Distribution of predator and prey through time and space. *Journal of Zoology*, **270**, 25–30.

Williams, K., De Robertis, A., Berkowitz, Z., Rooper, C. & Towler, R. (2014). An underwater stereo-camera trap. *Methods in Oceanography*, **11**, 1–12.

Wolf, C. & Ripple, W.J. (2016). Prey depletion as a threat to the world's large carnivores. *Royal Society Open Science*, **3**, 160252.

Wolkovich, E.M., Cook, B.I., McLauchlan, K.K. & Davies, T.J. (2014). Temporal ecology in the Anthropocene. *Ecology Letters*, **17**, 1365–1379.

SUMMARY

Camera trapping has become a prime source of information about wild terrestrial mammals over the recent years, particularly for rare and elusive species and in challenging habitats. Key inference from camera-trapping encompasses species habitat associations, density and abundance estimations, or species interactions, amongst others. The rapid development of those methods and the large amount of data collected entail new challenges in terms of data management and analysis. The aim of this thesis was to contribute to the development of new methods for managing (Chapter 2) and analysing (Chapter 3 and 4) camera trapping data and to thus increase the efficiency and effectiveness of the use of camera-trapping data for practitioners both in academia and conservation.

Camera-trapping can generate vast volumes of data over short periods of time, making efficient yet flexible data management imperative. In my first manuscript (**Chapter 2**), I developed a free and open-source R package for camera trap data management, `camtrapR`. It is the first such toolbox in the popular R language and was designed to offer a comprehensive and flexible workflow from data acquisition to creating input for well-developed downstream analytical tools, e.g. in occupancy or spatial capture-recapture frameworks. The package has quickly gained popularity and is now being used worldwide in scientific and conservation work, while it is still being actively maintained and developed.

Species occurrence data from camera-traps can be combined with habitat information at camera traps via occupancy models in order to identify habitat associations of species while explicitly accounting for imperfect detection. The spatial scale at which habitat information are collected (grain and extent) will influence results heavily. In my second manuscript (**Chapter 3**), I assessed the influence of spatial scale on estimates of species-habitat relationships by varying the spatial resolution and extent of habitat covariates used in single-species occupancy models for six mammal species from Sabah, Malaysian Borneo. Habitat data from high-resolution (5-m RapidEye) satellite imagery had considerably higher model support than lower resolution data (≥30 m). Likewise, habitat data from patches of 50 meters around camera traps had higher model support than smaller (10 m) or larger (100 – 500 m) habitat patches. This study was the first to use 5-m RapidEye imagery in occupancy models and demonstrated the potential of such high-resolution satellite imagery for obtaining more realistic species-habitat associations in occupancy modelling, particularly in

heterogeneous landscapes. The flexibility high-resolution satellite imagery offer in defining suitable spatial scales further add to their value.

Species distributions in space and time are not only shaped by habitat preferences, but also by interactions between species, such as predator-prey relationships or various forms of competition. Discovering such spatiotemporal interactions in camera trapping data is challenging due to the sparseness and randomness of camera trapping data and further exacerbated by a lack of systematic comparisons of methods to assess such interactions. Therefore, in my third manuscript (**Chapter 4**)**,** I developed a method to flexibly simulate camera trapping records of two interacting species. These simulated data are used for the first comparative assessment of the statistical power and robustness of a suite of statistical tests for spatiotemporal interactions. Linear models were the most powerful and flexible method to discover such interactions. Nevertheless, only strong interactions could be detected reliably with any of the methods tested. This novel simulation approach and the recommendations given can serve as guidelines for practitioners wishing to assess interactions between or within species from camera trapping data.

ZUSAMMENFASSUNG

Kamerafallen haben sich in den letzten Jahren zu einer der wichtigsten Datenquellen über wildlebende terrestrische Säugetiere entwickelt, insbesondere für seltene und schwer beobachtbare Arten sowie in herausfordernden Habitaten. Wichtige Rückschlüsse, welche aus Kamerafallendaten gewonnen werden können, sind unter anderem Habitatassoziationen von Arten, Schätzungen von Dichte und Abundanz, oder Interaktionen zwischen Arten. Die rasante Entwicklung dieser Methoden und die enormen Datenmengen, die dabei entstehen, hatten neue Herausforderungen hinsichtlich Datenverwaltung und –analyse zur Folge. Das Ziel dieser Arbeit war, zur Entwicklung von neuen Methoden zum Verwalten (Kapitel 2) und Analysieren (Kapitel 3 und 4) von Kamerafallendaten beizutragen und damit sowohl Effizienz als auch die Effektivität der Nutzung von Kamerafallendaten in Wissenschaft und Naturschutzarbeit zu verbessern.

Da Kamerafallenstudien in kurzer Zeit riesige Datenmengen produzieren können, ist effizientes und flexibles Kamerafallendatenmanagement zwingend erforderlich. In meinem ersten Manuskript (**Kapitel 2**) habe ich ein frei verfügbares und quelloffenes R-Paket für die Verwaltung von Kamerafallendaten entwickelt, `camtrapR`. Es ist das erste derartige Paket in der weitverbreiteten Programmiersprache R, und es wurde konzipiert, um einen umfassenden und flexiblen Arbeitsfluss von der Datenerhebung bis zum Bereitstellen von Daten für weitergehende Analysen zu gewährleisten, z.B. mit Occupancy- oder Spatial Capture-Recapture-Methoden. Das Paket wird weiterhin gepflegt und weiterentwickelt, hat schnell an Popularität gewonnen und wird weltweit in Wissenschaft und Naturschutzarbeit genutzt.

Daten über das Vorkommen von Arten aus Kamerafallen kann mit Habitatinformationen an Kamerafallenstandorten mit Hilfe von Occupancy-Modellen kombiniert werden, um Habitatassoziationen von Arten zu identifizieren und gleichzeitig für die unvollständige Detektierbarkeit von Arten zu korrigieren. Das räumliche Ausmaß (scale), in dem Habitatinformationen gesammelt werden, beeinflusst die Modellergebnisse erheblich. In meinem zweiten Manuskript (**Kapitel 3**) habe ich den Einfluss des räumlichen Ausmaßes von Habitatdaten auf die Abschätzung von Habitatassoziationen von Arten anhand von sechs Säugetierarten aus einem Kamerafallendatensatz aus Sabah, Borneo, Malaysia untersucht. Das geschah, indem ich die räumliche Auflösung und die Ausdehnung von Habitatinformationen in Occupancy-Modellen für die individuellen Arten variiert habe. Habitatinformationen aus hochauflösenden Satellitenbildern (5-m

RapidEye) hatten deutlich höheren Modellsupport als niedrig aufgelöste Daten (≥30 m). Habitatdaten mit einem Radius von 50 m um Kamerafallen hatten gleichermaßen höheren Modellsupport als Daten aus kleineren (10 m) oder größeren (100 – 500 m) Radien. Dies war die erste Studie, die 5-m RapidEye Satelllitendaten in Occupancy-Modellen verwendet und demonstriert den Eignung dieser hochauflösenden Satellitendaten, insbesondere in heterogenen Landschaften mit Hilfe von Occupancy-Modellen zu realistischeren Habitatassoziationen zu gelangen. Die Flexibilität, mit der geeignete räumliche Ausdehnungen von Habitatdaten festgelegt werden können, ist ein weiterer Vorteil dieser Daten.

Die Verbreitung von Arten in Raum und Zeit hängt nicht nur von Habitatpräferenzen ab, sondern auch von Interaktionen zwischen Arten, etwa in Räuber-Beute Beziehungen oder Konkurrenz zwischen oder innerhalb von Arten. Solche Beziehungen in Kamerafallendaten zu identifizieren ist herausfordernd aufgrund der Spärlichkeit und Zufälligkeit in Kamerafallendaten, und weiter erschwert durch das Fehlen eines systematischen Vergleiches von Methoden, um solche Interaktionen zu untersuchen. Deswegen habe ich in meinem dritten Manuskript (**Kapitel 4**) eine Methode entwickelt, mit der sich Kamerafallendaten zweier interagierender Arten flexibel simulieren lassen. Diese simulierten Kamerafallendaten wurden verwendet für die erste vergleichende Bewertung der statistischen Teststärke (power) und Robustheit einer Reihe von statistischen Tests zur Untersuchung räumlich-zeitlicher Interaktionen. Lineare Modelle hatten die höchste Teststärke und sind die flexibelste Methode, um solche Interaktionen festzustellen. Dennoch konnten mit allen untersuchten Methoden nur starke Interaktionen zwischen Arten zuverlässig nachgewiesen werden. Dieser neuartige Simulationsansatz und die daraus folgenden Empfehlungen können als Richtlinien für die Untersuchung von Interaktionen zwischen Arten oder innerhalb von Arten in Kamerafallendaten dienen.

CURRICULUM VITAE

For reasons of data protection, the curriculum vitae is not published in the electronic version.

SELBSTÄNDIGKEITSERKLÄRUNG

Hiermit versichere ich, dass ich die vorliegende Doktorarbeit eigenständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Berlin, 14.9.2017

Jürgen Niedballa