Journal of Applied Ecology

Searle Charlotte E. (Orcid ID: 0000-0002-9861-1429)
Macdonald David W (Orcid ID: 0000-0003-0607-9373)
Loveridge Andrew J (Orcid ID: 0000-0002-3493-6419)

Corresponding Author Mail id : charlotte.searle@zoo.ox.ac.ukss

Random forest modelling of multi-scale, multi-species habitat associations within KAZA transfrontier conservation area using spoor data

Author names and affiliations

Charlotte E. Searle[1,*], **Żaneta Kaszta**[1], Dominik T. Bauer[1], M. Kristina Kesch[1,2], Jane E. Hunt[1], Roseline Mandisodza-Chikerema[3], Michael V. Flyman[4], David W. Macdonald[1], Amy J. Dickman[1], Andrew J. Loveridge[1], Samuel A. Cushman[5]

[1] Wildlife Conservation Research Unit, Department of Zoology, Recanati-Kaplan Centre, University of Oxford, Oxford, UK

[2] Frankfurt Zoological Society, Bernhard-Grzimek-Allee 1, 60316 Frankfurt, Germany

[3] Zimbabwe Parks and Wildlife Management Authority, PO Box CY140, Causeway, Harare, Zimbabwe

[4] Ministry of Environment, Natural Resources Conservation and Tourism, Private Bag BO199, Gaborone, Botswana

[5] US Forest Service, Rocky Mountain Research Station, Flagstaff, AZ, USA

*** Corresponding author**

Joint first authorship (Searle, Kaszta)

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/1365-2664.14234

Abstract

1. As landscape-scale conservation models grow in prominence, assessments of how wildlife utilise multiple-use landscapes are required to inform effective conservation and management planning. Such efforts should incorporate multi-species perspectives to maximise value for conservation, and should account for scale to accurately capture species-environment relationships.
2. We show that the random forest machine learning algorithm can be used to model large-scale sign-based data in a multi-scale framework. We used this method to investigate scale-dependent habitat associations for 16 mammal species of high conservation importance across the southern Kavango Zambezi (KAZA) Transfrontier Conservation Area in Botswana and Zimbabwe.
3. Our findings revealed substantial variation in factors shaping habitat use across species, and illustrate that different species often have divergent responses to the same environmental and anthropogenic factors, and differ in the scales at which they respond to them. For all variables across all species, scale optimisation most often selected our largest scale.
4. Precipitation, soil nutrients, and vegetation appeared to be the most important factors determining mammal distributions, likely through their associations with food resources for herbivores and, in turn, prey availability for carnivores.
5. Anthropogenic pressures also had an important influence, with many species selecting against areas with high cattle density. The variety of relationships with human density indicated that species vary in their tolerance of humans. We found a consistent positive relationship with areas under high protection, and negative relationship with unprotected and less-strictly protected areas.
6. *Policy implications*: Through a novel application of random forest modelling to spoor data from 16 mammal species, this study highlights the importance of adopting a multi-scale, multi-species approach for decision-making processes that depend on understanding wildlife distributions and habitat associations, such as protected area and corridor prioritisation. The findings identify changing rainfall patterns and increasing livestock numbers as emerging trends that may impact wildlife distributions, both within sub-Saharan Africa and on a global scale. Wildlife management authorities should use modelling exercises and adaptive management to ensure protected area networks remain fit for purpose under anticipated changes in rainfall under climate change, and explore initiatives that promote coexistence of wildlife and livestock.

Keywords

Habitat associations, multi-scale modelling, random forest, transfrontier conservation area, wildlife distributions, African mammals, multi-species, machine learning, southern Africa, spoor

1. Introduction

Protected areas (PAs) play a key role in safeguarding wildlife from the threats of land conversion and over-exploitation (Brondizio *et al.*, 2019). However, failing to also maintain connectivity among and within these protected patches of core habitat can lead to the severance of important migratory and dispersal routes, increased genetic isolation, compromised resilience to diseases and population bottlenecks, and reduced opportunities to shift ranges in response to climate change (Zeller, McGarigal and Whiteley, 2012). A lack of intact habitat outside PAs can also exacerbate edge effects around PA boundaries, which can have harmful impacts that reverberate throughout populations (Loveridge *et al.*, 2010).

As a result, there is a growing consensus that landscape-scale conservation models are required to secure viable wildlife populations into the future (Ripple *et al.*, 2014). While critical areas of habitat within individual countries can be secured through coordinated national land management planning, many areas of importance for wildlife do not adhere to political boundaries, and instead straddle multiple countries (Farhadinia, Rostro-Garcia and Feng, 2020). In this context, transfrontier conservation areas (TFCAs) provide an institutional framework that can help secure both core habitat and critical linkages.

It is critical that we improve our understanding of how wildlife populations use these wider mosaic landscapes so this information can be used to inform conservation and management plans (Cushman *et al.*, 2006; Zeller, McGarigal and Whiteley, 2012). In particular, identifying which factors underpin space use by wildlife populations is likely to be a central component of PA, corridor, and buffer zone prioritisation – exercises that will become increasingly necessary in the face of rapid ongoing changes in climate and rising anthropogenic disturbance, and the considerable costs associated with maintaining already chronically-underfunded PAs (Lindsey *et al.*, 2018).

To effectively make decisions in this context, policymakers should strive to incorporate multi-species perspectives, as maintaining healthy wildlife communities is critical to safeguard functional ecosystems (CBD, 2010). Sign-based surveys allow information to be collected simultaneously on multiple species, and can be conducted across large scales rapidly and at relatively low cost. Sign-based data can therefore be a powerful tool to develop species distribution models and assess habitat associations at the community level (Zeller *et al.*, 2018).

Landscape-level assessments of habitat use should also seek to account for scale, as this is widely recognised as a central component of species-environment relationships (McGarigal *et al.*, 2016). Failing to account for the scale at which species respond to environmental variables can undermine efforts to disentangle species-habitat relationships, and may result in incorrect inferences being used to guide conservation and management interventions (Everatt *et al.*, 2015). Despite this, many studies investigating habitat use fail to incorporate multiple scales – and of those that do, fewer still make use of scale optimisation procedures to empirically determine the most relevant scale for each covariate of interest (McGarigal *et al.*, 2016).

Identifying factors associated with habitat use for multiple species across multiple scales is a complex task (Holland, Bert and Fahrig, 2004). In recent years, however, machine learning approaches have emerged as a powerful tool to model spatially complex and temporally dynamic systems, as they are able to disentangle complex and non-intuitive interactions from large datasets much more readily than traditional modelling approaches (Evans *et al.*, 2011). One model that has become an important part of the ecological toolset is the random forest algorithm (Breiman, 2001), a powerful machine-learning algorithm which has recently been shown to outperform many other predictive modelling methods (Cushman and Wasserman, 2018).

In this study, we use the random forest machine learning algorithm to investigate scale-dependent habitat associations for 16 mammal species – six carnivores and ten herbivores – across the southern Kavango Zambezi TFCA in Botswana and Zimbabwe, using data from large-scale presence-only surveys. This is the first time spoor data has been modelled using random forest. Of the 16 study species, six are classified as *Endangered*, *Vulnerable* or *Near Threatened* on the IUCN Red List, while nine are thought to have decreasing population trends (IUCN, 2020). We combine these species-specific insights to build a multi-species view of the key drivers of habitat use across the study landscape, and identify potential future trends that may impact wildlife distributions.

2. Materials and Methods

This research was carried out with permission from the Governments of Botswana and Zimbabwe, under research permits granted by the Ministry of Environment, Natural Resources Conservation and Tourism in Botswana (Permit number: EWT 8/36/4 XXIII (15)), and by the Parks and Wildlife

Management Authority in Zimbabwe (Permit numbers: REF:DM/Gen/(T) 23(1)(c)(ii): 12/2012, 08/2013, 51/2014, 10/2015). Ethical approval was not required for the research.

2.1. Study area

The Kavango Zambezi Transfrontier Conservation Area (KAZA TFCA) is the world's largest terrestrial TFCA, covering an area of approximately 520,000 km^2 in southern Africa (Southern African Development Community, 2018). The KAZA TFCA spans the boundaries of five countries, encompassing 36 National Parks (NPs) and a host of other land uses, including community-managed Wildlife Management Areas (WMAs), trophy hunting areas, unprotected land and communal areas. See Appendix S1 in Supporting Information for a description of land use designations in the study area.

Our study area consists of approximately 30,000 km^2 in the southern KAZA TFCA, across Botswana and Zimbabwe (Fig. 1). We collected data in Hwange NP & Zambezi NP in Zimbabwe, and Nxai Pan NP in Botswana, as well as areas adjacent to Kazuma Pan NP in Zimbabwe, and Chobe NP & Moremi Game Reserve in Botswana. We also collected data in a number of less-strictly protected WMAs, Safari Areas, Conservancy Ranches, and Forest Areas & Reserves, in addition to unprotected land.

The study area receives an average of 500-700 mm of rainfall per year, and is generally water- and nutrient-poor, although there are a large number of artificial waterholes across parts of the landscape (British Geological Survey, 2018b, 2018a). Vegetation is dominated by bushland savannah and woodland, interspersed with patches of grassland. Trophy hunting took place in Zimbabwe at the time of data collection, across approximately 14% of the total study area.

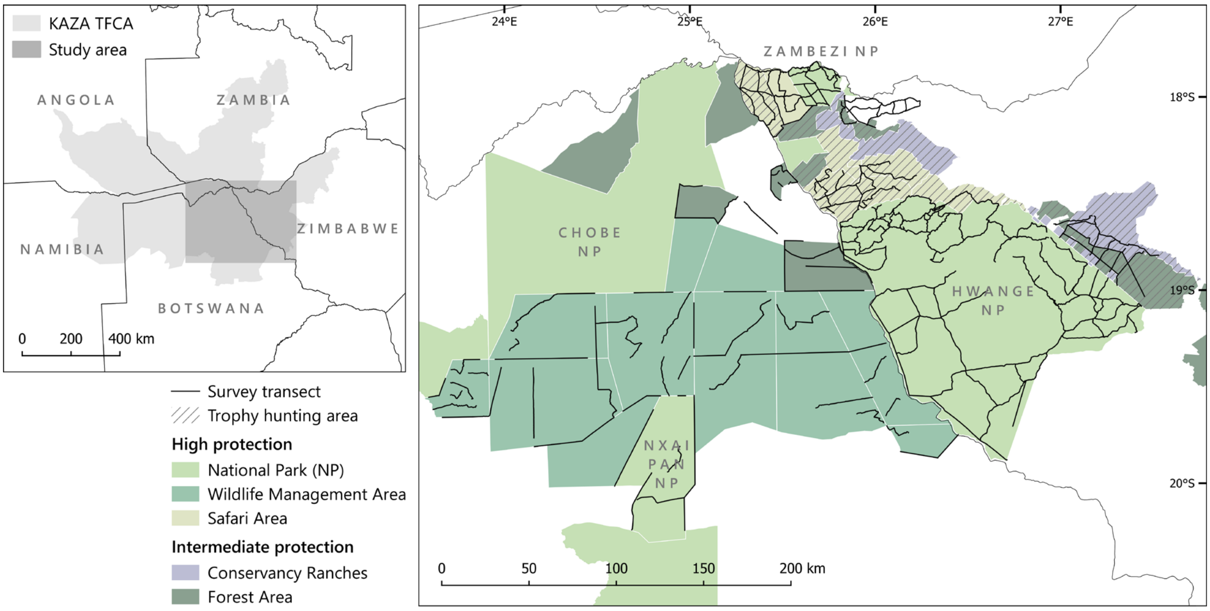

Figure 1: Left: the study area within the KAZA TFCA in southern Africa. Right: land use types in the study landscape and survey transects. PAs are shown only in the study countries (Botswana & Zimbabwe).

2.2. Data collection

We collected presence data for mammal species across the study area between May 2012 and October 2015. Surveys were carried out by driving transects at a low speed along roads, fire-breaks and boundary cutlines (Fig. 1), and recording any fresh spoor (tracks) of all medium and large mammals (see Results for a complete list of species ultimately included in the study). As highly-skilled local trackers assisted with spoor detection and identification, and identifications were verified through consensus among team members, incorrect identification of tracks was highly unlikely. Surveys were started at first light to avoid disturbance and maximise visibility.

2.3. Data preparation

We collected species presence data along 4,824 km of transects, most of which were surveyed on two occasions. We divided these transects into 96,474 segments of 50 m for analysis, which we consider an appropriate size to allow us to localise presences at a relatively fine scale. Although adjacent 50m segments are not independent, random forest models are highly robust to spatial autocorrelation (Evans *et al.*, 2011), and autocorrelation was reduced as a result of our subsamples being scattered across the landscape.

In each segment we noted a presence (1) or absence (0) of each species. Due to large number of absences, we calculated density of the presence locations for each species using a kernel density estimator with 500 m search radius in ArcGIS version 10.6 (ESRI, 2016). While a relatively high number of absences remained in the data after this conversion, simulations have shown that random forest models produce accurate predictions when the ratio of presences and absences to reflect their actual prevalence in the landscape, rather than being in relatively equal proportions (Cushman *et al.*, 2017).

As some transects were surveyed multiple times, segments were also assigned a measure of survey effort, reflecting the number of times that segment was surveyed.

2.4. Multi-scale variables

We selected 15 variables that we hypothesised may influence the distribution of mammals across the study area (Table 1). In addition to these, we included longitude and latitude to account for spatial structure (Evans *et al.*, 2011; as per Liu *et al.*, 2013), and survey effort. We employed a multi-scale approach for all variables except latitude, longitude, and effort, to identify the most appropriate scale at which to include each variable in our final models for each species. All raster layers were converted to the same projection, aligned, and resampled to the same extent at 50 m resolution. To test for selection of variables at different scales, we calculated a focal mean at varying circular moving windows of 250, 500, 1000, 2000, 4000, and 8000 m radius in ArcGIS version 10.6 (ESRI, 2016). For variables based on point data, we produced multi-scale raster layers using the Point Density tool with a circular radius equal to each scale. The final multi-scale variable values were then extracted by sampling each variable at each scale from the midpoint of all 50 m transect segments.

2.5. Random forest modelling

We fitted random forest models (Breiman, 2001) to the species detection density data using R package *randomForest* (Liaw and Wiener, 2002). For each species, scale optimisation was carried out by running random forest model selection on all variables using the rf.modelSel function in R package *rfUtilities* (Evans and Murphy, 2019). For each variable, we selected the scale with the highest MIR (Model Improvement Ratio; Evans and Murphy, 2019). The model selection process was then repeated with only the scale-optimised variables for that species, to produce a final random forest model consisting of the best and most parsimonious combination of variables at their best scales. Model performance was assessed via the percentage of variance explained in the spoor density data (Liaw and Wiener, 2002).

We investigated the relationship between the final model variables and detection density for each species by producing smoothed splines using the plsmo function in package *Hmisc* (Harrell Jr, 2020). The importance of different variables in the final model for individual species was assessed by ranking MIR values. The importance of variables across all species was assessed by calculating the proportion of species with each variable in its final model, and the average rank (based on MIR) of each variable across all species. The frequency of scales was assessed by calculating the proportion of variables across species at each spatial scale.

Table 1: Variables hypothesised to affect distribution of the study species across the southern KAZA TFCA.

	Variable	Code	Original resolution	Description	Units	Reference
1	Longitude	X	-	X coordinate (UTM35S) of pixel	-	-
2	Latitude	Y	-	Y coordinate (UTM35S) of pixel	-	-
3	Effort	Effort	-	Number of times segment was surveyed	-	This study
4	Carbon	C	250 m	Soil organic carbon content in g per kg at 5 cm depth (ORCDRC_M_sl2)	dg / kg	Hengl, Mendes de Jesus, et al. (2017)
5	Nitrogen	N	250 m	Soil nutrient maps of sub-Saharan Africa	cg / kg	Hengl, Leenaars, et al. (2017)
6	Vegetation cover (EVI)	EVI	250 m	MODIS/Terra Vegetation Indices (MOD13Q1), averaged from the midpoint of each month within the survey period	EVI	Didan (2015)
7	Leaf area index (LAI)	LAI	500 m	MODIS/Terra+Aqua Leaf Area Index (MCD15A2H), averaged from the midpoint of each month within the survey period	m^2 / m^2	Myneni et al. (2015)
8	Canopy cover (VCF)	VCF	250 m	MODIS/Terra Vegetation Continuous Fields (MOD44B), averaged from each year within the survey period	%	Dimiceli et al. (2015)
9	Precipitation	P	30 s (~1 km)	WorldClim (Version 2), monthly precipitation averaged from each month within the survey period	mm	Fick & Hijmans (2017)
10	Permanent water points	WP	-	Vector layer of permanent water points (pans, manmade pumps)	-	Trans-Kalahari Predator Programme (TKPP), unpublished data
11	Permanent rivers	R	-	Vector layer of major rivers	-	TKPP, unpublished data
12	Cattle density	CD	5 m (~8 km)	Gridded livestock of the world (GLW3) Global cattle distribution in 2010 (5 minutes of arc)	animals per pixel	Gilbert et al. (2018)
13	Human density	HD	3s (~100 m)	WorldPop 2013 population UN adjusted	population per pixel	Linard et al. (2012)
14	Settlements	S	-	Vector layer of settlements	-	TKPP, unpublished data
15	No protection	NP	-	Rasterised shapefile with value of (1) within all unprotected land, and (0) within PAs	-	TKPP, unpublished data
16	Intermediate protection	IP	-	Rasterised shapefile with value of (1) within Conservancy Ranches, Forest Areas, and Forest Reserves, and (0) elsewhere	-	TKPP, unpublished data
17	High protection	HP	-	Rasterised shapefile with value of (1) within National Parks, Wildlife Management Areas, and Safari Areas, and (0) elsewhere	-	TKPP, unpublished data
18	Trophy hunting	H	-	Rasterised shapefile with value of (1) within areas where trophy hunting was taking place during the study period, and (0) elsewhere	-	TKPP, unpublished data

3. Results

Restricting the data to medium- and large-bodied mammal species with sufficient detections (> 400) resulted in 16 study species: six carnivores (African wild cat; *Felis silvestris*; black-backed jackal, *Canis mesomelas*; caracal, *Caracal caracal* & serval, *Leptailurus serval* [grouped together due to their highly similar spoor]; leopard, *Panthera pardus*; lion, *Panthera leo*; and spotted hyaena, *Crocuta crocuta*), and ten herbivores (African buffalo, *Syncerus caffer*; African elephant, *Loxodonta africana*; common duiker, *Sylvicapra grimmia*; common warthog, *Phacochoerus africanus*; giraffe, *Giraffa camelopardalis*; greater kudu, *Tragelaphus strepsiceros*; impala, *Aepyceros melampus*; plains zebra, *Equus quagga*; sable antelope, *Hippotragus niger*; and steenbok, *Raphicerus campestris*). Maps of presence points for all species can be found in Appendix S2.

3.1. Scale optimisation

The scale optimisation procedure revealed substantial variation in the scales at which different species responded to variables (Tables 2B & 3). For all variables across all species, scale optimisation most often selected our largest scale (8000 m; 84% of scale optimised variables across all species, and 81% of variables in the final models; Tables 2B & 3), and the second largest scale (4000 m) was the second most frequently selected (9% of scale optimised variables across all species, and 11% of variables in the final models; Tables 2B & 3).

In the final species models, eight of the 15 multiscale variables appeared at more than one scale (Tables 2B & 3). The remaining seven multiscale variables were exclusively selected at the largest scale.

3.2. Model performance

Performance of selected models was high across all species, with more than 98% of variance explained for all species (Table 2).

3.3. Variable importance

The total number of variables in the final model for each species ranged from 8 to 18. Variables differed in their importance across species (Tables 2A & 2B), and those that appeared in a high proportion of final models generally also ranked highly in importance (based on MIR) across all species (Table 2A). The most and least important variables were largely consistent across both carnivores and herbivores

(Table 3). Smoothed plsmo splines of all variables in the final model for each species can be found in Appendix S3; select plots showing some of the most important species-variable relationships are shown in Fig. 2.

Table 2: Percentage of variance explained alongside the (A) importance (based on MIR, Model Improvement Ratio) and (B) optimised scale of all variables selected for the final random forest model for each species. Cells are coloured according to the MIR of that variable for that species at the optimised scale; variables with higher importance are shaded green, while those with lower importance are shaded yellow. Species have been grouped into carnivores and herbivores. See table 1 for variable codes.

(A)

Species	% variance explained	X	Y	Effort	C	N	EVI	LAI	VCF	P	WP	R	CD	HD	S	NP	IP	HP	H
African wild cat	98.25	0.847	0.742			0.677	0.570	0.583	0.603	1.000				0.639					
Black-backed jackal	98.51	0.562	0.673		0.507	0.497		0.377	0.377	1.000				0.404					
Caracal & Serval	98.09	0.959	0.949	0.233	0.714	0.998	0.871	0.775	0.742	1.000	0.148	0.007	0.517	0.977	0.072	0.203	0.151	0.277	0.366
Leopard	98.17	0.994	0.827		1.000	0.886		0.791	0.909	0.980				0.734					
Lion	98.02	0.503	0.620	0.183	0.451	0.381	0.376	0.360	0.725	1.000	0.259		0.356	0.374				0.217	
Spotted hyaena	98.46	0.858	0.668		0.762	0.822				0.986			1.000	0.737				0.641	
African buffalo	98.42	1.000	0.595		0.600		0.532	0.513		0.625			0.547	0.546					
						0.471	0.478	0.407	0.357	0.784	0.540	0.011	0.776	0.437	0.088	0.208	0.152	0.472	0.170
						0.583		0.587		1.000			0.563			0.604			
						0.917	0.906	0.793		1.000			0.825						
						0.600	0.571	0.555	0.593	1.000	0.304	0.033	0.982	0.654	0.067	0.222	0.196	0.703	0.233
						0.837				1.000			0.730	0.705				0.606	
						0.493	0.345	0.371	0.654	1.000	0.153	0.020	0.306	0.345	0.032	0.111	0.092	0.270	0.169
						0.486	0.393	0.419	0.532	1.000	0.431	0.020	0.500	0.364	0.047	0.144	0.095	0.420	0.148
						0.459	0.460		0.489	1.000				0.506					
						0.093	0.032		0.016	0.092	1.000		0.041			0.027	0.015	0.039	0.033

(B)

Species	% variance explained	X	Y	Effort	C	N	EVI	LAI	VCF	P	WP	R	CD	HD	S	NP	IP	HP	H
African wild cat	98.25	1	1			8000	8000	8000	8000	4000				8000					
Black-backed jackal	98.51	1	1		8000	8000		8000	8000	1000				8000					
Caracal & Serval	98.09	1	1	1	8000	8000	8000	8000	8000	8000	8000	8000	8000	8000	8000	8000	8000	8000	8000
Leopard	98.17	1	1		8000	8000		8000	8000	8000				8000					
Lion	98.02	1	1	1	8000	8000	8000	8000	4000	4000	4000		1000	8000				4000	
Spotted hyaena	98.46	1	1		8000	8000				8000			250	2000				4000	
African buffalo	98.42	1	1		8000		8000	8000		8000			2000	8000					
African elephant	99.40	1	1	1	8000	8000	8000	8000	4000	4000	8000	8000	250	8000	8000	8000	8000	8000	8000
Common duiker	98.89	1	1	1		8000		4000		8000			8000			8000			
Common warthog	98.34	1	1		8000	8000	8000	8000		8000			8000						
Giraffe	98.75	1	1	1	8000	4000	8000	8000	8000	8000	8000	8000	250	4000	8000	8000	8000	250	8000
Greater kudu	98.81	1	1		8000	8000				8000			8000	8000				4000	
Impala	99.03	1	1	1	8000	8000	8000	8000	4000	2000	8000	8000	8000	8000	8000	8000	8000	4000	8000
Plains zebra	98.87	1	1	1	8000	8000	8000	8000	250	4000	4000	8000	8000	8000	8000	8000	8000	500	8000
Sable antelope	98.59	1	1		8000	8000	8000		8000	250				8000					
Steenbok	98.78	1	1		8000	8000	8000		8000	8000	8000		8000			8000	8000	4000	8000

Figure 2: Smoothed plsmo splines showing the relationship between final model variables and detection density; the selected plots show the relationship with monthly precipitation (mm) for zebra (top left) and lion (bottom left), and cattle density (animals per pixel) for elephant (top right) and giraffe (bottom right).

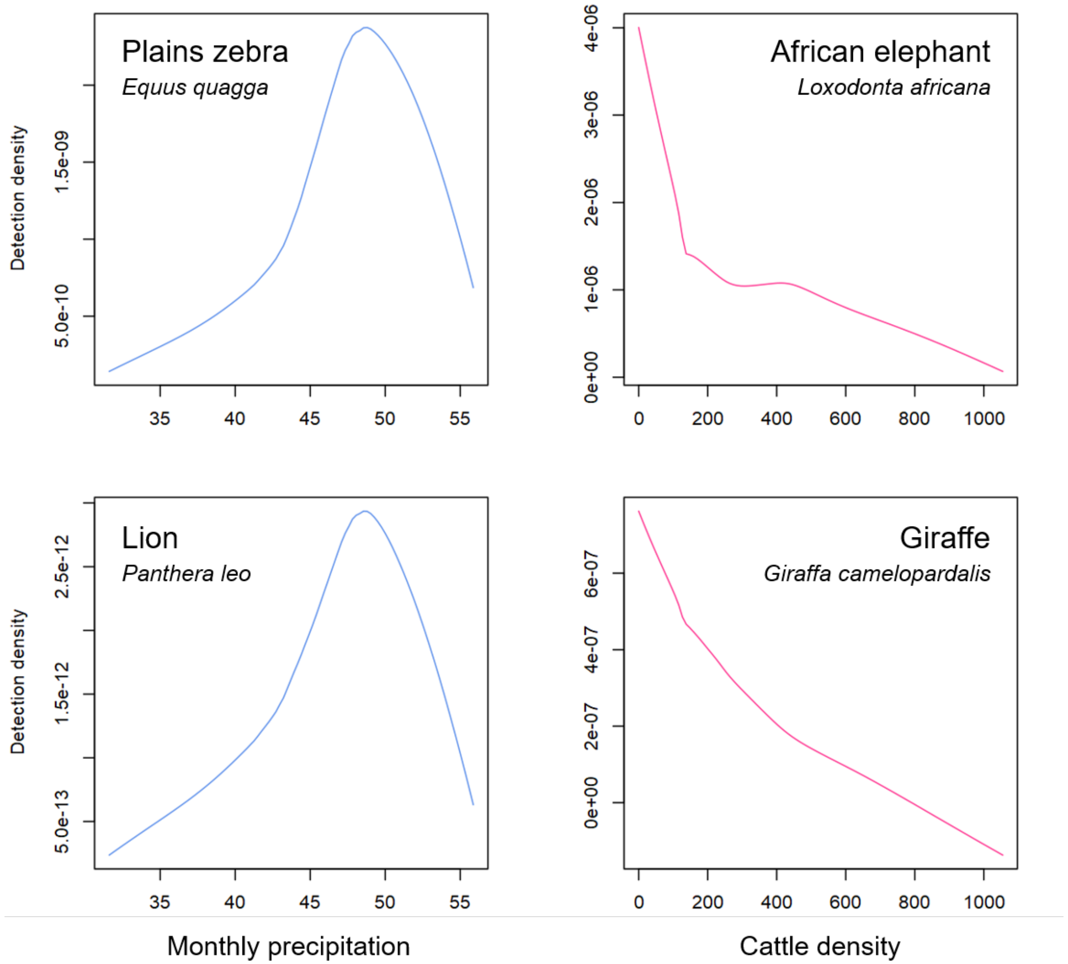

Table 3: Variable importance across final random forest models for all species. The number of scales at which each variable was selected during scale optimisation and in the final models (out of a possible total of 6) across all species; and the average and most frequent scale at which the variable was selected. Importance is indicated by the proportion of the 16 study species for which each variable appeared in the final model, separated into carnivores and herbivores; and the average rank (based on MIR) of each variable for each species, from 1 (most important) to 18 (least important).

		Optimised scale			Scale in final models			% of species with variable in final model			Average rank
Variable	Code	Number	Mean	Mode	Number	Mean	Mode	All	Carnivores	Herbivores	
Precipitation	P	5	5703	8000	5	5703	8000	1.00	1.00	1.00	1.4
Longitude (UTM35S)	X	-	-	-	-	-	-	1.00	1.00	1.00	3.2
Latitude (UTM35S)	Y	-	-	-	-	-	-	1.00	1.00	1.00	4.5
Nitrogen	N	2	7750	8000	2	7733	8000	0.94	1.00	0.90	5.5
Carbon	C	1	8000	8000	1	8000	8000	0.88	0.83	0.90	5.1
Human density	HD	3	7375	8000	3	7231	8000	0.81	1.00	0.70	7.8
Cattle density	CD	4	5734	8000	4	4979	8000	0.75	0.50	0.90	7.0
Leaf area index (LAI)	LAI	3	7281	8000	2	7667	8000	0.75	0.83	0.70	9.0
Canopy cover (VCF)	VCF	3	5547	8000	3	6205	8000	0.69	0.83	0.60	8.1
Vegetation cover (EVI)	EVI	1	8000	8000	1	8000	8000	0.69	0.50	0.80	8.4
High protection	HP	5	5109	8000	4	4083	4000	0.56	0.50	0.60	10.3
Water points	WP	2	7250	8000	2	6857	8000	0.44	0.33	0.50	11.8
Effort	E	-	-	-	-	-	-	0.44	0.33	0.50	12.1
No protection	NP	1	8000	8000	1	8000	8000	0.44	0.17	0.60	13.1
Trophy hunting	H	1	8000	8000	1	8000	8000	0.38	0.17	0.50	13.3
Intermediate protection	IP	1	8000	8000	1	8000	8000	0.38	0.17	0.50	15.6
Settlement points	S	1	8000	8000	1	8000	8000	0.31	0.17	0.40	16.9
Rivers	R	3	7375	8000	1	8000	8000	0.31	0.17	0.40	17.9

Longitude and latitude had high importance in our species distribution models, appearing in the final models for all species (Tables 2A & 3).

Our results suggest that, alongside geographical location, variables relating to climate, soil quality, and vegetation were the most important factors determining where mammals were found in the study landscape across southern KAZA: precipitation, soil nutrients (carbon and nitrogen), and vegetation indices (VCF, LAI, EVI) all appeared in the final model for more than half of the study species (Table 3).

Average monthly precipitation appeared in the final model for all species (average rank: 1.4; Table 3), and was the most important variable (MIR > 0.78) for nearly all species considered (Table 2A). For most species this relationship was curvilinear, with detection density increasing with precipitation until a peak at around 50 mm per month, after which detection decreased with increasing rainfall.

Biogeochemical variables were also highly important, with nitrogen appearing in 94% (average rank: 5.5) and carbon in 88% (average rank: 5.1) of final models across all species (Table 3). Both variables showed a generally positive relationship with detection density in most of the final models in which they were of high importance, but with detections dropping off after a peak for many species.

Of the variables relating to vegetation, LAI and VCF appeared to be more important for carnivores (both appearing in 83% of final models, versus 70% and 60% among herbivores, respectively), and EVI more important for herbivores (80% of final models, versus 50% among carnivores; Table 3).

Anthropogenic variables were important predictors for most species in the study area. Both human population density and cattle density showed high importance, each appearing in more than three quarters of final species models (Table 3). Human density appeared to be a more important factor for carnivores than herbivores (appearing in 100% versus 70% of final models), while cattle density was of higher importance for herbivores than carnivores (appearing in 90% versus 50% of final models; Table 3).

The relationship between detection density and human density varied substantially between species. Cattle density, in contrast, showed a consistent negative relationship with detection density for all but one species (Fig. 2).

Management variables were generally less important than climatic, biogeographic, vegetation, and anthropogenic variables (Tables 2A & 3). Protection status was much more important for herbivores than carnivores: no protection (MIR < 0.60, average rank: 13.1), intermediate protection (MIR < 0.20, average rank: 15.6), and trophy hunting (MIR < 0.37, average rank: 13.3) all appeared in only 17% of final carnivore models, versus 60%, 50%, and 50% for herbivores (Tables 2A & 3). High protection showed greater importance across all species (average rank: 10.3), appearing in 50% of final carnivore models and 60% of final herbivore models (Table 3).

When protection variables did appear in the final model for a species, they exhibited consistent relationships with detection densities: unprotected areas and areas under intermediate protection were negatively associated with detection density, while areas under high protection had a positive association.

Trophy hunting had a more variable influence on detections between species, and did not appear to be linked to how heavily different species are hunted: although greater kudu, buffalo, zebra, sable antelope, leopard, elephant and lion have been documented as the seven most hunted species in Zimbabwe (Barnett and Patterson, 2006), of these species, hunting only appeared as a variable in the final models for elephant (MIR: 0.17) and zebra (MIR: 0.15; Table 2A), showing a negative association with detection density for both species (Fig. 2).

The least important variables were settlement points (MIR < 0.09, average rank: 16.9) and rivers (MIR < 0.03, average rank: 17.9; Table 2A), which each appeared in only 31% of final models (Table 3).

4. Discussion

In this study, we used the random forest algorithm to investigate multi-scale, multi-species habitat associations across part of Africa's largest transfrontier conservation area. As far as we are aware, this is the first time machine learning has been used to model spoor data in a multi-scale framework. It is also one of the first studies to assess multi-scale habitat use for such a large and diverse mammal assemblage: our 16 study species range in size from around 5 kg (African wild cat) to 5,000 kg (elephant; Estes, 1991) and feature members of four taxonomic orders (including six carnivorans, eight even-toed ungulates, one odd-toed ungulate, and one proboscidean). Of these species, six are classified as *Endangered*, *Vulnerable* or *Near Threatened* on the IUCN Red List (*Endangered*: elephant; *Vulnerable*: leopard, lion, giraffe; *Near Threatened*: buffalo, zebra), and nine are thought to have decreasing population trends (leopard, lion, spotted hyaena, African wild cat, buffalo, common duiker, warthog, giraffe, zebra; IUCN, 2020).

One of the main limitations of this study is that data were collected by surveying along roads and fire cutlines, and are therefore not an unbiased representation of the entire study area. However, simulations have shown that, while models trained with spatially representative datasets out-performed models trained with spatially non-representative datasets in standard metrics of model performance, spatially non-representative models produced superior predictions of species-environment relationships (Chiaverini *et al.*, 2021). Our ecological inferences should thus not be compromised by the non-representative sampling design, which is typical of many ecological assessments. Another key limitation is that this study does not account for seasonal dynamics in species distributions, as the size of the study area meant that data collection had to be carried out across different months of multiple years. As such, this study presents insights into species distributions across seasons, and further research should be carried out into seasonal dynamics of species distributions in KAZA.

4.1. Factors determining mammal distributions

The prominence of longitude and latitude in all final species models reflects the importance of considering spatial structure in species distributions, particularly when assessments are being carried out across a very large area (Evans *et al.*, 2011). The strong latitudinal gradients likely reflect a combined influence of precipitation gradients and gradients in human impacts, which co-vary across the system in a cline of both increasing precipitation and human footprint from southwest to northeast.

Our results also suggest that variables relating to climate, soil quality, and vegetation are the most important factors determining mammal distribution across southern KAZA, in line with established literature on African mammal ecology (Coe, Cumming and Phillipson, 1976; East, 1984).

Average precipitation over the survey period was the most important predictor of habitat use for nearly all species considered. This is unsurprising given the aridity of the study area and its large extent, which enables spatial variation in precipitation – which affects nearly all aspects of ecological systems, and is the dominant driver of vegetation structure and density (Zhu and Southworth, 2013) – to affect the distribution of species in different ways depending on their ecological niches.

By considering only mean precipitation during all survey months, this analysis did not take into account seasonal rainfall patterns, which are known to play an important role in shaping wildlife movement patterns in KAZA (Naidoo et al., 2016). As such, modelling how species' distributions vary across different times of year would reveal valuable insights into how these seasonal dynamics shape species' habitat use, and would be an important component of corridor planning exercises in KAZA.

Soil carbon and nitrogen were also important variables predicting habitat use. Soil nutrients are generally higher in more productive ecosystems (Zhou *et al.*, 2011); for herbivores, these variables are therefore likely to be important predictors of distribution due to their impact on food availability (e.g. (Kaszta *et al.*, 2016). In turn, many African carnivore species are known to make habitat use decisions based on prey availability (Everatt *et al.*, 2015; Searle *et al.*, 2020). As we did not explicitly include information on prey in our carnivore models, it is likely that climatic, biogeochemical, and vegetation variables ranked highly in our carnivore models as they acted as proxies for prey. This suggests that environmental variables that are important for prey can be used to model carnivore distribution.

Although important in predicting detection densities, vegetation indices showed less consistent relationships across species, likely due to the range of habitats and vegetation structures preferred by different species. This highlights the importance of management and land planning efforts that seek to maintain varied habitat types for supporting diverse wildlife populations.

Anthropogenic pressures also appeared to have an important influence on habitat use. The majority of species considered showed a strong negative association with cattle density. Overgrazing can severely degrade soil nutrients, leaving areas used by cattle with reduced plant biomass and species diversity, and dominated by plant species preferred by livestock (Zhou *et al.*, 2011). As a result, while many ungulates compete directly with livestock for food, even those that do not can be left with little to consume when cattle densities are particularly high. Carnivores may select against areas with a high cattle density as they lack sufficient prey – although livestock can provide an alternative food source (Winterbach *et al.*, 2013) – but could also be killed or pushed out of these areas by harmful encounters with livestock herders, such as poisonings. In line with this, large carnivore range contractions have been shown to be significantly more likely in regions with high cattle density (Wolf and Ripple, 2017). However, cattle are also a strong indicator of human presence, and the effects of livestock and humans on species can be difficult to disentangle.

Human density was also an important predictor of habitat use. The negative association found for a number of species aligned with our expectations, as human population density has been established as a key cause of mammal species becoming threatened with extinction (McKee *et al.*, 2013). In contrast, the positive association with human density observed among black-backed jackal, giraffe, greater kudu, impala, leopard, and spotted hyaena may indicate an ability for these species to persist in areas of higher human density, provided there is sufficient habitat. However, it is more likely that this positive relationship is an artefact of the large number of human settlements adjacent to some of the most highly-productive and well-protected parts of the study landscape.

One important consideration in interpreting these results is that the vast majority of survey transects were conducted in areas without cattle or human presence. As a result, the identified effects of cattle and human density on mammal distributions in the study landscape are likely to be a result of edge effects acting on wildlife populations around the periphery of PAs, rather than direct impacts in areas where wild species coexist alongside livestock and humans. In addition, although they were the best sources of information available for our analysis and have been previously used in studies of this kind (e.g. Weise *et al.*, 2017), the layers representing cattle and human densities were calculated at a global or continental scale and are unlikely to be fully representative of fine-scale livestock and human presence in the study area.

Variables relating to protection status showed consistent relationships across species: habitat use was negatively associated with unprotected and less-strictly protected areas, and positively associated with highly-protected areas. Although protection status appeared to be more important for herbivores than carnivores, high protection was similarly important across both groups, highlighting the value of strictly protected areas as refugia for biodiversity. However, as less than 5% of our total survey effort was conducted in completely unprotected land, we are unable to make clear inferences about the importance of protected versus unprotected land in this study. It is likely that wildlife densities are much lower in unprotected areas than PAs across the KAZA landscape, given the role of protection in reducing illegal activities (such as bushmeat poaching; Loveridge et al., 2020), and the well-established link between effective protection and persistence of mammal populations (Karanth et al., 2010). In light of this, and the mounting pressures on wildlife areas outside PAs across Africa (Newmark, 2008), continued protection of core wildlife areas should remain a priority in the KAZA TFCA.

Trophy hunting was of intermediate importance for most of our study species, and showed greater importance among herbivores than carnivores. The relationship between trophy hunting and habitat use was negative for all herbivores with the variable in their final model.

Variables relating to permanent surface water showed intermediate importance in our final models for most species, and appeared to be more important for herbivores than carnivores. Although we expected water availability to have greater importance in shaping distribution of mammals in the water-limited southern KAZA TFCA, this may be a consequence of the limitations of our water variables. In particular, although the landscape includes a large number of seasonal rivers that are an important source of water for wildlife in the months that they flow (British Geological Survey, 2018a, 2018b), it would not have been possible to accurately capture seasonal river flows over such a large area. As a result, our rivers variable only included perennial rivers. Nevertheless, most species with variables representing permanent surface water in the final model showed a positive relationship with these variables.

4.2. Importance of a multi-scale, multi-species approach

We identified a number of spatial scales at which species respond to different variables, and at which different species responded to the same variables, in line with previous multi-scale habitat use studies (Khosravi, Hemami and Cushman, 2019; Ashrafzadeh *et al.*, 2020; Penjor *et al.*, 2021). This highlights the importance of adopting a multi-scale approach for species distribution modelling, as it can reveal divergent patterns of decision-making both among and within species (McGarigal *et al.*, 2016).

Our results suggest that the studied species generally make habitat use decisions at relatively broad scales in southern KAZA. This is consistent with previous studies which have found that mammals with large home ranges respond to habitat variables at broader spatial scales (e.g. Mateo-Sánchez, Cushman and Saura, 2014; Macdonald *et al.*, 2018; Khosravi, Hemami and Cushman, 2019).

We also illustrate that different species can have divergent responses to the same variables. This underscores the importance of adopting a multi-species approach for critical decision-making processes that depend on understanding wildlife distributions and habitat associations, such as PA expansion & prioritisation and corridor planning. Integrating multi-species, multi-scale thinking into these processes should help decision-makers to prioritise areas of habitat that maximise value for as many species as possible.

4.3. Potential future trends affecting mammal distributions

The importance of precipitation in predicting the distributions of all species highlights the significant impacts that altered rainfall patterns could have on wildlife distributions in water-limited areas like KAZA. Rainfall is predicted to decrease across southern Africa under multiple climate change scenarios, with significant drying projected over parts of Botswana (Moise and Hudson, 2008). Our models suggest that these changes could have substantial impacts on wildlife distributions across the KAZA TFCA, with potential implications for the landscape's long-distance ungulate migration (Naidoo et al., 2016).

Similar shifts in rainfall patterns are anticipated across much of the world, including many of the world's most biodiverse regions (IPCC, 2013). As such, management authorities in wildlife areas, including those in KAZA, must work to ensure that current PA networks will remain fit for purpose under anticipated changes in rainfall patterns and other impacts of climate change, through modelling exercises and adaptive management. A particularly important consideration will be the need to maintain connectivity between habitat patches across these changing conditions (Khosravi *et al.*, 2021).

The suggested negative impact of cattle density on mammals in the study area hints at another concerning trend for global wildlife populations. Global livestock numbers are expected to increase substantially in the coming years (FAO, 2009), with food demand for livestock products in sub-Saharan Africa projected to nearly double from 2010 to 2050 (van Vuuren *et al.*, 2009). Not only will the intensification of livestock production required to meet this demand increase the direct competition faced by wild herbivores and contribute to habitat degradation, but it is also likely to require conversion of habitat to support grassland expansion for grazing. The synergy of projected increases in aridity and simultaneous increases in cattle densities and extent of the landscape utilized for pastoral activities may be a harbinger of a biodiversity crisis in southern Africa.

In light of these projections, efforts must be made to ensure protection is enforced across global PA networks, to prevent encroachment by livestock and illegal grazing in strictly protected areas, and safeguard these areas' remaining natural habitat. At the same time, management authorities should strive to enact directives that promote the coexistence of wildlife and livestock, such as through controlled grazing and other sustainable pastoral practices, which have been shown to regenerate vegetation in overgrazed areas (Western *et al.*, 2020). Initiatives of this kind are likely to be particularly important in large conservation landscapes like the KAZA TFCA which encompass mosaics of protected and communal land, and will be a critical component of efforts to align conservation with human development.

Given the relative coarseness of livestock data used in this study and its resultant limitations, and the rapidly-evolving situation regarding livestock numbers in sub-Saharan Africa, we also recommend the collation of wildlife and livestock data from aerial surveys carried out by KAZA partner countries. Such up-to-date information would be highly valuable to inform effective management decisions and deliver recommendations on acceptable livestock densities to sustain the ecosystem, particularly in light of climate change impacts.

Acknowledgements

We would like to thank the Governments of Botswana and Zimbabwe for supporting this study, and the respective wildlife authorities for granting the research permits necessary for fieldwork; in Botswana, the Ministry of Environment, Natural Resources Conservation and Tourism (Permit number: EWT 8/36/4 XXIII (15)) and, in Zimbabwe, the Parks and Wildlife Management Authority (Permit numbers: REF:DM/Gen/(T) 23(1)(c)(ii): 12/2012, 08/2013, 51/2014, 10/2015). We thank the field staff of the Hwange Lion Project and TKPP for their assistance with data collection, and the Victoria Falls Wildlife Trust for their support of this project. Funding for the surveys was provided by the Robertson Foundation, the Recanati-Kaplan Foundation, and a Darwin Initiative for Biodiversity Grant DAR17-031 (2009-2012). DPhil scholarship funding for CS is provided by the University of Oxford & Natural Environment Research Council (NERC) Doctoral Training Partnership in Environmental Research NE/L002612/1.

Conflict of Interest

The authors have declared no conflicts of interest.

Authors' contributions

Charlotte Searle, Żaneta Kaszta, Sam Cushman, Andy Loveridge, and David Macdonald conceived the idea and designed the methodology; Dominik Bauer, Kristina Kesch, Andrew Loveridge and Jane Hunt collected the data; Charlotte Searle and Żaneta Kaszta analysed the data with input from Samuel Cushman; Charlotte Searle wrote the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Data Availability Statement

Data available via the Dryad Digital Repository https://doi.org/10.5061/dryad.hmgqnk9kg (Searle & Kaszta *et al.*, 2022).

References

Ashrafzadeh, M. R. *et al.* (2020) 'A multi-scale, multi-species approach for assessing effectiveness of habitat and connectivity conservation for endangered felids', *Biological Conservation*. Elsevier Ltd, 245, p. 108523. doi: 10.1016/j.biocon.2020.108523.

Barnett, R. and Patterson, C. (2006) *Sport hunting in the Southern African Development Community (SADC) region: An overview*.

Breiman, L. (2001) 'Random Forests', *Machine Learning*, 45, pp. 5–32.

British Geological Survey (2018a) *Hydrogeology of Botswana*, *Earthwise*. Available at: http://earthwise.bgs.ac.uk/index.php/Hydrogeology_of_Botswana (Accessed: 8 April 2018).

British Geological Survey (2018b) *Hydrogeology of Zimbabwe*, *Earthwise*. [Water Research Commission]. Available at: http://earthwise.bgs.ac.uk/index.php/Hydrogeology_of_Zimbabwe (Accessed: 8 April 2018).

Brondizio, E. S. *et al.* (2019) *Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. IPBES Secretariat.*

CBD (2010) *Decision adopted by the Conference of the Parties to the Convention on Biological Diversity at its tenth meeting X/2. The Strategic Plan for Biodiversity 2011-2020 and the Aichi Biodiversity Targets*.

Chiaverini, L. *et al.* (2021) 'Effects of non-representative sampling design on multi-scale habitat models: Flammulated owls in the Rocky Mountains.', *Ecological Modelling*. Elsevier B.V., 450(February), p. 109566. doi: 10.1016/j.ecolmodel.2021.109566.

Coe, M. J., Cumming, D. H. and Phillipson, J. (1976) 'Biomass and production of large African herbivores in relation to rainfall and primary production', *Oecologia*, 22, pp. 341–354.

Cushman, S. A. *et al.* (2006) 'Gene flow in complex landscapes: testing multiple hypotheses with causal modeling.', *The American naturalist*. The University of Chicago Press , 168(4), pp. 486–99. doi: 10.1086/506976.

Cushman, S. A. *et al.* (2017) 'Multiple-scale prediction of forest loss risk across Borneo', *Landscape Ecology*. Springer Netherlands, 32(8), pp. 1581–1598. doi: 10.1007/s10980-017-0520-0.

Cushman, S. A. and Wasserman, T. N. (2018) 'Landscape applications of machine learning: Comparing random forests and logistic regression in multi-scale optimized predictive modeling of American marten occurrence in northern Idaho, USA', in *Machine Learning for Ecology and Sustainable Natural Resource Management*. Springer International Publishing, pp. 185–203. doi: 10.1007/978-3-319-96978-7_9.

Didan, K. (2015) 'MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006 [Data set].

NASA EOSDIS Land Processes DAAC'. doi: 10.5067/MODIS/MOD13Q1.006.

Dimiceli, C. *et al.* (2015) 'MOD44B MODIS/Terra Vegetation Continuous Fields Yearly L3 Global 250m SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC'. doi: 10.5067/MODIS/MOD44B.006.

East, R. D. (1984) 'Rainfall, soil nutrient status and biomass of large African savanna mammals', *African Journal of Ecology*, 22, pp. 245–270.

ESRI (2016) 'ArcGIS Desktop: Release 10.6.1. Redlands, CA: Environmental Systems Research Institute.'

Estes, R. (1991) *The behavior guide to African mammals*. Berkeley: University of California Press.

Evans, J. S. *et al.* (2011) 'Modeling species distribution and change using random forest', in *Predictive Species and Habitat Modeling in Landscape Ecology: Concepts and Applications*. Springer New York, pp. 139–159. doi: 10.1007/978-1-4419-7390-0_8.

Evans, J. S. and Murphy, M. A. (2019) 'Package "rfUtilities"'.

Everatt, K. T. *et al.* (2015) 'The influence of prey, pastoralism and poaching on the hierarchical use of habitat by an apex predator', *African Journal of Wildlife Research*, 45(2), pp. 187–196. doi: 10.3957/056.045.0187.

FAO (2009) *The state of food and agriculture: Livestock in the balance*. Rome, Italy.

Farhadinia, M. S., Rostro-Garcia, S. and Feng, L. (2020) 'Big cats in borderlands: challenges and implications for transboundary conservation of Asian leopards', *Oryx*. doi: 10.1017/S0030605319000693.

Fick, S. E. and Hijmans, R. J. (2017) 'WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas', *International Journal of Climatology*. John Wiley & Sons, Ltd, 37(12), pp. 4302–4315. doi: 10.1002/joc.5086.

Gilbert, M. *et al.* (2018) 'Global distribution data for cattle, buffaloes, horses, sheep, goats, pigs, chickens and ducks in 2010', *Scientific Data*. Nature Publishing Groups, 5(1), pp. 1–11. doi: 10.1038/sdata.2018.227.

Harrell Jr, F. E. (2020) 'Package "Hmisc"'. Available at: https://cran.uib.no/web/packages/Hmisc/Hmisc.pdf (Accessed: 9 January 2021).

Hengl, T., Leenaars, J. G. B., *et al.* (2017) 'Soil nutrient maps of sub-Saharan Africa: Assessment of soil nutrient content at 250 m spatial resolution using machine learning', *Nutrient Cycling in Agroecosystems*. Springer Netherlands, 109(1), pp. 77–102. doi: 10.1007/s10705-017-9870-x.

Hengl, T., Mendes de Jesus, J., *et al.* (2017) 'SoilGrids250m: Global gridded soil information based on machine learning', *PLOS ONE*. Edited by B. Bond-Lamberty. Public Library of Science, 12(2), p. e0169748. doi:

10.1371/journal.pone.0169748.

Holland, J. D., Bert, D. G. and Fahrig, L. (2004) 'Determining the spatial scale of species' response to habitat', *BioScience*. American Institute of Biological Sciences, 54(3), pp. 227–233. doi: 10.1641/0006-3568(2004)054[0227:DTSSOS]2.0.CO;2.

IPCC (2013) *Climate change 2013: The physical science basis. Working Group I contribution to the fifth assessment report of the Intergovernmental Panel on Climate Change.*

IUCN (2020) *The IUCN Red List of Threatened Species. <https://www.iucnredlist.org>*.

Karanth, K. K. *et al.* (2010) 'The shrinking ark: Patterns of large mammal extinctions in India', *Proceedings of the Royal Society B: Biological Sciences*. Royal Society, 277(1690), pp. 1971–1979. doi: 10.1098/rspb.2010.0171.

Kaszta, Z. *et al.* (2016) 'Bulk feeder or selective grazer: African buffalo space use patterns based on fine-scale remotely sensed data on forage quality and quantity', *Ecological Modelling*, 323, pp. 115–122. doi: 10.1016/j.ecolmodel.2015.12.006.

Khosravi, R. *et al.* (2021) 'Prey availability modulates predicted range contraction of two large felids in response to changing climate', *Biological Conservation*. Elsevier Ltd, 255(January), p. 109018. doi: 10.1016/j.biocon.2021.109018.

Khosravi, R., Hemami, M.-R. and Cushman, S. A. (2019) 'Multi-scale niche modeling of three sympatric felids of conservation importance in central Iran', *Landscape Ecology*. Springer Netherlands, 34(10), pp. 2451–2467. doi: 10.1007/s10980-019-00900-0.

Liaw, A. and Wiener, M. (2002) 'Classification and regression by RandomForest', *R News*, 2/3. Available at: https://www.researchgate.net/publication/228451484 (Accessed: 14 December 2020).

Linard, C. *et al.* (2012) 'Population distribution, settlement patterns and accessibility across Africa in 2010', *PLOS ONE*. Edited by G. J.-P. Schumann. Public Library of Science, 7(2), p. e31743. doi: 10.1371/journal.pone.0031743.

Lindsey, P. A. *et al.* (2018) 'More than $1 billion needed annually to secure Africa's protected areas with lions', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 115(45), pp. E10788–E10796. doi: 10.1073/pnas.1805048115.

Liu, C. *et al.* (2013) 'Species distribution modelling for conservation planning in Victoria, Australia', *Ecological Modelling*. Elsevier B.V., 249(December 2011), pp. 68–74. doi: 10.1016/j.ecolmodel.2012.07.003.

Loveridge, A. J. *et al.* (2010) 'African lions on the edge: Reserve boundaries as "attractive sinks"', in *Biology and* *conservation of wild felids*, pp. 283–304.

Loveridge, A. J. *et al.* (2020) 'Evaluating the spatial intensity and demographic impacts of wire-snare bush-meat poaching on large carnivores', *Biological Conservation*. Elsevier, 244, p. 108504. doi: 10.1016/j.biocon.2020.108504.

Macdonald, D. W. *et al.* (2018) 'Multi-scale habitat selection modeling identifies threats and conservation opportunities for the Sunda clouded leopard (Neofelis diardi)', *Biological Conservation*. Elsevier, 227, pp. 92–103. doi: 10.1016/J.BIOCON.2018.08.027.

Mateo-Sánchez, M. C., Cushman, S. A. and Saura, S. (2014) 'Connecting endangered brown bear subpopulations in the Cantabrian Range (north-western Spain)', *Animal Conservation*, 17(5), pp. 430–440. doi: 10.1111/acv.12109.

McGarigal, K. *et al.* (2016) 'Multi-scale habitat selection modeling: a review and outlook', *Landscape Ecology*, 31(6), pp. 1161–1175. doi: 10.1007/s10980-016-0374-x.

McKee, J. *et al.* (2013) 'Human population density and growth validated as extinction threats to mammal and bird species', *Human Ecology*, 41, pp. 773–778. doi: 10.1007/s10745-013-9586-8.

Moise, A. F. and Hudson, D. A. (2008) 'Probabilistic predictions of climate change for Australia and southern Africa using the reliability ensemble average of IPCC CMIP3 model simulations', *Journal of Geophysical Research*. Blackwell Publishing Ltd, 113(D15113). doi: 10.1029/2007JD009250.

Myneni, R., Knyazikhin, Y. and Park, T. (2015) 'MCD15A2H MODIS/Terra+Aqua Leaf Area Index/FPAR 8-day L4 Global 500m SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC'. doi: 10.5067/MODIS/MCD15A2H.006.

Naidoo, R. *et al.* (2016) 'A newly discovered wildlife migration in Namibia and Botswana is the longest in Africa', *Oryx*. Cambridge University Press, 50(01), pp. 138–146. doi: 10.1017/S0030605314000222.

Newmark, W. D. (2008) 'Isolation of African protected areas', *Frontiers in Ecology and the Environment*. John Wiley & Sons, Ltd, 6(6), pp. 321–328. doi: 10.1890/070003.

Penjor, U. *et al.* (2021) 'Prioritizing areas for conservation outside the existing protected area network in Bhutan: The use of multi-species, multi-scale habitat suitability models', *Landscape Ecology*. Springer Netherlands, 36(5), pp. 1281–1309. doi: 10.1007/s10980-021-01225-7.

Ripple, W. J. *et al.* (2014) 'Status and ecological effects of the world's largest carnivores', *Science*, 343(1241484), pp. 1–11. doi: 10.1126/science.1241484.

Searle, C. E. *et al.* (2020) 'Drivers of leopard (Panthera pardus) habitat use and relative abundance in Africa's largest transfrontier conservation area', *Biological Conservation*. Elsevier, 248, p. 108649. doi: 10.1016/j.biocon.2020.108649.

Searle, C.E. and Kaszta, Z. *et al*. (2022) Data from: Random forest modelling of multi-scale, multi-species habitat associations within KAZA transfrontier conservation area using spoor data. Dryad Digital Repository. https://doi.org/10.5061/dryad.hmgqnk9kg

Southern African Development Community (2018) *Transfrontier Conservation Areas*.

van Vuuren, D. P. *et al.* (2009) 'Outlook on agricultural change and its drivers', in McIntyre, B. D. et al. (eds) *International Assessment of Agricultural Knowledge, Science and Technology for Development (IAASTD): Agriculture at a Crossroads, global report*. Washington, DC, USA: Island Press, pp. 255–305.

Weise, F. J. *et al.* (2017) 'The distribution and numbers of cheetah *(Acinonyx jubatus)* in southern Africa', *PeerJ*. PeerJ Inc., 5, p. e4096. doi: 10.7717/peerj.4096.

Western, D. *et al.* (2020) 'Conservation from the inside-out: Winning space and a place for wildlife in working landscapes', *People and Nature*, 2(2), pp. 279–291. doi: 10.1002/pan3.10077.

Winterbach, H. E. K. *et al.* (2013) 'Key factors and related principles in the conservation of large African carnivores', *Mammal Review*, 43, pp. 89–110. doi: 10.1111/j.1365-2907.2011.00209.x.

Wolf, C. and Ripple, W. J. (2017) 'Range contractions of the world's large carnivores', *Royal Society Open Science*, 4(7), p. 170052. doi: 10.1098/rsos.170052.

Zeller, K. A. *et al.* (2018) 'Are all data types and connectivity models created equal? Validating common connectivity approaches with dispersal data', *Diversity and Distributions*. Edited by J. Bolliger. Blackwell Publishing Ltd, 24(7), pp. 868–879. doi: 10.1111/ddi.12742.

Zeller, K. A., McGarigal, K. and Whiteley, A. R. (2012) 'Estimating landscape resistance to movement: a review', *Landscape Ecology*. Springer Netherlands, 27(6), pp. 777–797. doi: 10.1007/s10980-012-9737-0.

Zhou, Z. Y. *et al.* (2011) 'Dynamics of vegetation and soil carbon and nitrogen accumulation over 26 years under controlled grazing in a desert shrubland', *Plant and Soil*. Springer, 341(1–2), pp. 257–268. doi: 10.1007/s11104-010-0641-6.

Zhu, L. and Southworth, J. (2013) 'Disentangling the relationships between net primary production and precipitation in southern Africa savannas using satellite observations from 1982 to 2010', *Remote Sensing*. Multidisciplinary Digital Publishing Institute, 5(8), pp. 3803–3825. doi: 10.3390/rs5083803.