High-resolution species-distribution model based on systematic sampling and indirect observations: a case-study of the wild ass in the Negev Desert

Oded Nezer, Department of Environmental Engineering, Technion – Israel Institute of Technology. nezer.oded@gmail.com

Shirli Bar-David, Mitrani Department of Desert Ecology, Jacob Blaustein Institutes for Desert Research, Ben Gurion University of the Negev. shirlibd@bgu.ac.il

Tomer Gueta, Department of Environmental Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel. tomer.gu@gmail.com

Yohay Carmel*, Department of Environmental Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel. yohay@technion.ac.il. Phone: 972-4-8292609.

*Corresponding author.

Abstract

Species Distribution models (SDMs) are often limited by the use of coarse-resolution environmental variables and by the number of observations needed to calibrate SDMs. This is particularly true in the case of elusive animals. Here, we developed a SDM by combining three elements: a database of explanatory variables, mapped at a fine resolution; a systematic sampling scheme; and an intensive survey of indirect observations. Using MaxEnt, we developed the SDM for the population of the Asiatic wild ass (*Equus hemionus*), a rare and elusive species, at three spatial scales: 10, 100, and 1000 m per pixel. We used indirect observations of feces mounds. We constructed 14 layers of explanatory variables, in five categories: water, topography, biotic conditions, climatic variables and anthropogenic variables. Woody vegetation cover and slopes were found to have the strongest effect on the wild ass distribution and were included as the main predictors in the SDM. Model validation revealed that an intensive survey of feces mounds and high-resolution predictor layers resulted in a highly accurate and informative SDM. Fine-grain (10 m and 100 m) SDMs can be utilized to: 1) characterize the variables influencing species distribution at high resolution and local scale, including anthropogenic effects and geomorphologic features; 2) detect potential population activity centers; 3) locate potential corridors of movement and possible isolated habitat patches. Such information may be useful for the conservation efforts of the Asiatic wild ass. This approach could be applied to other elusive species, particularly large mammals.

Keywords: *Equus hemionus;* habitat preferences; Faeces; MAXENT; SDM; spatially explicit model.

Introduction

Spatially explicit Species Distribution Models (SDMs) are commonly used for purposes of conservation, environmental planning, and wildlife management programs (Guisan et al. 2013). SDM models quantify the relationships between the distribution and demography of a species and the environment (Peterson 2011). SDMs allow us to study species distribution in large areas and even in remote habitats, where logistic and financial restrictions preclude direct observations (Duff and Morrell 2007). They may be particularly useful for assessing the success of reintroduction activities (Manel et al. 1999; Manel et al. 2001). Understanding habitat characteristics and distribution determinants of reintroduced species is important in order to ensure the protection of landscape components that are critical for the long-term persistence of these species in the wild.

The use of environmental variables to explain and predict species distribution is not trivial, since these relationships are complex, and a large number of variables are involved (Guisan and Zimmermann 2000; Radosavljevic and Anderson 2014). It is well known that the variables that affect the distribution of a species change with the change of observation scale (Blank and Carmel 2012; Crawley and Harral 2001; Kent et al. 2011; Stauffer and Best 1986). Coarse-scale distribution models may be preferred in, for example, bio-geographic studies. In contrast, fine-scale distribution models can depict local scale phenomena such as essential corridors and animal passages and effects of roads and rivers, which coarse-scale models cannot detect. Thus, fine-scale distribution models may be preferable for conservation planning and management (Hess et al. 2006). Yet, in most studies, the selection of resolution is a consequence of the availability and quality of data pertaining to the specific study area, which is typically the limiting factor in distribution studies (Elith et al. 2006;

Hess et al. 2006). Data layers used in such studies are typically derived from global databases where 1 km^2 is considered the finest resolution.

Presence/absence information is thought to be preferred to presence-only information in SDMs. However, presence/absence information is more difficult or impossible to obtain than Presence-only information (Kent et al. 2011; Pearce and Boyce 2006; Tsoar et al. 2007). However, presence-only data may be subject to large errors due to small sample size and biased samples (Graham et al. 2004; Phillips and Elith 2013). A systematic data-collection survey, designed to collect data at precise locations should largely reduce these biases.

Indirect observations, and in particular dung surveys, are common non-invasive approaches for obtaining information about the presence of species and habitat selection. They are particularly useful when the studied species is hard to find due to its elusive behavior, rarity, or habitat (Fernandez et al. 2006; Vina et al. 2010). The use of indirect observations in SDMs requires a clear connection between the presence of the species and the feces (Gallant et al. 2007; Kays et al. 2008; Perinchery et al. 2011). Systematic dung survey, conducted in sites selected to represent the entire range of environmental conditions in a region, can be an appropriate solution to sampling-bias problems (Fernandez et al. 2006; Norris 2014; Vina et al. 2010).

Here, we developed a species distribution model for the population of the Asiatic Wild Ass (*Equus hemionus*), a rare and elusive species that was reintroduced into the Negev Desert in Israel. We combined three elements in order to overcome the obstacles in developing SDMs: a database of spatial layers of explanatory variables, mapped at a very fine resolution; a systematic sampling scheme; and an intensive survey of indirect observations, presence of feces mounds. This approach led to important insights regarding the habitat preferences of this species.

Methods

Study species

The Asiatic wild ass is an endangered species (Moehlman et al. 2008). In the past, the Syrian wild ass (*E. h. hemippus*) subspecies was found in the Middle East, and became extinct in the wild at the beginning of the 20th century (Groves 1986; Saltz et al. 2000; Schulz and Kaiser 2013). In 1968, a breeding core was established in Israel using individuals from the subspecies *E.h. onager* and *E. h. kulan* brought from Iran and Turkmenistan, respectively. In 1982 the Israel Nature and Parks Authority initiated a reintroduction program of the Asiatic wild ass (from the breeding core of these two subspecies (Saltz et al. 2000). The first individuals were released near Ein-Saharonim in Makhtesh Ramon (Fig. 1). By 1993, three additional releases were conducted at this site and two more in the Paran streambed (Saltz and Rubenstein 1995). A total of 38 individuals were released. The wild ass population expanded its range in the Negev Desert and the Arava valley (Saltz and Rubenstein 1995), and the current population is estimated at more than 250 individuals (Renan et al. 2015).

Study area

The study area extends over approximately 3,000 km^2 in the central part of the Negev Desert (Fig. 1). The area is arid and characterized by high daytime temperatures (on average 33°C) and relatively low night-time temperatures (on average 12°C). Mean annual precipitation ranges between 30 mm and 150 mm (Stern et al. 1986). Elevation ranges between 50-1033 m, and the area has a complex geomorphological structure. The bedrock is mainly hard limestone, resulting in a cliffy landscape and leveled floodplains. The majority of the area is drained by two main ephemeral streambeds (wadis) – Nekarot and Paran. There are several latitudinal geological faults in the

region that create a steep terraced landscape. Flash floods are a common phenomenon after rain events. The flash floods fill water holes in the streambeds, which may hold for a few months. There are very few natural water sources that provide water year-round. Vegetation is mostly limited to streams and their surroundings and generally located on the banks. Vegetation in the streams is mostly of a Saharo-Arabian origin, with a Sudanian component in the Arava (Danin 1999). It is dominated by three native Acacia tree species, *Acacia raddiana, A. tortilis,* and *A. pachyceras*.

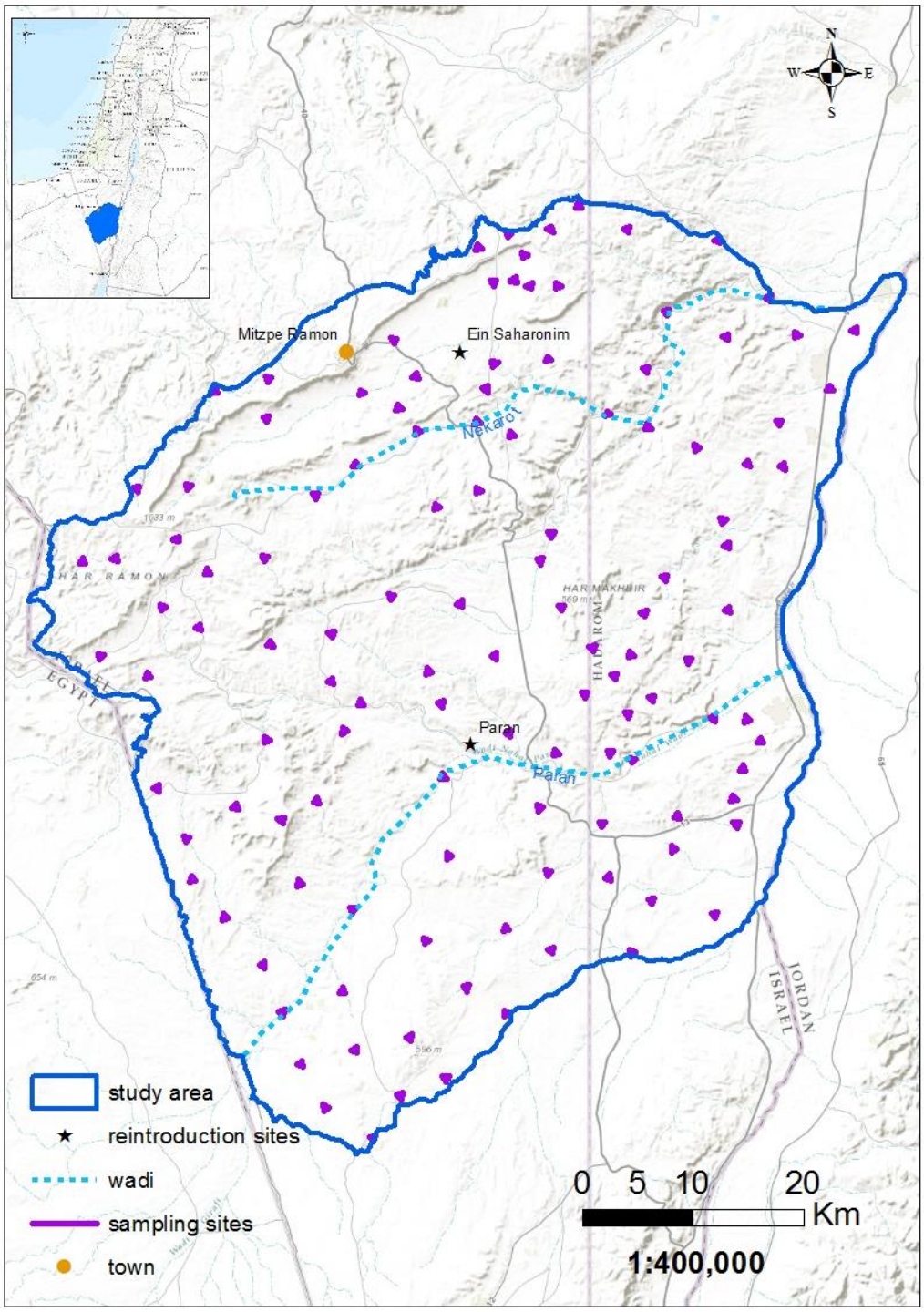


Figure 1. The study region, reintroduction and sampling sites in the Negev Desert, Israel.

Data collection

We selected 122 sampling sites using an approximate systematic sampling scheme (Fig. 1) in order to capture the full range of conditions found in the study region. To

ensure accurate representation of the environmental conditions in the sampling sites, we stratified the sampling locations according to three environmental parameters: distance from permanent water sources, altitude, and mean temperature of the hottest month. Based on our prior knowledge of the study species and a literature review (Henley and Ward 2006; Henley et al. 2007; Saltz and Rubenstein 1995; Saltz et al. 1999), we considered these environmental variables to have a high potential for explaining wild ass distribution. These variables were represented by GIS layers and combined into a three banded composite, on which we performed K-means unsupervised classification using ERDAS IMAGINE V 9.1. The objective of this classification was to divide the study region into polygons with similar combinations of these variables (Carmel and Stoller-Cavari 2006). The 122 sampling sites were systematically distributed among these polygons.

In each sampling site we conducted a feces survey. Fecal droppings of wild ass constitute a straightforward indicator for species presence because they are deposited frequently, and remain visible in the desert environment for several months (up to about a year). The survey in each site was composed of three 500 m belt transects arranged as an equilateral triangle with a total length of 1500 m, and divided into 150 survey units of 10X10 m per site. One of the triangle sides was always laid on a dry river-bed nearest to the point defined as the center of the sampling site. We recorded observations at a distance of 5 m on either side of the transect, where detection probability of feces was 100%. The exact location of feces mound (droppings as well as dung piles) observed on the transect were recorded using a GPS at a spatial accuracy of 4 m. The number of feces mounds within each 10 m pixel was recorded. During January 2009 to June 2009 we surveyed 122 sites and explored 150 units per

site, with a total sampled area of 183 ha. For the ‘presence-only’ SDM, we classified each unit as ‘present’ if one or more feces mounds were found in that unit.

Data analysis

Explanatory variables

We devoted extensive efforts to create a high resolution digital data set of environmental variables. We generated 14 spatial layers (Table 1), from which the model predictors were derived. These layers pertain to five main categories (Table 1): vegetation (one variable), topography (4), climate (2), anthropogenic variables (5), and distance from water (2 variables). The vegetation layer was derived from a complex processing of an aerial photo (Appendix 1). Topography was derived from a DEM of the area, at an original resolution of 10 m. Climate layers had an original resolution of 1 km, and were up-scaled to a 10 m resolution. Distance to—layers were constructed using Euclidean distance to specific elements on the map at an original resolution of 10 m. In order to reduce multicollinearity, correlation coefficients were calculated between each pair of variables; in pairs with a high correlation (>0.65 or <-0.65, Pearson correlation), one of the variables was eliminated from the model. A map of each explanatory variable appears in Appendix 2.

Table 1. Predictors used in the distribution model of wild ass. Stars (*****) indicate variables eliminated from the model due to high correlation with other variables.

#	Category	Description	Retrieval information
1	Vegetation	Percentage of woody vegetation cover (shrubs and trees with a radius greater than 0.2m). Each 10 m cell represents an averaged vegetation cover over a 100m radius.	Manual digitization from orthophoto.
2	Topography	Altitude above sea level	Generated from contour dataset retrieved from Survey of Israel (MAPI)
3		Slope (between 0-90 degrees)	Generated from Altitude using ArcMap 10

4		Aspect (between 0-360 degrees)	Generated from Altitude using ArcMap 10
5		Cumulative drainage	Generated from slope using ArcMap 10
6	Climate	*****Mean annual precipitation	Retrieved from the GIS Lab at the Hebrew University of Jerusalem
7		*****Mean temperature in August	
8	Anthropogenic factors	*****Distance from roads	Generated in ArcMap 10
9		*****Distance from reintroduction sites	Generated in ArcMap 10
10		Distance from military bases and settlements	Generated in ArcMap 10
11		Military training sites (binary)	Manual digitization from orthophoto.
12		Nature reserve (binary)	Manual digitization from orthophoto.
13	Water	Distance from all permanent water sources including springs and leaking pipes	Generated in ArcMap 10
14		*****Distance from watering holes	Generated in ArcMap 10

Statistical model

We used the "Maximum Entropy" model MAXENT V3.3.1 (Kumar et al. 2009; Phillips et al. 2006; Phillips and Dudik 2008). We selected this model out of a large number of possible models, since it was ranked in several comparative studies as one of the most effective models for predicting species distribution on the basis of presence-only data (Elith et al. 2006; Elith et al. 2011; Jeschke and Strayer 2008; Phillips et al. 2006; Radosavljevic and Anderson 2014). The MAXENT algorithm operates on a set of constraints that describes what is known from the sample of the target distribution (i.e., the presence data). Maxent characterizes the background environment with a set of background points from the study region. However, unlike the case of presence–absence data, the species occurrence at these background points is unknown. MAXENT predicts the probability distribution across all cells in the study area based on the presence data and, to prevent over-fitting, employs maximum

entropy principles and regularization parameters (Phillips et al. 2006). MAXENT produces two outputs: a probabilistic distribution map describing the establishment probability of the species in a specific site and the relative weight of each explanatory variable. Distribution maps of the Asiatic wild ass were obtained by applying MAXENT models to all cells in the study region, using a logistic link function to yield a habitat suitability index between zero and one (Phillips and Dudik 2008). We ran the model in three spatial resolutions: 10 m, 100 m and 1 km, with 10^6, 10^5 and 10^4 background points respectively. Recommended values were used for the convergence threshold (10^{-5}), maximum number of iterations (500), and regularization multiplier (1). Response functions were constrained to only three feature types: Linear, Threshold and Hinge.

In order to estimate the percent contribution of each environmental variable, in each iteration of the training algorithm, the increase in *Regularized gain* is added to the contribution of the corresponding variable. In order to estimate the permutation importance of each environmental variable, in turn the values of the corresponding variable on training presence and background data are randomly permuted. The model is reevaluated on the permuted data, and the resulting drop in *Training AUC* is normalized to percentages. In order to estimate if occurrence data of the wild ass are spatially autocorrelated, we calculated Moran's *I* Index (Moran 1950) for each spatial resolution separately (10 m, 100 m and 1 km).

Model validation

We validated the model using: **1**) MAXENT's five performance measures and **2**) a cross-validation procedure. MAXENT model generates three gain measures and two AUC measures. Gain measures the goodness of fit of a models, it represents the likelihood of presence records compared to background records (Phillips 2005). A

gain of 1.6 means that an average presence location has a relative probability of $e^{1.6}$, which is five times higher than an average background point. *Regularized training gain* accounts for the number of predictors in the model to address overfitting; *Unregularized training gain* has no compensation for the number of predictors in the model; and *Test gain* is calculated from presence records held out to test the model. AUC is the area under the curve of the receiver operating characteristic (ROC) plot. ROC curves are widely used for validating SDMs and for comparing between models (Elith et al. 2006; Hernandez et al. 2006; Marmion et al. 2009). The AUC values range between 0 and 1, where 1 represents perfect prediction ability of the model and 0.5 represents prediction that is no better than random. *Training AUC* calculates AUC using the training data; and *Test AUC* calculates AUC using the test data. A cross-validation procedure was used to estimate errors around predictive performance on held-out data (Elith et al. 2011). Occurrence data are randomly split into a number of equal-size groups (folds), and models are created leaving out each fold in turn. The left-out folds are then used for evaluation. Cross-validation uses all of the data for validation. A 10-folds cross-validation procedure was used for the 10 m and 100 m models, and a 5-folds cross-validation procedure was used for the 1 km model.

Results

We recorded a total of 3,232 feces mounds in 18,300 survey units (10 m cells). Feces mounds were found in 115 of the 122 sampling sites. The number of mounds per site ranged from 0 to 124. Five potential explanatory variables were eliminated from the model (Table 1) due to high correlation coefficient (>0.65 or <-0.65, Pearson correlation, see Appendix 3), leaving nine variables in the model. Three of these spatial data layers, namely vegetation, slope, and altitude, were considered as the most

influential explanatory variables by the MAXENT algorithm, accounting together for ~85% of the cumulative relative contribution (Table 2). Woody vegetation density was found to have the strongest effect on the Asiatic wild ass distribution (Table 2; Appendix 4). The response curve of woody vegetation cover (Appendix 5) showed an increasing presence of the animals with increasing vegetation cover, leveling off sharply at the saturation point (>72% coverage). Slope was the second most important variable (Table 2) and was inversely related to wild ass distribution (Appendix 5). In slopes steeper than 20°, no feces mounds were found. Altitude was the third most important variable, with 12% relative contribution. The other six explanatory variables that were included in the model had a lower effect on the distribution of wild ass, together accounting for ~15% of the relative contribution to the model, Table 2.

The performance of the three models (10 m, 100, and 1 km) differed markedly. The 10 m model yielded the highest averaged values in all five performance measures (Table 3), indicating high predictive capacity. The 1 km model yielded the lowest values in all measures, with extremely low values for *Test gain* (-0.02) and *Test AUC* (0.67), suggesting poor predictive capacity at this scale. The cross-validation procedure revealed high consistency between the different runs, since standard deviation values were relatively low (Table 2 and 3).

Table 2. Percent contribution and permutation importance of the predictor variables for the 10 m resolution MAXENT model for wild ass. See *Statistical model* section in the Methods for an explanation. Standard deviation is shown in parentheses.

Explanatory variable	Relative contribution in % (± Std)	Permutation importance in % (± Std)
Vegetation	54.5 (0.44)	47.83 (0.93)
Slope	18.04 (0.41)	28.83 (1.31)
Altitude (DEM)	11.97 (0.41)	9.45 (0.73)

Distance from all permanent and temporary water sources	6.76 (0.26)	6.44 (0.32)
Distance from military bases and settlements	5 (0.21)	4.43 (0.36)
Cumulative drainage	1.66 (0.14)	1.24 (0.15)
Nature reserve	1.39 (0.13)	0.95 (0.18)
Aspect	0.36 (0.07)	0.49 (0.08)
Military training sites	0.33 (0.06)	0.33 (0.09)

Table 3. The averaged MAXENT performance measures calculated using a 10-folds or a 5-folds cross-validation procedure. Standard deviation is shown in parentheses.

	Models		
Model performance measures:	**10 m model**	**100 m model**	**1 km model**
Regularized training gain	1.63 (0.01)	1.16 (0.01)	0.64 (0.03)
Unregularized training gain	1.94 (0.01)	1.43 (0.01)	0.97 (0.04)
Test gain	1.82 (0.06)	1.26 (0.12)	-0.02 (0.07)
Training AUC	0.93 (0)	0.9 (0)	0.85 (0.01)
Test AUC	0.92 (0.01)	0.88 (0.01)	0.67 (0.02)

Occurrence data at a 10 m resolution had a relatively low spatial autocorrelation (Moran's I Index of 0.13), while the 100 m and 1 km resolutions had higher values (0.38 and 0.22 respectively).

The probabilistic distribution map was heterogeneous and informative at the very fine scale of 10 m, and the fine scale of 100 m (Fig. 2A-B), and much less informative at the scale of 1 km (Fig. 2C). The strong effect of streambeds on the species distribution was apparent at the two finer scales: areas of high probability of presence were in streambeds (wadis) characterized by woody vegetation and moderate terrain. The high resolution allowed detection of the following trends and phenomena at: [1] Possible convenient movement corridors in a matrix of unsuitable environment, which

enable landscape connectivity among sites (Fig. 4A). [2] Isolated local sites/areas of high suitability for the wild ass (high-quality habitat "islands") situated within broad areas of low quality habitat (Fig. 4B). [3]. Human-induced local entities that affect the distribution, e.g., the influence of roads on the quality of proximate habitats (Fig. 4C, see discussion for details). [4] Important geomorphologic features that affect the distribution, e.g., streambeds (Fig. 4D).

In contrast to the high variability visualized at fine scale, this map did not show regional trends or gradients at the scale of the study area. Sites with very high and very low probabilities of wild ass presence were found near each other throughout the entire study area; however, in several areas, a spatial continuity of high value sites was noticeable: Makhtesh Ramon (A), Paran streambed (B), the upper part of Nekarot streambed (C), and the Lotz potholes (Borot Lotz) (D) (Fig. 3). These areas have the potential to serve as activity centers for the population. A spatial continuum of sites with low suitability for the Asiatic wild ass also was discernable (Fig. 3, points E-H).

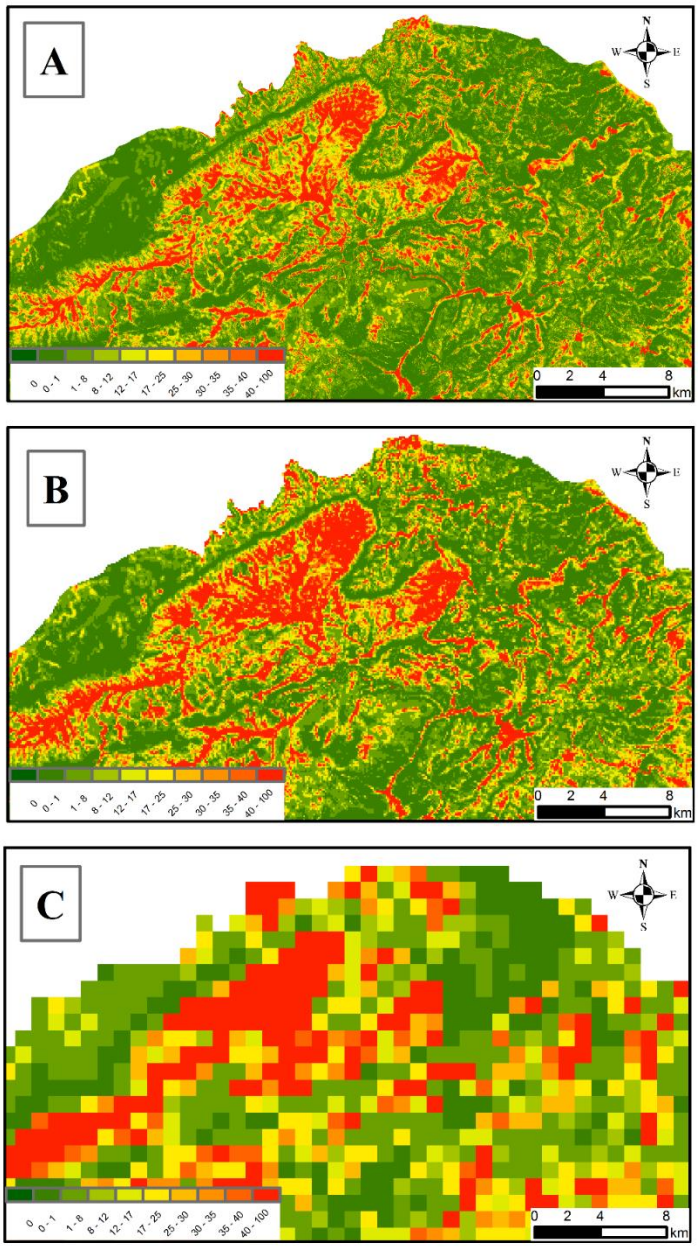


Figure 2: A comparison between the northern regions of the probabilistic distribution maps of the three models. (A) 10 m resolution model, (B) 100 m resolution model, and (C) 1 km resolution model.

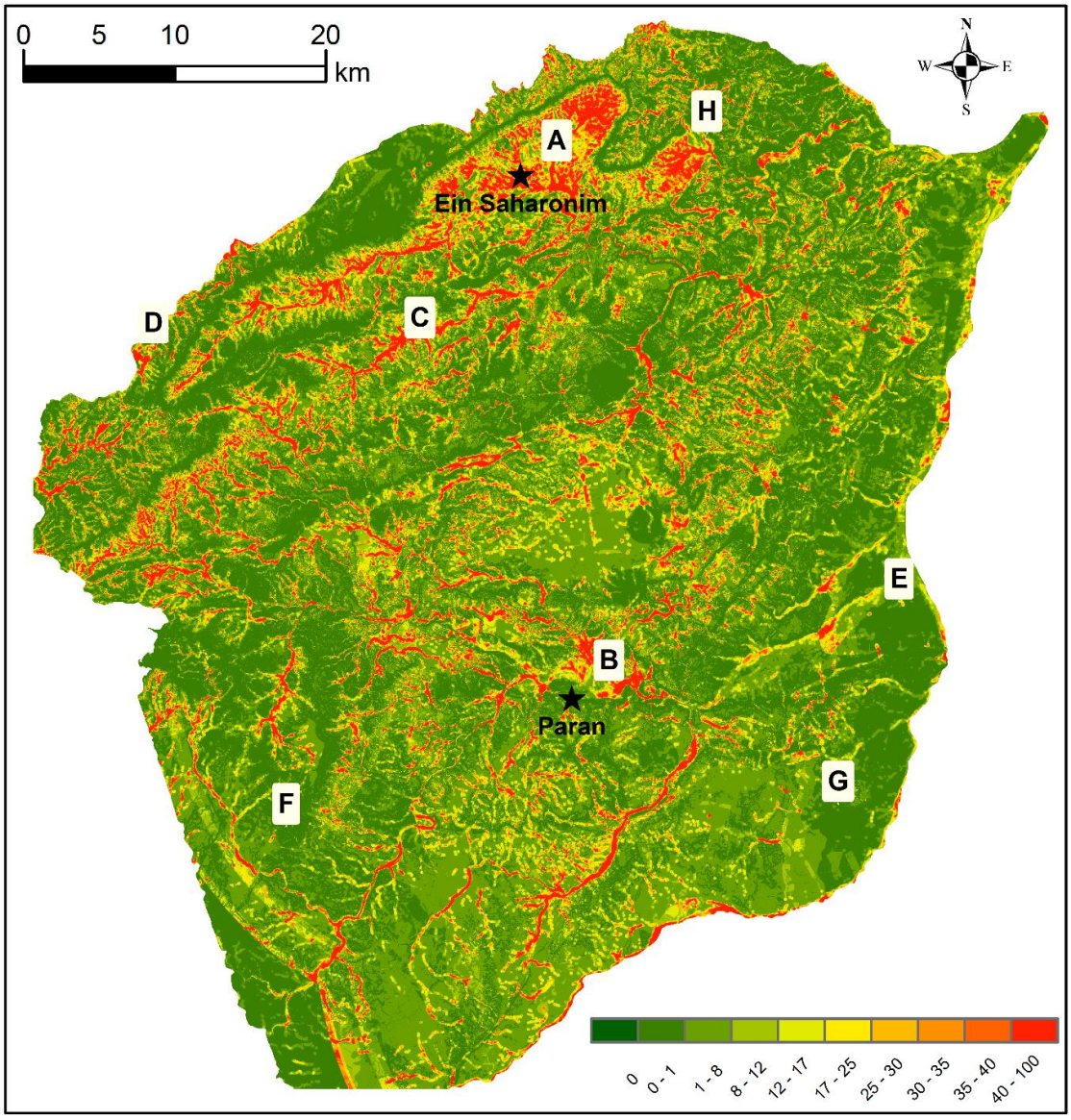


Figure 3: Probabilistic distribution maps of a 10 m resolution model for the Asiatic wild ass in the Negev. Potential wild ass activity centers: Makhtesh Ramon (A), Paran streambed (B), the upper part of Nekarot streambed (C) and the Lotz potholes (Borot Lotz) (D). A spatial continuum of sites with low suitability: the Paran Stream Estuary (E), the region south of Mount Karkom (F), Be'er Menuha (G), and the Eastern part of Makhtesh Ramon (H). Stars indicate reintroduction sites.

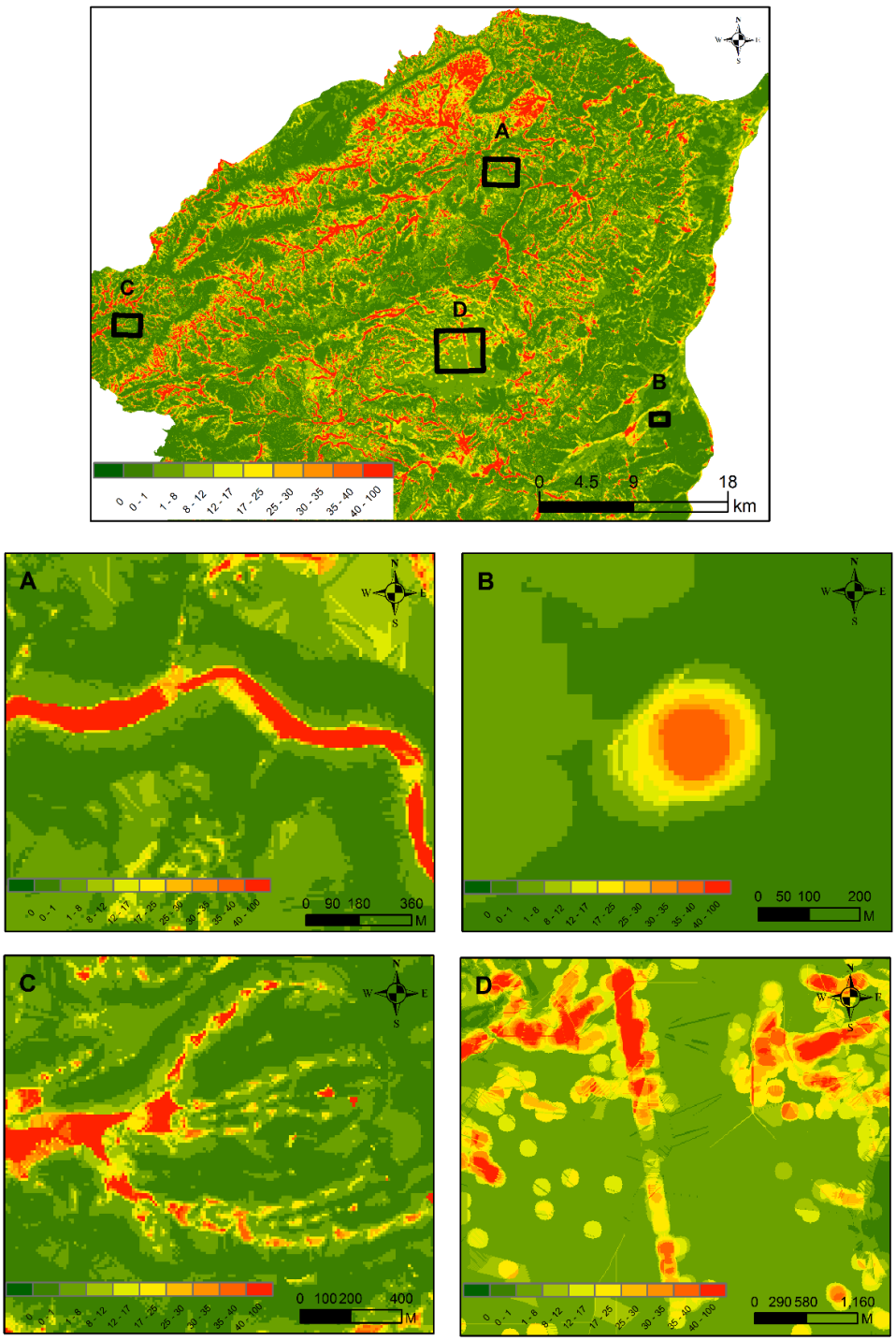


Figure 4. Detecting landscape features on the high-resolution map: (a) Potential movement corridors, (b) Isolated habitat patches, (c) Important geomorphologic features, (d) Anthropogenic effects on habitat quality (roads effect increased roadside vegetation). Colors represent predicted habitat suitability: from green, low suitability, to red, high suitability.

Discussion

In this study we combined three elements in order to develop a predictive distribution model for the wild ass, a rare and elusive animal: a database of spatial layers of explanatory variables, mapped at a very fine resolution; a systematic sampling scheme; and an intensive feces mound survey. The results indicate that this approach yields an accurate and informative model.

Factors affecting wild ass distribution

The most important variable in the model was the percentage of woody vegetation cover. Its relative contribution (54.5%) was much higher than that of the other variables. The importance of vegetation to the wild ass distribution is consistent with previous studies (Davidson et al. 2013; Giotto et al. 2015; Henley et al. 2007). The strong vegetation effect on the distribution is a result of its nutritional value (St-Louis and Côté 2014), the partial shade it offers, its value for hiding, and in arid areas the vegetation is a favorable microhabitat with reduced temperatures (Belsky et al. 1993).

The second most important variable in the model was slope (relative contribution of 18.04%). The Asiatic wild ass prefers moderate over steep terrain, and avoids steep slopes. This observation was supported by previous studies (Davidson et al. 2013; Giotto et al. 2015; Henley et al. 2007). The next variables in order of importance were altitude (11.97%) and distance from water sources (6.76%). The positive effect of altitude on wild ass distribution is probably related to lower temperatures associated with higher elevations. The Negev is a hyper- arid desert and we expected that distance from water sources would be a major predictor of wild ass distribution. Indeed, the water sources themselves were found to be centers of wild ass activity. However, the daily movement range of the wild ass can reach up to twenty km in each

direction (Saltz et al. 2000), and feces are distributed across most of the study area. This may explain why the distance from water source was not a major determinant of wild ass distribution.

Resolution

Model performance

Constructing models at various spatial resolutions and comparing between them enabled us to quantify the effect of resolution on SDM performance. Seemingly, model performance increased with increasing model resolution (Table 3). This finding contradicts a previous study (Guisan et al. 2007) of the effect of degrading model resolution on the performance of SDMs, who found that using finer cell sizes (from 1 km to 100 m, and from 10 km to 1 km) does not have a major effect on model predictions. Yet, our results suggest that when the effective resolution of the predictors was 10 m (10^2 m^2), the model provides useful insights regarding the species distribution that are not possible at coarser scales, as is elaborated in the following section.

AUC is one of the most commonly used statistics to characterize model performance (Yackulic et al. 2013), but its usage has been strongly criticized, particularly with presence-only data (Gueta and Carmel 2016; Jiménez-Valverde et al. 2013; Lobo et al. 2008; Yackulic et al. 2013), since it ignores the predicted probability values and the goodness-of-fit of the model (Yackulic et al. 2013). Corroborating these views, our 1 km model had a high *Training AUC* value (0.85) whereas the *Test gain* showed near zero predictive capability (Table 3). This reveals AUC's low informative value and its inadequacy as a performance index in presence-only modelling framework. Gain indices are more sensitive indicators of model performance (Gueta and Carmel 2016).

High-resolution spatial layers of explanatory variables

We invested considerable resources and effort to produce and obtain the layers of explanatory variables at a spatial resolution of 10 m wherever possible. For climatic variables, the original spatial resolution is 1 km. In contrast, the original resolution of the vegetation and topography layers was 10 m. Indeed, these two variables were the most important predictors in the 10 m model, somewhat less so in the 100 m model, and nearly meaningless in the 1 km model.

Distribution models of large mammals with large home ranges are typically constructed at resolutions of 100 – 10,000 m (e.g., (Bellamy et al. 2013), 2-6 orders of magnitude lower than the 10 m resolution of the present study. Apparently, the two predictors found to be the most important, vegetation and slope, appeared nearly meaningless at a resolution of 1000 m. The distribution map constructed at this coarse scale was not very informative.

High-resolution distribution map

The distribution map obtained by the model enabled us to examine the relative habitat suitability of each site for the wild ass at a fine resolution. The fine-grain image in Fig. 3 illustrates that low quality habitats are found within broad areas of suitable habitat, and vice versa. The high resolution of the map allowed the detection of four habitat components as important for the species' use of space (Fig. 4): (a) *Potential movement corridors* (Fig. 4A). Connectivity within the species' range is essential for the spatial, demographic and genetic dynamics of animal populations and their persistence over time (Colbert et al. 2001; Saccheri et al. 1998) and should be recognized as a high conservation priority (Beier et al. 2006). Identifying connectivity corridors is highly important for the protection of the species, since they may facilitate wild ass movements within a matrix of less suitable areas, enabling connectivity

between high-quality habitats (Fig. 2, points A-D). (b) *Isolated habitat patches* (Fig. 4B). Isolated sites or small fragments of high habitat quality within low quality areas may constitute potential ‘stepping stones’ sites that aid in connecting between activity centers. (c) *Important geomorphologic features* (Fig. 4C). The high-resolution map indicated clearly the importance of streambeds, including first order streams, in the distribution patterns of the wild ass. In coarser maps, the influence of the streambeds cannot be detected. (d) *Anthropogenic effect on distribution*. Anthropogenic features may influence distribution patterns of species and therefore it is important that they be identified (Valverde et al. 2008). For example, based on the high-resolution wild ass distribution model, roads were found to increase considerably the quality of habitats in the proximate areas of the roads (Fig. 4D). However, in a specific case, the road effect led to high density of roadside vegetation. The high vegetation quality, in turn, attracted wild asses to the proximity of the road, and several road-kills of wild asses were reported in this area, calling for roadside vegetation management (Asaf Tsoar, personal communication). This example illustrates the importance of the model as a tool to identify such potential negative anthropogenic effects.

Sampling

Systematic sampling of presence data

Many distribution models that are based on presence-only data suffer from inaccuracies, due to biased sampling (e.g., multiple observations near roads and accessible sites) and a distribution of observations that is unrepresentative of the range of environmental conditions in the study region (Barry and Elith 2006; Elith et al. 2011; Kramer-Schadt et al. 2013; Phillips and Elith 2013). In this study, we implemented an approximate systematic sampling scheme based on the spatial pattern of major environmental conditions in the study region, thus reducing the

aforementioned errors. A common problem in sampling rare species is a zero-inflated distribution of records. In order to reduce this problem, dry river beds were over-represented based on a prior knowledge that wild asses are usually found within riverbeds. Still, two-thirds of the samples were located off riverbeds. However, due to the dense network of riverbeds and the high density of sampling sites – only few areas were out of the reach of this sampling scheme (Fig. 1), and the possible bias was minimal.

Indirect observations for presence data

Predictive distribution models are usually based on direct observations. Creating a database of direct observations of an elusive organism with a small population, in a region with very limited accessibility is a very complicated task (Lozano et al. 2003; Sharp et al. 2001). Therefore, it was proposed that indirect observations (tracks, feces) may replace direct observations, when there is a clear connection between the presence of the species and the indirect observations (Fernandez et al. 2006; Kays et al. 2008; Perinchery et al. 2011; Vina et al. 2010). In this study, we relied on indirect observations using feces mounds as the basis for presence data. The major advantage of surveying feces mounds is that they remain in the field after the animal leaves, increasing the probability of recording activity in sites visited by the species. These factors are enhanced in a desert environment, since in arid regions the decomposition rate of the feces is slower, and mounds may last for long periods, in the case of the wild asses in the Negev up to a year. The large number of observations is a major component of the strength and reliability of a distribution model (Barry and Elith 2006).The feces surveys in our study led to a large number of observations. Obtaining a similar sized database using direct observations would have required a much greater, longer and costlier sampling effort.

Implications for conservation

SDMs can help to design conservation policies (Guisan and Zimmermann 2000). The endangered Asiatic wild ass can become a focus of conservation interest due to its impressive appearance, rarity, reintroduction process and its pivotal function in the Negev ecosystem (Polak et al. 2014). The SDM constructed in this study can serve to locate favorable high-quality patches, and potential future expansion directions of the species in the Negev Desert. It can also be used to locate potential routes and corridors among activity centers which are important to maintain connectivity within the population. Model predictions can then be validated by conducting field surveys (Davidson et al. 2013). This information can serve as the basis for developing conservation and management strategies for the wild ass. Specifically, the map enabled us to identify large continuous geographic areas of suitable habitat, which constitute potential activity centers. Three of the continuous areas identified in the map (central Makhtesh Ramon, Paran streambed, and Borot Lotz; Fig. 3) were confirmed in the field as significant activity centers, based on direct observations. Two of these sites – the Paran streambed and the central part of Makhtesh Ramon overlap with the reintroduction sites. However, distance from the reintroduction sites was not found to be a significant factor affecting species distribution in the statistical model. Each one of the three activity centers contains a permanent water source. The model further enabled us, in a parallel study, to identify areas with low landscape connectivity among activity centers (Gueta et al. 2014). These areas were suggested to limit gene flow, leading to the relative isolation of a subpopulation and to the development of population genetic structure in the reintroduced wild ass population (Gueta et al. 2014). Limited gene flow among activity centers may further affect the

population genetic diversity (Renan et al. 2015) which is essential for the population's long-term viability (Hughes et al. 2008).

The distribution model can also be used to locate a potential direction for expanding the wild ass range, by projecting the model onto additional areas (Bar-David et al. 2008). It is important to identify areas of potential spatial expansion, in order to ensure the protection and maintenance of landscape connectivity, which is essential for the species' distribution and, hence, for its persistence in the wild.

Acknowledgement

We would like to thank David Saltz, Alan R. Templeton, and Amos Bouskila for their contributions to this study; Itamar Giladi for providing insightful comments that greatly improved the manuscript. This research was supported by the United States-Israel Binational Science Foundation Grant 2011384 awarded to S. B-D, A. R. Templeton and A. Bouskila. GIS layers were provided by the GIS Department of the Israel Nature and Parks Authority. This is publication <xxx> of the Mitrani Department of Desert Ecology.

References

Bar-David S, Saltz D, Dayan T (2005) Predicting the spatial dynamics of a reintroduced population: the Persian fallow deer Ecological Applications 15:1833-1846

Bar-David S, Saltz D, Dayan T, Shkedy Y (2008) Using spatially expanding populations as a tool for evaluating landscape planning: The reintroduced persian fallow deer as a case study Journal for Nature Conservation 16 164

Barry S, Elith J (2006) Error and uncertainty in habitat models Journal of Applied Ecology 43:413-423

Beier P, Penrod K, Luke C, Spencer W, Cabañero C (2006) South coast missing linkages: Restoring connectivity to wildlands in the largest metropolitan area in the united states. In: Crooks KR, Sanjayan M (eds) Connectivity Conservation. Cambridge University Press, Cambridge, United Kingdom, pp 555-586

Bellamy C, Scott C, Altringham J (2013) Multiscale presence-only habitat suitability models: fine-resolution maps for eight bat species Journal of Applied Ecology 50:892-901

Belsky A, Mwonga S, Amundson R, Duxbury J, Ali A (1993) Comparative effects of isolated trees on their undercanopy environments in high- and low-rainfall savannas Journal of Applied Ecology 30:143-155

Blank L, Carmel Y (2012) Woody vegetation patch types affect herbaceous species richness and composition in a mediterranean ecosystem Community Ecology 13:72-81

Carmel Y, Stoller-Cavari L (2006) Comparing environmental and biological surrogates for biodiversity at a local scale Israel Journal of Ecology and Evolution 52:11-27

Colbert T et al. (2001) High-throughput screening for induced point mutations Plant Physiology 126:480-484

Crawley M, Harral J (2001) Scale dependence in plant biodiversity Science 291:864-868

Danin A (1999) Desert rocks as plant refugia in the near east Botanical Review 65:93-170

Davidson A, Carmel Y, Bar-David S (2013) Characterizing wild ass pathways using a non-invasive approach: Applying least-cost path modelling to guide field surveys and a model selection analysis Landscape Ecology 28:1465

Duff A, Morrell T (2007) Predictive occurrence models for bat species in california Journal of Wildlife Management 71:693-700

Elith J et al. (2006) Novel methods improve prediction of species' distributions from occurrence data Ecography 29:129-151

Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ (2011) A statistical explanation of MaxEnt for ecologists Diversity and distributions 17:43-57

Fernandez N, Delibes M, Palomares F (2006) Landscape evaluation in conservation: Molecular sampling and habitat modeling for the Iberian lynx Ecological Applications 16:1037-1049

Gallant D, Vasseur L, Berube C (2007) Unveiling the limitations of scat surveys to monitor social species: A case study on river otters Journal of Wildlife Management 71:258-265

Giotto N, Gerard JF, Ziv A, Bouskila A, Bar-David S (2015) Space-Use Patterns of the Asiatic Wild Ass (Equus hemionus): Complementary Insights from Displacement, Recursion Movement and Habitat Selection Analyses PLoS ONE 10 doi:10.1371/journal.pone.0143279

Graham C, Ferrier S, Huettman F, Moritz C, Peterson A (2004) New developments in museum-based informatics and applications in biodiversity analysis Trends in Ecology & Evolution 19:497-503

Groves C (1986) The taxonomy, distribution, and adaptations of recent equids. In: RH M, HP U (eds) Equids in the Ancient World. Ludwig Reichert Verlag, Wiesbaden,

Gueta T, Carmel Y (2016) Quantifying the value of user-level data cleaning for big data: A case study using mammal distribution models Ecological Informatics

Gueta T, Templeton A, Bar-David S (2014) Development of genetic structure in a heterogeneous landscape over a short time frame: The reintroduced asiatic wild ass Conservation Genetics 15:1231

Guisan A, Graham C, Elith J, Huettmann F (2007) Sensitivity of predictive species distribution models to change in grain size Diversity and Distributions 13:332-340

Guisan A et al. (2013) Predicting species distributions for conservation decisions Ecology letters 16:1424-1435

Guisan A, Zimmermann N (2000) Predictive habitat distribution models in ecology Ecological Modelling 135:147-186

Henley S, Ward D (2006) An evaluation of diet quality in two desert ungulates exposed to hyper-arid conditions African Journal of Range and Forage Science 23:185-190

Henley S, Ward D, Schmidt I (2007) Habitat selection by two desert-adapted ungulates Journal of Arid Environments 70:39-48

Hernandez P, Graham C, Master L, Albert D (2006) The effect of sample size and species characteristics on performance of different species distribution modeling methods Ecography 29:773-785

Hess GR, Bartel RA, Leidner AK, Rosenfeld KM, Rubino MJ, Snider SB, Ricketts TH (2006) Effectiveness of biodiversity indicators varies with extent, grain, and region Biological Conservation 132:448-457

Hughes A, Inouye B, Johnson M, Underwood N, Vellend M (2008) Ecological consequences of genetic diversity Ecology Letters 11:609

Jeschke J, Strayer D (2008) Usefulness of bioclimatic models for studying climate change and invasive species Annals of the New York Academy of Sciences 1134:1-24

Jiménez-Valverde A, Acevedo P, Barbosa AM, Lobo JM, Real R (2013) Discrimination capacity in species distribution models depends on the representativeness of the environmental domain Global Ecology and Biogeography 22:508-516

Kays R, Gompper M, Ray J (2008) Landscape ecology of eastern coyotes based on large-scale estimates of abundance Ecological Applications 18:1014-1027

Kent R, Bar-Massada A, Carmel Y (2011) Multiscale analyses of mammal species Composition–Environment relationship in the contiguous USA PloS One 6:e25440

Kramer-Schadt S et al. (2013) The importance of correcting for sampling bias in MaxEnt species distribution models Diversity and Distributions 19:1366-1379 doi:10.1111/ddi.12096

Kumar S, Spaulding S, Stohlgren T, Hermann K, Schmidt T, Bahls L (2009) Potential habitat distribution for the freshwater diatom didymosphenia geminata in the continental US Frontiers in Ecology and the Environment 7:415-420

Lobo JM, Jiménez-Valverde A, Real R (2008) AUC: a misleading measure of the performance of predictive distribution models Global ecology and Biogeography 17:145-151

Lozano J, Virgos E, Malo A, Huertas D, Casanovas J (2003) Importance of scrub-pastureland mosaics for wild-living cats occurrence in a mediterranean area: Implications for the conservation of the wildcat felis silvestris Biodiversity and Conservation 12:921-935

Manel S, Dias J, Buckton S, Ormerod S (1999) Alternative methods for predicting species distribution: An illustration with himalayan river birds Journal of Applied Ecology 36:734-747

Manel S, Williams H, Ormerod S (2001) Evaluating presence-absence models in ecology: The need to account for prevalence Journal of Applied Ecology 38:921-931

Marmion M, Parviainen M, Luoto M, Heikkinen R, Thuiller W (2009) Evaluation of consensus methods in predictive species distribution modelling Diversity and Distributions 15:59-69

Moehlman P, Shah N, Feh C (2008) Equus hemionus. IUCN. http://www.iucnredlist.org/details/full/7951/0. Accessed August 2016 2016

Moran PA (1950) Notes on continuous stochastic phenomena Biometrika 37:17-23

Norris D (2014) Model thresholds are more important than presence location type: Understanding the distribution of lowland tapir (*Tapirus terrestris*) in a continuous Atlantic forest of southeast Brazil Tropical Conservation Science 7:529-547

Pearce J, Boyce M (2006) Modelling distribution and abundance with presence-only data Journal of Applied Ecology 43:405-412

Perinchery A, Jathanna D, Kumar A (2011) Factors determining occupancy and habitat use by Asian small-clawed otters in the Western Ghats India Journal of Mammalogy 92:796-802

Peterson AT (2011) Ecological niches and geographic distributions (MPB-49). vol 49. Princeton University Press,

Phillips S, Anderson R, Schapire R (2006) Maximum entropy modeling of species geographic distributions Ecological Modelling 190:231-259

Phillips S, Dudik M (2008) Modeling of species distributions with maxent: New extensions and a comprehensive evaluation Ecography 31:161-175

Phillips SJ (2005) A brief tutorial on Maxent AT&T Research

Phillips SJ, Elith J (2013) On estimating probability of presence from use-availability or presence-background data Ecology 94:1409-1419

Polak T, Gutterman Y, Hoffman I, Saltz D (2014) Redundancy in seed dispersal by three sympatric ungulates: A reintroduction perspective Animal Conservation 17:565

Radosavljevic A, Anderson RP (2014) Making better Maxent models of species distributions: complexity, overfitting and evaluation Journal of Biogeography 41:629-643 doi:10.1111/jbi.12227

Renan S, Greenbaum G, Shahar N, Templeton A, Bouskila A, Bar-David S (2015) Stochastic modelling of shifts in allele frequencies reveals a strongly polygynous mating system in the re-introduced asiatic wild ass Molecular Ecology 24:1433

Saccheri I, Kuussaari M, Kankare M, Vikman P, Fortelius W, Hanski I (1998) Inbreeding and extinction in a butterfly metapopulation Nature 392:491-494

Saltz D, Rowen M, Rubenstein D (2000) The effect of space-use patterns of reintroduced asiatic wild ass on effective population size Conservation Biology 14:1852-1861

Saltz D, Rubenstein D (1995) Population-dynamics of a reintroduced asiatic wild ass Equus hemionus herd Ecological Applications 5:327-335

Saltz D, Schmidt H, Rowen M, Karnieli A, Ward D, Schmidt I (1999) Assessing grazing impacts by remote sensing in hyper-arid environments Journal of Range Management 52:500-507

Schulz E, Kaiser TM (2013) Historical distribution, habitat requirements and feeding ecology of the genus Equus (Perissodactyla) Mammal Review 43:111-123 doi:10.1111/j.1365-2907.2012.00210.x

Sharp A, Norton M, Marks A, Holmes K (2001) An evaluation of two indices of red fox vulpes vulpes abundance in an arid environment Wildlife Research 28:419-424

St-Louis A, Côté SD (2014) Resource selection in a high-altitude rangeland equid, the kiang (Equus kiang): Influence of forage abundance and quality at multiple spatial scales Canadian Journal of Zoology 92:239-249 doi:10.1139/cjz-2013-0191

Stauffer D, Best L (1986) Nest-site characteristics of open-nesting birds in riparian habitats in iowa The Wilson Bulletin 231-242

Stern E, Gardus Y, Meir A, Krakover S, Tzoar H (1986) Atlas of the Negev. Keter Publishing House, Jerusalem

Tsoar A, Allouche O, Steinitz O, Rotem D, Kadmon R (2007) A comparative evaluation of presence-only methods for modelling species distribution Diversity and Distributions 13:397-405

Valverde AJ, Lobo J, Hortal J (2008) Not as good as they seem: The importance of concepts in species distribution modelling Diversity and Distributions 14:885-890

Vina A, Tuanmu M, Xu W, Li Y, Ouyang Z, DeFries R, Liu J (2010) Range-wide analysis of wildlife habitat: Implications for conservation Biological Conservation 143:1960-1969

Yackulic CB, Chandler R, Zipkin EF, Royle JA, Nichols JD, Campbell Grant EH, Veran S (2013) Presence-only modelling using MAXENT: when can we trust the inferences? Methods in Ecology and Evolution 4:236-243

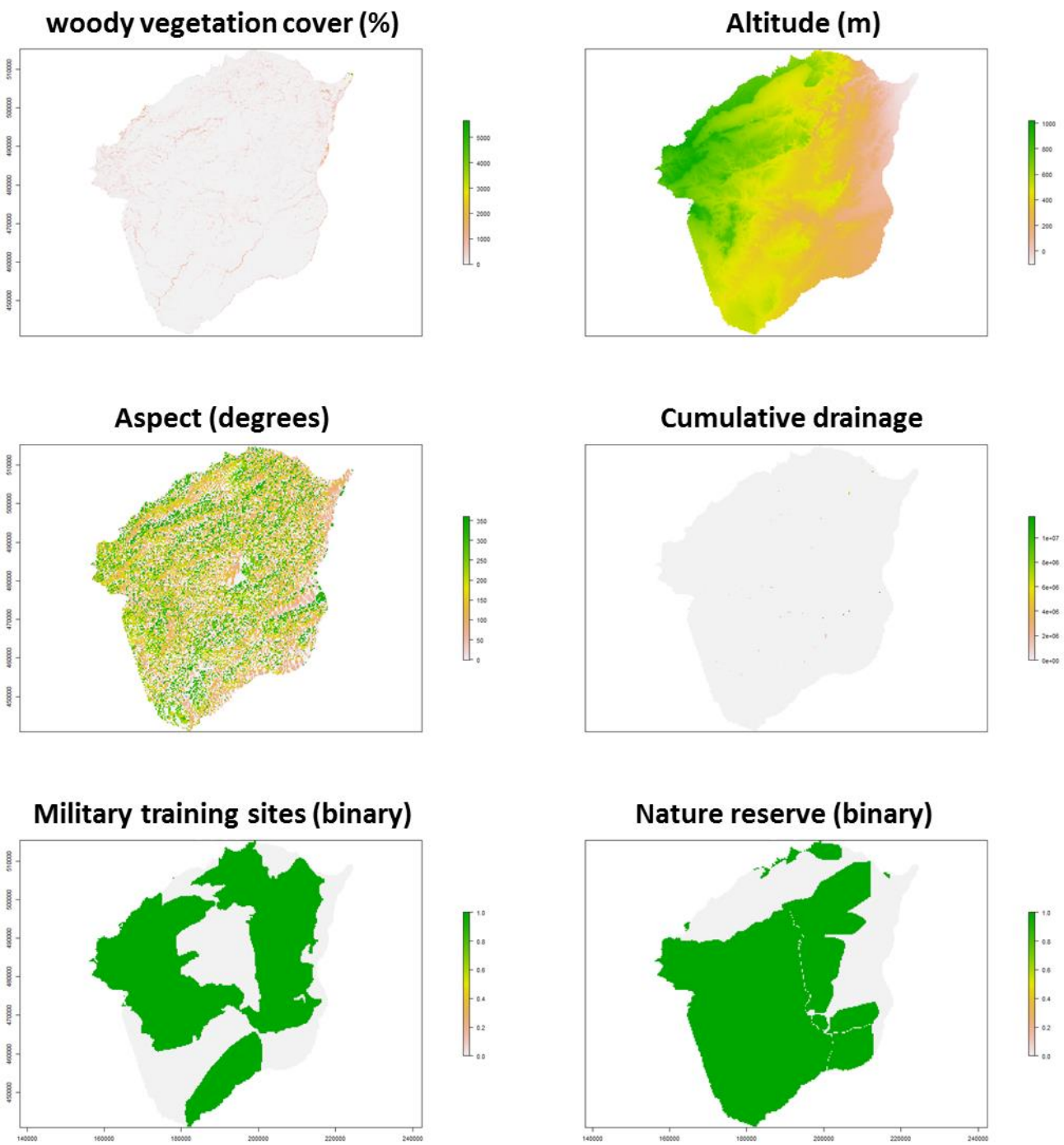
Supplementary material

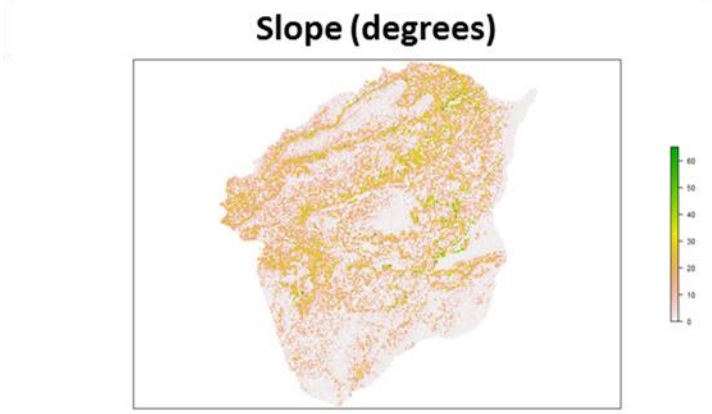
Appendix 1: Vegetation map.

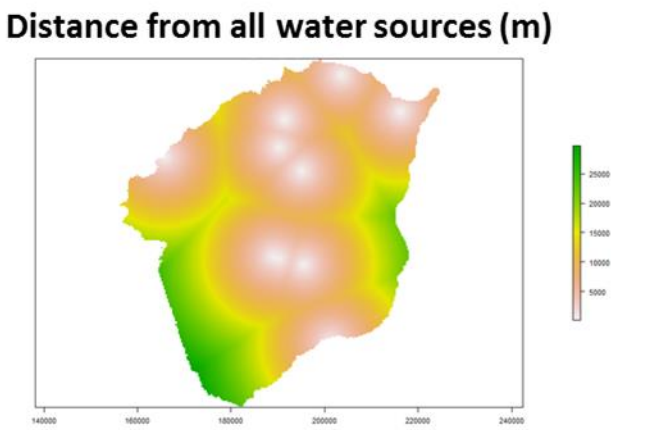
The vegetation layer was derived from a combination of computerized classification and manual interpretation of high-resolution (2 m pixel) aerial photographs of the area, taken in 2008 (two years prior to the feces survey). The classification scheme included two classes, woody vegetation and all other (mostly bare area). For some parts of the study area, an unsupervised classification provided sufficient results. However, in other parts, the dark bedrock and woody vegetation had very similar DN values. We found that only manual classification based on structural clues yielded satisfying results. The manual classification of these areas was laborious, but resulted in excellent accuracy. Assessment of the accuracy of the vegetation layer was conducted using manual interpretation of >500 control points on an air photo of the same area from a previous year. The overall accuracy of the binary vegetation classification was 0.96. This value is probably a conservative estimate (since some of the errors could be attributed to the control photo). The 2 m layer was downscaled to a resolution of 10 m, recording the proportion of woody vegetation in each 10 m cell. The resulting layer was processed using a moving window analysis, where each 10 m cell represented the averaged vegetation cover over a 100 m radius. The rationale for this was that the animal selects its location viewing the vegetation around it, not only its immediate location (Bar-David et al. 2005). A preliminary comparison between these two vegetation layers showed that they had very similar effect on the distribution model, and we selected the smoothed layer for the final models.

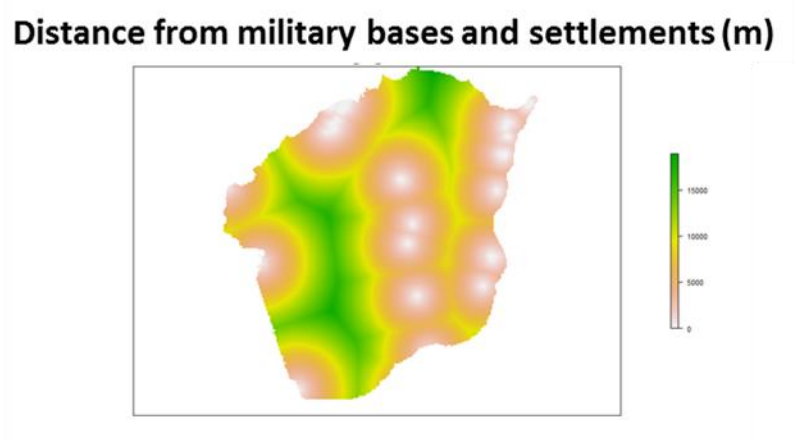
Appendix 2: Explanatory variables

Figure S1. Maps of the explanatory variables.




Slope (degrees)
60
50
40
30
20
10
0


Distance from all water sources (m)
25000
20000
15000
10000
5000
140000
160000
180000
200000
220000
240000


Distance from military bases and settlements (m)
15000
10000
5000
0

681

Appendix 3: Multicollinearity matrices

Table S2. Multicollinearity matrices for the variables used in the three models. (A) 10 m resolution. (B) 100 m resolution. (C) 1000 m resolution. Stars (*) indicate variables eliminated from the model due to high correlation with other variables.

#	Description	1	2	3	4	5	6	7	8	9	10	11	12	13
1	Aspect													
2	Altitude (DEM)	0.14												
3	* Distance from reintroduction sites	-0.02	-0.03											
4	* Distance from roads	0.04	0.14	-0.11										
5	Distance from all water sources	-0.02	0.09	0.56	0.05									
6	* Distance from watering holes	-0.09	-0.42	0.76	-0.05	0.69								
7	Distance from military bases and settlements	0.04	0.28	-0.14	0.83	0.05	-0.20							
8	Military training sites (binary)	0.02	0.21	-0.05	0.22	0.13	-0.03	0.27						
9	Cumulative drainage	-0.01	-0.02	-0.01	0.00	-0.01	0.00	0.00	0.00					
10	* Mean temperature in Aug	-0.13	-0.97	0.01	-0.11	-0.09	0.42	-0.28	-0.17	0.02				
11	* Mean annual precipitation	0.13	0.69	-0.17	0.18	-0.17	-0.46	0.25	-0.07	-0.01	-0.72			
12	Nature reserve (binary)	0.07	0.14	-0.02	0.25	-0.12	-0.21	0.18	-0.10	0.00	-0.14	0.19		
13	Slope	0.35	0.15	-0.05	0.12	-0.11	-0.13	0.09	0.02	-0.01	-0.13	0.19	0.19	
14	Vegetation	-0.08	-0.06	0.10	-0.07	-0.06	0.00	-0.07	-0.12	0.02	-0.01	0.06	0.00	-0.12

(A)

#	Description	1	2	3	4	5	6	7	8	9	10	11	12	13
1	Aspect													
2	Altitude (DEM)	0.19												
3	* Distance from reintroduction sites	-0.03	-0.04											
4	* Distance from roads	0.05	0.13	-0.12										
5	Distance from all water sources	-0.03	0.09	0.56	0.04									
6	* Distance from watering holes	-0.12	-0.42	0.76	-0.05	0.69								
7	Distance from military bases and settlements	0.05	0.28	-0.14	0.83	0.05	-0.20							
8	Military training sites (binary)	0.03	0.21	-0.06	0.23	0.13	-0.03	0.28						
9	Cumulative drainage	-0.04	-0.07	-0.03	0.00	-0.02	-0.02	0.00	-0.02					
10	* Mean temperature in Aug	-0.18	-0.97	0.01	-0.11	-0.09	0.43	-0.28	-0.18	0.05				
11	* Mean annual precipitation	0.17	0.69	-0.17	0.18	-0.17	-0.46	0.25	-0.07	-0.02	-0.73			
12	Nature reserve (binary)	0.09	0.15	-0.02	0.25	-0.12	-0.21	0.18	-0.10	0.02	-0.15	0.19		
13	Slope	0.41	0.18	-0.06	0.15	-0.13	-0.16	0.11	0.02	-0.04	-0.15	0.24	0.23	
14	Vegetation	-0.11	-0.06	0.10	-0.08	-0.06	0.00	-0.07	-0.13	0.07	-0.01	0.06	0.00	-0.14

(B)

Pearson correlation coefficients above 0.65 or less than -0.65

(C)

#	Description	1	2	3	4	5	6	7	8	9	10	11	12	13
1	Aspect													
2	Altitude (DEM)	0.38												
3	* Distance from reintroduction sites	-0.07	-0.04											
4	* Distance from roads	0.09	0.12	-0.15										
5	Distance from all water sources	-0.03	0.08	0.55	0.01									
6	* Distance from watering holes	-0.24	-0.43	0.75	-0.07	0.69								
7	Distance from military bases and settlements	0.10	0.27	-0.16	0.83	0.03	-0.21							
8	Military training sites (binary)	0.06	0.23	-0.08	0.24	0.13	-0.04	0.29						
9	Cumulative drainage	-0.09	-0.18	-0.10	0.00	-0.07	-0.05	-0.01	-0.03					
10	* Mean temperature in Aug	-0.37	-0.97	0.01	-0.08	-0.09	0.43	-0.27	-0.18	0.16				
11	* Mean annual precipitation	0.35	0.70	-0.17	0.18	-0.18	-0.48	0.24	-0.06	-0.06	-0.73			
12	Nature reserve (binary)	0.19	0.18	-0.04	0.26	-0.14	-0.24	0.20	-0.08	0.06	-0.17	0.22		
13	Slope	0.56	0.26	-0.11	0.23	-0.20	-0.25	0.18	0.04	0.00	-0.23	0.36	0.35	
14	Vegetation	-0.14	-0.10	0.21	-0.16	-0.10	0.03	-0.14	-0.26	0.11	0.00	0.07	-0.01	-0.18

Pearson correlation coefficients above 0.65 or less than -0.65

Appendix 4: Jackknife test of variable importance (a MAXENT output)

Figure S1. Results of the *Training gain* jackknife test of variable importance.

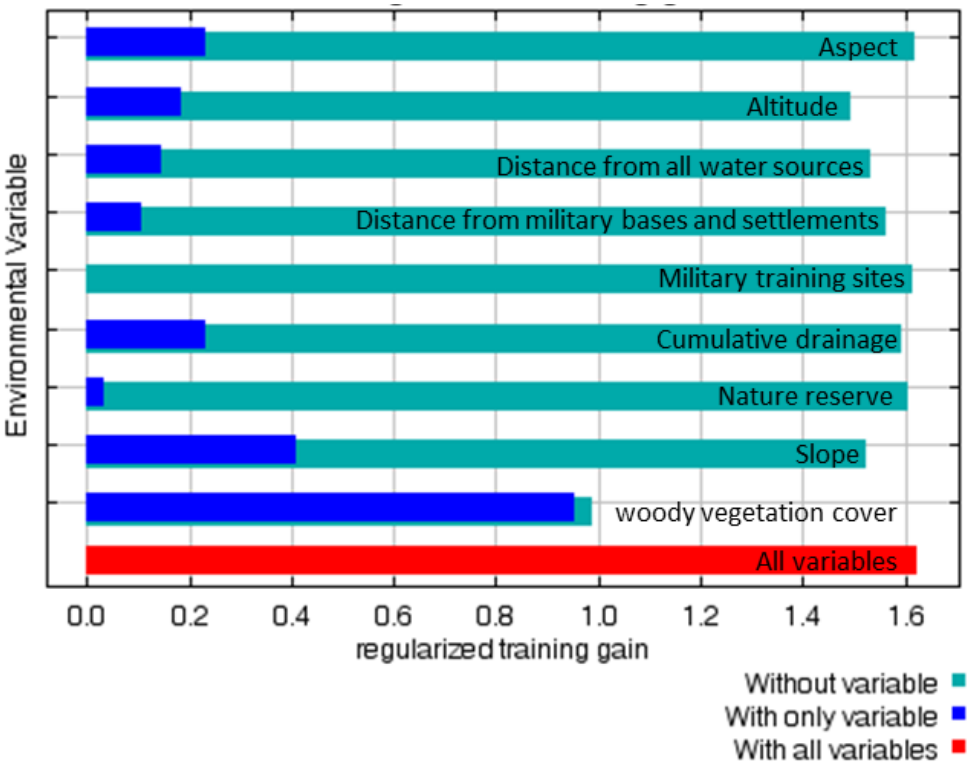


Figure S2. Results of the *Test gain* jackknife test of variable importance.

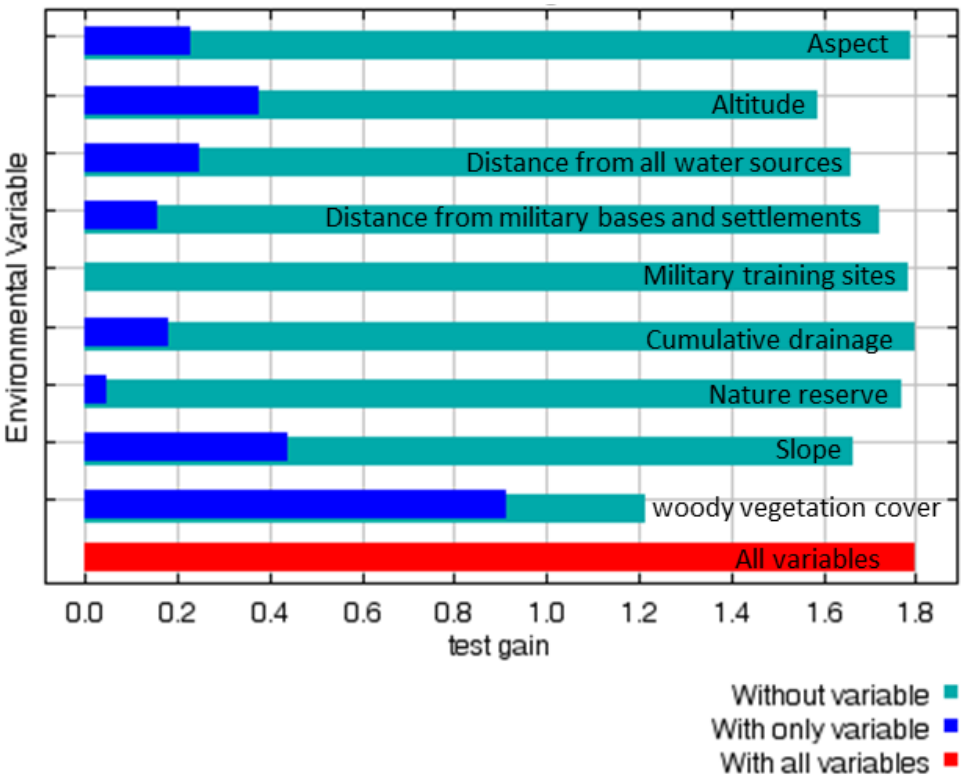
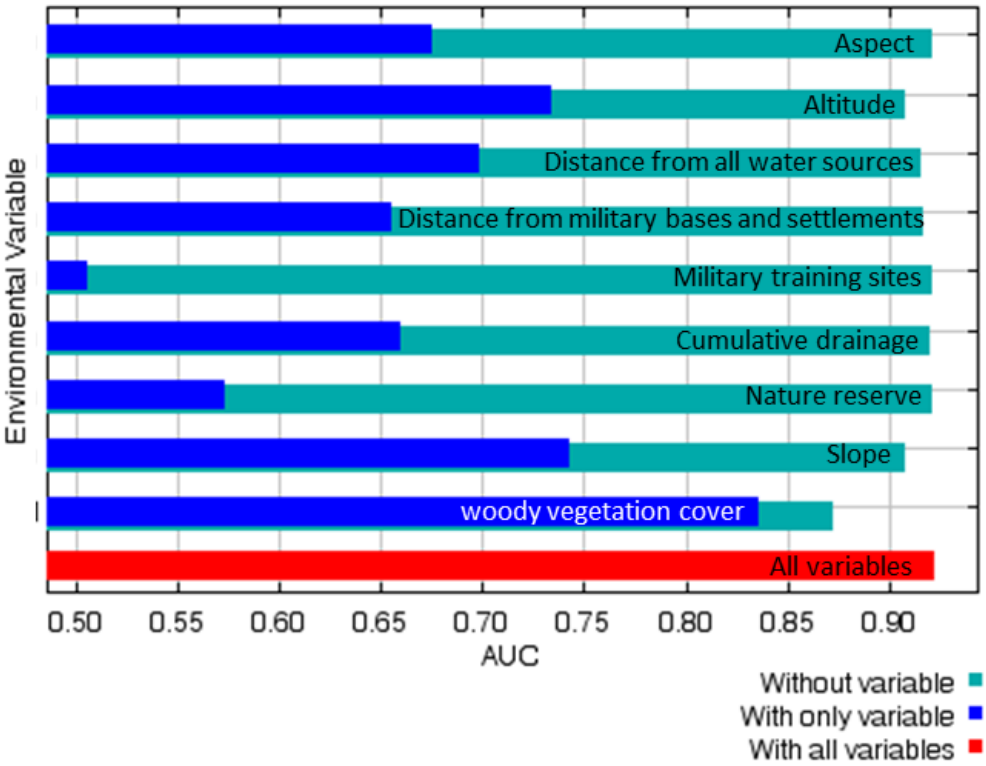


Figure S3. Results of the *Test AUC* jackknife test of variable importance.



The environmental variable with highest Training\Test gain and Test AUC when used in isolation is Vegetation, which therefore appears to have the most useful information by itself. The environmental variable that decreases the Training\Test gain and Test AUC the most when it is omitted is Vegetation, which therefore appears to have the most information that isn't present in the other variables.

Appendix 5: Response curves

Figure S4. Marginal response curves of the predicted probability of wild ass occurrences for predictor variables that contributed substantially to the distribution models. The y-axis is the predicted probability of suitable conditions as given by the logistic output format, with all other variables set to their average value over the set of presence localities. (A) Response of wild ass to woody vegetation cover. (B) Response of wild ass to slope.