SPATIAL ANALYSIS OF CHIMPANZEE HABITAT QUALITY BASED ON VEGETATION FOOD AVAILABILITY IN GOMBE NATIONAL PARK, TANZANIA

BY

Ying Zhong

April 24th, 2015

Advisors

Dr. Jennifer Swenson

Dr. Steffen Foerster

Master's Project submitted in partial fulfillment of the

requirements for the Master of Environmental Management degree in

the Nicholas School of the Environment of

Duke University

EXECUTIVE SUMMARY

Chimpanzee (*Pan troglodytes*) is an Endangered species listed in IUCN Red List with a most recently estimated total population ranging approximately from 172700 to 299700. In east Africa, habitat loss and degradation resulting from agriculture development is the major cause of population decline. Gombe National Park is one among few remaining habitats for chimpanzees in East Africa. The park has been well-protected and maintained high vegetation cover compared to its surrounding environment.

Given its importance to the chimpanzee's diet, the vegetation cover defines the spatial variation of chimpanzee habitat quality. The aim of this project is to evaluate chimpanzee habitat quality for Gombe National Park by mapping the spatial variation of vegetation food availability, for the ultimate goal of supporting chimpanzee habitat conservation planning and studies of chimpanzee feeding behavior and ecology. Two sections of work were done to achieve this goal: vegetation species distribution modeling and vegetation cover mapping:

Vegetation Species Distribution Modeling

Vegetation food availability is a significant determinant of chimpanzee habitat quality. Different food resources can be of significant difference in indicating habitat quality due to chimpanzee's feeding preference. Therefore, a species-level investigation of the forest cover is especially important for a small scale habitat in order to assess the spatial variation of chimpanzee habitat quality.

In this project, I used MaxEnt model to predict the distribution of 24 chimpanzee vegetation food species in Gombe National Park with a fine spatial resolution of 10 m. Furthermore, a chimpanzee habitat quality surface was generated based on vegetation food availability by overlaying the distribution extents of 24 important vegetation species after weighting each of them with chimpanzee's feeding preference. Further, biological relevance validations which correlate habitat quality with chimpanzee feeding time using linear regression models where performed to evaluate how good the model outputs capture information that influence chimpanzee feeding behavior.

The overall performance of species distribution modeling is relatively high with an average AUC of 0.85 and an average overall accuracy of 0.78. Significant correlation between predicted chimpanzee habitat quality and chimpanzee feeding time also indicates high predication accuracy of the habitat quality. Overall, this project proves that MaxEnt is capable in predicting spatial variation of vegetation species

distribution in a small area with a fine resolution of 10 m, which has seldom been investigated in precious studies, and proposes a method to evaluate vegetation food availability at the species level and to assess chimpanzee habitat quality based on vegetation food availability.

Vegetation Cover Mapping

Remote sensing technique is an alternative approach to characterize habitat quality by identifying vegetation cover in a broad scale. However, it is usually limited to general distinction of forest and non-forest or broad vegetation types rather than species level classification. To address this limitation, my goal of the second section of this project was to develop an innovative method of producing vegetation cover map which incorporates vegetation species composition information in each vegetation cover class, and to generate such a biologically meaningful vegetation cover map for Gombe National Park.

The method I used to achieve this goal is a combination of traditional remote sensing classification and ecological data mining technique. Specifically, I performed a cluster analysis of the vegetation survey data for the purpose of generating vegetation cover classes based on their similarity in vegetation species composition, and further used the clustered survey data to train a supervised classification algorithm – Maximum Likelihood Classification. Additionally, an innovative semi-automatic post-processing workflow was proposed and used to correct for misclassification resulting from spectral noises in high resolution images, which combines an automatic sieve and clump process, an automatic "buffer zone majority" correction, and a manual cloud and shadow correction. The result of this section of work is a vegetation cover map of Gombe National Park showing seven vegetation cover classes named after their dominant vegetation species with high spatial coherency.

The combination of cluster analysis and remote sensing technique advanced the vegetation cover classification to the species-level classification. Further, the cluster analysis method provides an alternative way of vegetation class schema design where little local knowledge of vegetation assemblages is needed. Moreover, the semi-automatic post-processing improved the efficiency and accuracy of post-processing of vegetation cover classification.

In conclude, this project produced 24 vegetation habitat suitability maps and distribution extent maps, an overall chimpanzee habitat quality surface map, and a vegetation cover map of Gombe National Park. These products provide useful spatial information to support vegetation studies, chimpanzee behavioral studies, and habitat evaluation and conservation planning in Gombe National Park.

ACKNOWLEDGEMENTS

It was more than 15 years ago for the first time I heard about Dr. Jane Goodall's story with chimpanzees and Gombe National Park, when I was still a little girl living thousands miles away from Africa. Since then Dr. Goodall has become my heroine, and the Gombe National Park in Africa has been an attractive mystery deep in the heart. I would have never imagined working so closely with that little forest in Africa if I haven't come to the Nicholas School and wasn't introduced to this most exciting project.

I thank Dr. Steffen Forester, my advisor and co-worker, who has always been inspiring and encouraging me to pursue rigorous, innovative, and advanced science. I enjoyed every meeting with Steffen during the past one year and considered them to be the most fruitful moment of science every week.

I want to thank Dr. Jennifer Swenson, my advisor in the Nicholas School. Jennifer introduced me to the world of remote sensing, and her advice on the project methodologies and paper writing help me improve the quality of my work.

I'm very grateful to Dr. Dean Urban, my landscape ecology professor. His lectures have greatly expanded my interest and knowledge in ecology. I'm also grateful to John Fay, who connected me with this project, and helped me a lot on the GIS work. The Nicholas School of Environmental at Duke is a wonderful place for anyone who loves nature, because of its great people. I will never forget all the help I received from Dr. Mariano Gonzalez Roglich, Brenna Forester, Amanda Schwantes, and a lot of my friends I met at Duke.

There are five people without whom I would have achieved none of the accomplishments. They are my parents, and my sisters and brothers, the most loving and supporting people in the world. They are my wind when I fly, and my nest when I rest.

My special thank you goes to Junyan Liu, for his love and support on this miracle journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGUERS

CHAPTER 1. INTRODUCTION

Status of primate species risking extinction on the earth are putting primate conservation in high priority of the world's wildlife conservation work. With about half of the species listed as Threatened Species in IUCN Red List (Mittermeier et al. 2009), the primate is one of the mammal family most threatened by extinction (Schipper et al. 2008). Main threats contribute to primate population decline are habitat loss and degradation, hunting for bushmeat, and wildlife trading (Mittermeier et al. 2009).

Like many primate species, chimpanzee (*Pan troglodytes*) is one of the Endangered species in IUCN Red List, with a most recently estimated total population ranging approximately from 172700 to 299700 (Butynski 2003). Chimpanzees are mainly found from the southern Senegal to the western Uganda and western Tanzania, along the north of Congo River (Oates et al. 2008). Given its endangered status, and its importance for animal behavior studies, the studies and conservation work of chimpanzee have driven international concern and effort.

Main threats that contribute to chimpanzee mortality include human hunting, habitat loss and degradation, and disease transmission (Plumptre 2010). Particularly, in East Africa, habitat loss and degradation resulting from agriculture development is the major cause of population decline especially outside of protected area where large proportion of chimpanzee population live (Plumptre 2010). Within protected area or intact natural habitat, disease transmission is the largest concern to chimpanzee population conservation (Köndgen et al. 2008; Williams et al. 2008).

Gombe National Park is one among few remaining habitats for chimpanzees in East Africa. The park is located in western Tanzania, with a total area of about 35 km^2 (Figure 1). The land was protected since the 1940s as a Game Reserve built by the British government (Moreau 1945). It has gained international attention through Dr. Jane Goodall's ground-breaking behavioral studies of chimpanzees dating back to 1960, which led to the establishment of a National Park in 1968 for chimpanzee population and habitat protection. It was estimated that 100 chimpanzees were located in Gombe National Park in 2006 (Moyer et al. 2006). Nowadays two big chimpanzee communities, Kasekela and Mitumba, and one diminishing chimpanzee community, Kalande, are located in the park. The population of the Kasekela community was recorded as 60 in 1966, and reached 63 in 2008 (Pusey et al. 2008). The main cause to death in the Kasekela community is disease, which accounted for 58% of deaths with known cause (Williams et al. 2008). The population of the Mitumba community has fluctuated from 20 to 25 since 1996 (Pusey et al. 2008). Eleven individuals were recorded in the Kalande community in 2008 (Pusey et al. 2008).

Gombe National Park is a small protected area with heterogeneous topographic characteristics across the park. The elevation rises from west to east, with a range of 766 m to 1622 m (Figure 1). Seven permanent east-west running streams are located in the park, along with numerous temporary streams that vary in the length which carry water into the dry season months. Vegetation cover varies distinctly across the park along the N-S and E-W axes (Figure 1). Dense forests and shrub lands are mainly located in the central and northern parts of the park, especially concentrating in the east-west running valleys at lower elevations, while the eastern, higher elevations of the park are generally covered by open shrubs, grasslands, and bare lands (Pintea 2007). Compared to its severely deforested surrounding environment, the park has been well-protected and maintained high and increasing vegetation cover (Pintea 2007). Nishida et al. (1983) reported 631 vegetation species in the park and 141 of them were found to be eaten by chimpanzees.

Given its importance to the chimpanzee's diet, the vegetation cover defines the spatial variation of chimpanzee habitat quality. The aim of this project is to evaluate chimpanzee habitat quality for Gombe National Park by mapping the spatial variation of vegetation food availability for the ultimate goal of supporting chimpanzee habitat conservation planning and studies of chimpanzee feeding behavior and ecology. In chapter 2, I descried the application of recently advanced machine learning techniques to describe and predict chimpanzee vegetation food species distribution across the park landscape. In chapter 3, I described the application of ecological data mining and land cover classification to produce a biologically meaningful vegetation cover map.

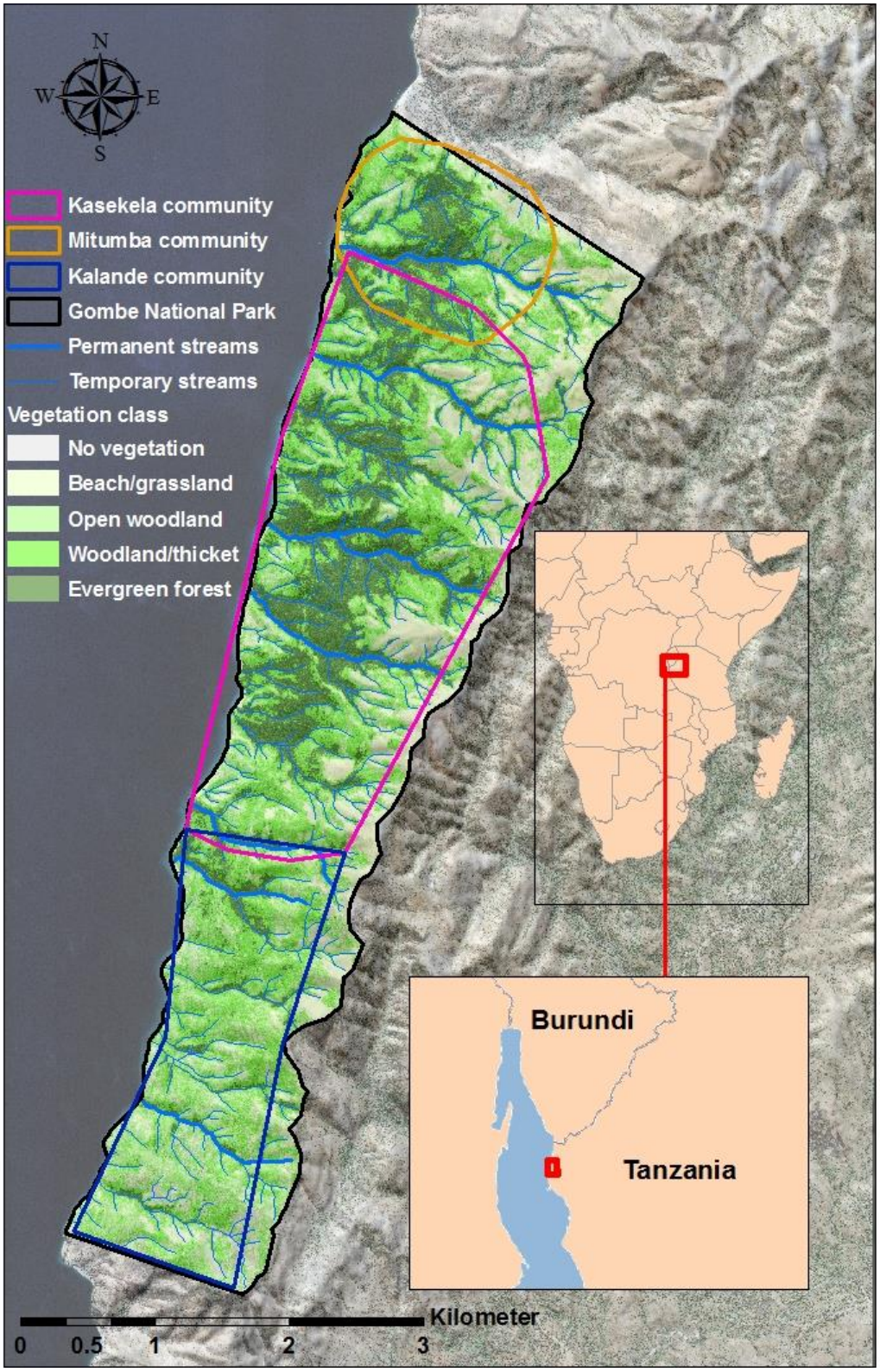


Figure 1. Gombe National Park, with outline of community boundaries and main vegetation types as assessed in 2000. The home range of the Kasekela community and the Mitumba community were drawn from the 99% Minimum Convex Polygon (MCP) extent, and the home range of the Kalande community is an estimated home range drawn based on local knowledge.

CHAPTER 2: CHIMPANZEE VEGETATION FOOD SPECIES DISTRIBUTION MODELING USING MAXENT MODEL

INTRODUCTION

Forest cover of primate habitat influences the habitat's carrying capacity. Carrying capacity is a widely used index for characterizing habitat quality (Begon et al. 1996; Krebs 1994). Further, food availability is a significant determinant of the carrying capacity. Large amount of studies have shown that the primate population is mainly mediated by food availability (Chapman and Chapman 1999; Struhsaker 1973; Altmann et al. 1977; Dittus 1977; Milton 1990). Particularly, in a small spatial scale, food availability was found to be the primary factor influencing primate population (Chapman and Chapman 1999). Other factors that were also found to affect primate population such as hunting, climate and disease (Plumptre 2010) are relatively homogeneous and thus less affect the spatial variation of habitat quality, unlike food availability which can greatly vary because of the heterogeneous vegetation cover in a small spatial scale. Therefore, evaluation of the food availability distribution is important for understanding the spatial variation of habitat quality. Additionally, because studies on primate population showed that availability of different food resources can be of significant difference in indicating habitat quality (Marshall 2010) due to various feeding preference, and given that chimpanzees rely heavily on vegetation foods in their diet, a species-level investigation of the forest cover is especially important for a small scale habitat in order to assess the spatial variation of chimpanzee habitat quality.

Traditional survey techniques have been used to estimate the availability of different food species and their relative abundance throughout Gombe National Park (Murray et al. 2006; Rudicell et al. 2010). These studies extrapolated the data obtained from vegetation surveys to the remaining landscape to estimate spatial variation in the availability of food (Murray et al. 2006; Rudicell et al. 2010). However, the small size of each plot and the diversity of forest composition over such a small area create considerable uncertainty in the estimation of food availability.

Species Distribution Models (SDMs) have been widely used to describe and predict the species distribution in ecology and conservation (Elith and Leathwick 2009). SDMs, such as Generalized Linear Models (GLMs) (Nelder and Baker, 1972), Maximum Entropy Model (MaxEnt) (Phillips et al. 2006), and Random Forests (Liaw and Winener 2002), simulate the distribution of species by linking the occurrence or abundance with the environmental properties. MaxEnt is a popular SDM for understanding current species distribution (Tinoco et al. 2009), forecasting future distribution under climate change scenarios (Elith et al. 2011), predicting invasive species distribution (Ward 2007), and for other landscape ecology applications. MaxEnt has superior predictive performance (Phillips et al. 2006), even with small sample

sizes (Wisz et al 2008), and the additional advantage of requiring only presence species dataset (Phillips et al. 2006). MaxEnt models species distribution by minimizing "the relative entropy" between the environmental data of presence sites and that of background landscape (Elith et al. 2011). It has been used to estimate the distribution and abundance of Chimpanzees in Congo Basin, where the prediction performance indexed with average area under the receiver operating characteristics (ROC) curve (AUC) (Bradley and Andrew 1997) was 0.653 with 10 predicting environmental variables and a spatial resolution of 10 km (Plumptre 2010).

MaxEnt model is often applied in a broad spatial scale, because of lack of high resolution spatial information. Some studies have simulated distribution with MaxEnt in spatial resolution of 30 m (Endries 2011; Amici et al. 2014; Keinath et al. 2010; Laporta et al. 2012). In this study, I explored the use of MaxEnt modeling to predict chimpanzee food species distribution and food availability variation over an unusually small area (35 km2) with a fine spatial resolution of 10 m. Using species distribution modeling method, I aimed to perform a species-level assessment of the spatial variation of vegetation food availability with the extensive field data of vegetation food species presence locality and high resolution spatial data of topographic conditions, vegetation cover condition, and stream locations processed with spatial analysis and remote sensing. Furthermore, I incorporated chimpanzee's feeding preference in the food availability assessment by applying the species-specific feeding proportion generated from long-term feeding observation records of chimpanzee community in Gombe National Park.

METHODS

MaxEnt Models

I used MaxEnt (version 3.3.3; Phillips et al. 2006) to model the distribution of main food species across the park. For a more technical description of the technique, see (Phillips et al. 2006; Phillips and Steven, 2006; Elith et al, 2011).

Vegetation Survey Datasets

Two vegetation surveys were conducted independently in 2003 and 2005-2007 (Wilson et al. 2009), and a phenology survey conducted around 2005. In the vegetation survey conducted in 2003, a total of 89 vegetation survey plots were randomly distributed across the park (Figure 2). The 2005 – 2007 vegetation survey was concentrated within the rainforest regions (Figure 2). Each of the vegetation plots was comprised by a larger 20 by 20 m plot and a nested 5 by 5 m subplot located in the southwest corner of the 20 m plot. Species and count of tress with DBH ≥ 10 cm within the larger plot, and species and count of vine, shrub, and trees with DBH < 10 cm within the subplot were recorded. The phenology survey recorded the vegetation species observed in each survey location. The phenology survey plots were distributed along trails and thus not randomly spread over the area (Figure 2).

Species Presence Dataset

Two vegetation survey datasets were used as basic vegetation species presence datasets for MaxEnt modeling. For species that have less than 40 presence records in the vegetation survey plots, I added the presence records in the phenology survey dataset to the total presence dataset. The phenology dataset was not considered as basic presence dataset because of its less random distribution pattern. Overall, species with more than 10 records of presence, a total of 24 species, were modeled in this study (Table 1). The smallest sample size of presence was 13. Wisz et al.'s study (2008) demonstrated that the AUC of MaxEnt prediction with a sample size of 10 could be larger than 0.65.

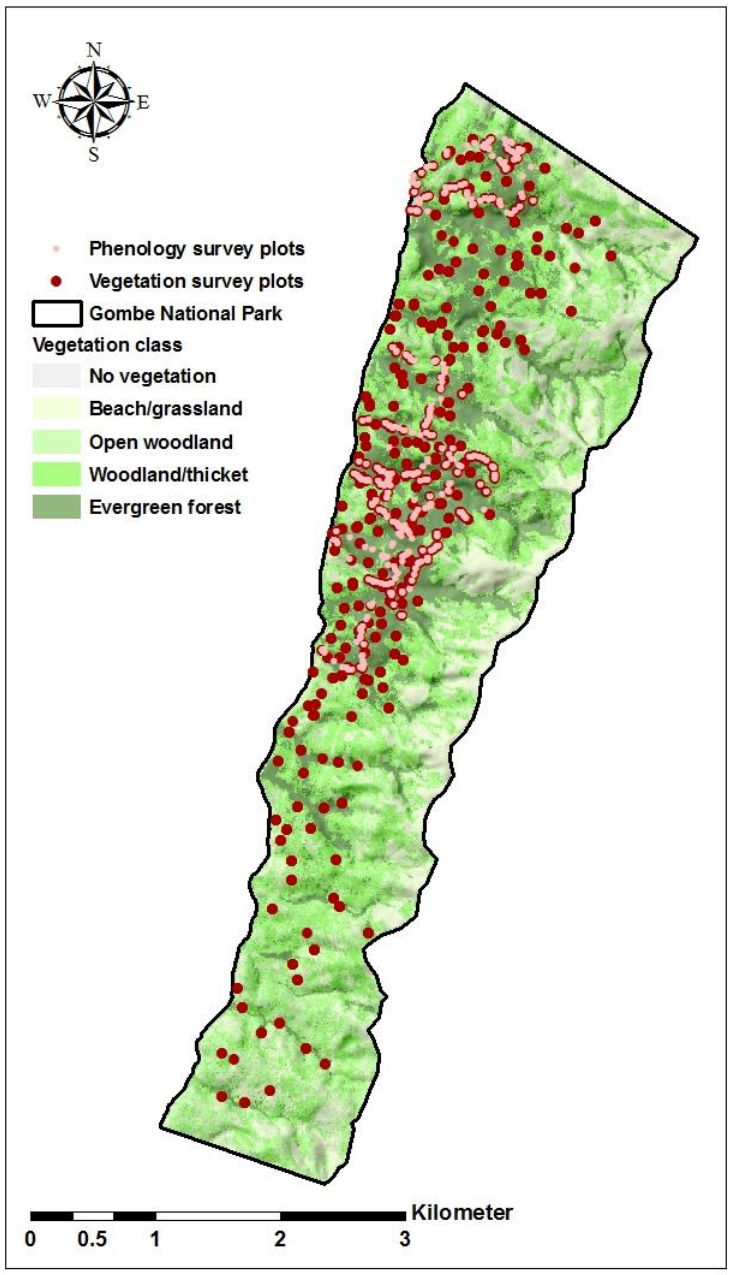


Figure 2. Locations of vegetation surveys and phenology survey.

Table 1. List of vegetation food species modeled with MaxEnt.

Vegetation Food Species	No. vegetation survey presence locations	No. phenology survey presence locations	No. Total presence
Annona senegalensis	36	0	36
Antiaris toxicaria	6	7	13
Antidesma venosum	34	0	34
Baphia capparidifolia	52	0	52
Canthium hispidum / venosum	36	19	55
Diplorhynchus condylocarpon	118	0	118
Elaies guineensis	25	32	57
Ficus sp.	23	37	60
Garcinia huillensis	28	17	45
Grewia platyclada	10	20	30
Harungana madagascariensis	10	15	25
Landolphia lucida	127	0	127
Mellera lobulata / Hypoestes verticillaris	74	0	74
Monanthotaxis poggei	167	0	167
Parinari curatellifolia	64	0	64
Pseudospondias microcarpa	49	0	49
Pterocarpus angolensis	19	29	48
Pterocarpus tinctorius	22	0	22
Saba comorensis var florida	83	0	83
Sabicea orientalis	11	19	30
Salacia leptoclada	29	0	29
Syzigium guineense	8	18	26
Uapaca nitida	3	20	23
Vitex fischeri	63	0	63

Environmental Variables

The environmental factors associated with presence locations were sampled from a set of environmental raster layers, and the background environmental data was sampled randomly from the background landscape where the presence information is unknown, which is considered as “pseudo absence”. I used eleven environmental variables for modeling. A 10 m resolution Digital Elevation Model (DEM) raster layer was developed by Pintea (2007) by digitizing topology maps and used here as elevation variable. From this layer, I derived aspect, slope, curve, Heat Load Index (McCune and Keon 2002), Surface Relief Ratio (Pike and Wilson 1971), and Compound Topographic Index (Gessler et al.

1995; Moore et al. 1993) using the ArcGIS Geomorphometry & Gradient Metrics Toolbox (Evans et al. 2014). In addition, I used two vegetation layers to constrain predictions to vegetated areas. First, I calculated a Normalized Difference Vegetation Index (NDVI) from Landsat Surface Reflectance Climate Data Records (U.S. Geological Survey 2014) collected on June 14, 2005. I resampled this layer from 30 m resolution to 10 m resolution in order to match the other environmental variables. Second, I used a vegetation map generated by Pintea (2006) to limit predictions for each species to an appropriate main vegetation type (i.e., forest species to forests, woodland species to woodlands, etc.). These vegetation classes included: beach/grassland, open woodland, woodland/thicket, and evergreen forest. Lastly, two distance surfaces, distance to the nearest temporal and permanent streams, were generated with ancillary vector data of temporary and permanent streams in the park. Distance to stream reflects the access to water for the plants. The raster display of the 11 environmental variables used as import data for MaxEnt model was presented in Appendix 1.

Soil and climate data are another two important environmental variables commonly used to predict vegetation distribution. However, the spatial resolution of soil data and climate data is very large. Furthermore, the soil information was derived from interpolation instead of on-site field surveys in the Gombe National Park area. Additionally, within such a small area, climate is highly homogeneous. Consequently, I didn't included soil and climate variables in the MaxEnt models in this project.

Sample Bias and Background Selection

Though vegetation survey locations were chosen as randomly as possible across the park, the inaccessibility of higher elevations resulted in the concentration of survey plots in the western parts of the park (Figure 2). To avoid biasing predictions (Elith et al 2011), I applied a mask that restricted the model training algorithm to sample background data only from the region that was covered by vegetation surveys or phenology transects (Figure 2). This mask was included as another environmental variable layer with only two constants, 0 and 1, to differentiate sampled and unsampled areas. The model, trained within the masked area, was then extrapolated and projected to the whole park to predict species distribution. Because some values of environmental variables in the unsampled areas exceeded the range of the training data, I implemented "clamping" in MaxEnt, which sets the exceeding value to the upper or lower limits of the training data (Phillips 2005).

Model Tuning and Model Evaluation

When modeling species distribution, the statistical model might be fitted too close to the training data, resulting on over-fitting and reduced performance when the model is projected to the whole study area. To minimize over-fitting and reduce model variances, I used a cross-validation method: the presence dataset was split into ten random sets, nine of which were used as training data and one that was reserved for model validation. This random sub-setting was repeated 10 times, and outputs of prediction across ten modeling replicates were averaged. The default output of MaxEnt is an averaged "logistic" output of ten duplicate models ranging from 0 to 1. A higher value of logistic output indicates higher habitat suitability of the simulated species (Elith et al. 2011). Therefore, the logistic outputs of MaxEnt are also interpreted as prediction of vegetation habitat suitability. The final logistic output – the vegetation habitat suitability map, is the average of ten models for each species.

The MaxEnt output provides an average cross-validated AUC, meaning the average of AUCs of ten models. However, because MaxEnt calculated AUC with "pseudo absence", the result may be biased by the possibly presence on those locations. Therefore, I recalculated AUC with true absence dataset derived from the vegetation survey dataset. The survey plots where a targeted species was not recorded were taken as absence sites for that species. Further, unlike MaxEnt calculating AUC with the withheld testing presence dataset, I used all the presence and absence records of a species to calculate its AUC. The calculated was performed with the "ROCR" R package (Sing et al. 2005; R Core Team 2014).

To gain vegetation species distribution extent maps of 24 vegetation species, the logistic outputs of 24 species were turned into binary outputs, 0 and 1 (0 represented absence and 1 represented presence), individually, by applying thresholds which respectively maximize the sum of model sensitivity and specificity. Sensitivity is the proportion of actual presence that was predicted as presence by the model, and specificity is the proportion of actual absence that was predicted as absence. Then I calculated the overall accuracy of the tuned binary output from the confusion matrix. The sensitivity and specificity were calculated with the "ROCR" R package (Sing et al., 2005; R Core Team 2014).

Habitat Quality Surface

Though an absolute abundance of vegetation food availability was not estimated in species distribution modeling, the spatial variation of food species distribution extent reflects the variation of food

availability. The number of food species that occur in any given area is unlikely to be an ideal proxy of quality, as species vary considerably in their relative importance as dietary items. Therefore, I created a final chimpanzee habitat quality surface by summing up the distribution extent of 24 food species after weighting each of them with its feeding proportion in chimpanzee's diet (Table 2). The feeding proportion was calculated from the proportion of feeding time on each species from long-term behavioral records of focal individuals in the Kasekela community (Goodall 1986) collected in the decade of 2000. The raster calculation was performed with the raster package in R (Hijmans and Etten 2014). The sum of feeding proportion of 24 modeled vegetation food species in the Kasekela community was about 75.14% of all vegetation food species of that community in 2000 decade.

Table 2. Feeding proportion of 24 vegetation food species.

Vegetation species	**Feeding proportion (%)**
Annona senegalensis	0.039
Antiaris toxicaria	0.254
Antidesma venosum	0.495
Baphia capparidifolia	2.935
Canthium hispidum venosum	0.499
Diplorhynchus condylocarpon	1.488
Elaeis guineensis	5.701
Ficus.sp	5.6
Garcinia huillensis	1.653
Grewia platyclada	1.188
Harungana madagascariensis	1.574
Landolphia lucida	9.19
Mellera lobulata/Hypoestes verticillaris	0.598
Monanthotaxis poggei	6.415
Parinari curatellifolia	11.307
Pseudospondias microcarpa	4.865
Pterocarpus angolensis	0.176
Pterocarpus tinctorius	5.362
Saba comorensis var florida	10.483
Sabicea orientalis	0.545
Salacia leptoclada	0.102
Syzigium guineense	1.147
Uapaca nitida	0.531
Vitex fischeri	2.996

Biological relevance validations of model outputs

To assess whether the Chimpanzee habitat quality surface map captures information that influences chimpanzee feeding behavior, I conducted validation analysis based on linear regression models which made use of extensive behavioral records on feeding behavior (Goodall 1986). Specifically, I tested whether chimpanzees spent more time feeding at sites predicted to have a high overall chimpanzee habitat quality (Model 1 in Table 3), and whether the chimpanzees fed on a specific vegetation species in a location that was predicted to have higher habitat suitability for that vegetation species (Model 3 in Table 3), or have higher average presence at that area (Model 5 in Table 3).

Feeding locations were recorded with an uncertainty of 100 m. In order to account for the 100 m uncertainty of the feeding dataset, I generated 100 by 100 m "fishnets" – a grid of sampling areas across the study area, and calculated the total feeding time within each fishnet, and feeding time on each of the 24 individual vegetation species within each fishnet. I also calculated the mean overall chimpanzee habitat quality value, and the mean suitability and mean presence/absence values of individual vegetation species for each 100 m fishnet. The presence/absence values were then converted from 1/0 binary factor to a continuous 0-1 scale numeric factor. To be consistent with the source data used for our MaxEnt modeling and overall chimpanzee habitat quality surface calculation, I only used feeding data collected in the decade of 2000 for the validation of overall habitat quality surface. However, because feeding records on individual vegetation species are limited in the decade of 2000, the feeding records collected from 1973 to 2013 were used for validation of vegetation habitat suitability maps and presence/absence maps, accepting the assumption that the distribution of those 24 modeled vegetation species didn't changed significantly in the past 40 years.

The Kasekela chimpanzee community was used as the study object in this case given that it was the largest and well-studied chimpanzee community in Gombe National Park. In order to reduce spatial autocorrelation effects, 25% of the fishnets, a total of 431 of plots, within the home range of the Kasekela community were sampled randomly for linear regression model tests (Figure 3).

Further, given that the feeding time recorded in a large proportion of samples was zero, meaning that no feeding records were found within those sample plots, three extra linear regression models were run with only random samples that had feeding times larger than zero, for the purpose of removing impacts of zero feeding time records (Model 2, 4, and 6 in Table 3). The sample size for this linear regression

model between overall chimpanzee habitat quality and nonzero feeding time was 317, given that 114 out of 431 random samples had no feeding records in the decade of 2000. The sample size of each species is shown in Table 4.

Additionally, to approach normal distribution of the feeding time as a response variable in linear regressions, I log-transformed the feeding time after adding 0.01 for the purpose of avoiding zero value.

Table 3. Linear regression models for biological relevance validations.

Model ID	Response variable	Predictor
1	Feeding time	Overall Chimpanzee habitat quality
2	Nonzero feeding time	
3	Feeding time	Vegetation species habitat suitability
4	Nonzero feeding time	
5	Feeding time	Vegetation species mean presence
6	Nonzero feeding time	

Table 4. Nonzero feeding time sample size of 24 species.

Vegetation Food Species	Nonzero feeding time Sample size
Annona senegalensis	32
Antiaris toxicaria	22
Antidesma venosum	75
Baphia capparidifolia	188
Canthium hispidum / venosum	61
Diplorhynchus condylocarpon	146
Elaies guineensis	120
Ficus sp.	237
Garcinia huillensis	94
Grewia platyclada	68
Harungana madagascariensis	86
Landolphia lucida	192
Mellera lobulata / Hypoestes verticillaris	105
Monanthotaxis poggei	225
Parinari curatellifolia	249
Pseudospondias microcarpa	162
Pterocarpus angolensis	74
Pterocarpus tinctorius	227
Saba comorensis var florida	208
Sabicea orientalis	73

Salacia leptoclada	36
Syzigium guineense	55
Uapaca nitida	59
Vitex fischeri	78

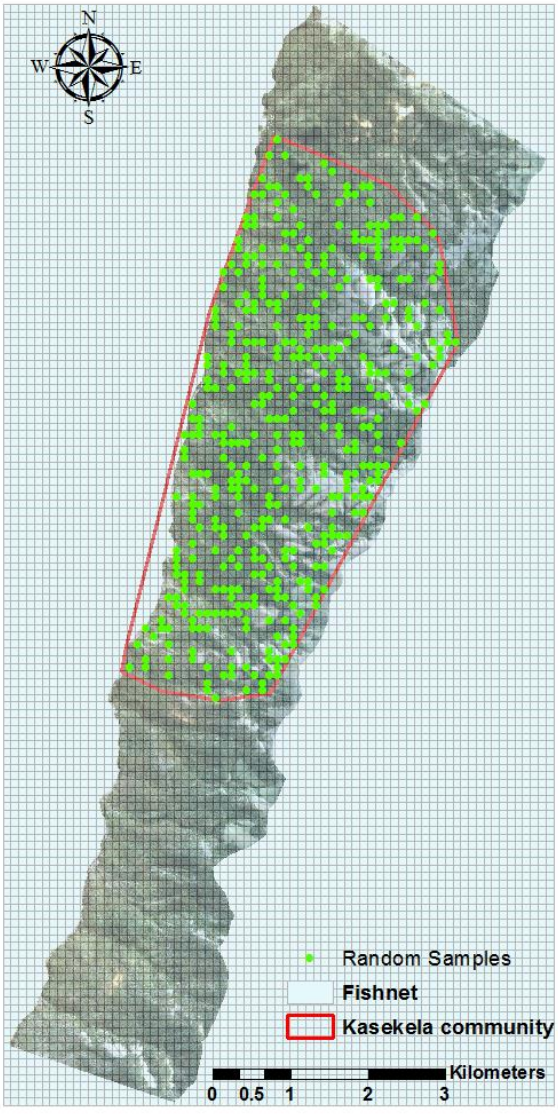


Figure 3. Random samples within Kasekela community for linear regression models. The random sample plots were the center of 25% random 100 by 100 m fishnets within the Kasekela community.

RESULTS

Vegetation Species Habitat Suitability and Distribution Extent

Overall, a rich spatial dataset was created consisting of: the vegetation habitat suitability maps, the logistic outputs of MaxEnt, and the vegetation distribution extent maps of 24 vegetation food species, the tuned binary outputs, and a chimpanzee habitat quality map (Figure 4). Appendix 2 is a complete dataset of vegetation habitat suitability maps and distribution extent maps.

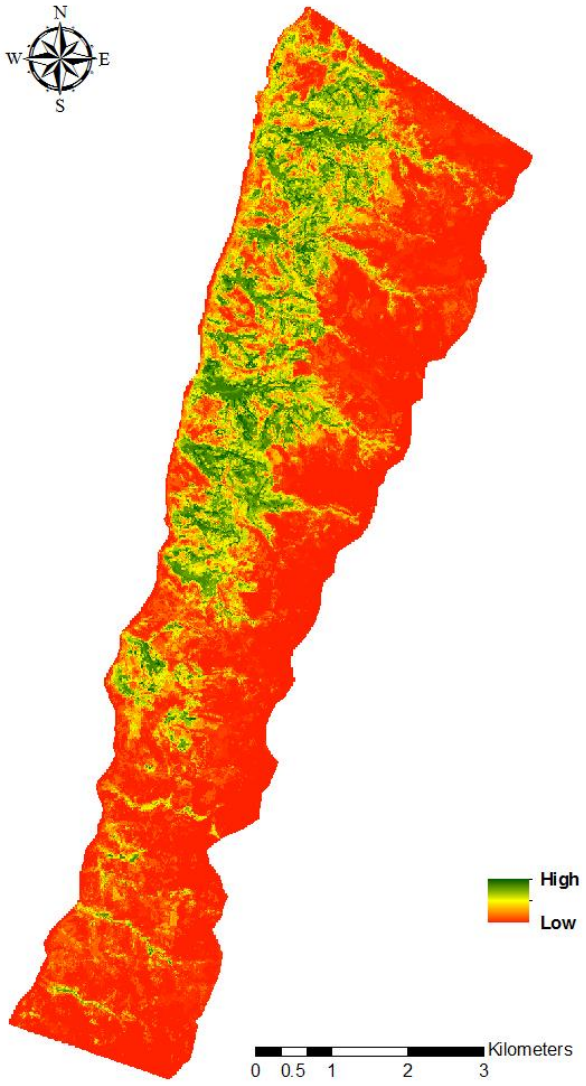


Figure 4. Chimpanzee habitat quality surface.

The AUCs of MaxEnt logistic outputs and overall accuracy assessment results of binary output calculated with true absence were listed in Table 5. The average AUC of the logistic outputs is 0.85, and the average overall accuracy of the tuned binary outputs is 0.78.

Table 5. Model evaluation of MaxEnt logistic output and binary output.

Species	AUC	Overall Accuracy	#Presence	#Absence
Annona senegalensis	0.804	0.744	36	202
Antiaris toxicaria	0.973	0.963	13	232
Antidesma venosum	0.688	0.706	34	204
Baphia capparidifolia	0.882	0.807	52	186
Canthium hispidum / venosum	0.712	0.716	55	202
Diplorhynchus condylocarpon	0.88	0.803	118	120
Elaeis guineensis	0.863	0.703	57	213
Ficus sp.	0.882	0.844	60	215
Garcinia huillensis	0.766	0.576	45	210
Grewia platyclada	0.896	0.864	30	228
Harungana madagascariensis	0.834	0.672	25	228
Landolphia lucida	0.862	0.807	127	111
Mellera lobulata / Hypoestes verticillaris	0.835	0.71	74	164
Monanthotaxis poggei	0.865	0.861	167	71
Parinari curatellifolia	0.753	0.752	64	174
Pseudospondias microcarpa	0.917	0.866	49	189
Pterocarpus angolensis	0.855	0.723	48	219
Pterocarpus tinctorius	0.866	0.769	22	216
Saba comorensis var florida	0.779	0.714	83	155
Sabicea orientalis	0.917	0.802	30	227
Salacia leptoclada	0.841	0.761	29	209
Syzigium guineense	0.891	0.887	26	230
Uapaca nitida	0.992	0.977	23	235
Vitex fischeri	0.771	0.71	63	175

Many vegetation food species were predicted to be located in the west part of the park where elevation was relatively low, streams were abundant, and rainforests were located, including Annona senegalensis (Figure 5, a), Antiaris toxicaria, Baphia capparidifolia, Canthium hispidum/venosumc, Elaies guineensis,

Ficus.sp, Garcinia huillensis, Grewia platyclada, Landolphia lucida, Mellera lobulated/Hyposestes verticllaris, Monanthotaxis poggei, Pseudospondias microcarpa, and Saba comrensis var florida.

Particularly, distributions of some species were strongly driven by stream locations. Baphia capparidifolia (Figure 5, b), Elaies guineensis, Ficus.sp, Garcinia huillensis, Grewia platyclada, Mellera lobulata/Hypoestes) verticillaris, Pseudospondias microcarpa, Salacia leptoclada, and Vitex fischeri were all predicted to be located along valleys and streams.

Some species were predicted to have relatively smaller variation of vegetation habitat suitability across the park, including Antidesma venosum (Figure 5, c), Diplorhynchus condylocarpon, Harungana madagascariensis, and Parinari curatellifolla.

Particularly, five vegetation species were predicted to present distinctive distribution pattern with other species: Pterocarpus angolensis was predicted to be spreading out the park except for the rainforest area (Figure 8); Sabicea orientali, Syzigium guineense, and Uapaca nitida prefer middle-west of the park, while Sabicea orientalis and Syzigium guineense have very similar distribution extent; the distribution pattern of Pterocarpus tinctorius is heavily influenced by the topographic factors such as aspect.

The vegetation suitability maps and distribution extent maps of 24 vegetation species are shown in Appendix 2.

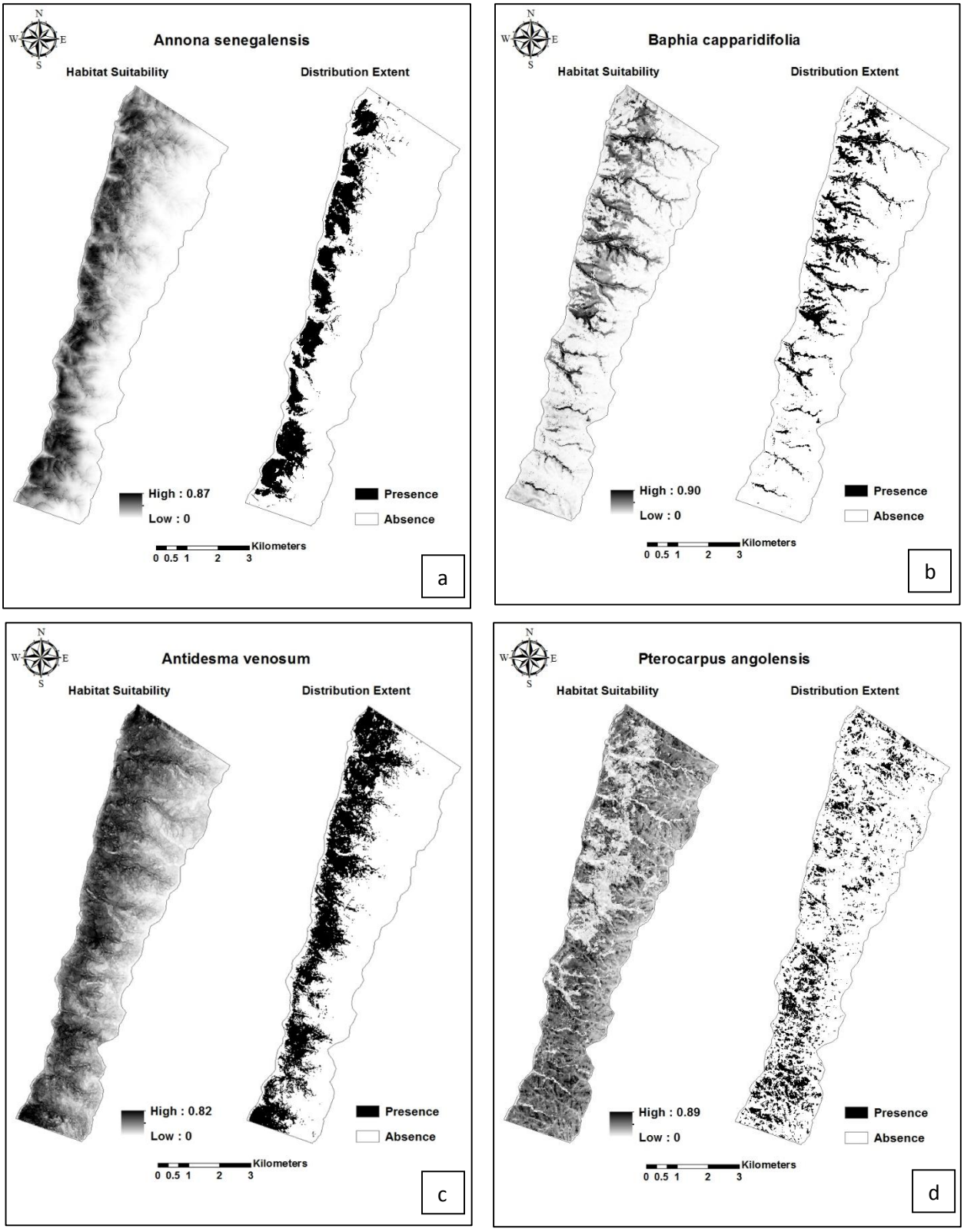


Figure 5. Examples of four vegetation habitat suitability and distribution extent maps.

Biological relevance validations

Overall the chimpanzee habitat quality surface is a significant predictor for the log-transformed chimpanzee feeding time (p-value < 0.05) with an adjusted R2 of 0.19. While removing the samples with zero feeding time, the p-value of overall chimpanzee habitat quality surface became 0.05, and adjusted R2 was smaller than 0.01.

Except for Antiaris toxicaria and Pterocarpus angolensis, vegetation habitat suitability and the mean presence value are both significant predictors to the feeding time on all other vegetation species (Table 6 and Table 8). After removing the samples with zero feeding time, vegetation habitat suitability remained significant for 12 vegetation species, while mean presence remained significant for 13 vegetation species (Table 7 and Table 9).

Table 6. Statistics of linear regression models between vegetation habitat suitability and feeding time on 24 vegetation species.

Vegetation Food Species	**Adjusted R^2**	**Suitability P-value**	**Suitability Coefficient**
Annona senegalensis	0.08	< 0.05	3.01
Antiaris toxicaria	0.00	0.27	0.80
Antidesma venosum	0.15	< 0.05	7.68
Baphia capparidifolia	0.27	< 0.05	14.09
Canthium hispidum / venosum	0.05	< 0.05	3.82
Diplorhynchus condylocarpon	0.11	< 0.05	9.59
Elaies guineensis	0.30	< 0.05	14.83
Ficus sp.	0.34	< 0.05	19.55
Garcinia huillensis	0.17	< 0.05	7.52
Grewia platyclada	0.16	< 0.05	9.98
Harungana madagascariensis	0.07	< 0.05	6.13
Landolphia lucida	0.37	< 0.05	15.51
Mellera lobulata / Hypoestes verticillaris	0.20	< 0.05	9.30
Monanthotaxis poggei	0.37	< 0.05	15.84
Parinari curatellifolia	0.22	< 0.05	14.48
Pseudospondias microcarpa	0.31	< 0.05	16.17
Pterocarpus angolensis	0.00	0.20	1.68
Pterocarpus tinctorius	0.07	< 0.05	8.78
Saba comorensis var florida	0.48	< 0.05	18.04
Sabicea orientalis	0.03	< 0.05	3.40
Salacia leptoclada	0.05	< 0.05	3.04
Syzigium guineense	0.09	< 0.05	4.83
Uapaca nitida	0.11	< 0.05	7.78
Vitex fischeri	0.10	< 0.05	6.82

Table 7. Statistics of linear regression models between vegetation habitat suitability and nonzero feeding time on 24 vegetation species.

Vegetation Food Species	Adjusted R^2	Suitability P-value	Suitability Coefficient	Sample size
Annona senegalensis	0.03	0.16	2.46	32
Antiaris toxicaria	0.03	0.21	-1.72	22
Antidesma venosum	-0.01	0.99	-0.02	75
Baphia capparidifolia	0.06	< 0.05	1.91	188
Canthium hispidum / venosum	0.02	0.15	-1.36	61
Diplorhynchus condylocarpon	0.01	0.12	1.09	146
Elaies guineensis	0.21	< 0.05	4.06	120
Ficus sp.	0.18	< 0.05	3.89	237
Garcinia huillensis	0.00	0.31	1.10	94
Grewia platyclada	0.08	< 0.05	2.17	68
Harungana madagascariensis	0.07	< 0.05	2.59	86
Landolphia lucida	0.09	< 0.05	2.82	192
Mellera lobulata / Hypoestes verticillaris	0.20	< 0.05	3.16	105
Monanthotaxis poggei	0.10	< 0.05	2.79	225
Parinari curatellifolia	0.09	< 0.05	3.10	249
Pseudospondias microcarpa	0.18	< 0.05	3.66	162
Pterocarpus angolensis	0.00	0.27	0.97	74
Pterocarpus tinctorius	0.00	0.29	0.63	227
Saba comorensis var florida	0.20	< 0.05	3.99	208
Sabicea orientalis	0.00	0.40	0.62	73
Salacia leptoclada	-0.01	0.39	0.98	36
Syzigium guineense	0.00	0.39	0.64	55
Uapaca nitida	0.11	< 0.05	1.75	59
Vitex fischeri	0.03	0.06	1.87	78

Table 8. Statistics of linear regression models between mean presence and feeding time on 24 vegetation species.

Vegetation Food Species	Adjusted R^2	Mean Presence P-value	Mean Presence Coefficient
Annona senegalensis	0.04	< 0.05	1.33
Antiaris toxicaria	0.00	0.93	0.07
Antidesma venosum	0.10	< 0.05	3.97
Baphia capparidifolia	0.20	< 0.05	6.94
Canthium hispidum / venosum	0.02	< 0.05	1.78
Diplorhynchus condylocarpon	0.08	< 0.05	4.81
Elaies guineensis	0.25	< 0.05	6.32
Ficus sp.	0.17	< 0.05	10.10
Garcinia huillensis	0.14	< 0.05	3.57
Grewia platyclada	0.10	< 0.05	4.92

Harungana madagascariensis	0.06	< 0.05	2.56
Landolphia lucida	0.36	< 0.05	8.24
Mellera lobulata / Hypoestes verticillaris	0.17	< 0.05	4.30
Monanthotaxis poggei	0.37	< 0.05	7.29
Parinari curatellifolia	0.11	< 0.05	5.63
Pseudospondias microcarpa	0.25	< 0.05	10.26
Pterocarpus angolensis	0.00	0.09	0.98
Pterocarpus tinctorius	0.09	< 0.05	4.89
Saba comorensis var florida	0.34	< 0.05	9.30
Sabicea orientalis	0.03	< 0.05	1.66
Salacia leptoclada	0.02	< 0.05	1.26
Syzigium guineense	0.07	< 0.05	4.05
Uapaca nitida	0.09	< 0.05	5.41
Vitex fischeri	0.09	< 0.05	3.45

Table 9. Statistics of linear regression models between mean presence and nonzero feeding time on 24 vegetation species.

Vegetation Food Species	**Adjusted R^2**	**Mean Presence P-value**	**Mean Presence Coefficient**	**Sample size**
Annona senegalensis	0.03	0.18	0.81	32
Antiaris toxicaria	-0.03	0.55	-0.94	22
Antidesma venosum	-0.01	0.93	-0.04	75
Baphia capparidifolia	0.06	< 0.05	0.97	188
Canthium hispidum / venosum	0.00	0.31	-0.54	61
Diplorhynchus condylocarpon	0.02	< 0.05	0.66	146
Elaies guineensis	0.16	< 0.05	1.85	120
Ficus sp.	0.14	< 0.05	2.22	237
Garcinia huillensis	0.00	0.40	0.37	94
Grewia platyclada	0.06	< 0.05	1.05	68
Harungana madagascariensis	0.04	< 0.05	0.89	86
Landolphia lucida	0.08	< 0.05	1.19	192
Mellera lobulata / Hypoestes verticillaris	0.19	< 0.05	1.41	105
Monanthotaxis poggei	0.09	< 0.05	1.13	225
Parinari curatellifolia	0.04	< 0.05	0.99	249
Pseudospondias microcarpa	0.17	< 0.05	2.13	162
Pterocarpus angolensis	0.00	0.36	0.40	74
Pterocarpus tinctorius	0.00	0.34	0.28	227
Saba comorensis var florida	0.21	< 0.05	1.78	208
Sabicea orientalis	-0.01	0.46	0.27	73
Salacia leptoclada	-0.01	0.49	0.38	36
Syzigium guineense	-0.02	0.95	0.03	55
Uapaca nitida	0.11	< 0.05	1.28	59
Vitex fischeri	0.03	0.08	0.77	78

DISCUSSION

The overall prediction performance of the MaxEnt model is relatively satisfactory, though the performance varied across different vegetation species. While no apparent correlations between prediction accuracy and presence sample size were found, most of the vegetation species that had low habitat suitability prediction accuracy presented a common habitat suitability and distribution extent prediction pattern: these species were predicted to be wide-spread in the park, comparing to other species with higher prediction accuracy and more concentrated predicted home range. For example, five vegetation species with worst prediction accuracy, Antidesma venosum, Parinari curatellifolia, Pterocarpus angolensis, Harungana madagascariensis, and Pterocarpus tinctorius, were all predicted to be found in a wide range of the park (Appendix 2). Therefore, a likely reason that explains the poor prediction performance for these vegetation species is that it is more difficult to capture the environmental variable differences between habitat and non-habitat given their wide ecological niches.

The biological validations of chimpanzee habitat quality surface, vegetation habitat suitability, and vegetation mean presence by linear models correlating with feeding time also indicate that the models were fairly successful in predicting the spatial variation of vegetation food availability for chimpanzee feeding. However, the significance of vegetation species distributions in predicting chimpanzee feeding time decreased after removing records with zero feeding time. Two potential reasons may explain the decrease of significance: the model predictions were more capable in predicting the presence of vegetation food than in predicting the abundance of vegetation food with the vegetation habitat suitability outputs; or the assumption, chimpanzee spend longer feeding on a site with higher food abundance, is violated, as if ignores other factors which may affect chimpanzee feeding behavior, such as competitions between individuals and communities.

Despite the relatively weaker capacity in capturing the distribution pattern of species with wide ecological niche and unknown capability in indicating abundance of vegetation individuals, it is still shown that MaxEnt is effective in capturing spatial variation of vegetation distribution in a small area (35 km2) with fine spatial resolution of 10 m, which has seldom been investigated in previous studies.

The species distribution modeling method provided a species-level investigation of the vegetation food availability for chimpanzees in Gombe National Park. Furthermore, in contrast to traditional method of food abundance evaluation which extrapolates the vegetation cover of survey plots evenly to the whole

park, species distribution modeling reflects more accurate spatial variation of the vegetation food availability. Area with higher abundance of important vegetation food species was assumed to be a habitat with high quality for the chimpanzees in the park with a high carrying capacity. The method that overlaid the feeding-time-weighted vegetation species distributions generated an overall chimpanzee habitat quality surface reflecting food availability. The consistency between the habitat area with high quality predicted by this surface and the home range of chimpanzee communities (Figure 2), as well as the significant correlation between overall chimpanzee habitat quality surface values and the chimpanzee feeding demonstrated the credibility of the habitat quality modeling method proposed in this project.

Limitations of this project have been observed, and further studies that would adopt methods proposed in this project should focus on addressing these limitations: 1) Soil was not used as one environmental variable in the MaxEnt models of this project. In further studies, accurate and high resolution soil data should be included if it is available, because soil layer has significant influence on plants. 2) Chimpanzee's feeding proportion on vegetation species was adopted as proxy of chimpanzee's feeding preference. However, this assumption may be biased by the fact that chimpanzees may not have access to their most preferred food species, possibly because of overfeeding, and therefore had to feed longer on less preferred but more abundant food species. An alternative method to elicit chimpanzee feeding preference could be based on expert knowledge or results of other chimpanzee feeding behavior studies. 3) Since feeding data on individual vegetation food species was limited, this project used feeding data collected in 40 years to test the correlation between chimpanzee feeding time and the habitat quality generated based on 2005 image. However, vegetation cover was very likely to change over 40 years in Gombe National Park, even though in a small scale. A more precise assessment of the correlation would be only using data close to 2005.

CONCLUSION

Although MaxEnt tends to perform better in species distribution modeling for species that have a narrow ecological niche, the overall high accuracy of MaxEnt model prediction and the significant correlation between MaxEnt modeled vegetation habitat quality and chimpanzee habitat quality shows that MaxEnt was capable in predicting spatial variation of vegetation species distribution in a small area with a fine resolution of 10 m. Barriers impeding the application of MaxEnt model at high spatial resolution include the large geographical uncertainties of field survey location, and the limited availability of high resolution data of the study area, such soil data source (unavailable for Gombe National Park). If limitations are addressed properly, the vegetation food availability and chimpanzee habitat quality evaluation methods proposed in this project could be used in further vegetation and primate studies as well as practical conservation projects.

CHAPTER 3: VEGETATION COVER CLASSIFICATION FOR GOMBE NATIONAL PARK

INTRODUCTION

Remote sensing technique is an alternative approach to characterize habitat quality, which have been widely used in the recent decades especially for vegetation cover evaluation. In contrast to field survey method which simply extrapolates survey plot data to the whole landscape and species distribution modeling method which links presence with environmental conditions, remote sensing technique relies on the distinction of forest cover characteristics through electromagnetic radiation detection. For example, Peck et al. (2010) applied remote sensing technique to delineate forest larger than a species-specific cover proportion criteria which make the forest capable to hold primate population. In a previous study of land cover changes in Gombe National Park, Pintea (2007) evaluated the forest cover change using 1972 Landsat Multispectral Scanner (Landsat MSS) image and 1999 Landsat Enhanced Thematic Mapper Plus (ETM+) image. Additionally, his supervised vegetation classification using a 2000 IKONOS image classified the whole landscape into evergreen forest, thicket woodland, open woodland, and beach/grassland. The result showed that the central and northern parts of the park had higher canopy cover than the rest.

Despite of its superior capacity of quick identification of vegetation cover in a broad scale, remote sensing is usually limited to general distinction of forest and non-forest or broad vegetation types rather than species-level classification. As discussed above, different vegetation species could be of very different importance to habitat quality indication. Therefore, the lack of species information holds back the remote sensing technique in the application of habitat quality evaluation.

To address this limitation, my goal of the second section of this project was to develop an innovative method of producing vegetation cover map which incorporates vegetation species composition information in each vegetation cover class, and to generate such a biologically meaningful vegetation cover map for Gombe National Park.

In traditional supervised classification, the vegetation class schema design depends on prerequisite knowledge of the local landscape, which means that the researcher has been informed of how many vegetation classes exist in the landscape before performing classification, and aided by sample data of vegetation cover for each class. In the case of unsupervised classification, the categories of land cover are created from the spectral clustering and separating characteristics of the image pixels, which is then combined with local landscape knowledge to define the auto-clustered groups into vegetation classes.

Both of these two methods requires prerequisite knowledge of local vegetation cover condition. Further, none of these vegetation class schema incorporate the vegetation species composition information. Moreover, given that the number of existing vegetation types and the species composition in typical vegetation communities at Gombe National Park are unknown, traditional class schema methods design methods were not applicable in this project.

Cluster analysis is a conventional ecological data mining technique which aims to identify the grouping patterns of a dataset (Kaufman et al. 2009). It's a common technique for making classification based on characteristic similarity and dissimilarity between data points (Romesburg 2004). Knollová et al. (2004) for example, classified vegetation associations based on species composition using cluster analysis and compared the result with geographically delimited association. To produce biologically meaningful vegetation cover map, the cluster analysis technique is useful in generating vegetation class groups based on vegetation species composition for this project.

METHODS

The vegetation classification consists of four major processes: class schema design, supervised classification, post processing, and accuracy assessment.

Cluster Analysis

I first performed a cluster analysis based on the species composition information obtained from vegetation survey plots to group these plots into different vegetation classes, and then used these classified plots as training samples for a supervised classification. This approach was based on the hypothesis that plots similar in species composition were also similar in surface spectral characteristic, because the reflected electromagnetic radiation is predominately affected by vegetation composition. The dataset of vegetation surveys conducted in 2003, with a total of 89 vegetation plots, were used as training and validation data for the vegetation classification. This vegetation survey was more randomly distributed across the park and sampled different vegetation cover types evenly.

The dissimilarity of species composition of vegetation in the survey plots was indexed with the "Bray-Curtis" dissimilarity index (Bray and Curtis 1957). It was calculated using the "ecodist" R package (Goslee and Urban 2007). Further, the "Bray-Curtis" dissimilarity distances was modified in order to deal with the problem that dissimilarities saturate to the value of 1 when large proportion samples have no common in species composition. Therefore, the distance of 1, meaning sample plots have no common in species composition, were replaced with the shortest paths via other samples. The modification was done using the "vegan" R package (Oksanen et al. 2014). Species that appeared in less than 5% of plots were removed to eliminate noise. In the end, 27 vegetation species in 89 sample plots were used in the cluster analysis (Appendix 3). Furthermore, although cluster analysis provides the statistically optimal clusters which maximize the within-cluster homogeneity, in order to evaluate the proper cluster level, namely the number of the clusters to be retained, a Mantel's test of distance matrix was conducted to compare the within-cluster heterogeneity relative to among-cluster heterogeneity for cluster levels from 2 to 10. In the Mantel's test, elements were assigned a value of 0 if two samples were in the same group and otherwise were assigned to 1. A higher Mantel's test statistics implies a high among-group heterogeneity relative to within-group heterogeneity. The data analysis was performed on R Program (R Core Team 2014).

Supervised classification

Supervised classification is a traditional pixel-based remote sensing classification technique. It generates spectral clusters of land cover classes based on the training samples provided by the researcher, and then classified the rest of the pixels into established classes using certain classification techniques and decision rules such as Minimum Distance classification or Maximum Likelihood Classification (MLC).

The remote sensing image used for supervised classification in this project was a QuickBird image collected in 2009. The spatial resolution of the image is 2.4 m. The four spectral bands of the image are Blue (450-520 μm), Green (520-600 μm), Red (630-690 μm), and Near Infrared (760-900 μm). Atmospheric correction was done and the image was processed into surface reflectance by the image provider.

The grouping of 89 vegetation plots for classification was based on the cluster analysis results. The plots grouped in the same cluster were divided into two sets of samples equally, one of them was used as training samples and the other was used as accuracy assessment samples. The vegetation classes were named after the dominant species, which were defined as the most abundant species occurring in the plot samples of that group. For example, a group of which the most abundant species is Landolphia lucida will be named as Landolphia lucida forest. Additionally, three more classes were created: bare land, cloud, shadow. Since the bare land of the park was not sampled in the vegetation survey, training samples of bare land was created manually, and likewise, training samples were also generated for clouded area and the cloud shadow area.

The classification technique applied in this study is Maximum Likelihood Classification (MLC). MLC is a parametric classification method which assumes that the pixel values in each class are normally distributed, and assign the pixels to a class where they reach their largest probability of truly being in that class (Richards 1999). MLC is superior to many other classification techniques in terms of its capability to incorporate the variability information within one class. The classification was performed with the “Maximum Likelihood Classification” function in ENVI.

Post Processing

In the case of high resolution image classification, the classification output image is particularly mottled because a high level of detail was captured by the image. For example, tree gaps in open shrub land and

shadows of big trees, may be captured by the image and incorrectly classified into bare land and shadow. Therefore, in order to improve spatial coherency, the classification output was sieved and clumped to remove small isolated parcels. Assuming that in the park, vegetation cover is homogeneous within 10 m, the minimum group threshold for sieving was 17, which means that the parcels that are smaller than 16 pixels, 92.16 m2, were reclassify to “unclassified”. Further, classes were clumped with an operator size of 5 by 5 pixels, equaling 12 m by 12 m.

Though the “unclassified”, “cloud”, and “shadow” classes were created during the classification process, they were not useful as final classes. Therefore, parcels belonging to these three classes should be reclassified into other vegetation classes. For this purpose, I invented the “buffer zone majority” method which classifies isolated parcels to the majority of their 10 m buffer zone vegetation classes. I assumed that it was highly likely that parcels would have the same vegetation cover within a 10 meter adjacent surrounding.

This method worked well for small unclassified parcels, cloud and shadow areas where the cloud or shadow edge was clear. However, it is less effective for cloud and shadow areas where the cloud and shadow edge was fuzzy. At the edge of the cloud and shadow, if the clouds were thin and the vegetation was not completely covered, the landscape was not classified into cloud or shadow but into another vegetation class. However, the classified vegetation class was not reliable because the electromagnetic radiation reflectance was contaminated by the cloud or shadow. By using the “buffer zone majority” method, the buffer zone, which was the edge of the cloud and shadow, didn’t provide the correct vegetation class information. Therefore, further cloud and shadow correction was necessary. After applying the “buffer zone majority” method, those problematic parcels were then manually identified and reclassified to the proper vegetation classes, in most case the surrounding vegetation class, by interpreting the QuickBird image. The “buffer zone majority” method was operated in ArcGIS 10.0 (Desktop 2011).

Accuracy Assessment

Half of the vegetation field plots were held back as accuracy assessment samples. Additionally, a set of bare land plots were manually created for accuracy assessment of the bare land class. The ratio of total area of bare land accuracy assessment plots to the total area of vegetation class accuracy assessment plots is equal to the areal ratio of vegetated area and bare lands in the park.

An overall accuracy, as well as the producer accuracy and user accuracy were calculated from the confusion matrix generated from “Confusion Matrix Tool” in ENVI. “Producer accuracy” means that the proportion of pixels that belongs to a particular class is accurately classified into that class in the vegetation map, and “user accuracy” means the proportion of pixels that was classified into a particular interested class truly belongs to that class on the ground.

RESULTS

Cluster analysis

By comparing the Mantel's test statistics, a cluster level of 7 has the highest Mantel's test statistics (Figure 6). In other word, statistically 7 groups maximize the among-group heterogeneity relative to within-group heterogeneity according to the Mantel's test. However, the Mantel's test statistics of 7-10 is very close (Figure 6). In order to retain more vegetation classes, I adopted the 10 cluster level instead of the 7 cluster level. Among the groups of 10-cluster level, group 4 only has one vegetation plot sample. This group was taken off to reduce noise. Further, group 4, group 9, and group 10 only have two sample plots in each. After evaluating the pairwise spectral separability statistics (Appendix 4), the separability statistics between group 10 and group 8, between group 9 and group 3, and between group 5 and group 7 are smaller than 1, therefore, group 10 was merged to group 8, group 9 was merge to group 3, and group 5 was merge to group7. The merged classes were named by the dominant species of the two original groups. At the end of this process a total of 6 vegetation classes were maintained.

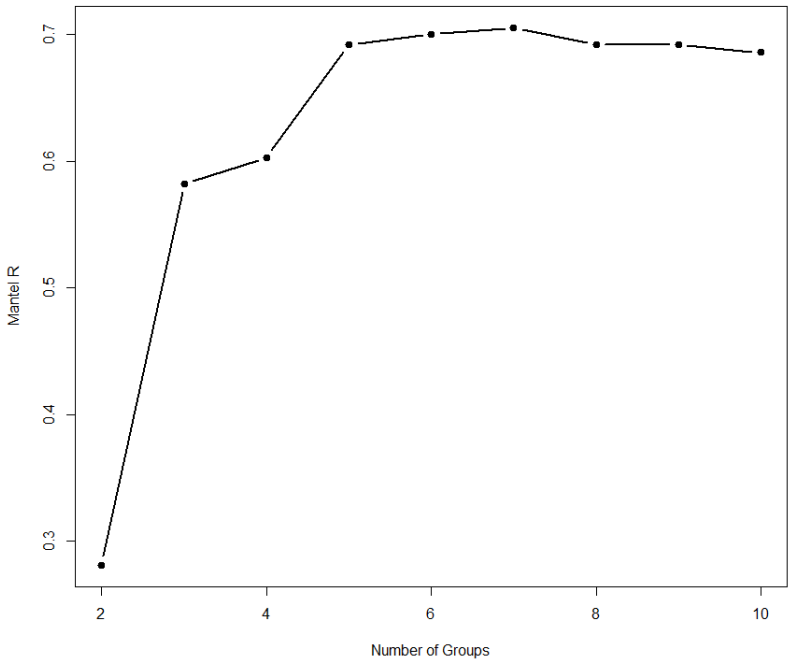


Figure 6. Mantel's test statistics of cluster level 2-10. X-axes is the number of clusters to retain, and the Y-axes is the Mantel's test statistics.

Supervised classification

The raw supervised classification output image was mottled with isolated pixel-sized vegetation classes (Figure 7, a). After sieving and clumping, the image was more coherent (Figure 7, b). After applying the "buffer zone majority" method, the "unclassified", "cloud", and "shadow" parcels were classified into the majority of surrounding vegetation classes. Though the "buffer zone majority" method was effective in reclassifying the unclassified pixels resulting from sieve and clump operation, tree shadow parcels, and big cloud and shadow parcels with clear edge cut, it didn't work effectively for several clouded areas and shadow areas where the edge was not clear. With the "buffer zone majority" method, these areas were reclassified into the vegetation class that was the majority of their surroundings, which were not ground-truth because the buffer zone classification was also affected by the cloud and shadow edge. Three major areas were spotted to be jeopardized by this problem (Figure 8). The problematic parcels, including the cloud and shadow area and edge affected were manually reclassified to the most possible vegetation class, according to the high resolution QuickBird image and the vegetation classes near them. The final vegetation classification map is shown in Figure 9.

The overall accuracy of the final classification is 62.39%, and the Kappa Coefficient is 0.53. The producer accuracy and user accuracies vary across different vegetation classes (Table 10). "Bare land" has the highest accuracy. Spectral characteristics of bare land are relatively monotonous and also clearly separated from that of vegetation areas, making the classification task is easier. The "Landdophia lucida rainforest" class also has a high accuracy, especially the user accuracy. Both "Annona senegalensis/Brachystegia spp. Shrubland" class and "Diplorhynchus condylocarpon shrubland" class have higher value of producer accuracy than of user accuracy. Three vegetation classes, "Uapaca nitida/Harungana madagascariensis shrubland", "Annona senegalensis/ Parinari curatellifolia shrubland", and "Mellera lobulata/Hypoestes verticillaris rainforest" have low accuracy.

Table 10. Supervised classification accuracy.

Class	Producer accuracy	User accuracy
Bare land	100	93.89
Landolphia lucida rainforest	65.63	82.77
Annona senegalensis/Brachystegia spp. shrubland	62.50	44.05
Diplorhynchus condylocarpon shrubland	61.89	53.75
Annona senegalensis/ Parinari curatellifolia shrubland	41.88	14.12
Uapaca nitida/Harungana madagascariensis shrubland	11.81	5.26
Mellera lobulata/Hypoestes verticillaris rainforest	4.26	54.00

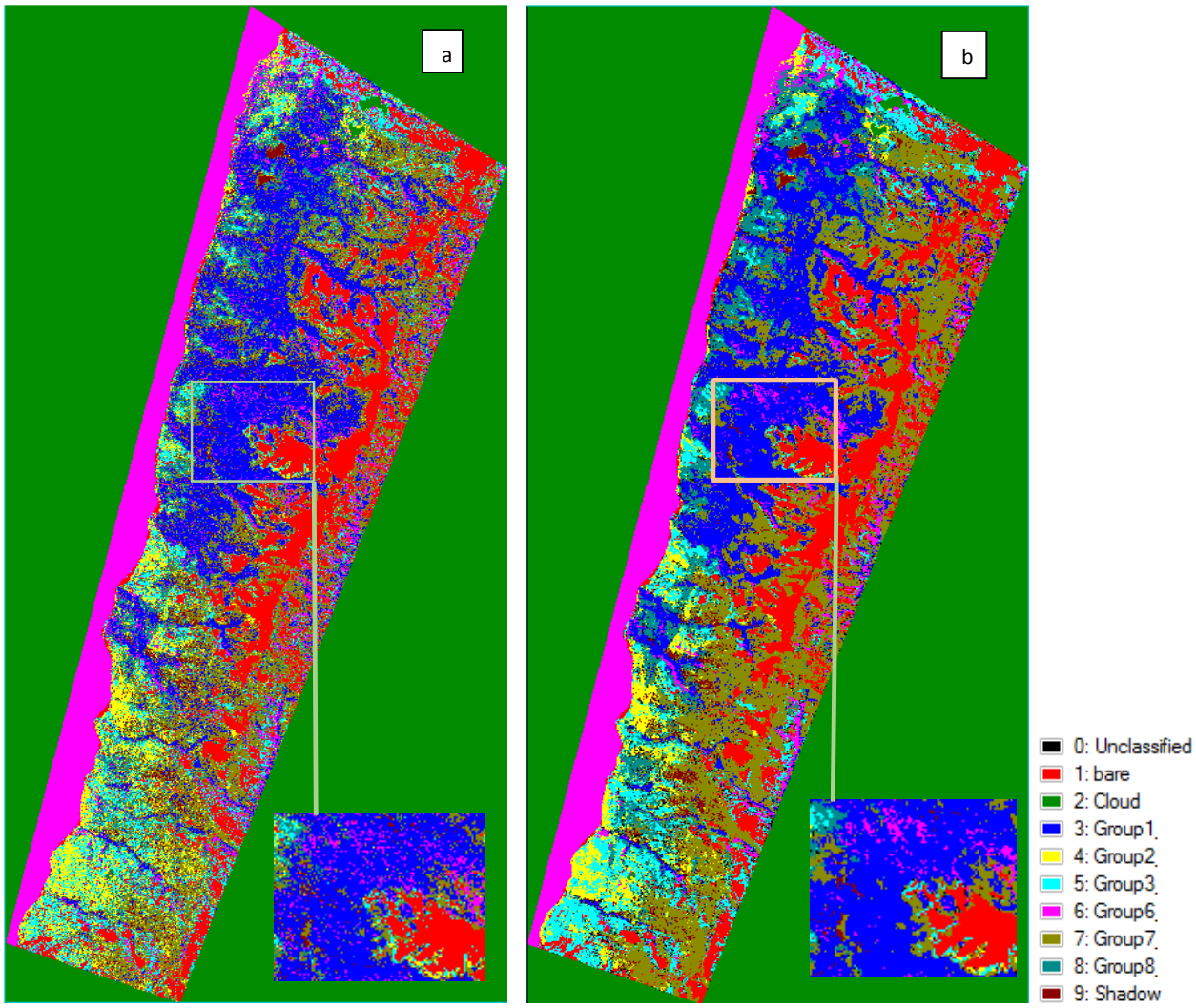


Figure 7. a) Raw supervised classification output and b) sieved and clumped classification output.

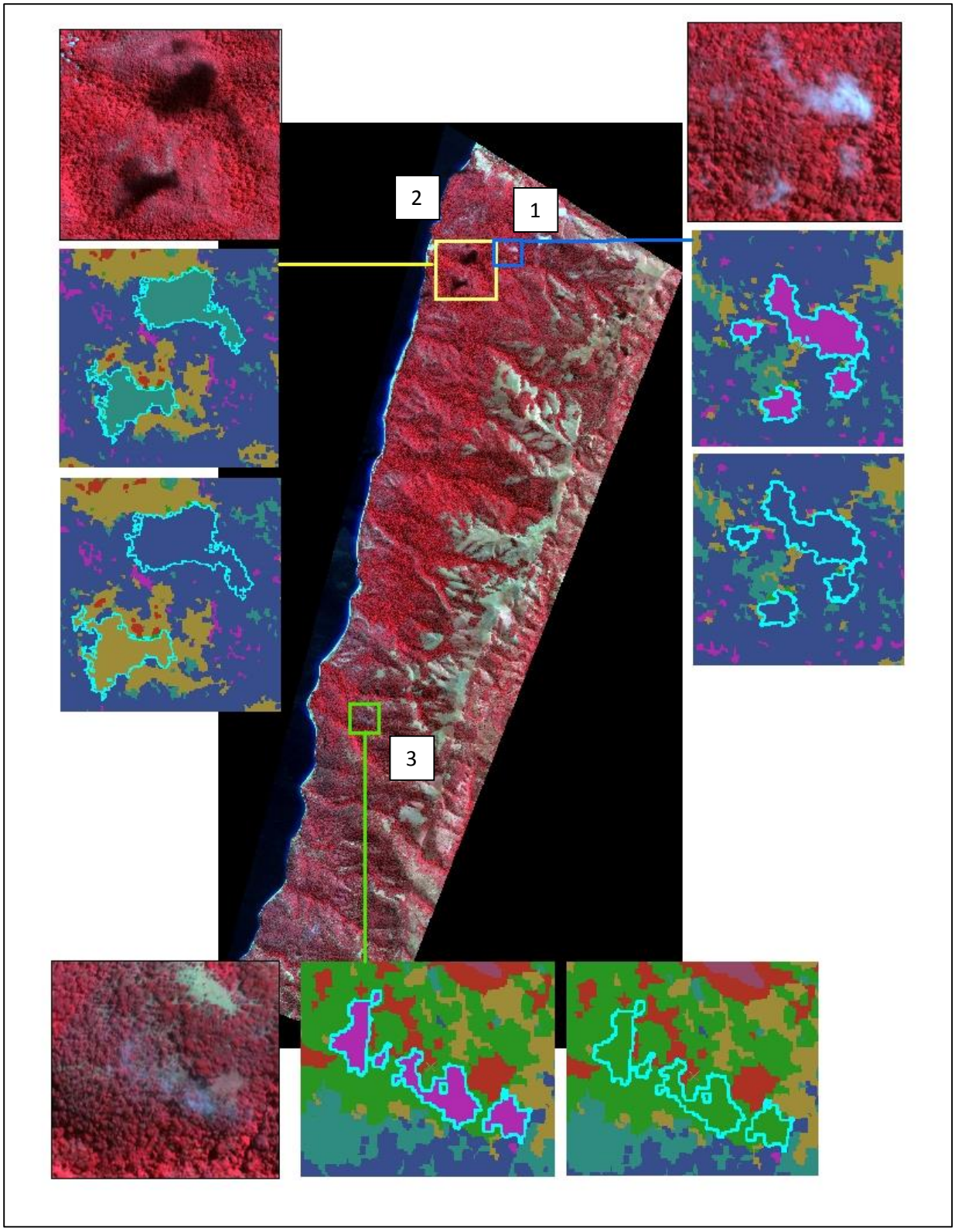


Figure 8. Cloud and Shadow correction. The background is false color image of Gombe National Park. The small portions show three major areas affected by cloud and shadow. The parcels that are incorrectly classified were highlighted with cyan border (1, 2, 3). The images closer to the small false color images are classification image before cloud and shadow correction, and the images further are classification after cloud and shadow modification. Different vegetation classes are shown in different colors.

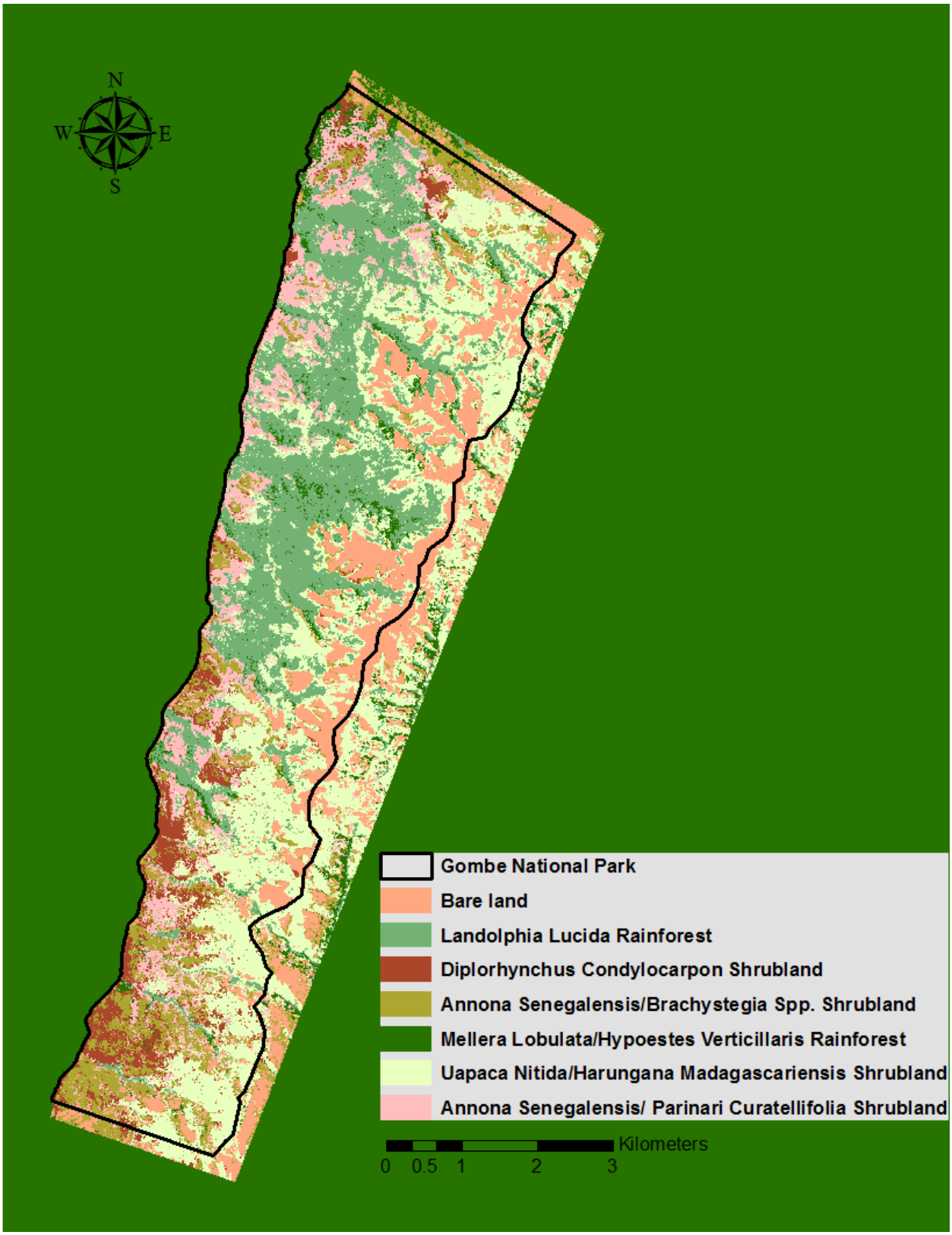


Figure 9. Final vegetation class map of Gombe National Park.

DISCUSSION

The vegetation class schema design was based on a species composition cluster analysis. By using species composition clusters as vegetation classes in the remote sensing image classification, the hypothesis is that plots that are similar in species composition should also have similar electromagnetic radiation reflectance. According to the separability statistics and classification results, the species composition clusters were not effectively separated in spectral characteristics. Particularly, when two clusters were different in species composition but presented similar forest structure, such as the "Landolphia lucida rainforest" and "Mellera lobulata/Hypoestes verticillaris rainforest", they were similar in spectral characteristics and therefore the classification task was difficult.

Improvements to the vegetation classification could be made by adding more informational spectral bands to the classification, in addition to the 4 spectral bands of QuickBird image in combination with DEM, slope, aspects and other topographic informational bands, and vegetation index bands such as Ratio Vegetation Index (RVI) and Normalized Difference Vegetation Index (NDVI). Ancillary data is often used in addition to spectral bands to improve classification accuracy (Eiumnoh and Rajendra 2000; Rozenstein and Kamineli 2011). The application of ancillary data is usually based on rules deriving from expert knowledge or data pattern driven from automatic data mining techniques which requires no priori focal knowledge. Therefore, rule-based classification, such as Classification and Regression Tree (CART) analysis (Lawrence and Andrea Wright 2001), or random forest classification could be used to improve the classification accuracy of Gombe National Park.

Additionally, the classification could also be improved by adding sample plots which may help to build a more representative and distinctive spectral signature of each vegetation class. Moreover, stratifying the park landscape before conducting cluster analysis and supervised classification could also extract more vegetation class information from the landscape.

High resolution image such as QuickBird images captures lots spectral details, some of which are noises resulting from tree gaps and shadows, and clouds. Therefore, the raw classification image generated from QuickBird in this project was very mottled and contained large amount of small misclassified patches. The "sieve" and "clump" functions provided by ENVI largely cleaned out this patches and reclassified them to "unclassified" class. These patches and other "unclassified" patches generated from the classification process were useless for the user. Given the advantage of recent advanced spatial

analysis capacity of ArcGIS, the “majority buffer zone” method invented in this project is capable to reclassify these patches into the relatively accurate class efficiently.

CONCLUSION

The traditional methods of identifying vegetation cover characteristics using remote sensing images are usually restricted to detection of vegetation coverage or broad vegetation cover types. Therefore, advancing the vegetation specious composition detection capacity of remote sensing technique would largely broaden its application in habitat quality evaluation and conservation planning. With this goal, I combined the traditional remote sensing classification with ecological data mining technique to produce a vegetation cover map which reflects the vegetation species composition in each vegetation cover class in Gombe National Park. Specifically, I performed a cluster analysis of the vegetation survey data for the purpose of generating vegetation cover classes based on their similarity in vegetation species composition, and further used the clustered survey data to train a supervised classification algorithm – Maximum Likelihood Classification. The result of this method is a vegetation cover map of Gombe National Park showing seven vegetation cover classes named after their dominant vegetation species.

Moreover, the cluster analysis method provides an alternative way of vegetation class schema design where little local knowledge of vegetation assemblages is needed. The number of vegetation classes and the class definition were “mined” from the data pattern of the vegetation survey results. Though the supervised classification accuracy based on the cluster analysis results was not as successful as hoped, it’s reasonable to expect that the classification accuracy will be improved by adding additional bands in addition to spectral bands, such as topographic conditions and vegetation indexes, and by applying rule-based classification techniques and increasing classification training samples.

This project also innovated a semi-automatic post-processing workflow to correct for misclassification resulting from spectral noises in high resolution images, which consists of an automatic sieve and clump process, an automatic “buffer zone majority” correction, and a manual cloud and shadow correction. This method successfully created a more coherent and accurate vegetation cover map for Gombe National Park. This semi-automatic correction workflow could be adopted in future studies to improve the efficiency and accuracy of post-processing of vegetation cover classification.

CHAPTER 4. CONCLUSION

This project produced 24 vegetation habitat suitability maps and distribution extent maps, an overall Chimpanzee habitat quality surface map, and a vegetation cover map of Gombe National Park. These outcomes were achieved based on the long-term chimpanzee behavior field data collected for more than 40 year, the advancement in spatial analysis techniques and ecological application of machine learning techniques, and the availability of high quality remote sensing images.

The products of the vegetation species distribution provide useful spatial information for vegetation studies, chimpanzee habitat quality evaluation, chimpanzee behavioral studies, as well as for supporting habitat conservation planning in Gombe National Park. Further, this project successfully proved the capacity of MaxEnt model in predicting distribution pattern in a small area with fine resolution of 10 m, which would justify and promote the application of MaxEnt in small scale conservation planning projects. Moreover, the innovative methods and workflows proposed in this project to address limitations of traditional vegetation food abundance assessment and remote sensing classification of vegetation cover will improve the accuracy and efficiency of habitat quality evaluation once adopted in future practical conservation planning projects.

REFERENCES

Altmann, J., Altmann, S. A., Hausfater, G., & McCuskey, S. A. (1977). Life history of yellow baboons: physical development, reproductive parameters, and infant mortality. *Primates*, *18*(2), 315-330.

Begon, M., Harper, J. D., and Townsend, C. R., & (1996). Ecology: individuals, populations and communities.

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern recognition*, *30*(7), 1145-1159.

Bray, J. R., & Curtis, J. T. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecological monographs*, *27*(4), 325-349.

Butynski, T. M. (2003). The robust chimpanzee Pan troglodytes: taxonomy, distribution, abundance, and conservation status. *Status Survey and Conservation Action Plan: West African Chimpanzees*, 5-12.

Chapman, C. A., & Chapman, L. J. (1999). Implications of small scale variation in ecological conditions for the diet and density of red colobus monkeys.*Primates*, *40*(1), 215-231.

Dittus, W. P. (1977). The social regulation of population density and age-sex distribution in the toque monkey. *Behaviour*, *63*(3), 281-322.

Eiumnoh, A., & Shrestha, R. P. (2000). Application of DEM data to Landsat image classification: evaluation in a tropical wet-dry landscape of Thailand.*Photogrammetric Engineering and Remote Sensing*, *66*(3), 297-304.

Elith, J., & Leathwick, J. R. (2009). Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 677.

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*,*17*(1), 43-57.

Endries, M. (2011). Aquatic Species Mapping in North Carolina Using Maxent.

Desktop, A. (2011). Release 10. *Redlands, CA: Environmental Systems Research Institute*.

Evans J.S., Oakleaf J., Cushman S.A., Theobald D. (2014). An ArcGIS Toolbox for Surface Gradient and Geomorphometric Modeling, version 2.0-0. Available: http://evansmurphy.wix.com/evansspatial. Accessed: 2014 Dec 2nd.

Gessler, P. E., Moore, I. D., McKenzie, N. J., & Ryan, P. J. (1995). Soil-landscape modelling and spatial prediction of soil attributes. *International Journal of Geographical Information Systems*, *9*(4), 421-432.

Goodall, J. (1986). The chimpanzees of Gombe: patterns of behavior.

Goslee, S. C., & Urban, D. L. (2007). The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, *22*(7), 1-19.

Hijmans, R., & van Etten, J. (2014). raster: raster: Geographic data analysis and modeling. *R package version*, 2-2.

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis* (Vol. 344). John Wiley & Sons.

Keinath, D. A., Andersen, M. D., & Beauvais, G. P. (2010). Range and modeled distribution of Wyoming's species of greatest conservation need. *Report prepared by the Wyoming Natural Diversity Database, Laramie Wyoming for the Wyoming Game and Fish Department, Cheyenne, Wyoming and the US Geological Survey, Fort Collins, Colorado*.

Knollová, I., & Chytrý, M. (2004). Oak-hornbeam forests of the Czech Republic: geographical and ecological approaches to vegetation classification.

Köndgen, S., Kühl, H., N'Goran, P. K., Walsh, P. D., Schenk, S., Ernst, N., ... & Leendertz, F. H. (2008). Pandemic human viruses cause decline of endangered great apes. *Current Biology*, *18*(4), 260-264.

Krebs, C. J., Hickman, G. C., & Hickman, S. M. (1994). *Ecology: the experimental analysis of distribution and abundance* (Vol. 4). New York: HarperCollins College Publishers.

Laporta, G. Z., Ribeiro, M. C., Ramos, D. G., & Sallum, M. A. M. (2012). Spatial distribution of arboviral mosquito vectors (Diptera, Culicidae) in Vale do Ribeira in the South-eastern Brazilian Atlantic Forest. *Cadernos de Saúde Pública*,*28*(2), 229-238.

Lawrence, R. L., & Wright, A. (2001). Rule-based classification systems using classification and regression tree (CART) analysis. *Photogrammetric engineering and remote sensing*, *67*(10), 1137-1142.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest.*R news*, *2*(3), 18-22.

Marshall, A. J. (2010). Effect of habitat quality on primate populations in Kalimantan: gibbons and leaf monkeys as case studies. In *Indonesian primates* (pp. 157-177). Springer New York.

McCune, B., & Keon, D. (2002). Equations for potential annual direct incident radiation and heat load. *Journal of vegetation science*, *13*(4), 603-606.

Milton, K. (1990). Annual mortality patterns of a mammal community in central Panama. *Journal of Tropical Ecology*, *6*(04), 493-499.

Mittermeier, R. A., Wallis, J., Rylands, A. B., Ganzhorn, J. U., Oates, J. F., Williamson, E. A., ... & Schwitzer, C. (2009). Primates in peril: the world's 25 most endangered primates 2008-2010. *Primate Conservation*, *24*, 1-57.

Moore, I. D., Lewis, A., & Gallant, J. C. (1993). Terrain attributes: estimation methods and scale effects.

Moreau, R. E. (1945). The distribution of the chimpanzee in Tanganyika Territory. *Tanganyika Notes and Records*, *14*, 52-55.

Moyer, D., Plumptre, A. J., Pintea, L., Hernandez-Aguilar, A., Moore, J., Stewart, F., ... & Mwangoka, M. (2006). Surveys of chimpanzees and other biodiversity in Western Tanzania. *Unpublished report. Arlington, VA: United States Fish and Wildlife Service (USFWS)*.

Murray, C. M., Eberly, L. E., & Pusey, A. E. (2006). Foraging strategies as a function of season and rank among wild female chimpanzees (Pan troglodytes).*Behavioral Ecology*, *17*(6), 1020-1028.

Nelder, J. A., & Baker, R. J. (1972). Generalized linear models. *Encyclopedia of Statistical Sciences*.

Nishida, T., Wrangham, R. W., Goodall, J., & Uehara, S. (1983). Local differences in plant-feeding habits of chimpanzees between the Mahale Mountains and Gombe National Park, Tanzania. *Journal of Human Evolution*,*12*(5), 467-480.

Oates, J.F., Tutin, C.E.G., Humle, T., Wilson, M.L., Baillie, J.E.M., Balmforth, Z., Blom, A., Boesch, C., Cox, D., Davenport, T., Dunn, A., Dupain, J., Duvall, C., Ellis, C.M., Farmer, K.H., Gatti, S., Greengrass, E., Hart, J., Herbinger, I.,

Hicks, C., Hunt, K.D., Kamenya, S., Maisels, F., Mitani, J.C., Moore, J., Morgan, B.J., Morgan, D.B., Nakamura, M., Nixon, S., Plumptre, A.J., Reynolds, V., Stokes, E.J. & Walsh, P.D. 2008. *Pan troglodytes*. The IUCN Red List of Threatened Species. Version 2014.3. <www.iucnredlist.org>. Downloaded on 22 April 2015.

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Stevens, M. H. H., Oksanen, M. J., & Suggests, M. A. S. S. (2007). The vegan package. *Community ecology package*.

Phillips, S. (2005). A brief tutorial on Maxent. *AT&T Research*.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological modelling*, *190*(3), 231-259.

Peck, M., Thorn, J., Mariscal, A., Baird, A., Tirira, D., & Kniveton, D. (2011). Focusing conservation efforts for the critically endangered brown-headed spider monkey (Ateles fusciceps) using remote sensing, modeling, and playback survey methods. International journal of primatology, 32(1), 134-148.

Pike, R. J., & Wilson, S. E. (1971). Elevation-relief ratio, hypsometric integral, and geomorphic area-altitude analysis. *Geological Society of America Bulletin*,*82*(4), 1079-1084.

Pintea, L., Pusey, A., Bolstad, P., & Bauer, M. (2006). Remote sensing of chimpanzee habitat change in Gombe National Park, Tanzania: implications for behavioral research and conservation. *International Society of Primatology. Entebbe, Uganda. pp. Abst*, *49*.

Pintea, L. (2007). Applying remote sensing and GIS for chimpanzee habitat change detection, behaviour and conservation.

Plumptre, A. J. (2010). *Eastern Chimpanzee (Pan Troglodytes Schweinfurthii): Status Survey and Conservation Action Plan, 2010-2020*. IUCN.

Pusey, A. E., Wilson, M. L., & Anthony Collins, D. (2008). Human impacts, disease risk, and population dynamics in the chimpanzees of Gombe National Park, Tanzania. *American Journal of Primatology*, *70*(8), 738-744.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Richards, J. A., & Richards, J. A. (1999). *Remote sensing digital image analysis* (Vol. 3). Berlin et al.: Springer.Romesburg, C. (2004). *Cluster analysis for researchers*. Lulu. com.

Rozenstein, O., & Karnieli, A. (2011). Comparison of methods for land-use classification incorporating remote sensing and GIS inputs. *Applied Geography*,*31*(2), 533-544.

Rudicell, R. S., Jones, J. H., Wroblewski, E. E., Learn, G. H., Li, Y., Robertson, J. D., ... & Wilson, M. L. (2010). Impact of simian immunodeficiency virus infection on chimpanzee population dynamics. *PLoS pathogens*, *6*(9), e1001116.

Schipper, J., Chanson, J. S., Chiozza, F., Cox, N. A., Hoffmann, M., Katariya, V., ... & Hammond, P. S. (2008). The status of the world's land and marine mammals: diversity, threat, and knowledge. *Science*, *322*(5899), 225-230.

Sing, T., Sander, O., Beerenwinkel, N., & Lengauer, T. (2005). ROCR: visualizing classifier performance in R. *Bioinformatics*, *21*(20), 3940-3941.

Struhsaker, T. T. (1973). A recensus of vervet monkeys in the Masai-Amboseli Game Reserve, Kenya. *Ecology*, 930-932.

Tinoco, B. A., Astudillo, P. X., Latta, S. C., & Graham, C. H. (2009). Distribution, ecology and conservation of an endangered Andean hummingbird: the Violet-throated Metaltail (Metallura baroni). *Bird Conservation International*,*19*(01), 63-76.

U.S. Geological Survey (2014). Landsat Surface Reflectance Climate Data Records: U.S. Geological Survey Fact Sheet 2013–3117, 1 p., *http://dx.doi.org/10.3133/fs20133117*.

Ward, D. F. (2007). Modelling the potential geographic distribution of invasive ant species in New Zealand. *Biological Invasions*, *9*(6), 723-735.

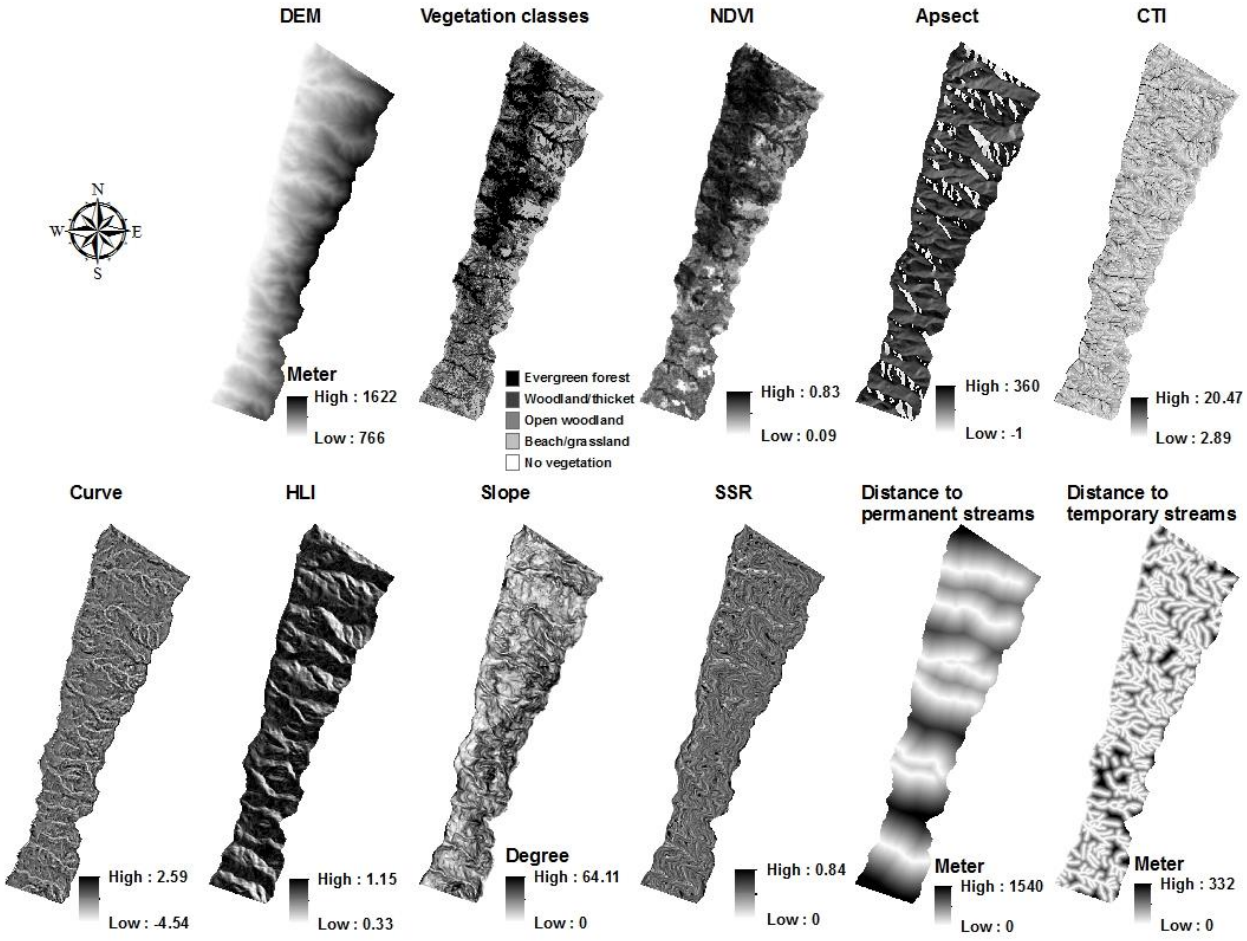
Williams, J. M., Lonsdorf, E. V., Wilson, M. L., Schumacher-Stankey, J., Goodall, J., & Pusey, A. E. (2008). Causes of death in the Kasekela chimpanzees of Gombe National Park, Tanzania. *American Journal of Primatology*, *70*(8), 766-777.

Wilson, M., Mjungu, D., Butoki, B., Pintea, L., Murray, C., Matama, H.. (2009) "Relative abundance of plant species within the home ranges of three chimpanzee communities in Gombe National Park, Tanzania." Unpublished report.

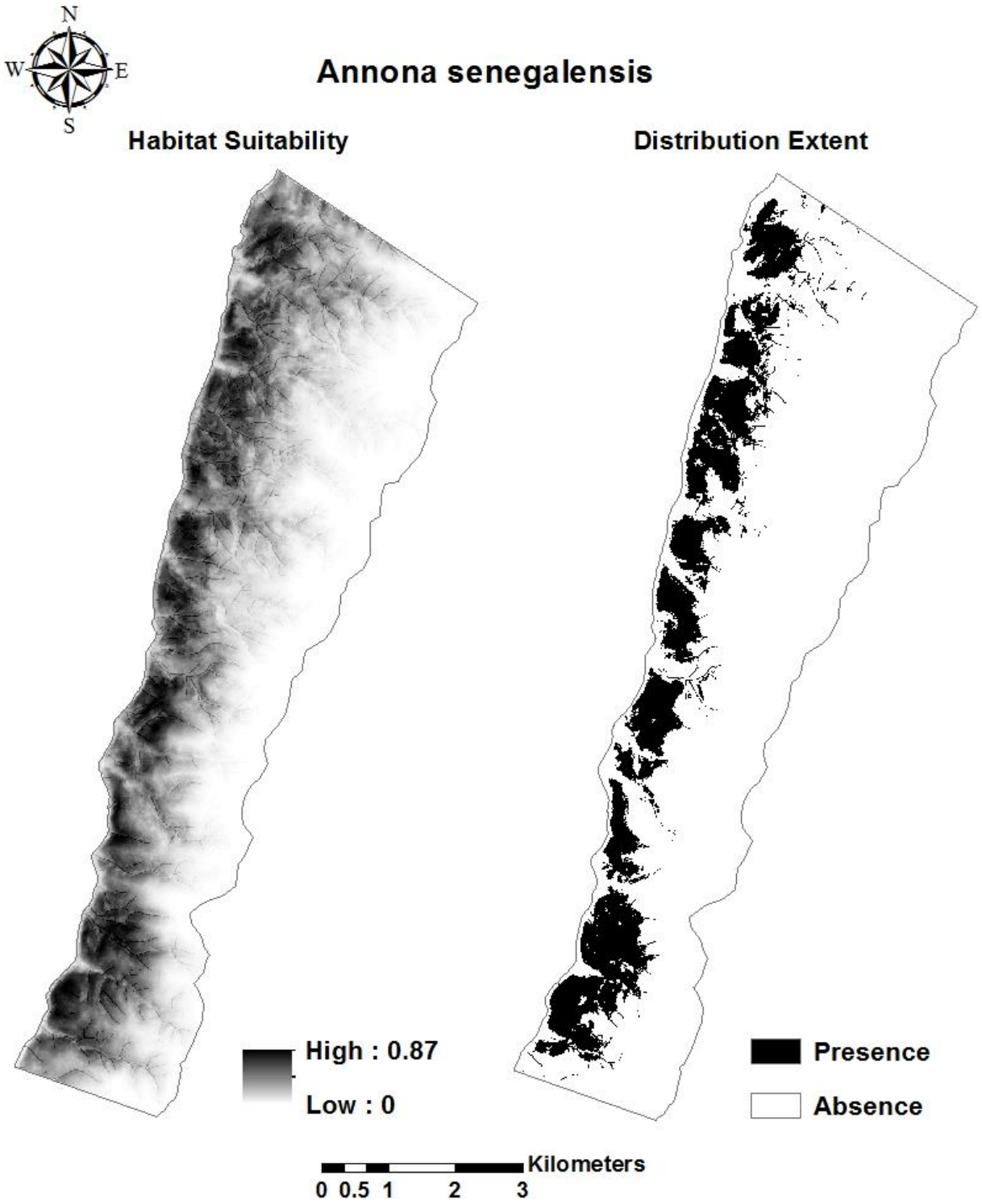
Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H., & Guisan, A. (2008). Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, *14*(5), 763-773.

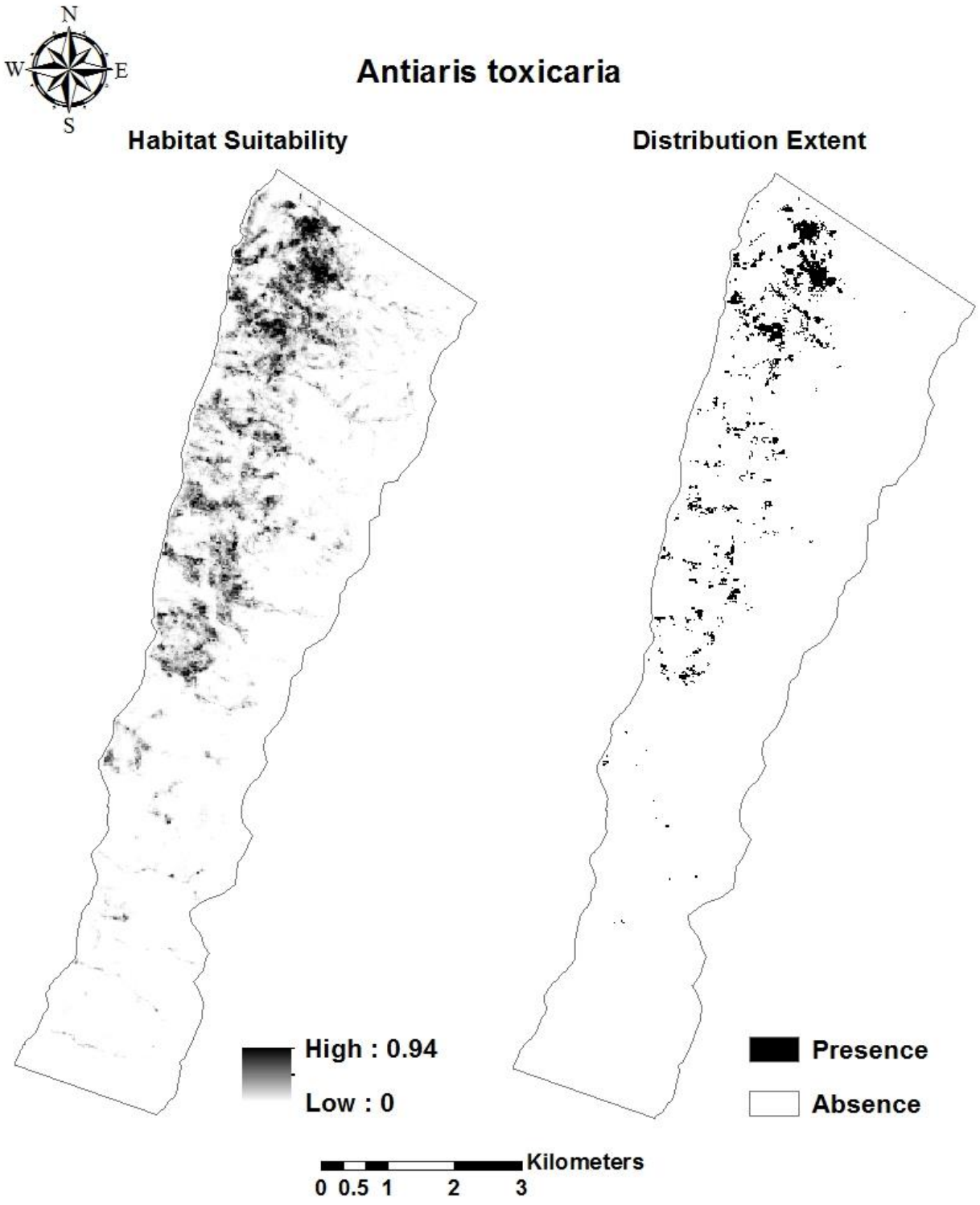
APPENDIX

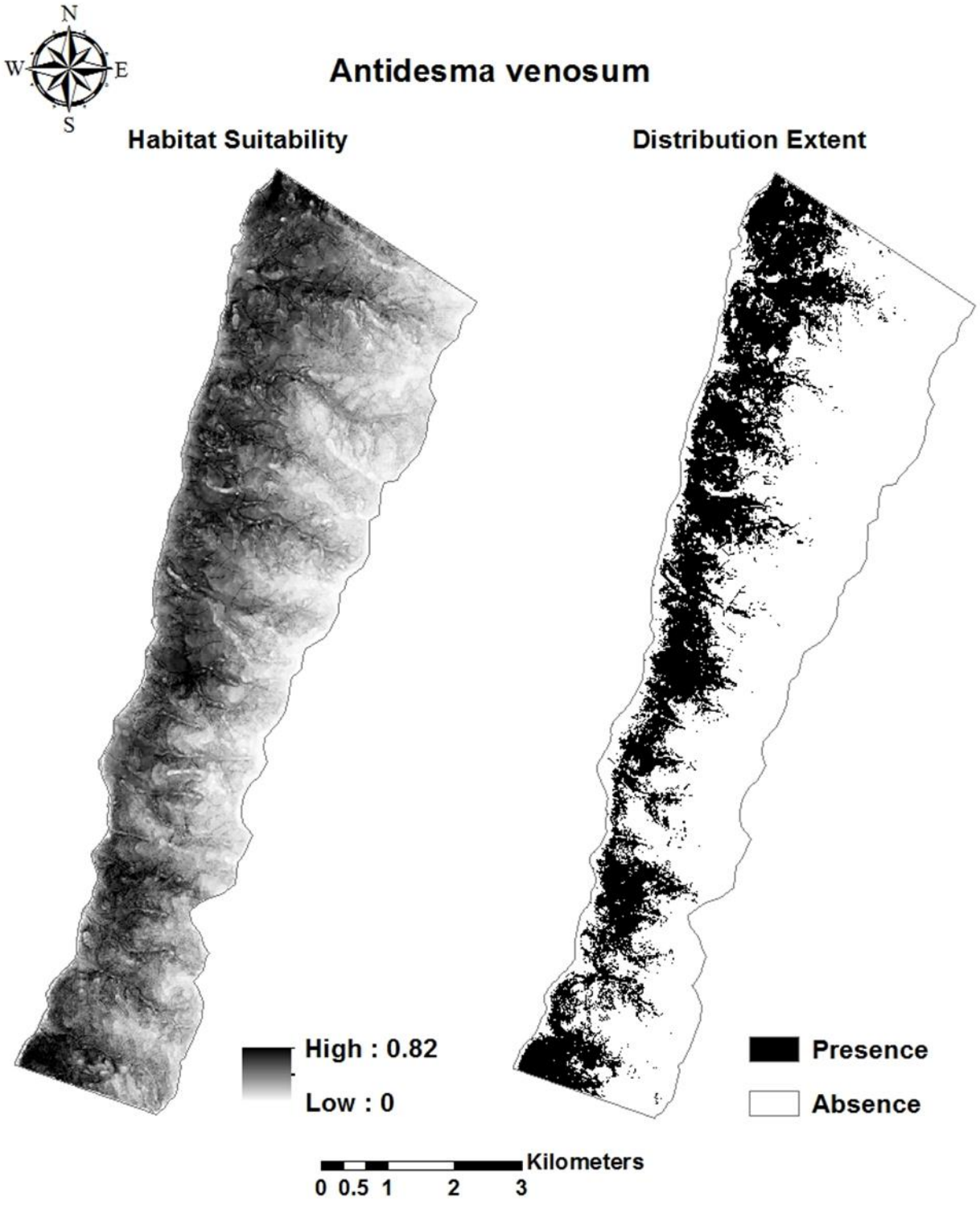
APPENDIX 1 *Raster layers of environmental variables*

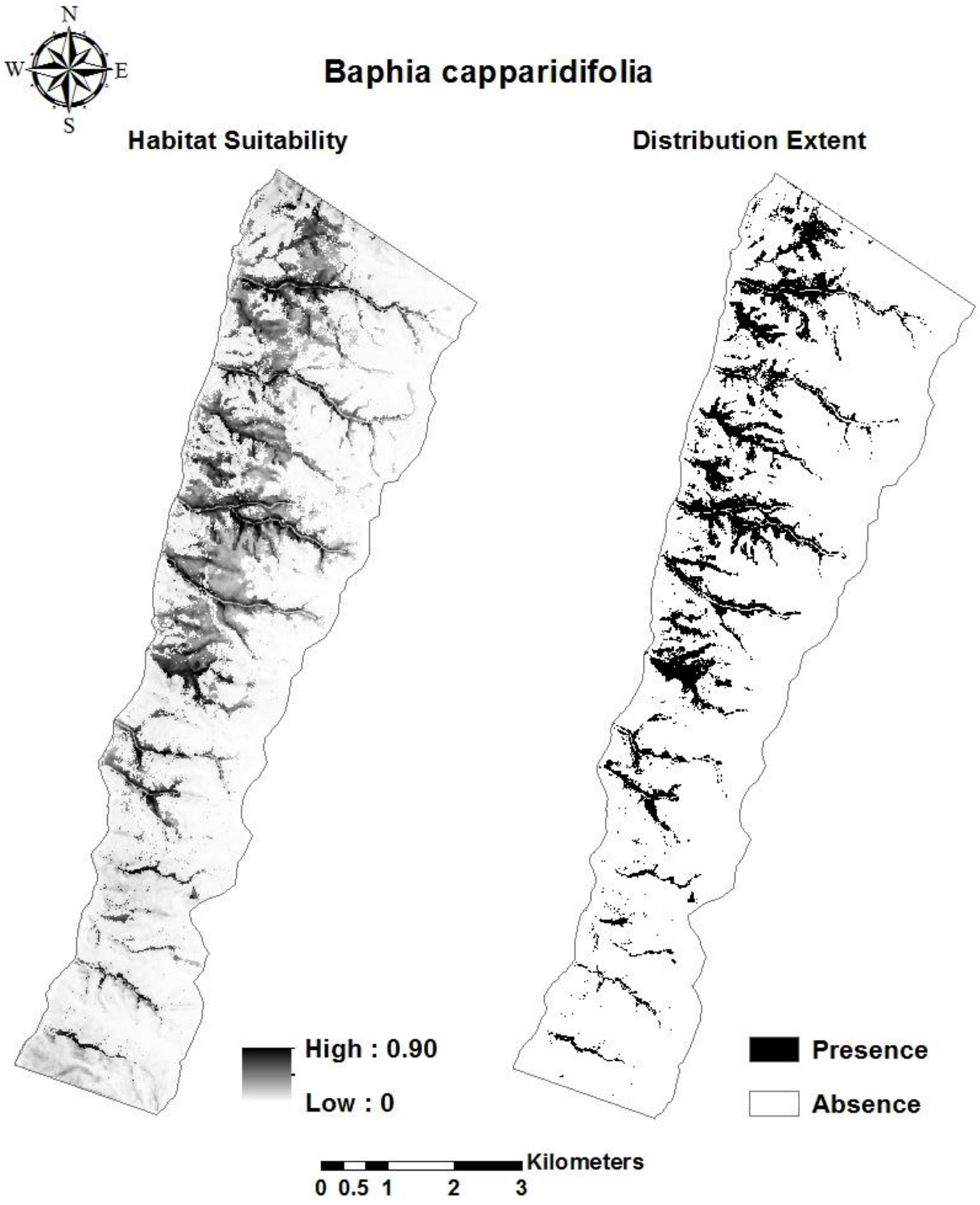


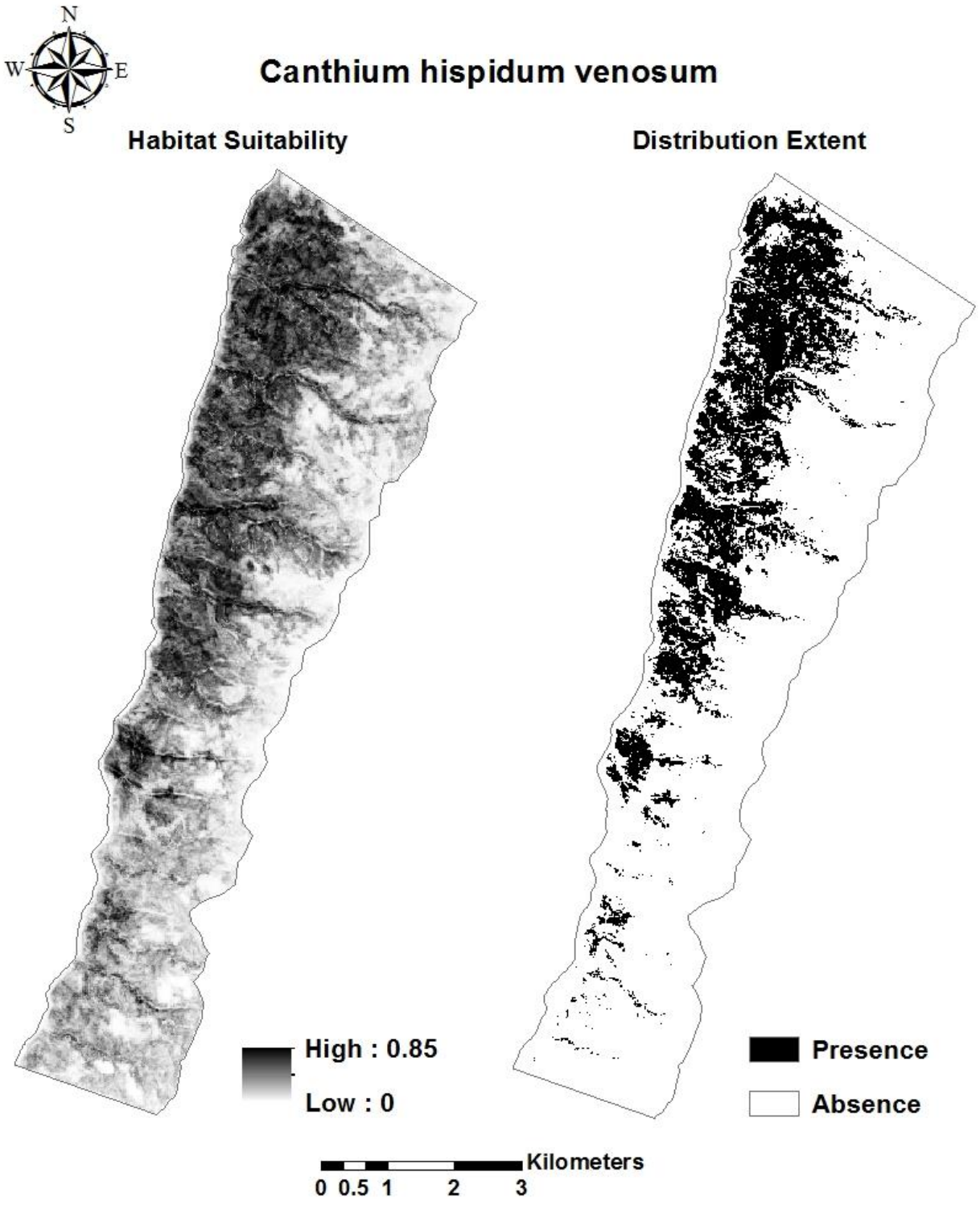
APPENDIX 2 *Vegetation species habitat suitability and distribution extent maps*

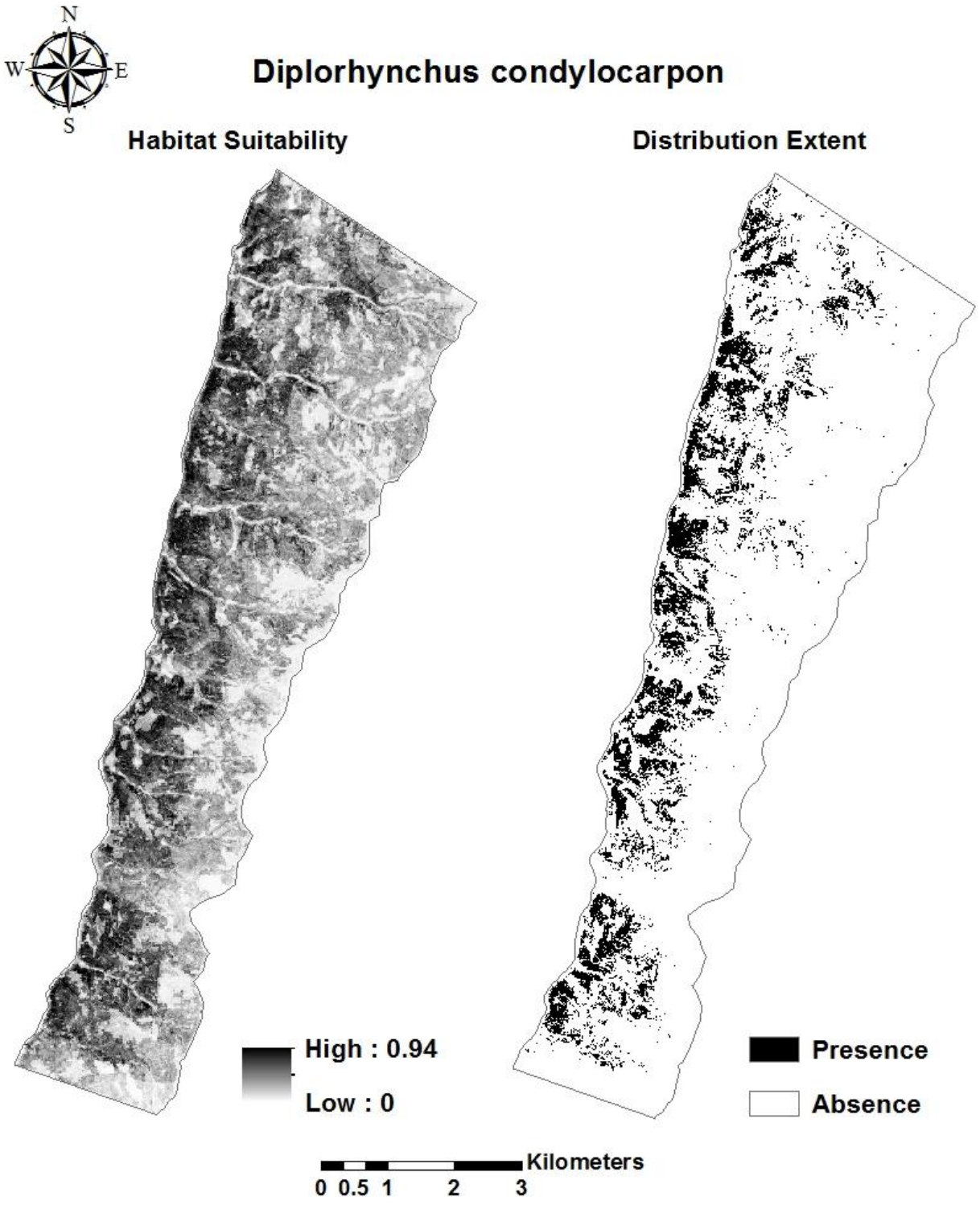


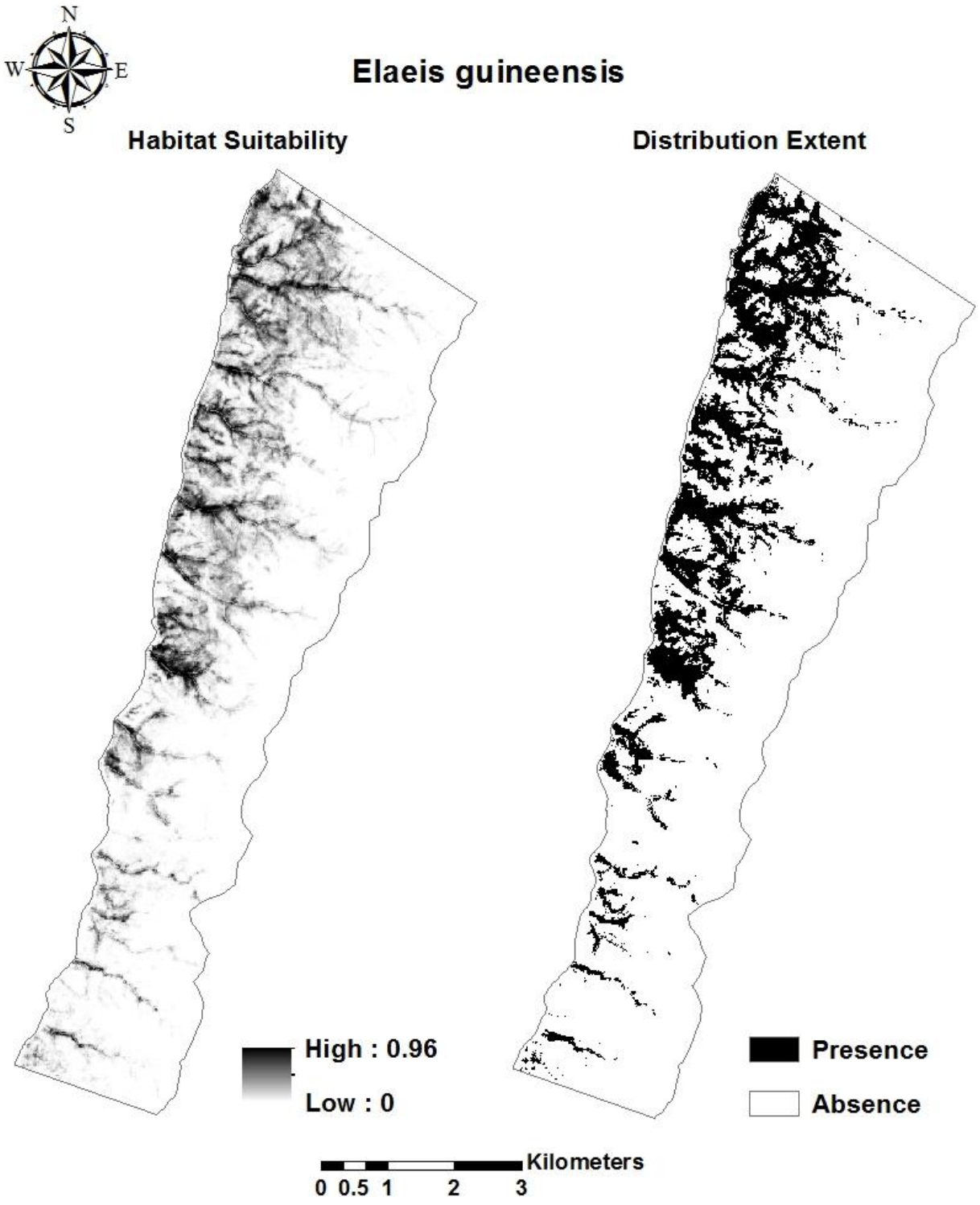

N
W
E
S
Antiaris toxicaria
Habitat Suitability
Distribution Extent
High : 0.94
Low : 0
Presence
Absence
Kilometers
0 0.5 1 2 3

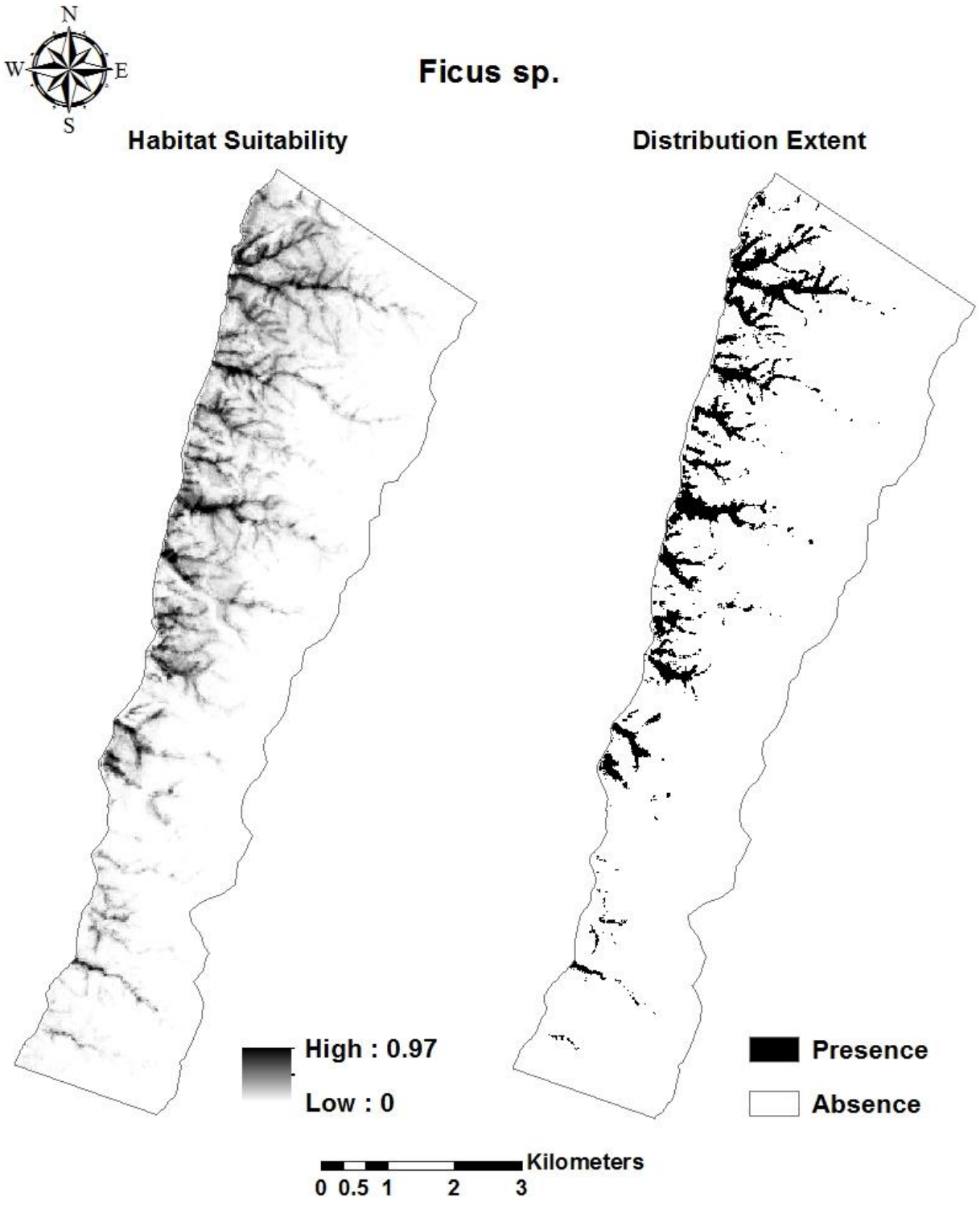

N
W
E
S
Antidesma venosum
Habitat Suitability
Distribution Extent
High : 0.82
Low : 0
Presence
Absence
Kilometers
0 0.5 1 2 3

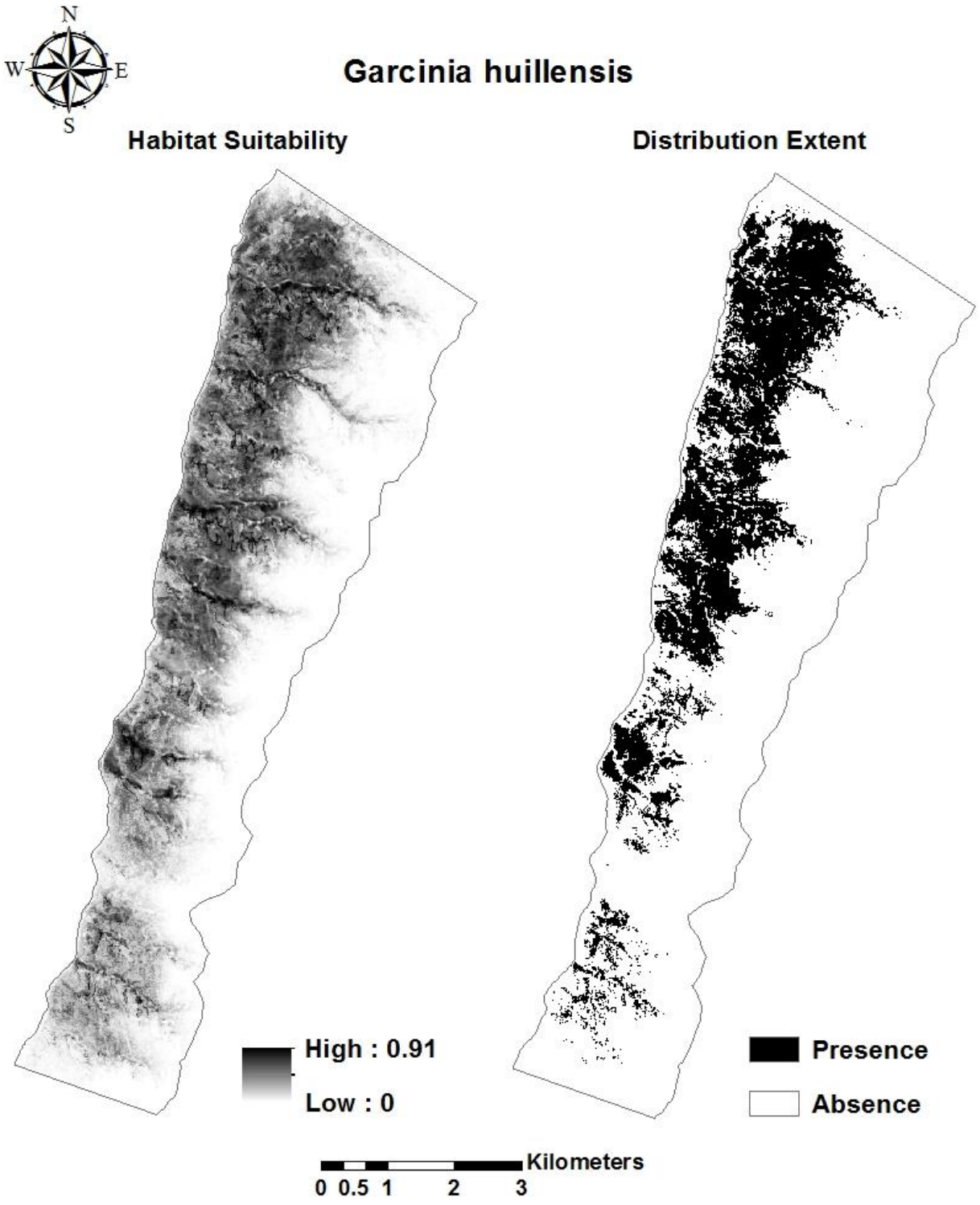

N
W
E
S
Baphia capparidifolia
Habitat Suitability
Distribution Extent
High : 0.90
Low : 0
Presence
Absence
Kilometers
0 0.5 1 2 3

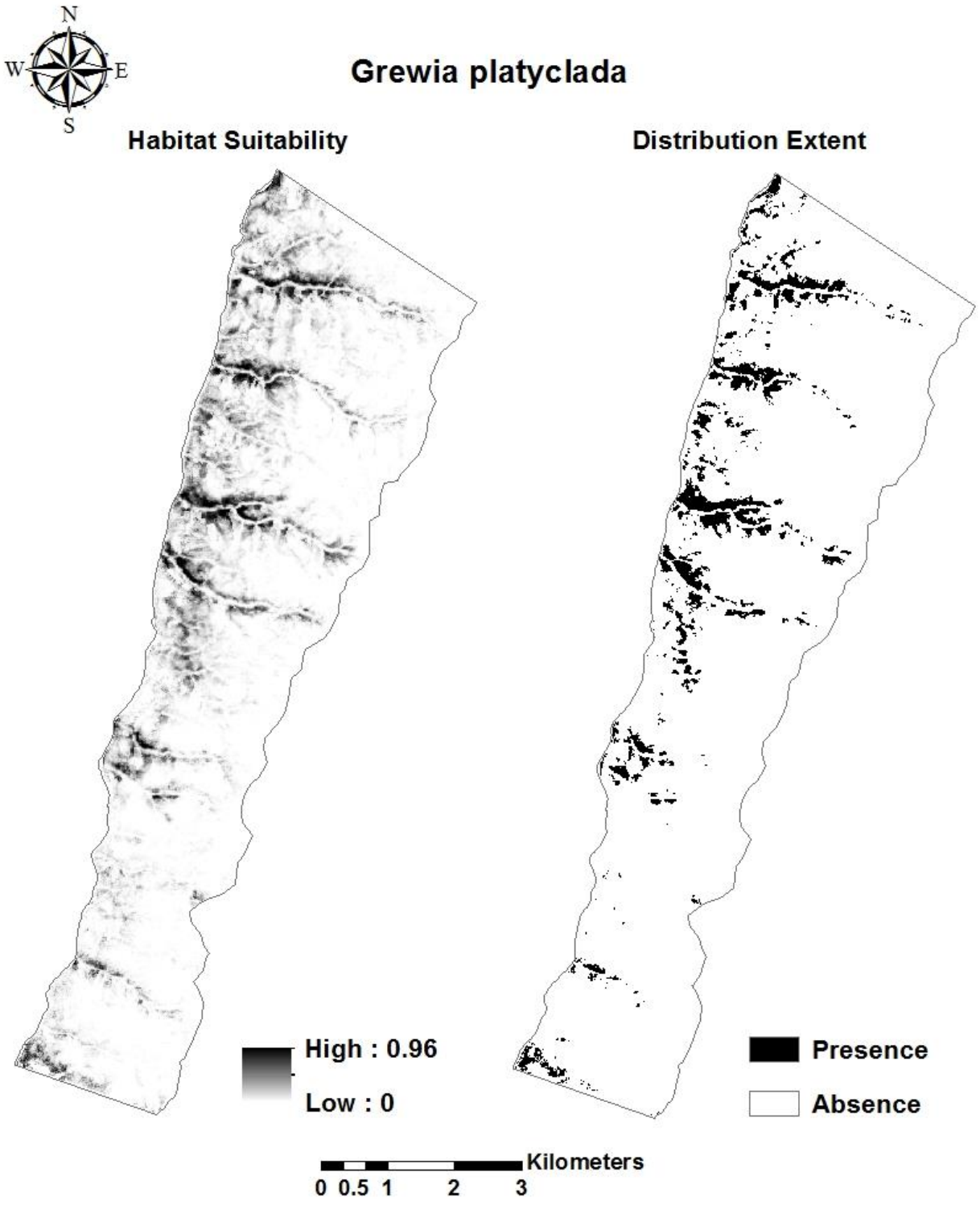

N
W
E
S
Canthium hispidum venosum
Habitat Suitability
Distribution Extent
High : 0.85
Low : 0
Presence
Absence
Kilometers
0 0.5 1 2 3

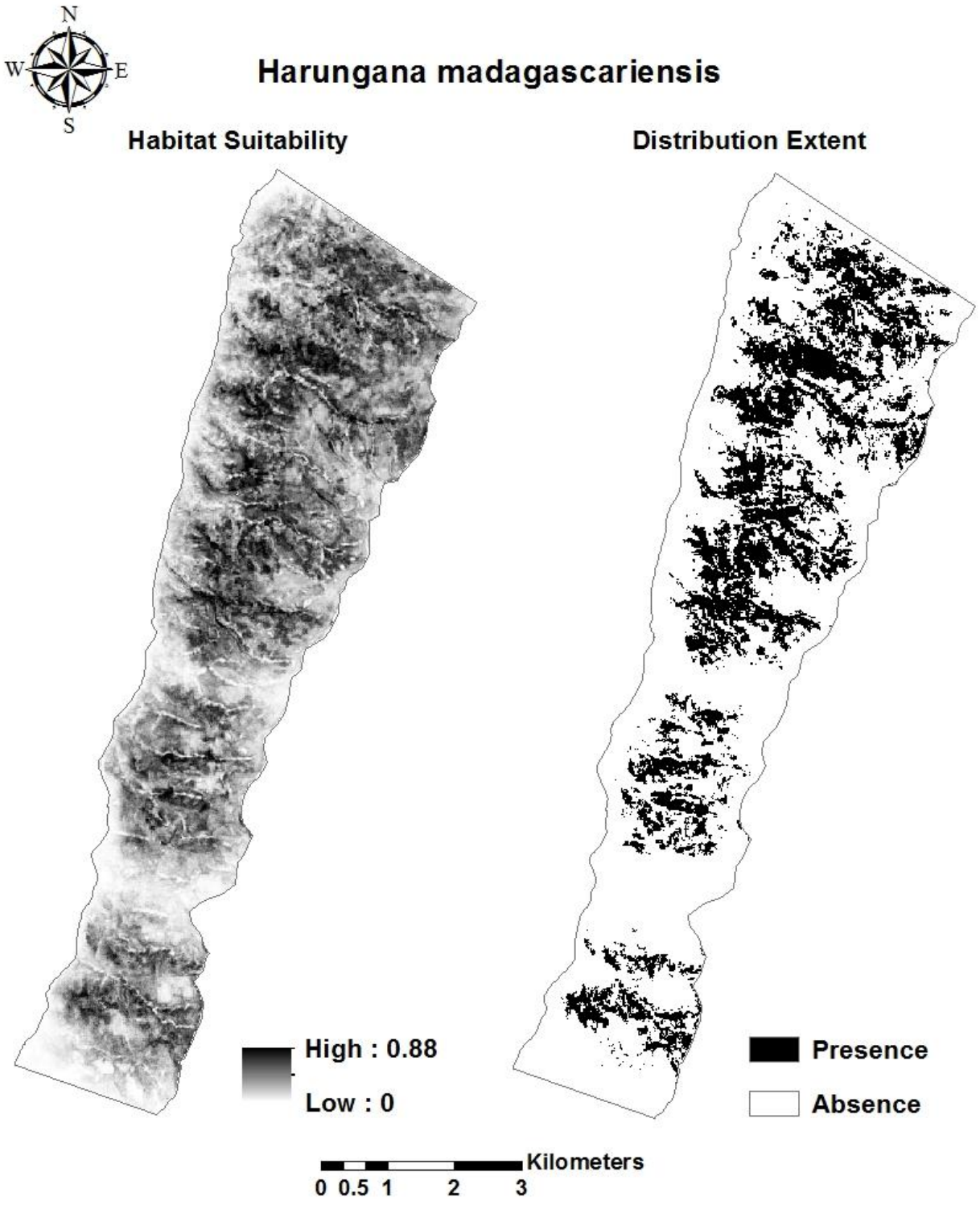

N
W
E
S
Diplorhynchus condylocarpon
Habitat Suitability
Distribution Extent
High : 0.94
Low : 0
Presence
Absence
Kilometers
0 0.5 1 2 3

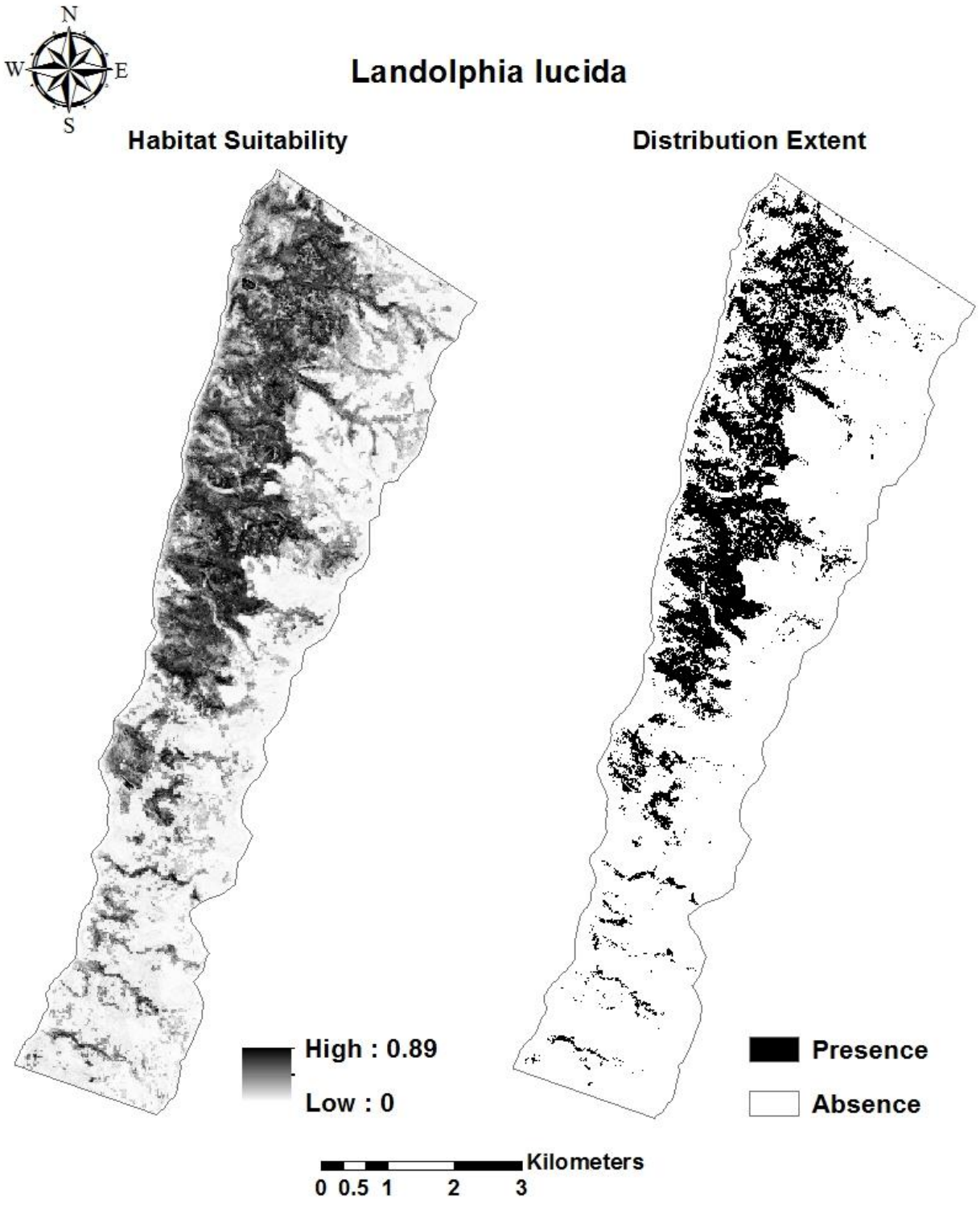

N
W
E
S
Elaeis guineensis
Habitat Suitability
Distribution Extent
High : 0.96
Low : 0
Presence
Absence
Kilometers
0 0.5 1 2 3

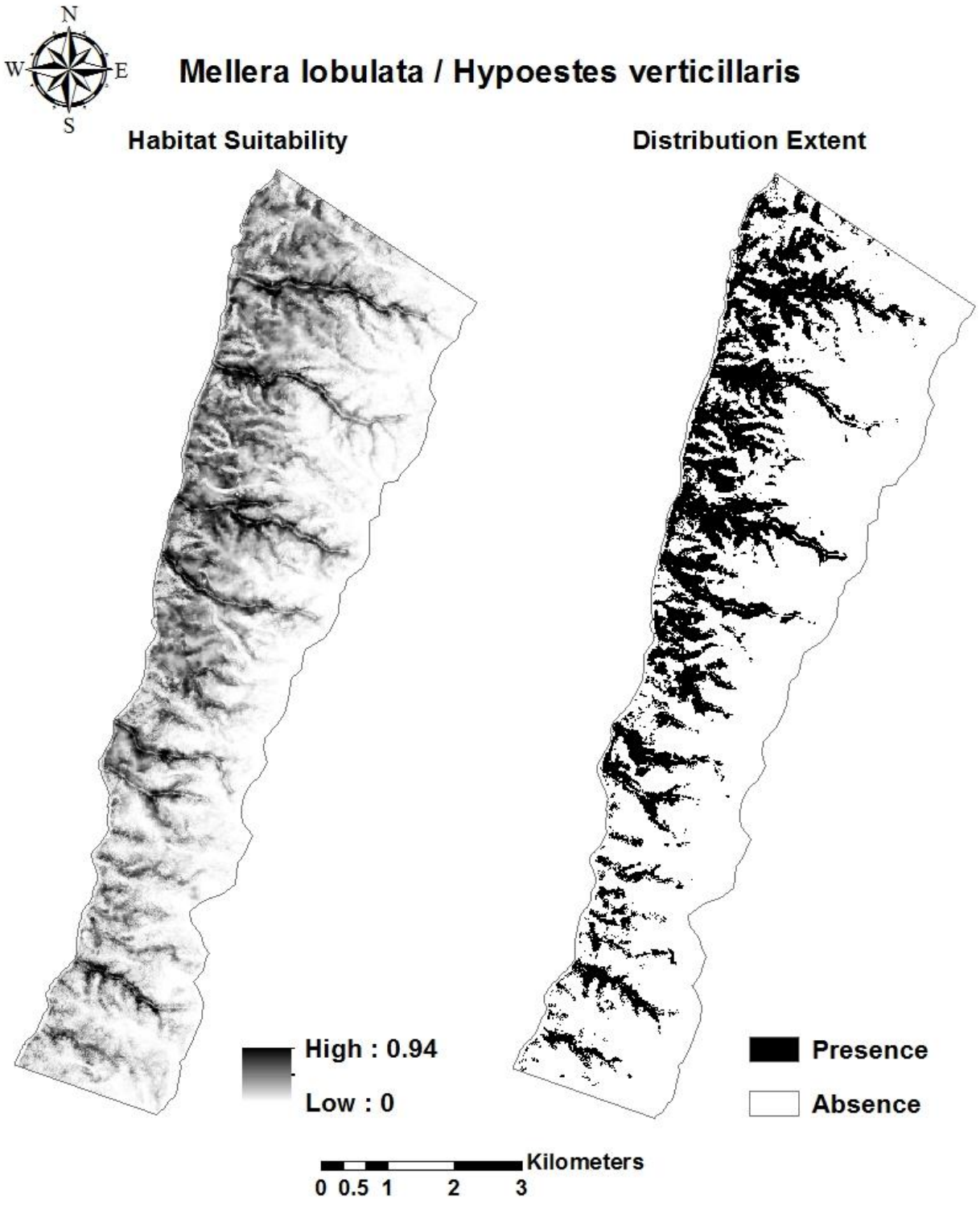

N
W
E
S
Ficus sp.
Habitat Suitability
Distribution Extent
High : 0.97
Low : 0
Presence
Absence
Kilometers
0 0.5 1 2 3

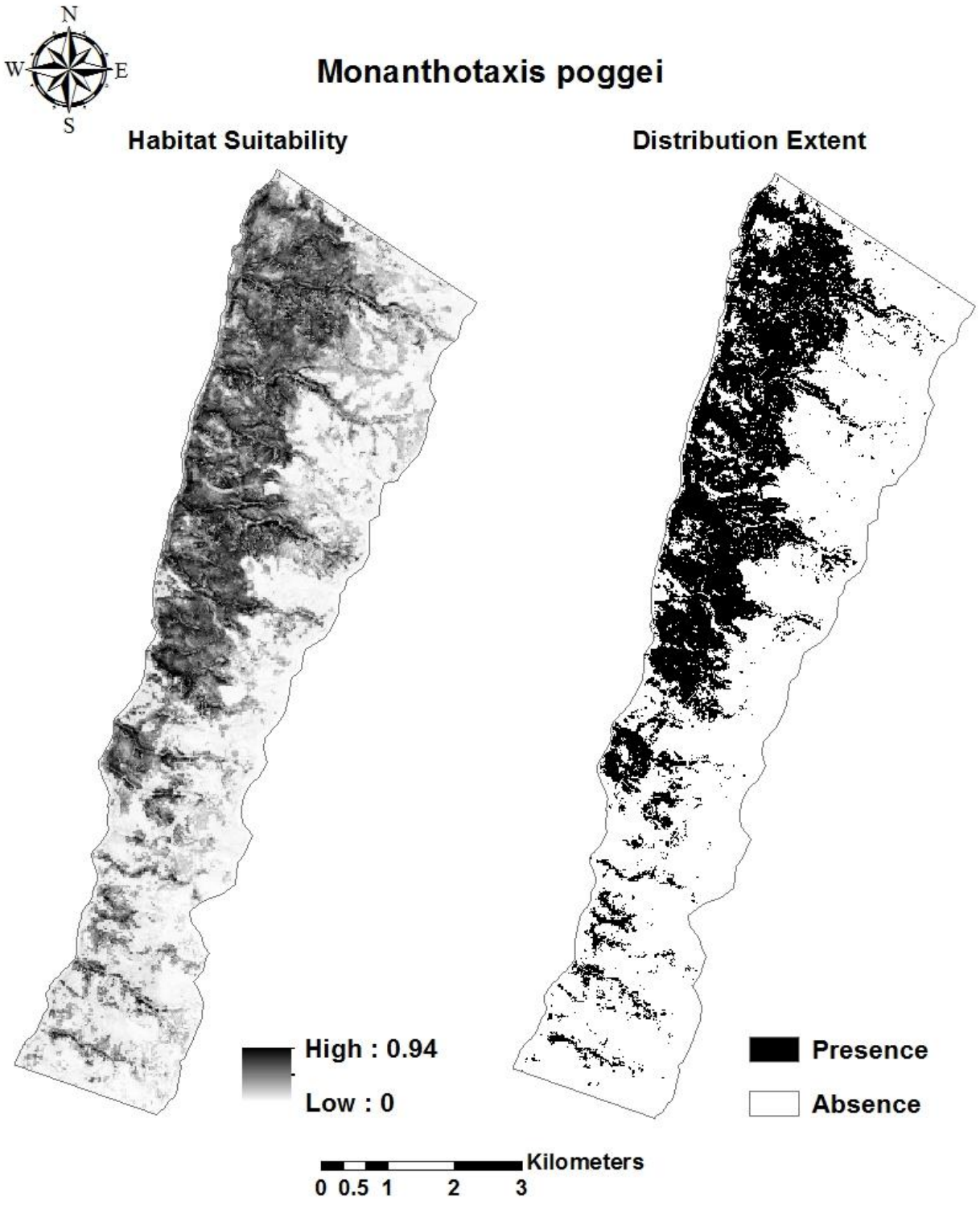

N
W
E
S
Garcinia huillensis
Habitat Suitability
Distribution Extent
High : 0.91
Low : 0
Presence
Absence
Kilometers
0 0.5 1 2 3

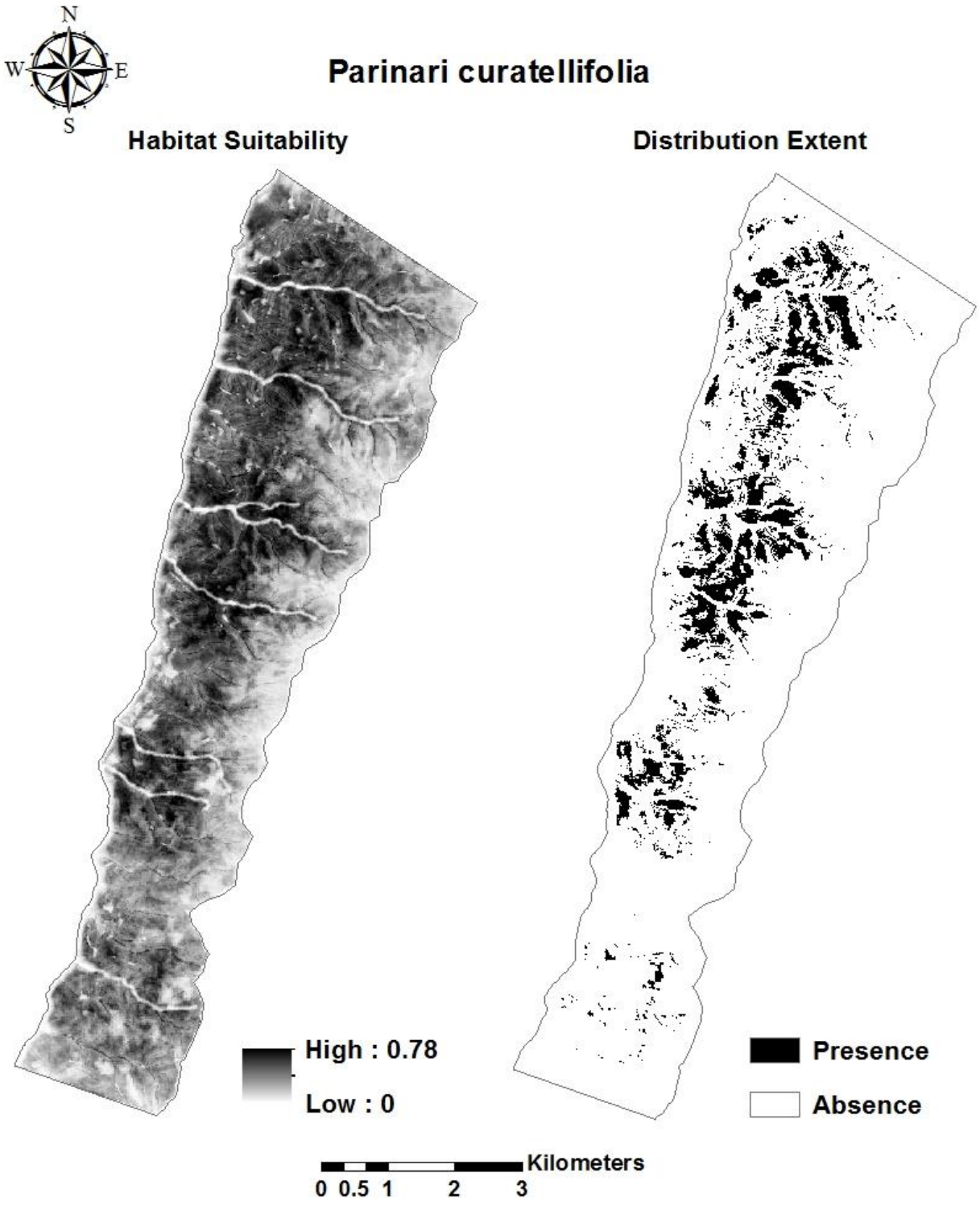

N
W
E
S
Grewia platyclada
Habitat Suitability
Distribution Extent
High : 0.96
Low : 0
Presence
Absence
Kilometers
0 0.5 1 2 3

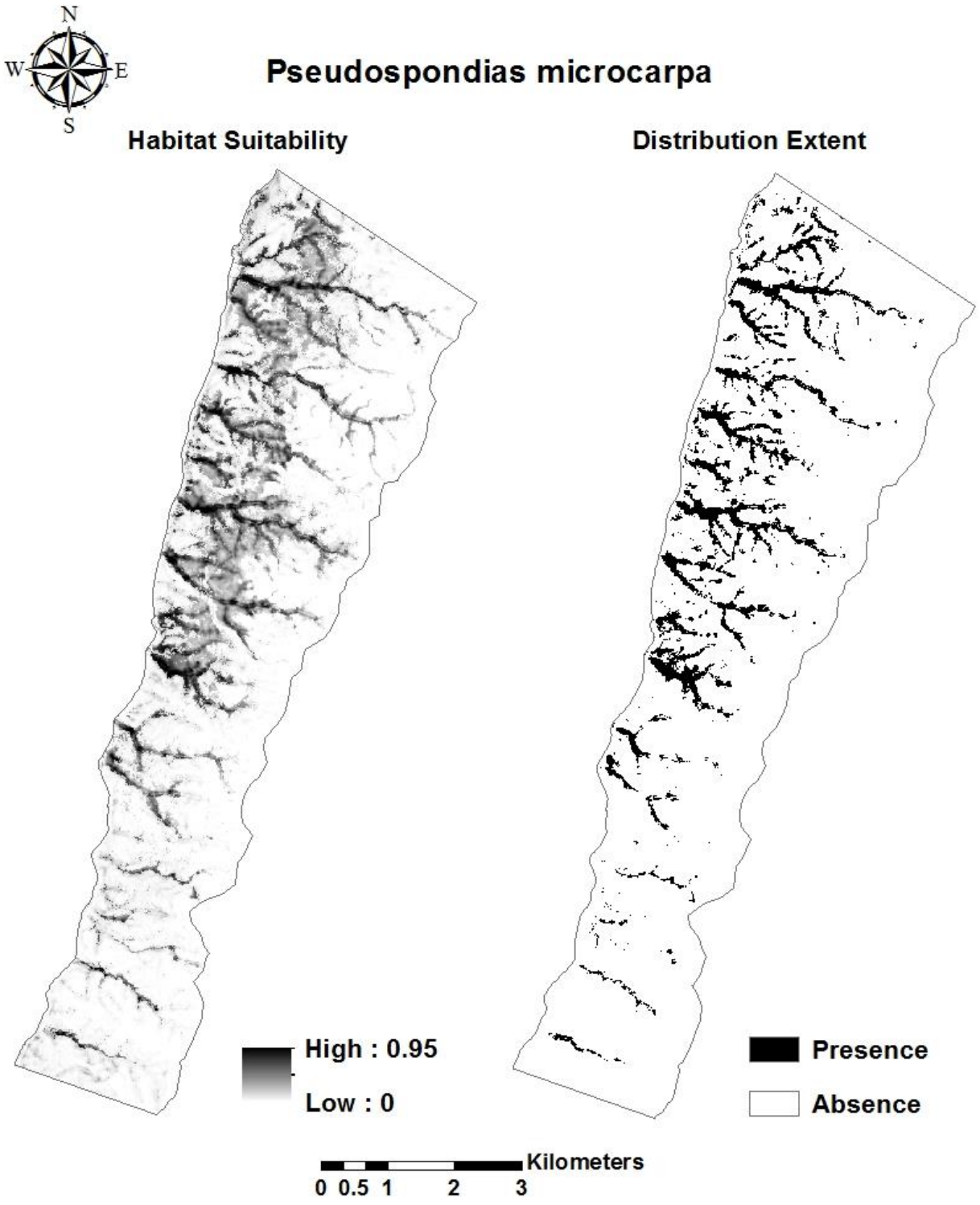

N
W
E
S
Harungana madagascariensis
Habitat Suitability
Distribution Extent
High : 0.88
Low : 0
Presence
Absence
Kilometers
0 0.5 1 2 3

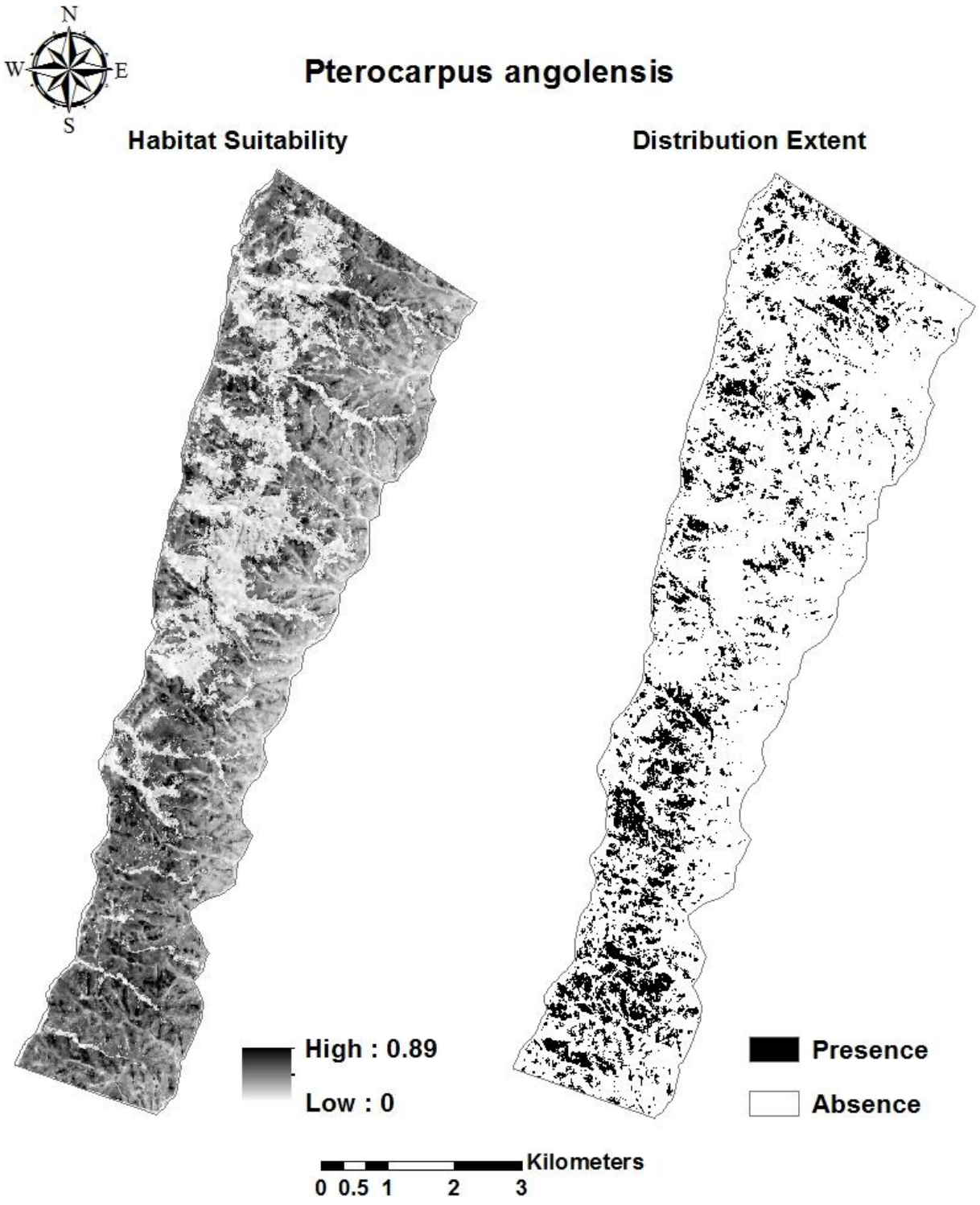

N
W
E
S
Landolphia lucida
Habitat Suitability
Distribution Extent
High : 0.89
Low : 0
Presence
Absence
Kilometers
0 0.5 1 2 3

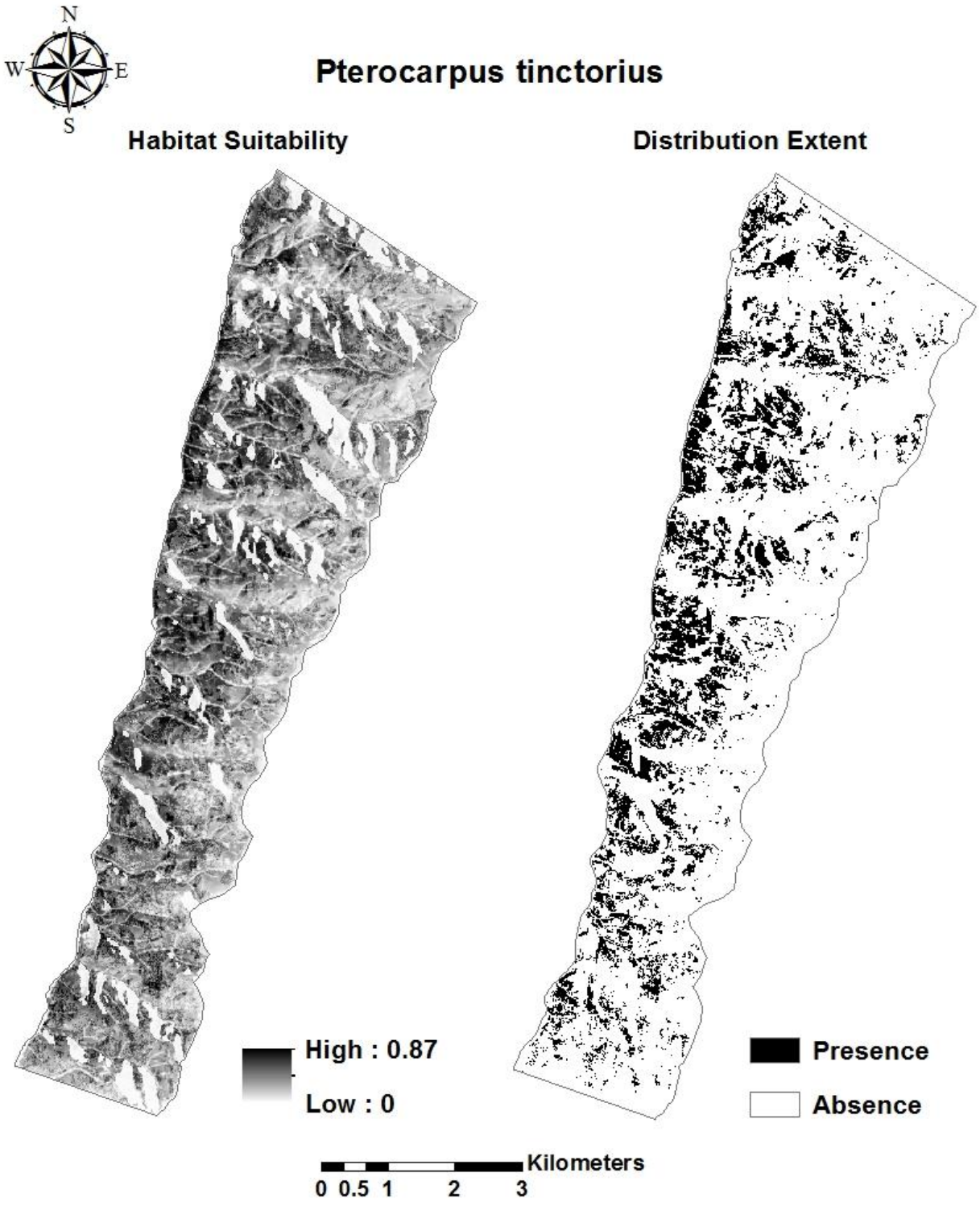

N
W
E
S
Mellera lobulata / Hypoestes verticillaris
Habitat Suitability
Distribution Extent
High : 0.94
Low : 0
Presence
Absence
Kilometers
0 0.5 1 2 3

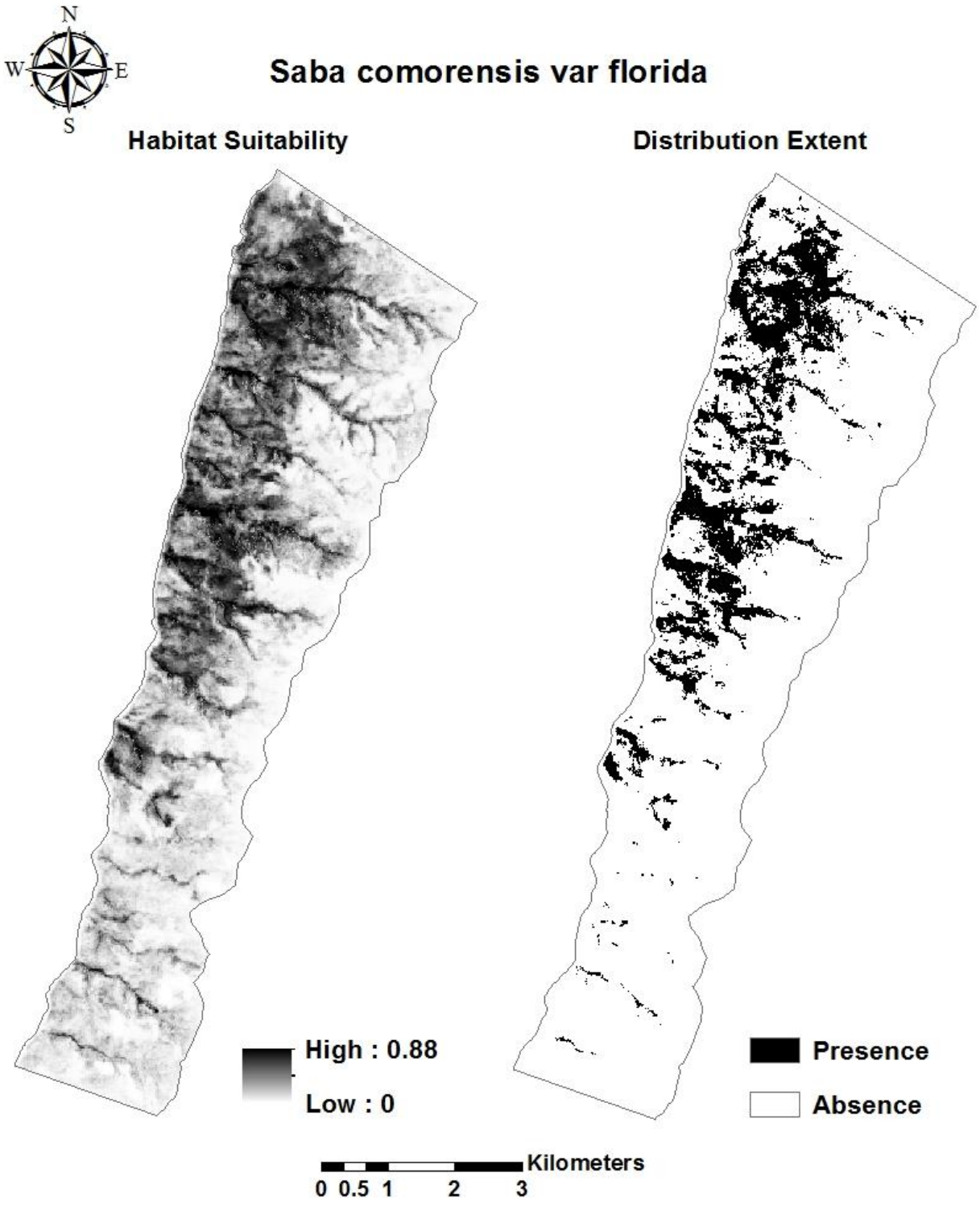

N
W
E
S
Monanthotaxis poggei
Habitat Suitability
Distribution Extent
High : 0.94
Low : 0
Presence
Absence
Kilometers
0 0.5 1 2 3

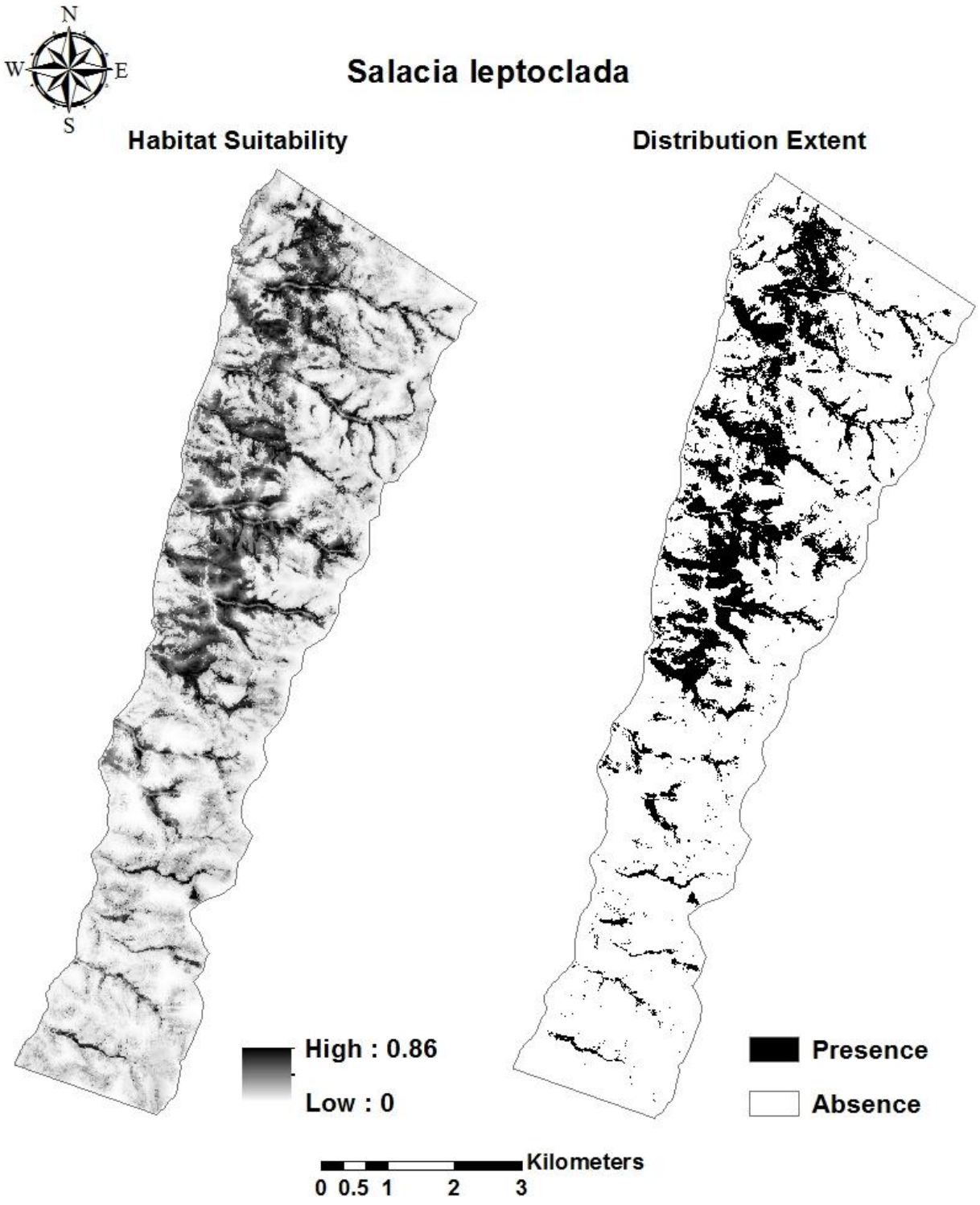

N
W
E
S
Parinari curatellifolia
Habitat Suitability
Distribution Extent
High : 0.78
Low : 0
Presence
Absence
Kilometers
0 0.5 1 2 3

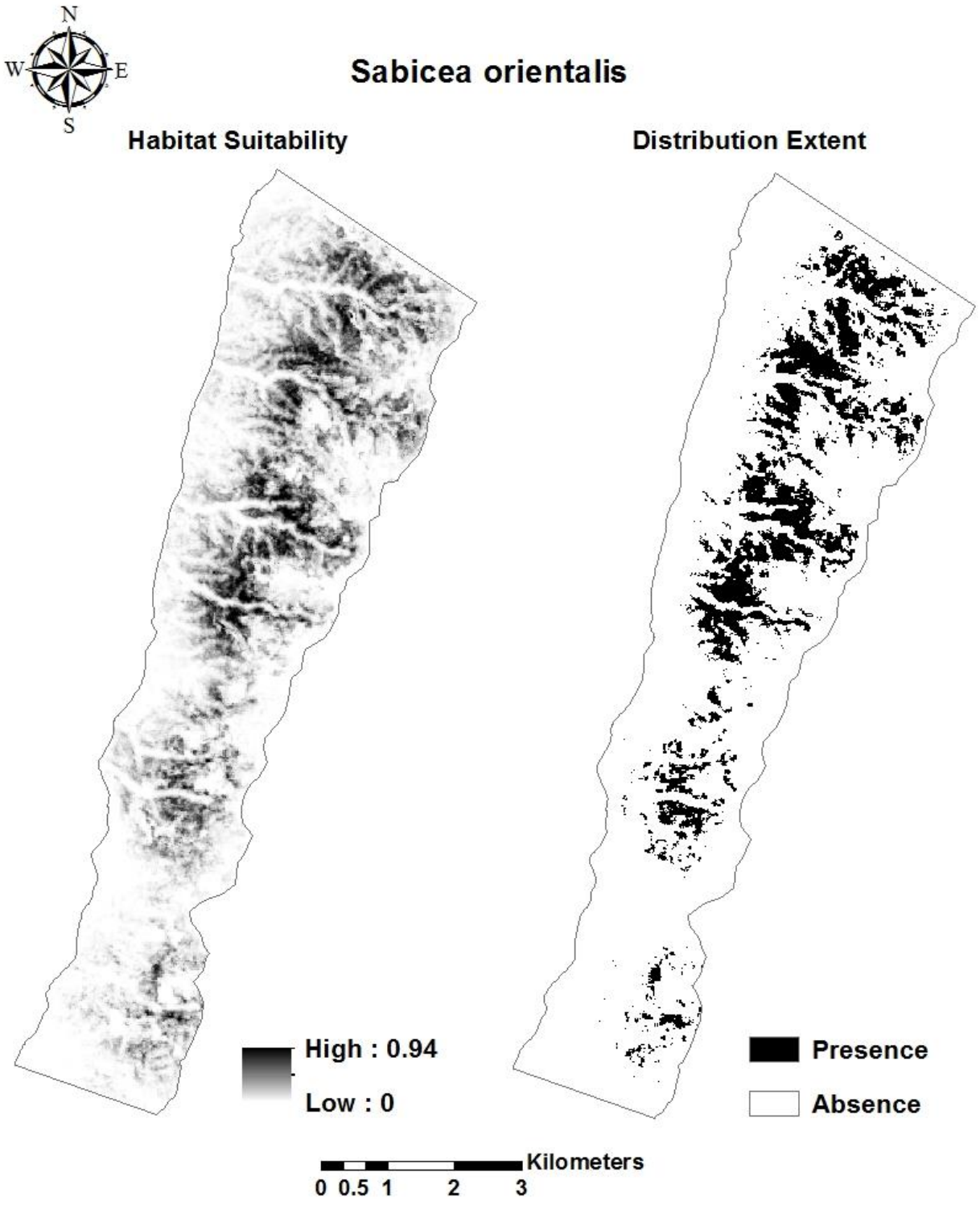

N
W
E
S
Pseudospondias microcarpa
Habitat Suitability
Distribution Extent
High : 0.95
Low : 0
Presence
Absence
Kilometers
0 0.5 1 2 3

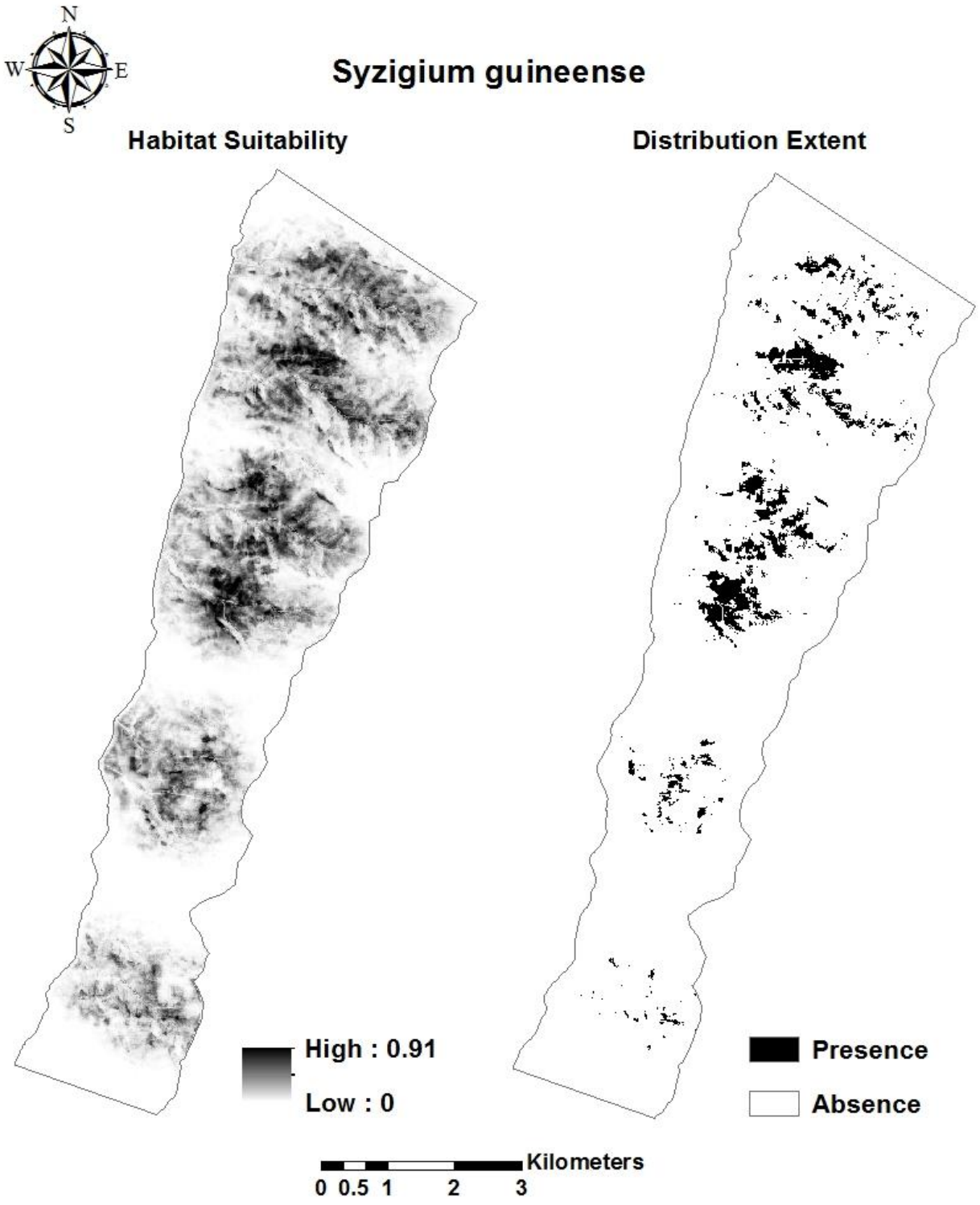

N
W
E
S
Pterocarpus angolensis
Habitat Suitability
Distribution Extent
High : 0.89
Low : 0
Presence
Absence
Kilometers
0 0.5 1 2 3

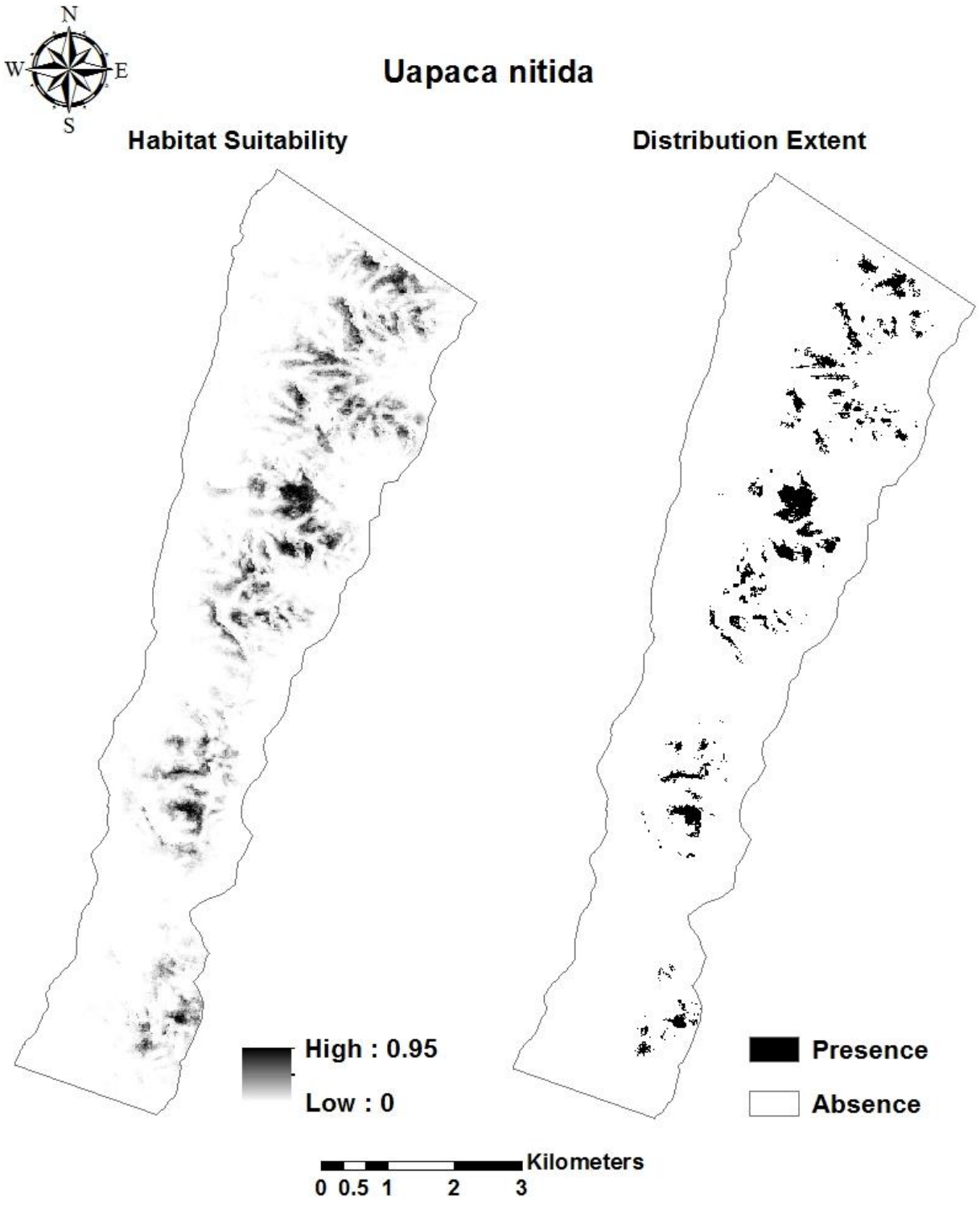

N
W
E
S
Pterocarpus tinctorius
Habitat Suitability
Distribution Extent
High : 0.87
Low : 0
Presence
Absence
Kilometers
0 0.5 1 2 3

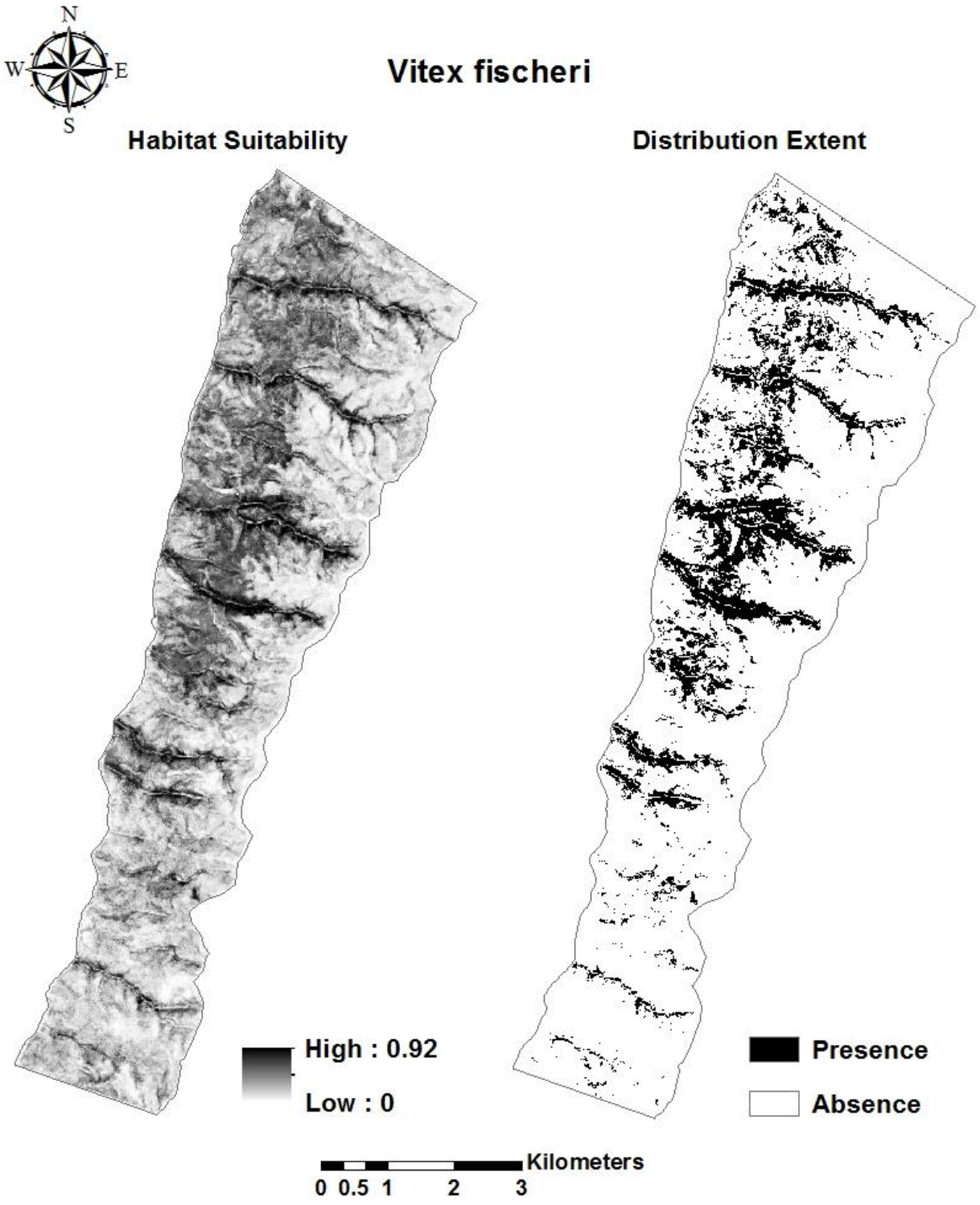

N
W
E
S
Saba comorensis var florida
Habitat Suitability
Distribution Extent
High : 0.88
Low : 0
Presence
Absence
Kilometers
0 0.5 1 2 3


N
W
E
S
Salacia leptoclada
Habitat Suitability
Distribution Extent
High : 0.86
Low : 0
Presence
Absence
Kilometers
0 0.5 1 2 3


N
W
E
S
Sabicea orientalis
Habitat Suitability
Distribution Extent
High : 0.94
Low : 0
Presence
Absence
Kilometers
0 0.5 1 2 3


N
W
E
S
Syzigium guineense
Habitat Suitability
Distribution Extent
High : 0.91
Low : 0
Presence
Absence
Kilometers
0 0.5 1 2 3


N
W
E
S
Uapaca nitida
Habitat Suitability
Distribution Extent
High : 0.95
Low : 0
Presence
Absence
Kilometers
0 0.5 1 2 3


N
W
E
S
Vitex fischeri
Habitat Suitability
Distribution Extent
High : 0.92
Low : 0
Presence
Absence
Kilometers
0 0.5 1 2 3

APPENDIX 3 *Vegetation species considered in cluster analysis.*

Vegetation Species	**Vegetation Species**
Ampelocissus cavicaulis	*Mellera lobulata/Hypoestes verticillaris*
Annona senegalensis	*Monanthotaxis poggei*
Antiaris toxicaria	*Panicum maximum*
Antidesma venosum	*Parinari curatellifolia*
Baphia capparidifolia	*Pseudospondias microcarpa*
Brachystegia spp.	*Pterocarpus angolensis*
Canthium hispidum venosum	*Pterocarpus tinctorius*
Diplorhynchus condylocarpon	*Saba comorensis var florida*
Discorea sp.	*Sabicea orientalis*
Ficus.sp.	*Salacia leptoclada*
Garcinia huillensis	*Syzigium guineense*
Grewia platyclada	*Uapaca nitida*
Harungana madagascariensis	*Vitex fischeri*
Landolphia lucida	

APPENDIX 4 *Spectral separability statistics of 10 vegetation classes.*

Group2 and Group3 - 0.23996525	*Group1 and Group3 - 1.36579588*
Group1 and Group6 - 0.49261361	*Group1 and Group4 - 1.40685445*
Group2 and Group6 - 0.49445032	*Group4 and Group7 - 1.41562523*
Group3 and Group8 - 0.53434493	*bare and Group7 - 1.60114892*
Group2 and Group8 - 0.53558947	*Group5 and Shadow - 1.67604924*
Group8 and Group10 - 0.54777226	*Group1 and Group9 - 1.73800155*
Group3 and Group9 - 0.54823241	*Group5 and Group9 - 1.76353508*
Group6 and Group10 - 0.54998701	*bare and Group9 - 1.76467967*
Group2 and Group9 - 0.59852372	*Group2 and Shadow - 1.76527303*
Group1 and Group10 - 0.63998919	*Group1 and Shadow - 1.77268372*
Group7 and Group10 - 0.69917606	*Group6 and Shadow - 1.78688340*
Group2 and Group4 - 0.72187085	*Group4 and Group5 - 1.78818688*
Group6 and Group8 - 0.72732522	*Group3 and Shadow - 1.83362212*
Group3 and Group4 - 0.73620471	*bare and Group3 - 1.86397040*
Group7 and Group8 - 0.74376731	*bare and Group2 - 1.88895276*
Group4 and Group8 - 0.74673107	*Group8 and Shadow - 1.89683890*
Group3 and Group6 - 0.74991983	*Group7 and Shadow - 1.89992636*
Group6 and Group7 - 0.78640710	*bare and Group6 - 1.92960264*
Group3 and Group7 - 0.80519477	*Cloud and Group3 - 1.93277321*
Group5 and Group7 - 0.82288837	*Group4 and Shadow - 1.93840651*
Group5 and Group10 - 0.84659787	*Cloud and Group7 - 1.93867906*
Group2 and Group10 - 0.86329162	*Cloud and Group2 - 1.94357989*
Group2 and Group7 - 0.89727789	*Cloud and Group9 - 1.94852556*
Group7 and Group9 - 0.97665115	*Cloud and Group6 - 1.95041595*
Group3 and Group10 - 0.98125709	*Cloud and Group10 - 1.95869881*
Group4 and Group6 - 1.00221120	*Cloud and Group5 - 1.95992260*
Group8 and Group9 - 1.00876682	*Cloud and Group4 - 1.96230222*
Group1 and Group8 - 1.02600222	*Cloud and Group8 - 1.96413765*
Group4 and Group10 - 1.07221088	*Group10 and Shadow - 1.97126537*
Group4 and Group9 - 1.09418194	*bare and Group5 - 1.98113745*
Group6 and Group9 - 1.14227928	*bare and Group10 - 1.98339508*
Group5 and Group6 - 1.15105873	*bare and Group8 - 1.98547369*
Group1 and Group2 - 1.16151879	*Cloud and Group1 - 1.98744677*
Group1 and Group7 - 1.16561133	*Group9 and Shadow - 1.98990730*
Group5 and Group8 - 1.16935710	*bare and Cloud - 1.99746981*
Group3 and Group5 - 1.27015970	*bare and Group4 - 1.99785646*
Group2 and Group5 - 1.27578312	*bare and Group1 - 1.99809388*
Group1 and Group5 - 1.28365415	*bare and Shadow - 1.99953439*
Group9 and Group10 - 1.32394100	*Cloud and Shadow - 1.99998718*