Effects of variable selection on modelling habitat and potential distribution of the Andean bear in Bolivia

Ximena Velez–Liendo¹,³, Diederik Strubbe², and Erik Matthysen²

¹Centre of Biodiversity and Genetics, Faculty of Sciences and Technology, University of San Simon, Sucre and Parque La Torre, Cochabamba, Bolivia
²Evolutionary and Ecology Group, Department of Biology, University of Antwerp, Groenenborgerlaan 171, 2020 Antwerp, Belgium

Abstract: Species distribution models are used in ecology and conservation biology to draw inferences about the drivers of species’ ranges. However, poor conceptual background, environmental variable selection, and algorithm selection can contribute to misleading model predictions. We assessed the effects of environment variable selection and compared statistical performance and output maps of correlative resource- and biotope-based models for estimating the habitat and potential distribution of the Andean bear (Tremarctos ornatus) in Bolivia’s Tropical Andes. The resource-based approach estimated bear habitat using 7 resources associated with 3 ecological functions: feeding, shelter, and access to water. In contrast, the biotope model described the habitat by applying 11 environmental predictors related to topography, vegetation, and human activities. Both models performed equally well overall and better than random, with shelter as the most influential variable for the resource model and Yunga forest for the biotope model. However, discrepancies in the extent and arrangement of predicted bear distribution between models differed and emphasized the effect of variable selection, which could influence the delineation of conservation areas for this species. We suggest using a resource-based approach when modelling species distribution because of the more direct relationship to the species investigated and greater ease of interpreting results.

Key words: Andean bears, Bolivia, species distribution modelling, Tremarctos ornatus, variable selection

Species distribution models (SDM) attempt to reconstruct a species’ ecological requirements to estimate potential geographic distribution and habitat associations (Araújo and Guisan 2006, Peterson 2006). Use of SDMs has increased in recent years, with practical applications to conservation biology, ecology, and wildlife management (Martinez-Meyer et al. 2006, Newton-Cross et al. 2007, Gavashelishvili and Lukarevskiy 2008, LaRue and Nielsen 2008, Roth et al. 2008, Strubbe and Matthysen 2009). However, Kearney (2006) and Soberón (2007) discussed the potential for misleading model predictions when the conceptual framework of what is modeled (such as habitat or niche, selection of environmental variables, and modeling approach) is incorrect or poorly defined.

Habitat and niche are central, though sometimes inconsistently defined, concepts in ecology (Pearson and Dawson 2003, Mitchell 2005, Kearney 2006, Soberón 2007) and SDMs have been described as modelling a species habitat, fundamental niche, realized niche, or environmental niche (Kearney 2006), leading to controversy over what is actually modelled (Mitchell 2005). Generally defined as the place where an animal lives (Odum 1963, Morrison 1998), habitat is not restricted to a particular region, but describes the distributional response of a species to select environmental variables at a pre-determined scale of space and time (Whittaker et al. 1975). Although these variables are directly related to an organism’s presence, a mechanistic understanding of how its fitness is affected is not required (Kearney 2006). In contrast, the concept of niche necessarily implies the understanding of how these environmental variables (biotic and abiotic) affect an organism’s fitness (Kearney 2006). To integrate these ecological concepts into SDM, Kearney (2006) proposed to reserve the term ‘habitat modelling’ for descriptive
or correlative analyses of an organism’s environment, whereas niche modelling should be used for mechanistic analyses of how environmental factors influence a species’ fitness. In a correlative approach, the input is a set of species’ point occurrences and spatial environmental data. The output is a map of the probability of species’ occurrence throughout the study area. In a mechanistic approach, a species is modelled using functional traits (behavior, morphology, and physiology). The challenge is to link these traits to spatial environmental data in a mechanistic fashion (Kearney and Porter 2004). Model outputs are thus associated with environmental conditions that provide insight into possible causality of the observed relationship between species distribution and environmental data.

Variable selection is perhaps the most biological-based decision in SDM, yet, controversy still remains over which variables should be used to describe a species’ habitat. Often, researchers have associated species’ habitats with general vegetation or land-cover types, whereby species are assumed to prefer certain ‘habitat patches’ (i.e., preferred vegetation types) that occur within a matrix of unsuitable habitats (Schadt 2002, Mestre et al. 2007, Sattler et al. 2007). This view implies that a habitat is synonymous with a vegetation type or biotope (Dennis et al. 2003). However, species require a set of resources and conditions to complete their lifecycle (Dennis and Shreeve 1996, Weddell 2002), and although in some cases these resources and conditions may coincide with a particular vegetation type, this is generally not the case (Dennis et al. 2003). Within this context, Dennis et al. (2003) have called for the use of a bottom-up resource-based concept for habitat modelling. This approach considers the landscape as a continuum of overlapping resources and, although correlative, it aims to identify functional relationships between species and essential resources and conditions in their environment rather than using general vegetation types as surrogates (Vanreusel et al. 2007).

To test the effect of variable selection, we built a biotope-based model (BBM) that describes the environmental conditions where Andean bears (Tremarctos ornatus) occur, and a resource-based habitat model (RBM). The biotope-based model was constructed using general vegetation types, topography, and anthropogenic disturbance variables. In contrast, resource-based model variables defined bear habitat in terms of 3 ecological functions: access to food, shelter, and water. As there is no information available on the physiological ecology of Andean bears, the generation of a mechanistic niche modelling approach (sensu Kearney 2006) was not possible. Consequently, we used a correlative habitat modelling approach in which we relate bear presence data to environmental variables that indexed important ecological function. We then compared the performance and spatial output of these correlative habitat models using the Andean bear as the focal species. We describe the biological relevancy of each model and potential implications to Andean bear conservation.

**Study area**

The study area encompassed the eastern slope and inter-Andean valleys of the Bolivian Tropical Andes and comprised about 240,000 km² (Fig. 1). This area contains numerous microhabitats, each with diverse climate, topography, and geology. The western slope includes the Altiplano or Andean plateau, an area considered unsuitable habitat for Andean bears because of its phytogeographical characteristics and climate (Ibish et al. 2003). Geographically, the region can be divided in North and South sections (Ibish et al. 2003). The North spans from north of Titicaca Lake (15° 52′S–69°18′W) to the Amboro node (17°48′S–63°46′W) and is characterized by wide mountain ranges that run west to east and elevational gradients that range from 300 to 4800 m above sea level (asl). The vegetation of the North section is heavily influenced by the humidity of the Atlantic and the Amazon flora (Ibish et al. 2003). The South section is characterized by much lower mountains, running parallel in a north to south direction. It spans from the Amboro node to Tariquia (22°04′S 64°46′W) with elevations from 500 to 2500 m asl (Ibish et al. 2003). The vegetation in the South is much drier and is strongly influenced by the dry Chaco region and inter-Andean valleys. Despite strong anthropogenic impacts in the region, these valleys are still considered important centers of endemism (Lopez 2003, Larrea-Alcazar and Lopez 2005).

**Methods**

**Bear presence data**

We compiled 112 bear observations where the geographic locations of sightings or evidence of sign
(such as tracks) were reported in recent literature (Salazar and Anderson 1990, Rumiz et al. 1998, Azurduy 2000, Paisley 2001, Rios-Uzeda et al. 2006) and by personal communications with biologists, park rangers, and local residents ($n = 9$). Data were displayed on a 0.5 x 0.5 degree grid layer, and empty grid cells ($n = 29$) were selected to conduct further field work. During 2005–07 we conducted surveys for bear sign (food remains, claw marks on trees, ground and tree nests, footprints, scats, hairs, and...
direct observations) along 146 transects in 29 grid cells with an average transect length of 3.1 km (SD = 0.27 km) and average width of 14 m, covering an approximate area of 6.34 km². We used a modified transect methodology, which instead of surveying fixed transects at pre-established sites, surveyed only game trails along mountain ridgelines. This method maximizes efficiency by assessing locations with greater probabilities of sign (I. Goldstein, Wildlife Conservation Society, Merida, Venezuela, personal communication, 2007).

We first reviewed topographic maps to identify ridgelines and suitable access. At each site, we followed game trails characterized by low vegetation with evidence of frequent animal activity and searched for sign of bear activity. A location where sign of bear activity was recorded was considered a separate observation if the distance to the nearest record exceeded 1 km, which represents the median daily linear distance a bear traverses (Paisley 2001), and if more than one sign of bear presence was recorded. This last rule was applied to avoid confusion with other animals' signs such as tree marks made by jaguars (Panthera onca) or epiphytic bromeliads eaten by monkeys (Cebus apella). Finally, spatial autocorrelation was reduced by applying a buffer of 1 km around each record to the dataset.

Resource-based model (RBM) variables

Food. Based on feeding studies in Bolivia (Eulert 1995, Azurduy 2000, Paisley 2001, Rivadeneira 2001), the most common foods of Andean bears belong to the families Bromeliaceae, Ericaceae, Lauraceae, Poaceae, and Rosaceae. We obtained 3,679 geo-referenced occurrences in the study area of these families (Bromeliaceae 1,134; Ericaceae 627; Lauraceae 1406; Poaceae 170, and Rosaceae 342) from the Global Biodiversity Information Facility (GBIF) Data Portal (Appendix). To reduce potential inaccuracies, the GBIF database was evaluated by a local expert (N. De la Barra, Martin Cardenas Botanical Garden, Cochabamba, Bolivia). Plant family distributions were modeled as a function of 10 environmental variables, including 8 largely uncorrelated ($r < 0.70$) climate variables following Loiselle et al. (2008) and 2 topographical variables. Climate variables included annual mean temperature, mean diurnal range, isothermality, temperature seasonality, annual precipitation, precipitation of driest month, precipitation seasonality, and precipitation of warmest quarter, derived from the WORLDCLIM database (Version 1.4; http://www.worldclim.org; Hijmans et al. 2005). Topographical variables included elevation and slope derived from 90-m digital elevation model (DEM) data from the NASA Shuttle Radar Topographic Mission (SRTM; US Geological Survey 2004).

Shelter. We defined shelter as areas within forests where bears can find shade from midday heat, dens to give birth, and cover to build ground and tree resting nests (Velez-Liendo 1999, Azurduy 2000, Paisley 2001). We used forest cover from a multi–temporal Moderate Resolution Imaging Spectroradiometer (MODIS) Vegetation Continuous Fields (VCF) collection of forest cover change (2000–05; Mulligan 2007) to represent this variable. The raster-based MODIS-VCF data provides the percent of tree canopy cover obtained from monthly composites of 500-m resolution imagery (Hansen et al. 2003). We transformed MODIS–VCF data into a continuous raster layer (a layer where the phenomena have no clear boundaries and it is assumed that the value assigned to each cell is what is represented at the center of the cell) by applying the Distance module in ArcView version 3.2 (ESRI, Redlands, California, USA). Thus, the final layer represented distance to shelter.

Water. Although bears can rely on the succulent parts of bromeliads and bamboo when water is scarce, availability of water can modify movement patterns (Velez–Liendo 1999, Paisley 2001). We used the variable rivers to represent this requirement. This vector layer was obtained from the Digital Chart of the World (Danko 1992) and transformed into a continuous raster layer with a resolution of 30 arc-sec (about 1 km) using the Distance module in ArcView version 3.2. As with shelter, the final layer represented distance to river.

Biotope-based model (BBM) variables. Based on previous studies of habitat characterization (Velez-Liendo 1999, Paisley 2001, Sanchez-Mercado 2008, Sanchez-Mercado et al. 2008) and habitat use (Cuesta et al. 2003), we applied 11 variables associated with Andean bear habitat to build the biotope–based model. Topographic variables included: elevation (meters), slope (degrees), distance to rivers (see above), and 2 variables representing human influence: roads used by vehicles (Road) and trails used by humans only (Trail). Elevation was derived from a 90-m digital elevation model (DEM) created from 4 radar layers from the NASA Shuttle Radar Topographic Mission (SRTM; US Geological Survey 2004) and slope from a DEM.
applying the function Derive Slope from Surface Tools (version 1.6a; Jennes 2008) in ArcView version 3.2. We obtained vector data for rivers and trails from the Digital Chart of the World (Danko 1992) and transformed data into continuous raster layers to represent distance to water and trails, respectively. We included 6 vegetation variables corresponding to Ecoregions (Fig. 1): Sub–Andean Amazon forest, Yunga forest, Puna grasslands, Inter–Andean dry forest, Boliviano–Tucumano forest, and Chaco–Serrano forest. Detailed vegetation composition is described by Ibish et al. (2003). We obtained ecoregion data as independent polygons from NatureServe (2003), and we corrected these data based on Ibish et al. (2003), Navarro and Ferreira (2004), 4 LANDSAT TM+ mosaics (MDA Federal 2004), and high resolution images from GoogleEarth and National Aeronautics and Space Administration (NASA) WorldWind (2007). We transformed all variables to continuous raster layers and exported each as ASCII files for MaxEnt modelling.

**Model development.** We used maximum entropy algorithms (MaxEnt version 3.2.1; Phillips et al. 2006) to model Andean bear probability of occurrence using resource- and biotope-based parameters (Fig. 2). Based on a probabilistic framework, MaxEnt generates predictions using presence-only data. The probability of a species distribution is estimated based on the assumption that an incomplete empirical distribution can be approximated by finding the probability of maximum entropy. The estimated distribution, however, is subject to the constraint that the expected value for each environmental variable under the estimated distribution matches its empirical average (Loiselle et al. 2003). The MaxEnt distribution is then calculated by performing a series of iterations, in which the weights associated with the environmental variables are adjusted to maximize the average probability of the sample locations for the species (Buermann et al. 2008). The probability is displayed in terms of gain, a measure of the likelihood of the samples, and is used to estimate the distribution throughout the study area. Thus, the distribution expresses the suitability of each pixel as a function of the environmental variables for that pixel (Phillips et al. 2006).

Variable importance was evaluated by their relative percent contribution to both models. This value is calculated by the increase in regularized gain added to the corresponding variable at each iteration of the training algorithm. MaxEnt also computes

![Fig. 2. Steps and variable inclusion for a resource-based maximum entropy model (a) and biotope-based maximum entropy model (b) to estimate Andean bear probability of occurrence.](image)
response curves to illustrate how the contribution to the prediction is influenced by a particular environmental variable (Phillips et al. 2006). Finally, the MaxEnt output format is a logistic function representing the probability that the species is present in a particular pixel, based on the environmental conditions. It is scale-independent, with values ranging from 0 to 1, and pictured by default using a linear scale (Phillips 2008).

We performed resource and biotope models using a resolution of 30 arc-seconds of longitude and latitude (~1 km²). Occurrence data was randomly assigned into training (70%) and testing (30%) locations, with 100 iterations performed for each model. We used the recommended convergence threshold of $10^{-5}$ and maximum interactions of 500 values, but we used 100,000 background points instead of the default number of points ($n = 10,000$) because the study area had a large number of cells (>300,000). We also used a regularization multiplier of 3 to reduce over-fitting of the data. The selection of features (derived from environmental variables) was carried out automatically.

The predictive ability of both models was evaluated using the area under the receiver operating characteristic (ROC) curve AUC. Typically, a ROC curve is a plot of the sensitivity versus (1 – specificity) for a binary classifier system, where sensitivity is the fraction of all positive instances correctly classified, while specificity is the fraction of all negative instances correctly classified (Fawcett 2006). However, for presence-only data, presence is distinguished from background rather than from absences and the maximum achievable AUC value is less than 1 (Wiley et al. 2003). Swets (1998) categorized models with values >0.9 as highly accurate, 0.7–0.9 as useful, and <0.7 as poorly accurate. Models with a large number of variables often have greater AUC values (Baldwin and Bender 2008); thus, we performed a critical ratio test (Pearce and Ferrier 2000) to determine if these differences were significant at $\alpha = 0.05$. We then mapped probability of Andean bear presence using the biotope and resource model, and categorized probability of presence as high (>0.5), medium (0.3 to ≤0.5), and low (0.2 to <0.3).

Results
We compiled 565 presence-only records: 453 from field work and 112 from the literature review. We attributed the skewed distribution (Fig. 1) to the actual distribution of Andean bears rather than collection bias.

All 5 plant models used as input data for the resource model performed better than random, with AUC values >0.75 (Bromeliaceae = 0.877, Ericaceae = 0.953, Lauraceae = 0.938, Poaceae = 0.939, Rosaceae = 0.948).

Model performance and variable importance
The AUC evaluation of the performance of biotope-based model and resource-based model approaches were considered useful (0.823 and 0.851, respectively), with no difference between models ($z = 0.9462, P = 0.344$). The relative contributions of environmental variables to the resource model were: shelter, 31.8%; Ericaceae, 24.6%; Bromeliaceae, 15.22%; Lauraceae, 6.7%; Poaceae, 5.69%, and Rosaceae, 9.92%. The response curves computed by MaxEnt indicate that Andean bear habitat appears defined primarily by locations within the forest canopy and the occurrence of Bromeliaceae and Ericaceae (Fig. 3). On the other hand, the relative contributions of environmental variables to the biotope model were: Yunga forest, 33.3%; Chaco Serrano forest, 11.4%; elevation, 10.6%; Inter–Andean dry forest, 10.2%; Puna grasslands, 9.3%; Boliviano–Tucumano forest, 5.9%; Rivers, 4.9%; Road, 4.9%; Trail, 3.8%; Sub–Andean Amazon forest, 3.8%; and slope, 2.4%. Response curves indicate the probability of bear presence within the limits of Yunga forest, outside Chaco–Serrano forest, and between 1,000–3,000 m asl (Fig. 3).

Visualization of presence
Despite similarity in statistical performance, the models differed in extent and arrangement of potential distribution (Fig. 4). Overall model predictions agreed for 16.3% of the area. Areas with a high probability (>0.5) of bear presence in both models were predominantly in the North section. The resource model exhibited a more gradual change in distribution influenced by the convergence of greater probabilities of Bromeliaceae and Ericaceae presence. Conversely, in the biotope model these high probability patches were restricted to 4 large discrete areas corresponding largely to Yunga forest. Overall, 20.7% of high probability areas were similarly predicted by both models (Fig. 4).

In both biotope-based model and resource-based model maps, medium probability of occurrence
areas in the North section connected or buffered areas of high (>0.5) probability of occurrence. However, in the biotope model map, a division within the North patches was observed, corresponding to the presence of Inter–Andean dry forest. In the South section, predicted areas differed between models. The resource model showed parallel patches running north to south, corresponding to distance to shelter and to a lesser extent, Bromeliaceae (0.3–0.5 of probability of presence). In contrast, the biotope model predicted 2 areas both influenced by distance to water and restricted by the presence of Puna grasslands in the West. The area of agreement predicted by both models was 18.9%.

Finally, a matrix of low probability areas surrounded medium probability patches in both models, connecting and expanding both distributions. General patterns followed shelter distribution for the resource model, and avoidance of Chaco Serrano and Sub–Andean Amazon Forests for the biotope model. The area of agreement predicted by both models was 11.8%.

Discussion

Model performance

The resource and biotope models used performed better than random (AUC scores >0.8) in predicting the probability of occurrence of the Andean bear. However, output maps showed considerable differ-
ences between them, with only 16% of the study area similarly predicted. Phillips et al. (2006) advised that despite the reasonable performance of MaxEnt, environmental variables should only include those most relevant to the species. Although both models were built with this consideration, substantial differences in the extent and arrangement of predicted bear occurrence between models occurred. While these differences can be explained by the statistical performance of each variable in the overall distribution output, a critical issue is to discriminate which variables provided the greatest biological relevance to the overall models.

The more gradual gradient of probability of occurrence in the resource model map was a result of the distribution and contribution of shelter (31%), Ericaceae (25%), and Bromeliaceae (15%). Distribution of these variables followed natural gradients and, unlike biotope variables, were not mutually exclusive and showed considerable overlap. Shelter was a categorical variable encompassing all forest types. Because bear occurrence was predominantly recorded within forest canopy (Fig. 1), probability of bear occurrence was expected to be positively associated with this variable. However, the second and third variables in importance, Ericaceae and Bromeliaceae, may have constrained the influence of shelter and led to over-estimation of probability of bear occurrence in areas with high canopy cover, but were otherwise unsuitable (such as the Amazon Basin).

Unlike the resource model, the more clumped and restricted high probability areas from the biotope model output was a consequence of the distribution and influence of Yunga forest and, to a lesser extent, by Chaco Serrano forest and elevation. The importance of Yunga forest was expected, as more than 50% of bear presence points were recorded within this ecoregion (Fig. 1). Despite converting all categorical variables to continuous raster layers, the clumped output of the biotope model was caused by a boundary effect of the 2 main variable contributors, Yunga forest and Chaco-Serrano forest. Though our ecoregions system is the most common vegetation classification system for Bolivia, an important issue is whether the different ecoregions are perceived as different by Andean bears. For example, Yunga and Boliviano Tucumano forest are phytogeographically different (Ibish et al. 2003), but functionally similar (both are considered to contain similar resources) for the Andean bear.

The interpretation of variable importance in relation to ecological requirements of bears not only assisted in understanding model output, but in determining whether these predictors have biological significance. For instance, the biological relevance of shelter in the resource model could be explained by the extensive arboreal activity of Andean bears, including building nests or platforms and feeding on fruits, epiphytic bromeliads, and orchid pseudobulbs (Peyton 1999), whereas Bromeliaceae and Ericaceae are related to feeding behavior. Ericaceae are seasonal food sources representing up to 27% of the bear's diet during the late wet season (Paisley 2001). In contrast, Bromeliaceae are consumed year round as a diet supplement (Peyton 1980, Velez-Liendo 1999, Azurduy 2000, Paisley 2001) or occasionally as a primary food during portions of the wet season (Paisley 2001). In spite of low nutritional content, Bromeliads have played a fundamental dietary role in the evolution of the Andean bear from a carnivore ancestor (Sacco and Valkenburgh 2004, Christiansen 2007) to a species feeding almost exclusively on vegetation. In fact, the Andean bear is the only species that feeds on terrestrial species of bromeliads. Therefore, it was not surprising that the distribution of this species encompassed large extent areas where bromeliads are available.

Biotope model variables showed no direct relevance to the bear’s biology. Yunga forest is a continuous vegetation unit influenced by the humid Amazon basin. Thus, food is assumed to be available almost year-round and shelter is plentiful. The Chaco Serrano forest was associated with bear absence rather than presence, which could be related to poor forest cover. Elevation does not have a direct biological role on bear habitat, but rather in limiting the distribution of vegetation classes.

**Conservation implications**

The accelerated habitat degradation caused by human expansion and effects of climate change are causing an alarming loss of biodiversity worldwide (Goldstein et al. 2008). Conservation strategies for the Tropical Andes have focused on protecting habitats of umbrella species such as the Andean bear, for example, the landscape species program of the Wildlife Conservation Society (Coppolillo et al. 2004). However, incomplete datasets on species’ distributions have hindered these conservation programs, and modelling techniques to reconstruct
species’ ecological requirements and predict their distribution have become useful tools to generate such information.

The resource-based model and the biotope-based model performed better than random, yet with considerable differences in their predictions of suitable bear habitat. The main advantage of the biotope model was that predictors were easier to obtain and more accurate than the resource model. In contrast, the resource model included an additional potential source of error by using third-party datasets and plant models as data input. However, this error was reduced by using high-quality data and independent experts for data review. The biotope model showed a strong bias toward the North section of Bolivia’s Tropical Andes study area and predicted few areas with medium probability of bear presence, which could play an important role as corridors or buffer zones. If the biotope model were used to address conservation issues, it might limit efforts to protect Yunga forests and underestimate the potential presence of this species in the South section of the study area. In contrast, the resource model produced output that identified areas where essential resources such as bromeliads and shelter occur. This additional information absent in the biotope model could be used to guide or support the establishment or expansion of protected zones by selecting areas with specific food sources, shelter, or both. Furthermore, the output map also identified zones of potential bear presence where no bears were previously recorded, and these predictions could be used as baseline information to conduct additional surveys to determine species’ presence. The main drawback of the resource-based approach was selecting and acquiring environmental predictors. It requires knowledge of the species’ biology to determine which variables are most relevant, and data might be unavailable, restricted to specific locations, or inaccurate. These disadvantages might be more evident with species such as the Andean bear, where the biology is poorly known and ecological studies usually require long–term evaluations.

Conclusions

In recent years, the use of SDM has increased dramatically due to improved access to geographical and biological data and user-friendly statistical modelling software (MaxEnt, Garp, DivaGIS, Biomapper). Nevertheless, use of SDMs to address conservation issues requires the involvement of species experts from project conception through model construction, selection, and interpretation. Based on our results and use of SDM to address conservation and management issues, we recommend using a resource-based approach because of the more direct biological relationships modelled and an improved ability to interpret outputs based on species’ ecology.

Acknowledgments

This study was funded by The Russell E. Train Education for Nature Program–WWF Scholarship, International Association for Bear Research and Management–John Sheldon Bevins Memorial Foundation, Parks in Peril Programme–Conservation International and Centre of Research and Conservation, and the Royal Zoological Society of Antwerp. We thank M. Mulligan, Department of Geography, King’s College London, for providing the forest canopy layer. We appreciate the support provided by the Centre of Biodiversity and Genetics, University of San Simon in Bolivia and cooperation from the Belgian Flemish Interuniversity Council (VLIR-UDC).

Literature cited

Photogrammetric Engineering and Remote Sensing
58:1125–1128.
DENNIS, B., R.A. DESHARNAIS, J.M. CUSHING, S.M.
and Irish offshore islands: Ecology and biogeography.
EULERT, C. 1995. Evaluación del estado actual del jucumari
Tremarctos ornatus en el Parque Nacional Amboro, Santa
Cruz, Bolivia. Dissertation, Universidad Autónoma
Gabriel Rene Moreno, Santa Cruz, Bolivia. (In Spanish.)
GAVASHELISHVILI, A., AND V. LUKAREVSKY. 2008. Model-
ing the habitat requirements of leopard Panthera pardus
in west and central Asia. Journal of Applied
GOLDSTEIN, I., X. VELLEZ-LIENDO, S. PAISLEY, AND D.L.
GARSHIELIS. 2008. Tremarctos ornatus. In IUCN Red
Switzerland, and Cambridge, UK.
HANSEN, M.C., R.S. DEFFRIES, J.R.G. TOWNSEND, M.
Global percent tree cover at a spatial resolution of
500 meters: First results of the MODIS vegetation
HIMANS, R.J., S.E. CAMERON, J.L. PARRA, P.G. JONES,
and A. JARVIS. 2005. Very high resolution interpolated
climate surfaces for global land areas. International
IBISH, P.L., S.G. BECK, B. GERRMANN, AND A. CARRETERO.
Ibish and G. Merida, editors. Biodiversidad: La riqueza
vegetal de Bolivia. Estado de conocimiento y conservacio-
ón potencial de Bolivia: Una base para el análisis de
vacíos de conservación. Revista Boliviana de
Ecologı´a y Conservacio ´n del medio ambiente 15:1–40.
(In Spanish.)
JIMÉNEZ-MEYER, E., A.T. PETERSON, J.I. SERVIN, AND
L.F. KIFF. 2006. Ecological niche modelling and
prioritizing areas for species reintroductions. Oryx
MDA FEDERAL 2004. Landsat GeoCover 2000/ETM+
umd.edu/data/mosaic/, accessed 21 February 2012.
Modelling the distribution of the European polecat
Mustela putorius in a Mediterranean agricultural
MITCHELL, S.C. 2005. How useful is the concept of habitat?
MORRISON, M.L. 1998. Wildlife–habitat relationships:
Concepts and applications. University of Wisconsin,
Press, Madison, Wisconsin, USA.
MULLIGAN, M. 2007. VCF_change: A Google Earth
visualisation of MODIS VCF tree cover change globally
trees?, accessed 21 February 2012.
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION.
2007. World Wind 1.4. NASA Ames Research Centre,
 Moffett Field, California, USA. http://worldwind.arc.
nasa.gov/, accessed 21 February 2012.
NATURESERVE. 2003. Ecological systems of Latin American
and the Caribbean database. Version 1.01. Nature-
Serve, Washington DC, USA. www.natureserve.org/
getData/LACeologyData.jsp, accessed 21 February 2012.
NAVARRO, G., AND W. FERREIRA. 2004. Zonas de vegeta-
tción potencial de Bolivia: Una base para el análisis de
vacíos de conservación. Revista Boliviana de
Ecología y Conservación del medio ambiente 15:1–40.
(In Spanish.)
Modelling the distribution of badgers Meles meles:
Comparing predictions from field-based and remotely
derived habitat data. Mammal Review 37:54–70.
Appendix. Full listing of plant data sources retrieved from the GBIF network, all accessed on 29 April 2008. The URL for each source starts with http://data.gbif.org/datasets/resource/ and ends with the number listed as URL ending.

<table>
<thead>
<tr>
<th>Source</th>
<th>Plant family</th>
<th>URL ending</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAU Herbarium Database</td>
<td>Ericaceae, Poaceae, Rosaceae</td>
<td>224</td>
</tr>
<tr>
<td>Australian National Herbarium (CANB)</td>
<td>Poaceae</td>
<td>47</td>
</tr>
<tr>
<td>California State University, Chico</td>
<td>Bromeliaceae</td>
<td>737</td>
</tr>
<tr>
<td>Canadian Museum of Nature Herbarium</td>
<td>Rosaceae</td>
<td>123</td>
</tr>
<tr>
<td>CGN-PGR</td>
<td>Poaceae</td>
<td>1102</td>
</tr>
<tr>
<td>EURISCO, The European Genetic Resources Search Catalogue</td>
<td>Poaceae</td>
<td>1905</td>
</tr>
<tr>
<td>EURISCO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fairchild Tropical Botanic Garden Virtual Herbarium Darwin Core format</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harvard University Herbaria</td>
<td>Bromeliaceae, Ericaceae, Lauraceae, Poaceae, Rosaceae</td>
<td>1827</td>
</tr>
<tr>
<td>Herbario</td>
<td>Rosaceae</td>
<td>566</td>
</tr>
<tr>
<td>Herbario del Instituto de Ecología, A.C., México (IE-XAL)</td>
<td>Lauraceae</td>
<td>1597</td>
</tr>
<tr>
<td>Herbarium Berolinense</td>
<td>Bromeliaceae, Ericaceae, Lauraceae, Poaceae, Rosaceae</td>
<td>1095</td>
</tr>
<tr>
<td>Herbarium W</td>
<td>Bromeliaceae, Rosaceae</td>
<td>1496</td>
</tr>
<tr>
<td>Institut Botanic de Barcelona, BC</td>
<td>Rosaceae</td>
<td>299</td>
</tr>
<tr>
<td>Internation Botanical Collections (S)</td>
<td>Bromeliaceae, Ericaceae, Lauraceae, Poaceae, Rosaceae</td>
<td>1983</td>
</tr>
<tr>
<td>IPK Genebank</td>
<td>Poaceae</td>
<td>1851</td>
</tr>
<tr>
<td>Lund Botanical Museum (LD)</td>
<td>Ericaceae</td>
<td>1028</td>
</tr>
<tr>
<td>MEXU/Plantas Vasculares</td>
<td>Poaceae</td>
<td>780</td>
</tr>
<tr>
<td>Missouri Botanical Garden</td>
<td>Bromeliaceae, Ericaceae, Lauraceae, Poaceae, Rosaceae</td>
<td>621</td>
</tr>
<tr>
<td>National Herbarium Nederland - Leiden Branch</td>
<td>Bromeliaceae, Ericaceae, Lauraceae, Poaceae, Rosaceae</td>
<td>1085</td>
</tr>
<tr>
<td>Nordic Genetic Resources</td>
<td>Poaceae</td>
<td>1487</td>
</tr>
<tr>
<td>NSW herbarium collection</td>
<td>Poaceae</td>
<td>968</td>
</tr>
<tr>
<td>Paleobiology Database</td>
<td>Ericaceae, Rosaceae</td>
<td>563</td>
</tr>
<tr>
<td>Phanerogamie</td>
<td>Bromeliaceae, Ericaceae, Poaceae, Rosaceae</td>
<td>1506</td>
</tr>
<tr>
<td>Real Jardin Botanico (Madrid), Vascular Plant Herbarium (MA)</td>
<td>Bromeliaceae, Ericaceae, Lauraceae, Poaceae, Rosaceae</td>
<td>240</td>
</tr>
<tr>
<td>Royal Botanic Gardens, Kew</td>
<td>Bromeliaceae, Lauraceae, Poaceae, Rosaceae</td>
<td>629</td>
</tr>
<tr>
<td>Seeds in long term storage</td>
<td>Poaceae</td>
<td>1776</td>
</tr>
<tr>
<td>Staatliches Museum für Naturkunde</td>
<td>Bromeliaceae</td>
<td>1100</td>
</tr>
<tr>
<td>Stuttgart, Herbarium</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The System-wide Information Network for Genetic Resources (SINGER)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type herbarium, Gottingen (GOET).</td>
<td>Lauraceae, Poaceae, Rosaceae</td>
<td>1494</td>
</tr>
<tr>
<td>UCD Botanical Conservatory</td>
<td>Bromeliaceae</td>
<td>739</td>
</tr>
<tr>
<td>United States National Plant Germplasm System Collection</td>
<td>Bromeliaceae, Ericaceae, Poaceae, Rosaceae</td>
<td>1429</td>
</tr>
<tr>
<td>University of California Botanical Garden DIGIR provider</td>
<td>Bromeliaceae</td>
<td>1412</td>
</tr>
<tr>
<td>USU-UTC Specimen Database</td>
<td>Poaceae</td>
<td>1508</td>
</tr>
<tr>
<td>Vascular Plant Type Specimens</td>
<td>Bromeliaceae, Ericaceae, Lauraceae, Poaceae, Rosaceae</td>
<td>731</td>
</tr>
</tbody>
</table>