

## Predicting the habitat usage of African black rhinoceros (*Diceros bicornis*) using random forest models

Lucy Lush<sup>1,2\*</sup>, Martin Mulama<sup>3</sup> and Martin Jones<sup>2</sup>

<sup>1</sup>Centre for Environmental and Marine Sciences, University of Hull, Scarborough campus, Filey Road, YO11 3AZ., Scarborough, U.K., <sup>2</sup>Division of Biology and Conservation Ecology, Faculty of Science & Engineering, Manchester Metropolitan University, John Dalton Building, Chester Street, Manchester, M1 5GD, U.K. and <sup>3</sup>Ol Pejeta Conservancy, Private Bag, Nanyuki, 10400, Kenya

### Abstract

Species distribution models are often used in ecology to ascertain relationships between environmental variables and species presence. Modelling to understand this relationship can be used to aid conservation management strategies. In this paper, we applied the random forest classification method to predict habitat used by black rhino for browsing. The random forest model was created using detailed habitat data collected from Ol Pejeta Conservancy in Kenya. Variables from plots where rhino had been present were compared to those not used by rhino. Independent data were used to test the predictive accuracy of the rules generated. The model performed well with the independent test data, correctly classifying 69% of the sampling plots where black rhino were present. Important habitat features that affected rhino presence were browse availability and density of vegetation, with *Vachellia drepanolobium* (formerly *Acacia*) and *Euclea divinorum* being important components. The analysis also highlighted areas of potential high browse pressure, which should be the focus of continued monitoring and management.

**Key words:** black rhino, browsing, habitat, random forest, savannah

### Résumé

En écologie, on utilise souvent des modèles de distribution des espèces pour établir les relations entre des variables environnementales et la présence d'espèces. Pour comprendre cette relation, la modélisation peut être utilisée pour aider les stratégies de gestion de la conservation. Dans cet article, nous avons appliqué la méthode de classement par forêt aléatoire pour prédire quel habitat fréquente le rhino noir pour se nourrir. Le modèle de

classement par forêt aléatoire a été créé en utilisant des données détaillées sur l'habitat, collectées dans l'Aire de conservation d'Ol Pejeta, au Kenya. Les variables de parcelles où le rhino était présent furent comparées à celles d'autres, non fréquentées par le rhino. Des données indépendantes ont permis de tester la justesse prédictive des règles générées. Le modèle donnait de bons résultats avec les données indépendantes du test, classant correctement 69 % des parcelles échantillons où le rhino était présent. D'importantes caractéristiques de l'habitat qui affectaient la présence du rhino étaient la disponibilité de la nourriture adéquate et la densité de la végétation, dont *Vachellia drepanolobium* (anc. *Acacia*) et *Euclea divinorum* sont des éléments importants. L'analyse a aussi mis en lumière les zones où la pression de la consommation des rhinos risque d'être élevée, ce qui devrait être le point focal de la poursuite du suivi et de la gestion.

### Introduction

Modelling to understand the presence of animals based on habitat data has been utilized extensively by ecologists to aid conservation management strategies (Manel, Williams & Ormerod, 2001; Stockwell & Peterson, 2002). However, errors in the accuracy of predicting species presence have been found in various studies, resulting in inappropriate management strategies (Manel, Williams & Ormerod, 2001; Loiselle *et al.*, 2003; Araújo & Guisan, 2006). The ability to accurately predict suitable habitat for endangered species becomes more important as habitat fragmentation increases, and suitable habitat for many taxa is dramatically reduced.

The techniques most often used to model species distribution are artificial neural networks, discriminant

\*Correspondence: E-mail: llush@hotmail.co.uk

analysis and generalized linear models, such as, logistic regression (Barreto *et al.*, 1998; Özesmi & Özesmi, 1999; Liaw & Wiener, 2002; Labonne, Allouche & Gaudin, 2003; Edwards Jr. *et al.*, 2006). More recently, machine learning methods have been applied, including support vector machines and classification trees, which includes random forest (Debeljak *et al.*, 2001; Olivier & Wotherspoon, 2005; Drake, Randin & Guisan, 2006; Edwards Jr. *et al.*, 2006). Comparisons between these different types of distribution models have shown random forest to be a more accurate algorithm in predicting species habitat suitability (Stockwell & Peterson, 2002; Edwards Jr. *et al.*, 2006; Garzón *et al.*, 2006; Marmion *et al.*, 2009).

Random forest models (Breiman & Cutler, 2001) have been used extensively within bioinformatics as a predictive tool. The advantages to this algorithm are the ability to handle large data sets that contain various data types and deal with outliers without biasing results (Bell, 1999; Fielding, 2007). It can also cope with highly correlated variables, small sample sizes with larger numbers of predictor variables and nonlinear associations between response and predictor variables without the user having to specify a model structure (Boulesteix *et al.*, 2012). Therefore, random forest models were more suitable for this type of data set compared to other methods, such as generalized linear mixed models (GLMM) that require a model structure to be specified and a number of model assumptions to be met (Zuur *et al.*, 2009). Perhaps more significantly, it can estimate the importance of each variable as a classifier, which is a valuable tool to aid ecological studies (Garzón *et al.*, 2006; Fielding, 2007). However, it is still a relatively new technique; therefore, further studies to fully understand the theoretical aspect of the mathematical algorithms and how it deals with different data are needed (Boulesteix *et al.*, 2012).

The black rhino (*Diceros bicornis*, Linnaeus 1758) is a critically endangered species with a population of 4880 in Africa and, approximately, 594 in Kenya (Emslie, 2011). High levels of poaching have been the main cause of the decline, but increased density of herbivore browsing and changes to habitat structure caused by human encroachment have also reduced black rhino numbers (Walpole *et al.*, 2001; Rice & Jones, 2006). Therefore, being able to assess potential areas of high browsing pressure and understanding important habitat characteristics could be beneficial in black rhino conservation.

The aim of this study was to identify important habitat variables to predict habitat use of black rhino through applying the random forest model to an ecological situation. Classifying the important habitat factors for black rhino could enable more effective management of habitat and highlight areas of potential browsing pressure. This could aid assessment of carrying capacities of reserves, which has been prioritized in the National Rhino Strategy (Kenya Wildlife Service, 2012), not only for current populations, but for future translocations.

## Materials and methods

### Study area

The research was carried out at Ol Pejeta Conservancy, located in the Laikipia region of Kenya, from May until August 2007. Ol Pejeta Conservancy (OPC) was established to protect black rhino within a small enclosed 9700 ha reserve as part of Kenya Wildlife Service's rhino management plan. Numbers of black rhino had been increasing by a rate of 10% each year since 1994, which resulted in increased levels of competition and browsing pressure with other large herbivores (Brett, 1993). This reduced the potential carrying capacity of black rhino within the reserve to 40 (Birkett, 2002; Birkett & Stevens-Wood, 2005). Therefore, the decision was made to permanently enlarge the reserve to 36,500 ha, during March 2007, and incorporate the neighbouring ranch area (Patton *et al.*, 2010). The potential carrying capacity increased to 120 (Amin *et al.*, 2006) and a further 27 black rhino were translocated from Solio Game Reserve into the new area. This provided the opportunity to model factors that affected the habitats used for browsing and to predict how the newly released rhino would utilize the new habitat. Therefore, identifying areas of potential high browsing pressure, which is an important factor for black rhino conservation.

### Habitat data

A total of 41 plots were sampled within the reserve, where black rhino have been well established since 1989, and these formed the training data for the model. Within the ranch, where the translocated rhinos were released, a total of 26 plots were sampled. These plots provided the test data to investigate whether the model can predict the habitats

likely to be used by black rhino for foraging and where browsing pressure may be high.

Habitat plots were randomly chosen using the UTM gridlines that were spaced 2500 m apart both vertically and horizontally across the map, to ensure plots were spatially independent. Where these lines crossed, a plot was completed. This was repeated for both the reserve and the ranch. Each plot was 20 m × 20 m square, a tree was chosen at random as the centre tree, close to where the gridlines crossed, and a GPS coordinate was recorded. Using a tape measure and compass 10 m were walked north from the centre tree and 10 m south and then again west and east to mark out the plot. This method was found to be the most efficient method of collecting habitat data (Astbury, 2010). The inclination of the land was estimated using a clinometer and the aspect of the slope recorded using a compass. All trees and bushes within the plot were identified, and the following measurements recorded: height, height to the lower limit of the main canopy, diameter at breast height (DBH) of the trunk, or the largest stem if multi-stemmed, the diameter of the canopy, general shape of the canopy, and percentage amount of leaf cover for each tree and bush within the plot. The shape was determined in terms of the basic geometric shape, that is an inverted triangle or a cylinder. For the bushes, which were sometimes dense and consisted of more than one species of tree, the number of stems were counted and measurements taken for the entire clump.

#### Rhino and Elephant data

Competition with other mega herbivores, such as elephants (*Loxodonta africana* Blumenbach, 1797) could affect habitat selection by rhino, as studies have shown that rhino may avoid areas where they are present (Morgan, Mackey & Slotow, 2009). High densities of elephants also increase browsing pressure (Pradhan *et al.*, 2008). Therefore, evidence of elephant browsing damage was recorded alongside rhino browse data. Browse damage was estimated on each tree by counting the number of damaged branches. Rhino browse damage was easily identified as they produce a clean cut on the stems, whereas elephants break the stems and strip the bark (Birkett & Stevens-Wood, 2005).

#### Habitat classification

ArcView GIS 3.2 (ESRI, California, U.S.A.) was used to map the study area; shape files were obtained from OPC of

the reserve boundaries, roads, settlements and water sources. Rhino sightings data, collected daily by Ol Pejeta Rhino Patrols, were acquired for 43 rhinos from the reserve and the 27 translocated rhinos for June and July 2007. Those that were seen feeding as opposed to running or walking were mapped onto ArcView and vegetation plots that had rhino sightings within 100 m of the centre of the plot were denoted as present. This was used in conjunction with browse damage to assess areas used by rhinos for feeding. The distances from each plot to the nearest road, settlement and water source were measured and recorded to assess if proximity to these affected where rhinos browsed. Each plot was classified into one of six habitats that were previously used to classify vegetation at OPC. These were, grass closed (no trees), shrubs closed (*Euclea divinorum* only), shrubs open (dominated by more than 60% *E. divinorum*), trees closed (*Vachellia drepanolobium* only, formerly *Acacia*), trees open (dominated by more than 60% *V. drepanolobium*) and mixed when it was equal.

#### Browse availability

Browse availability was calculated using data on the height, diameter, shape and height to canopy. Two main canopy shapes were used; the cone shape (volume =  $1/3 \pi r^2 h$ ), that typified the shape of *V. drepanolobium*, and cylinder for *E. divinorum* (volume =  $\pi r^2 h$ ). Rhinos can only reach up to a maximum of two metres to browse (Muya & Ouge, 2000; Amin *et al.*, 2006); therefore, only heights up to 2 m were included. A chi-squared test was used to assess if rhinos browsed certain trees more than others, and a Mann–Whitney *U*-test was used to compare habitat variables between the ranch and reserve and test if rhinos used a particular habitat type. All figures were calculated per sampling plot (400 m<sup>2</sup>), because of the difference in total area sampled between the ranch and reserve.

#### Random forest modelling

The reserve data set was used as the training data to build a random forest classification model using the 'R' software package (R Development Core Team, 2007, version 2.5) and RATTLE (R Analytical Tool To Learn Easily, Williams, 2011). The default parameters were used to produce 500 trees with six variables at each split (square root of available variables) (Williams, 2011). Half and twice the default parameters were tested as suggested by Liaw & Wiener (2002) but it did not make any difference to the

output. Fifty-four predictor variables were used and the presence of rhino browse damage and feeding sightings were combined to produce the binary dependant variable that denoted the use of that area by rhinos for both the reserve and the ranch.

The analysis produced a 'variable importance' statistic, which indicated which variables had the most influence on the outcome of the predictions, and how good they were at classifying between absent and present (Garzón *et al.*, 2006). The mean decrease of accuracy at each variable was calculated from the difference between the out-of-bag (OOB) prediction accuracy, across all trees, against the accuracy when that variable was permuted (Liaw & Wiener, 2002). The difference was then averaged across all trees. The OOB error rate is the number of times a case is misclassified averaged over all trees (Fielding, 2007).

Random forest makes predictions from the data based on the class with the majority vote (Liaw & Wiener, 2002). In this case, it was between two classes, present and absent. It also calculated the probabilities for each class based on the votes each case received for the two classes across all trees (Liaw & Wiener, 2002). Therefore, you can identify which classes are more definite, close to 1 or 0, and those that are more indistinct, because they cluster around 0.5. To test the predictive accuracy of the model, the independent ranch data set was used on the random forest model that was created using the reserve data. Therefore, the splitting rules created at each tree node of the classification trees using the training data were then applied to the ranch data to test whether absence and presence were able to be classified using the same rules.

## Results

### *Comparison of reserve and ranch habitat*

Habitat variables between the reserve and ranch were all similar with no significant differences between them (all comparisons with  $P > 0.05$ ,  $df = 67$ , using Mann–Whitney  $U$ -test). Therefore, the translocated rhinos were using very similar habitat. The most common habitat type in the sampled areas was shrubs open (reserve = 48.8%, ranch = 34.6%). Rhino browse damage was considerably higher in the reserve than the ranch ( $U = 172.00$ ,  $n = 67$ ,  $P = 0.001$ ) and predominately found in 'trees open' (mean trees browsed = 7.33,  $SD = 3.32$ ) and 'shrubs closed' habitat (mean trees browsed = 6.33,  $SD = 2.89$ ). Black rhino appear to browse some tree species more than others

( $\chi^2 = 31.837$ ,  $df = 9$ ,  $P = 0.001$ ). A comparison of the percentage of damaged to undamaged trees showed that *Scutia myrtina* (mean = 17.89%,  $SD = 23.84\%$ ) was browsed the most by rhinos, followed by *E. divinorum* (mean = 10.58%,  $SD = 18.66\%$ ) and *V. drepanolobium* (mean = 10.24%,  $SD = 13.84\%$ ).

### *Rhino presence*

Rhino presence was found in 32 plots in the reserve and thirteen plots in the ranch compared to nine plots with no evidence of rhino presence in the reserve and thirteen plots in the ranch. Rhino presence was significantly different between habitat types ( $U = 35.0$ ,  $n = 41$ ,  $P \leq 0.001$ ) with 59.38% of rhino presence found in shrubs open habitats compared to 77.78% of rhino absence found in grass closed. Table 1 shows all of the habitat variables that differ significantly between plots where black rhino were present or absent suggesting that rhinos have a strong habitat association.

### *Random forest modelling*

The model created using the reserve data classified 31 plots with rhino present and seven plots where they were predicted to be absent. The separation was very accurate with only two plots predicted as 'absent' when it was actually 'present' and one plot predicted as present when it was absent (Table 2). The 'OOB' error rate was only 7.32%.

The variables that had higher importance in predicting rhino presence were the amount of browse available, total number of trees/shrubs and density of vegetation. The habitat type, amount of elephant damage and number of trees damaged by elephants were also important (Fig. 1). Results in Table 1 suggested that rhino presence was associated with higher tree density and browse availability, and with areas that have larger amounts of elephant damage. It was also evident from Fig. 1 that *V. drepanolobium* and *E. divinorum* have a more influential role in predicting the presence of rhino compared to other tree species that had lower mean decrease accuracy figures.

When the model was applied to the independent test data from the ranch, the percentage of cases correctly predicted was 69% with an overall error rate of 0.31. There were eight areas where rhinos were predicted to be present but no browse signs or sightings were observed (Fig. 2). Black rhinos appeared to be using most of the ranch except areas on the north-western side, which were predicted to be used for browsing. The probability figures

**Table 1** Comparison of habitat variables that are significantly different between the presence and absence of black rhino. Standard deviations are in brackets

	Rhino presence Mean (n = 45)	Rhino absence Mean (n = 22)	Mann–Whitney U-test (n = 41)
Total density of vegetation <sup>a</sup> (trees per m <sup>2</sup> )	0.18 (0.85)	0.06 (0.69)	138.0 $P \leq 0.001$
<i>V. drepanolobium</i> density (trees per m <sup>2</sup> )	0.07 (0.54)	0.03 (0.04)	235.5 $P = 0.001$
<i>E. divinorum</i> density (trees per m <sup>2</sup> )	0.03 (0.03)	0.01 (0.03)	208.5 $P \leq 0.001$
Number of trees and shrubs <sup>a</sup>	73.1 (34.3)	23.3 (27.7)	138.0 $P \leq 0.001$
Number of <i>V. drepanolobium</i>	28.2 (21.7)	11.9 (15.9)	235.5 $P = 0.001$
Number of <i>E. divinorum</i>	10.5 (12.3)	4.09 (12.3)	208.5 $P \leq 0.001$
Total browse availability <sup>a</sup> (m <sup>3</sup> )	$16.2 \times 10^5$ ( $15.1 \times 10^5$ )	$4.5 \times 10^5$ ( $6.4 \times 10^5$ )	171.0 $P \leq 0.001$
<i>V. drepanolobium</i> browse available (m <sup>3</sup> )	$2.9 \times 10^5$ ( $2.7 \times 10^5$ )	$2.1 \times 10^5$ ( $2.2 \times 10^5$ )	351.0 $P = 0.05$
Amount of <i>E. divinorum</i> browse available (m <sup>3</sup> )	$8.1 \times 10^5$ ( $9.8 \times 10^5$ )	$0.4 \times 10^5$ ( $1.1 \times 10^5$ )	166.0 $P \leq 0.001$
Amount of <i>S. myrtina</i> browse available (m <sup>3</sup> )	$1.5 \times 10^5$ ( $2.4 \times 10^5$ )	136.4 (243.7)	315.0 $P = 0.013$
Mean diameter of trees (cm)	113.3 (48.1)	67.1 (74.2)	311.0 $P = 0.014$
Mean diameter of <i>E. divinorum</i> (cm)	151.6 (85.2)	17.4 (35.4)	133.5 $P \leq 0.001$
Mean height of <i>E. divinorum</i> (cm)	154.4 (87.9)	19.8 (40.9)	136.0 $P \leq 0.001$
Mean DBH of <i>E. divinorum</i> (cm)	2.71 (2.49)	0.45 (1.10)	219.0 $P = 0.001$
Number of trees with elephant damage <sup>a</sup>	7.07 (3.61)	3.32 (2.16)	227.0 $P \leq 0.001$
Amount of elephant damage <sup>a</sup> (Number of branches)	18.49 (13.85)	7.36 (3.84)	254.0 $P \leq 0.001$
Distance to nearest road (metres)	350.5 (266.3)	276.8 (256.7)	346.0 $P = 0.047$

<sup>a</sup>Variables that are highly important in the RF model.

**Table 2** Confusion matrix for the random forest model

Predicted	Actual		Class error
	Absent (no evidence of rhino)	Present (rhino present)	
Absent	7	2	0.22
Present	1	31	0.03

for the likelihood of presence and absence reveal very distinctive classes, indicating that the model made strong predictions for both absence and presence (mean probability for present classes = 0.881, range = 0.52–1, absent classes = 0.008, range = 0.006–0.01). The probabilities can also be used to map the areas with the highest predictions in terms of rhino presence. Figure 2c shows which areas were predicted by random forest to have a >90% chance of rhino presence and therefore potential high browsing pressure.

## Discussion

The classification of the reserve plots into rhino present and rhino absent based on habitat data was very good, with a

low error rate. This suggests that the random forest model created should be a reliable predictor. When the independent test data from the ranch were run through the model it performed well, correctly predicting 69% of black rhino presence within the ranch. This is in line with other studies where random forest has been used to predict presence and was shown to perform better than other model types including; Generalized Linear Models, Generalized Additive Models and other classification and regression trees (Stockwell & Peterson, 2002; Edwards Jr. *et al.*, 2006; Garzón *et al.*, 2006; Prasad, Iverson & Liaw, 2006). Sites where presence was predicted but no signs were found might be areas of suitable habitat that could be occupied in the future. The rhino population had not reached the ranch's potential carrying capacity (Amin *et al.*, 2006); therefore, not all suitable habitats were likely being used by the current rhino numbers. Data collected by the rhino patrols revealed the translocated rhinos had dispersed across the ranch from the release sites and individuals had already begun to occupy certain areas. Grass closed habitats were the least used by rhino, which corresponds with the findings by Tatman, Stevens-Wood & Smith (2000) that found open grassland were the least favoured habitat.

The random forest variable importance measure indicated that browse availability and the density of vegetation



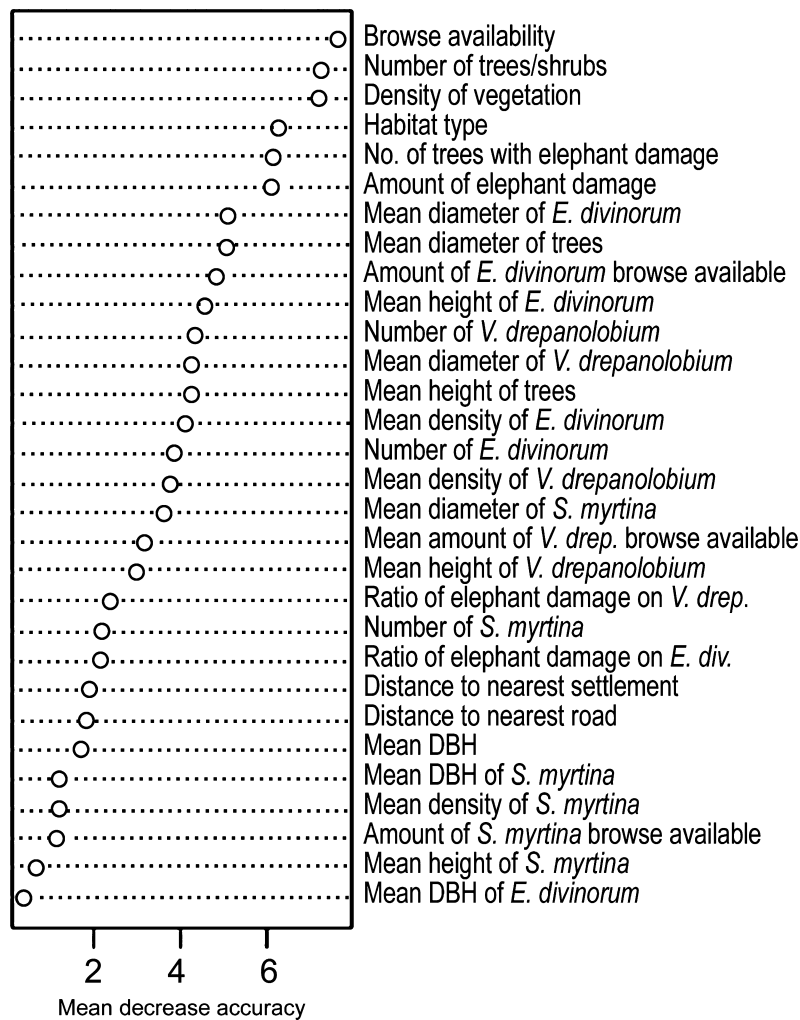
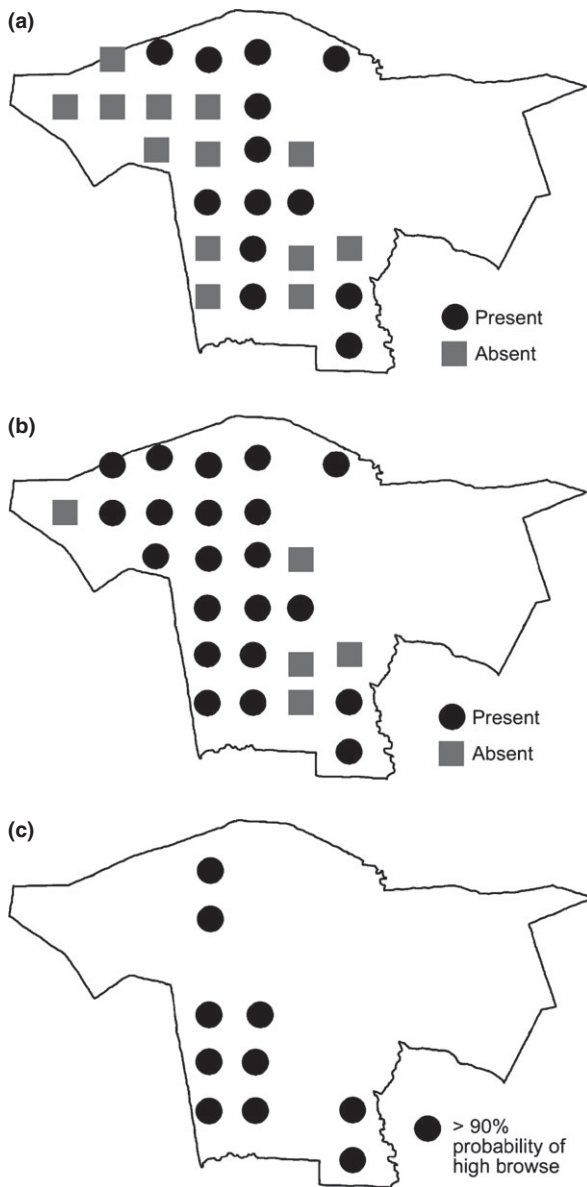


Fig 1 Variable importance plot generated by random forest algorithm. Showing the ranked importance of the model’s predictive factors, measured using the mean decrease accuracy

(particularly that of *V. drepanolobium* and *E. divinorum*) were important factors in the modelling. These were also found to be important in assessing carrying capacity (Amin *et al.*, 2006), therefore, highlighting the need to identify areas of potential high browse pressure. It also reflects the fact that *V. drepanolobium*, *E. divinorum* and the rarer *S. myrtina* supported the highest amounts of rhino browse. Rhino presence was predominately found within ‘trees open’ and ‘shrubs closed’ habitats. This has also been found by other studies (Kotze & Zacharias, 1993; Tatman, Stevens-Wood & Smith, 2000; Rice & Jones, 2006); however, Tatman, Stevens-Wood & Smith (2000) found that rhino generally avoided ‘trees open’ habitats. The

difference in findings could be explained by the signs used to measure habitat use. In their study, they did not include browse damage in the assessment, and as *V. drepanolobium* constitutes a high proportion of the black rhino’s diet (Oloo, Brett & Young, 1994; Birkett & Stevens-Wood, 2005) it is likely to be important in assessing habitats used by rhino for browsing. Therefore, the slight difference in results could be more indicative of the type of habitat used for different purposes. Kotze & Zacharias (1993) predicted that if ‘trees open’ changed to a closed habitat it would be detrimental to rhino, as trees would become taller and reduce access to browse. This emphasises the need to maintain levels of available browse by ensuring a mix of



**Fig 2** The recorded habitat use within the ranch, denoted by browse damage and rhino patrol sightings (a), and predicted habitat use (b) of black rhino, using the random forest model. Areas that are predicted to have high levels of rhino browse in the future (c)

*V. drepanolobium* and *E. divinorum* at the appropriate density, as well as *S. myrtina*. The distance to human disturbance had some importance in the habitats used by rhino, which was also found by other studies (Mukinya, 1973; Tatman, Stevens-Wood & Smith, 2000), but the

distance to water sources did not appear to greatly impact rhino presence.

Elephant browse damage was also an important indicator of rhino presence, that is both species exploited the same areas. This highlights potential competition between the species for forage (Pradhan *et al.*, 2008). Elephants are also capable of altering the habitat and reducing the number of trees creating more open habitats (Laws, 1970), which could be less suitable for rhino in the long term. Therefore, elephant populations within the newly opened ranch area need to be monitored closely.

## Conclusion

Random forest modelling utilized detailed vegetation type and structure data to predict habitats used by black rhino for browsing. It also highlighted areas of potential high browsing pressure that may need to be closely monitored, and areas that were likely to be used by rhino. Further research within the ranch could be implemented to continually monitor the levels of browse available and monitor the effects of elephants on vegetation structure. Whilst the use of random forest models within ecological studies is still in its infancy it does have the potential to be a useful predictive tool when using large datasets and applying it to ecological problems that could benefit future conservation management.

## Acknowledgements

I would like to thank Ol Pejeta Conservancy and their staff for their assistance and permission to carry out the research on the reserve, as well as supplying GIS data. The project was partly funded by The North of England Zoological Society studentship. A special thank you to Jes Graham, Rose Argall, Emma Tecwyn and Martina Ozan for their help in collecting many hours of tree data.

## References

- AMIN, R., OKITA-OUA, B., ADCOCK, K., EMSLIE, R.H., MULAMA, M. & PEARCE-KELLY, P. (2006) An integrated management strategy for the conservation of Eastern black rhinoceros *Diceros bicornis michaeli* in Kenya. *Int. Zoo Yearb.* **40**, 118–129.
- ARAÚJO, M.B. & GUISAN, A. (2006) Five (or so) challenges for species distribution modelling. *J. Biogeogr.* **33**, 1677–1688.

- ASTBURY, E. (2010) *Managing and monitoring endangered species with particular reference to the black rhino in Kenya*. Manchester Metropolitan University, PhD thesis.
- BARRETO, G.R., RUSHTON, S.P., STRACHAN, R. & MACDONALD, D.W. (1998) The role of habitat and mink predation in determining the status and distribution of water voles in England. *Anim Conserv.* **1**, 129–137.
- BELL, J.F. (1999) Tree-based methods. In: *Machine Learning Methods for Ecological Applications* (Ed. A. H. FIELDING). Kluwer Academic Publishers, Boston, Dordrecht, London.
- BIRKETT, A. (2002) The impact of giraffe, rhino and elephant on the habitat of a black rhino sanctuary in Kenya. *Afr. J. Ecol.* **40**, 276–282.
- BIRKETT, A. & STEVENS-WOOD, B. (2005) Effect of low rainfall and browsing by large herbivores on an enclosed savannah habitat in Kenya. *Afr. J. Ecol.* **43**, 123–130.
- BOULESTEIX, A.L., JANITZA, S., KRUPPA, J. & KÖNIG, I.R. (2012) Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *WIREs Data Mining Knowl. Discov.* **2**, 493–507.
- BREIMAN, L. & CUTLER, A. (2001) Random forests. *Mach. Learn.* **45**, 5–32.
- BRETT, R.A. (1993) *Conservation Strategy and Management Plan for the Black Rhinoceros (Diceros bicornis) in Kenya*. Kenya Wildlife Service, Nairobi. pp 1–105.
- DEBELJAK, M., DŽEROSKI, S., JERINA, K., KOBLER, A. & ADAMIČ, M. (2001) Habitat suitability modelling for red deer (*Cervus elaphus* L.) in South-central Slovenia with classification trees. *Ecol. Model.* **138**, 321–330.
- DRAKE, J.M., RANDIN, C. & GUISAN, A. (2006) Modelling ecological niches with support vector machines. *J. Appl. Ecol.* **43**, 424–432.
- EDWARDS, T.C. JR., CUTLER, D.R., ZIMMERMAN, N.E., GEISER, L. & MOISEN, G. (2006) Effects of sample survey design on the accuracy of classification tree models in species distribution models. *Ecol. Model.* **199**, 132–141.
- EMSLIE, R. (2011) *Diceros bicornis*. In: *IUCN 2012. IUCN Red List of Threatened Species*. Version 2012. Available at: [www.iucnredlist.org](http://www.iucnredlist.org) (Accessed on 14 October 2012).
- FIELDING, A. (2007) *Cluster and Classification Techniques for the Biosciences*. Cambridge University Press, Cambridge.
- GARZÓN, M.B., BLAZEK, R., NETELER, M., SÁNCHEZ DE DIOS, R., OLLERO, H.S. & FURLANELLO, C. (2006) Predicting habitat suitability with machine learning models: the potential area of *Pinus sylvestris* L. in the Iberian Peninsula. *Ecol. Model.* **197**, 383–393.
- Kenya Wildlife Service (2012) *Conservation and Management Strategy for the Black Rhino (D. b. michaeli) in Kenya*, (2012–2016), 5th edn. Kenya Wildlife Service, Nairobi, Kenya, pp. 57.
- KOTZE, D.C. & ZACHARIAS, P.J.K. (1993) Utilization of woody browse and habitat by the black rhino (*Diceros bicornis*) in western Itala game reserve. *Afr. J. Range Forage.* **10**, 36–40.
- LABONNE, J., ALLOUCHE, S. & GAUDIN, P. (2003) Use of a generalised linear model to test habitat preferences: the example of *Zingel asper*, an endemic endangered percid of the River Rhône. *Freshwater Biol.* **48**, 687–697.
- LAWS, R.M. (1970) Elephants as agents of habitat and landscape change in East Africa. *Oikos* **21**, 1–15.
- LIAW, A. & WIENER, M. (2002) Classification and regression by randomForest. *R. News* **2/3**, 18–22.
- LOISELLE, B.A., HOWELL, C.A., GRAHAM, C.H., GOERCK, J.M., BROOKS, T., SMITH, K.G. & WILLIAMS, P.H. (2003) Avoiding pitfalls of using species distribution models in conservation planning. *Conserv. Biol.* **17**, 1591–1600.
- MANEL, S., WILLIAMS, H.C. & ORMEROD, S.J. (2001) Evaluating presence-absence models in ecology: the need to account for prevalence. *J. Appl. Ecol.* **38**, 921–931.
- MARMION, M., PARVIAINEN, M., LUOTO, M., HEIKKINEN, R.K. & THUILLER, W. (2009) Evaluation of consensus methods in predictive species distribution modelling. *Divers. Distrib.* **15**, 59–69.
- MORGAN, S., MACKAY, R.L. & SLOTOW, R. (2009) A priori valuation of land use for the conservation of black rhinoceros (*Diceros bicornis*). *Biol. Conserv.* **142**, 384–393.
- MUKINYA, J.G. (1973) Density, distribution, population structure and social organization of the black rhinoceros in Masai Mara Game Reserve. *Afr. J. Ecol.* **11**, 385–400.
- MUYA, S.M. & OGUGE, N.O. (2000) Effects of browse availability and quality on black rhino (*Diceros bicornis michaeli* Groves 1967) diet in Nairobi National Park, Kenya. *Afr. J. Ecol.* **38**, 62–71.
- OLIVIER, F. & WOTHERSPOON, S.J. (2005) GIS-based application of resource selection functions to the prediction of snow petrel distribution and abundance in East Antarctica: comparing models at multiple scales. *Ecol. Model.* **189**, 105–129.
- OLOO, T.W., BRETT, R. & YOUNG, T.P. (1994) Seasonal variation in the feeding ecology of black rhinoceros (*Diceros bicornis* L.) in Laikipia, Kenya. *Afr. J. Ecol.* **32**, 142–157.
- ÖZESMI, S.L. & ÖZESMI, U. (1999) An artificial neural network approach to spatial habitat modelling with interspecific interaction. *Ecol. Model.* **116**, 15–31.
- PATTON, F.J., MULAMA, M.S., MUTISYA, S. & CAMPBELL, P.E. (2010) The effect of removing a dividing fence between two populations of black rhinos. *Pachyderm* **47**, 55–58.
- PRADHAN, N.M.B., WEGGE, P., MOE, S.R. & SHRESTHA, A.K. (2008) Feeding ecology of two endangered sympatric megaherbivores: Asian elephant *Elephas maximus* and greater one-horned rhinoceros *Rhinoceros unicornis* in lowland Nepal. *Wildlife Biol.* **14**, 147–154.
- PRASAD, A.M., IVERSON, L.R. & LIAW, A. (2006) Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* **9**, 181–199.
- R Development Core Team (2007) R: A Language and Environment for Statistical Computing. The R Foundation for Statistical Computing, Vienna, Austria. ISBN: 3-900051-07-0. Available at: <http://www.R-project.org/> (Accessed on 3 September 2014).
- RICE, M.B. & JONES, M. (2006) Characteristics of black rhinoceros (*Diceros bicornis*) bedding sites. *Afr. J. Ecol.* **44**, 452–457.



- STOCKWELL, D.R.B. & PETERSON, A.T. (2002) Effects of sample size on accuracy of species distribution models. *Ecol. Model.* **148**, 1–13.
- TATMAN, S.C., STEVENS-WOOD, B. & SMITH, V.B.T. (2000) Ranging behaviour and habitat usage in black rhinoceros, *Diceros bicornis*, in a Kenyan sanctuary. *Afr. J. Ecol.* **38**, 163–172.
- WALPOLE, M.J., MORGAN-DAVIES, M., MILLEDGE, S., BETT, P. & LEADER-WILLIAMS, N. (2001) Population dynamics and future conservation of a free-ranging black rhinoceros (*Diceros bicornis*) population in Kenya. *Biol. Conserv.* **99**, 237–243.
- WILLIAMS, G.J., (2011). *Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery (Use R!)*. Springer, New York & London.
- ZUUR, A.F., IENO, E.N., WALKER, N.J., SAVELIEV, A.A. & SMITH, G.M. (2009) *Mixed Effects Models and Extensions in Ecology with R*. Springer, New York, USA.

(Manuscript accepted 03 November 2014)

doi: 10.1111/aje.12192