Biodivers Conserv
DOI 10.1007/s10531-009-9615-5

ORIGINAL PAPER

Using habitat suitability models to sample rare species in high-altitude ecosystems: a case study with Tibetan argali

Navinder J. Singh · Nigel G. Yoccoz · Yash Veer Bhatnagar · Joseph L. Fox

Received: 9 May 2008 / Accepted: 19 February 2009

Abstract Models of the distribution of rare and endangered species are important tools for their monitoring and management. Presence data used to build up distribution models can be based on simple random sampling, but this for patchy distributed species results in small number of presences and therefore low precision. Convenience sampling, either based on easily accessible units or a priori knowledge of the species habitat but with no known probability of sampling each unit, is likely to result in biased estimates. Stratified random sampling, with strata defined using habitat suitability models [estimated in the resource selection functions (RSFs) framework] is a promising approach for improving the precision of model parameters. We used this approach to sample the Tibetan argali (*Ovis ammon hodgsoni*) in Indian Transhimalaya in order to estimate their distribution and to test if it can lead to a significant reduction in survey effort compared to random sampling. We first used an initial sample of argali feeding sites in 2005 and 2006 based on a priori selected vantage points and survey transects. This initial sample was used to build up an initial distribution model. The spatial predictions based on estimated RSFs were then used to define three strata of the study area. The strata were randomly sampled in 2007. As expected, much more presences per hour were obtained in the high quality strata compared to the low quality strata—1.33 obs/h vs. 0.080/h. Furthermore the best models selected on the basis of the prospective sample differed from those using the first a priori sample, suggesting bias in the initial sampling effort. The method therefore has significant implications for decreasing sampling effort in terms of sampling time in the field, especially when dealing with rare species, and removing initial sampling bias.

N. J. Singh · N. G. Yoccoz (✉) · J. L. Fox
Department of Biology, Faculty of Mathematics and Natural Sciences, University of Tromsø, 9037 Tromsø, Norway
e-mail: nigel.yoccoz@ib.uit.no

N. J. Singh
e-mail: n.singh@imperial.ac.uk

N. J. Singh · Y. V. Bhatnagar
Nature Conservation Foundation, 3076-5, IV Cross, Gokulam Park, Mysore 570002, India

Published online: 10 March 2009

Keywords Efficiency · Transhimalaya · Stratified random sampling · Sampling bias · Effort

Introduction

Developing effective methods for identifying factors shaping the distribution and abundance of rare species is of prime concern for conservation (Thompson 2004; Guisan et al. 2006). Crucial objective is to obtain an accurate estimate of use for a given level of total survey effort, or to achieve a desired level of precision with minimal effort (Mackenzie and Royle 2005). Accuracy means low bias and high precision—the former is usually achieved using some form of random sampling (e.g. Yoccoz et al. 2001; Williams et al. 2002). Simple random sampling, however, often leads to too few observations of rare, patchy species (Edwards et al. 2005). This prevents robust analyses (Green and Young 1993; Edwards et al. 2004) e.g. for use in explanatory or predictive models of suitable habitat or spatial distribution (Hill and Keddy 1992; Wiser et al. 1998). Convenience sampling may result into more observations but, because units are selected a priori with unknown selection probability, is likely to lead to biased models. Stratified random sampling has therefore been recommended as a way to improve precision (e.g. Williams et al. 2002).

An efficient approach to define strata is through predictive habitat distribution or ecological niche modelling (Guisan and Zimmermann 2000; Peterson 2001). These models statistically relate the geographic distribution of the species to their environment by estimating the effects of biotic and abiotic factors. These models have already been used in turn to sample rare lichen and plant species (Edwards et al. 2005; Guisan et al. 2006). Guisan et al. (2006) recommended also an iterative approach based on model-fitting and new sampling based on previous model. Studies focusing on animals have used concepts of resource selection (Manly et al. 1993; Schaefer and Messier 1995; Boyce et al. 2002; Walker et al. 2007). The units selected by animals are conceived as resources and predictor variables associated with these resource units may be elevation, forage quality/quantity, predation, competition, and disturbance. A resource selection function (RSF) is a model that yields values proportional to the probability of use of a resource unit (Manly et al. 1993). Thus, a proportional probability of use of the given habitat or resource can be estimated through RSF which can be used to stratify the sampling area. This sampling approach is therefore model-based in the sense that strata are defined on the basis of an RSF, but the sampling itself follows the rules of probability sampling with known probabilities of sampling study units (Yoccoz et al. 2001).

Many rare and endangered species such as the snow leopard (*Uncia uncia*), Tibetan argali (*Ovis ammon hodgsoni*), Tibetan antelope (*Pantholops hodgsonii*) and Tibetan gazelle (*Procapra picticaudata)* inhabit the remote regions of the Indian Transhimalayas and Tibetan plateau (Schaller 1977; 1998). There is an urgent need to assess the status of these species, but the regions were little explored by biologists until recently. Most of the available information is based on visual surveys in the existing and proposed reserves (Schaller 1977; Schaller and Gu 1994; Chundawat and Qureshi 1999; Harris et al. 1999; Bhatnagar and Wangchuk 2001; Harris and Loggers 2004; Harris et al. 2005). These surveys do not provide reliable estimates of population distribution. However, broad scale distribution maps have been prepared based on those surveys (Jackson and Hunter 1996; Shackleton 1997; Schaller 1998; Chundawat and Qureshi 1999; Harris et al. 2005) and provide an opportunity to focus sampling effort so as to refine the estimated distribution within the broader population range. This can be achieved by understanding the fundamental ecological requirements based on environmental characteristics of known occurrence sites (Kadmon and Heller 1998; Peterson et al. 1999).

We used the model-based sampling approach to analyse distribution of the Tibetan argali, a species with a patchy distribution over the Tibetan plateau (Schaller 1998; Bhatnagar and Wangchuk 2001; Harris et al. 2005). More specifically, we used resource selection functions to direct a stratified random sampling of the study area to improve the sampling of argali. We assessed particularly by how much the number of observations per unit effort was increased in high vs low intensity of use strata.

Materials and methods

Study area

The study area, known as Tso Kar basin, (*c*. 650 km^2) is situated in the Changthang region of Ladakh, India (32°15′N, 78°00′E)(Fig. 1). The altitude ranges from 4,500 to 6,371 m (Table 1). The area is comprised of rolling hills that enclose a basin and two lakes, a smaller freshwater lake (4 km^2) and a larger salt water lake (16 km^2). The high altitude cold desert ecosystems climate has temperatures oscillating from −40°C (Min. winter) to 25°C (Max. summer) and a mean annual precipitation of about 200 mm (Rawat and Adhikari 2005). The alpine and desert steppe vegetation is mainly composed of grasses (Poaceae), sedges (Cyperaceae), and short dicotyledonous forbs and shrubs (Rawat and Adhikari 2005). About 150 argali are believed to inhabit the basin region. Other wild ungulates found in the region include a population of over 300 kiangs (*Equus kiang*), and 50 blue sheep or bharal (*Pseudois naur*). About 18,000 livestock comprised of sheep, goats, yaks and horses use the area in winter. The key predator for the wildlife and livestock is the Tibetan wolf (*Canis lupus chanco*), which is found in small numbers. Snow leopard and lynx (*Lynx lynx*) are also present.

Implementation of iterative approach

The approach is described in Fig. 2. The initial sampling surveys were conducted during the summers 2005–2006 using predetermined vantage points and direct observations of feeding sites used by argali (sample I). The feeding sites were visited and geographic coordinates were recorded using GPS (position accuracy—up to 3 ± 1.5 m). The 80 observations made in 2005–2006 were used as the training data. The availability points were randomly sampled using Hawth's tools in Spatial ArcGIS program (ESRI Inc.). We used Use-Availability design (Johnson et al. 2006) to estimate Resource Selection Functions (RSFs) (Fig. 3a). We then categorized the predictions into 3 strata based on the RSF values (range 0–0.99) as lowest (0–0.2), medium (0.21–0.4) and highest (0.41–0.99) suitability (Fig. 3b). The strata were defined in order to achieve a compromise between a not too wide range of suitability and not too large difference in area. The lowest, medium and highest strata covered 68.6, 18.5 and 12.8% of the total area, respectively. As one of our goals here was to assess the effort needed to achieve a given number of observations, the number of transects was larger in the lowest stratum. Accordingly we randomly located 15, 5 and 4 transects in the lowest, medium and highest strata, respectively. We visited these transects in the field on a random basis and ensured that all transects were repeated five times during the sampling season. New occurrences, when found were recorded using GPS.

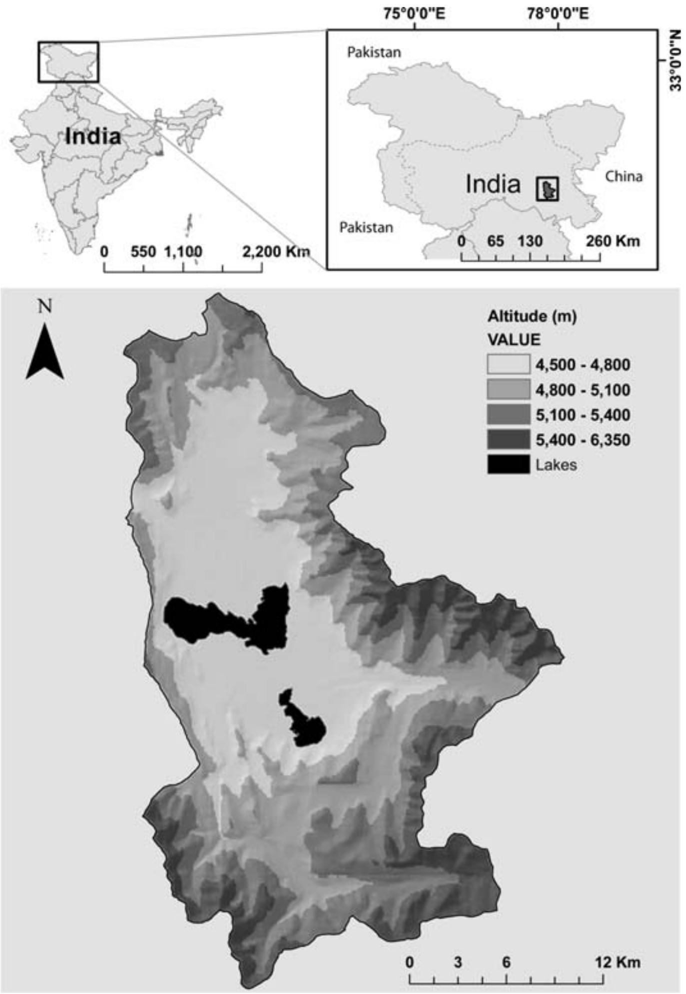

Fig. 1 Study area showing the range of altitude along with main water bodies

Table 1 Environmental predictor variables used in sampling and RSF modelling

Variable	Description	Observed range in 'use' dataset
Altitude	Calculated from the digital elevation model	4,633–5,573 m
NDVI	Estimated using Landsat ETM (30 m res.) satellite imagery as— $\text{ndvi} = \frac{\text{IR(band4)}-\text{R(band3)}}{\text{IR(band4)}+\text{R(band3)}}$	−0.14–0.04
Northness	Calculated from the digital elevation model, transformed to northness—cosine (aspect)	−1–1
Ruggedness (SARI)	Slope Aspect Ruggedness Index= $\frac{\text{SD of slope}*\text{variety of aspect}}{\text{SD of slope}+\text{variety of aspect}}*$	1.04–4.71
Slope	Calculated from the digital elevation model	0.84–29.71°

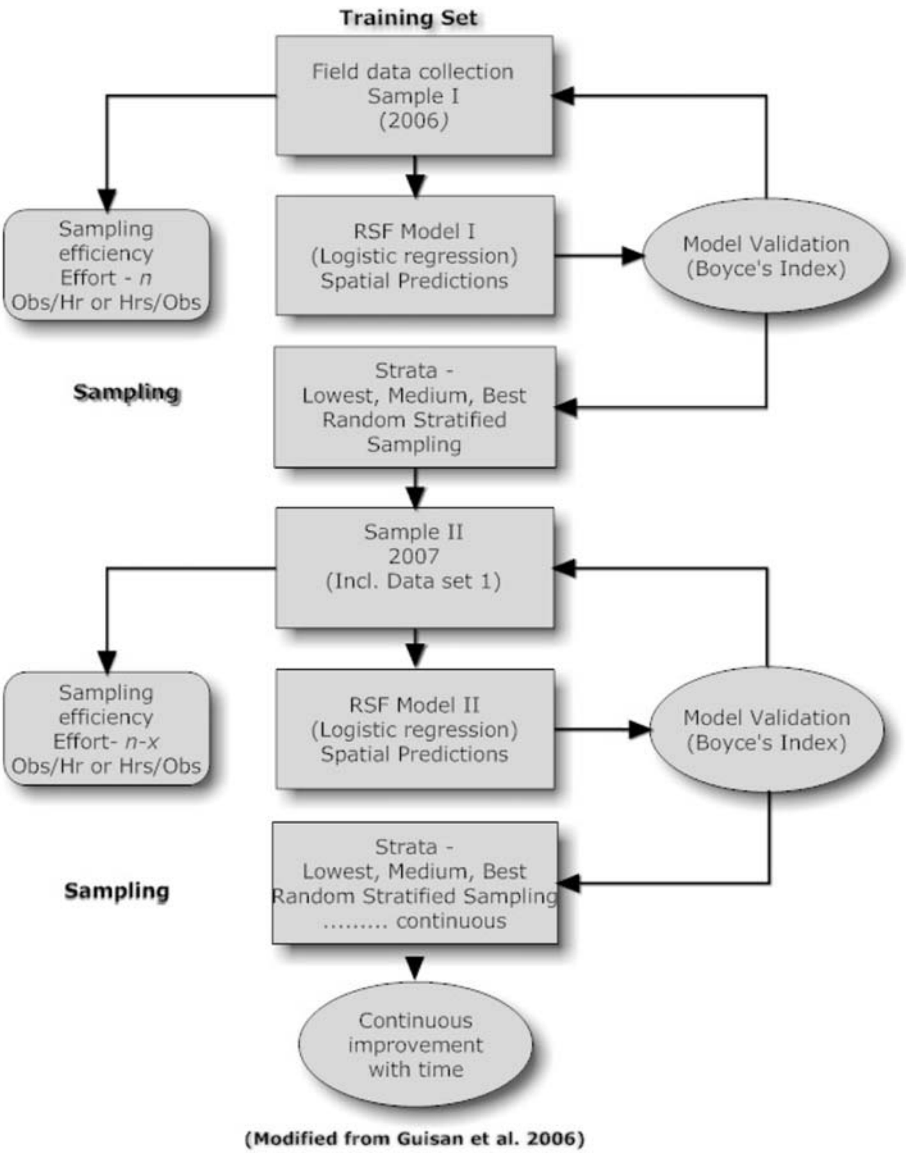

Fig. 2 Process flowchart for RSF based stratified random sampling used for sampling argali habitat selection (modified from Guisan et al. 2006)

Measurement of effort and sampling output

Initially (for sample I), we spent about 20 min at each of the vantage points sampled. In total we surveyed 13 vantage points. Each of the 13 vantage points was sampled 12 times during the sampling season with a radius up to 5 km. Total time spent sampling was estimated as the sum of minutes spent sampling at each vantage point. The total output for each survey point was converted to number of observations per hour and hours spent per observation.

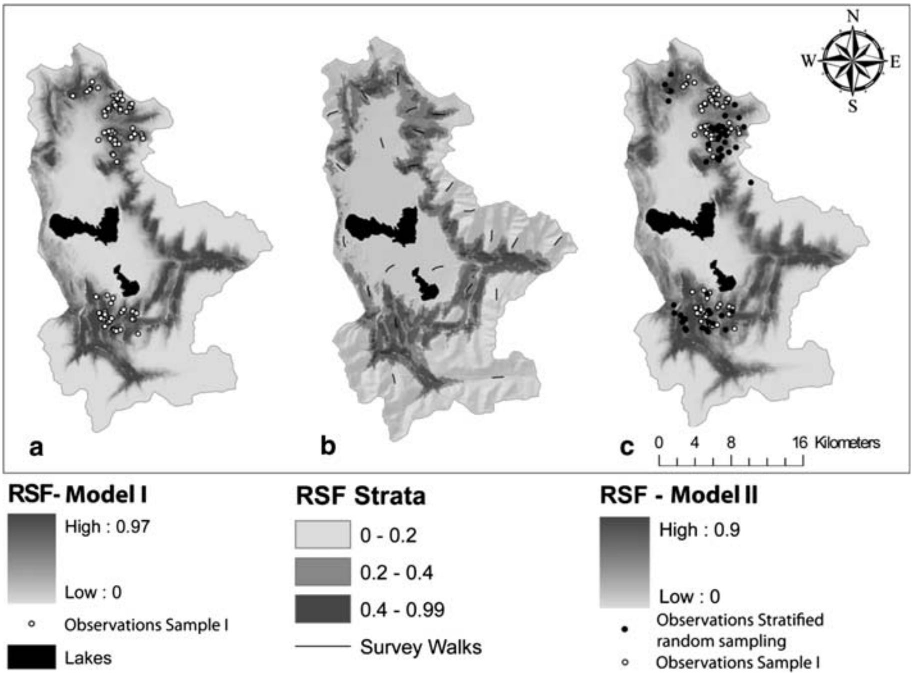

Fig. 3 **a** Spatial predictions from model I used to stratify the area based on habitat suitability, estimated using Resource Selection Function (RSF), **b** Strata classified for sampling with survey transects, **c** Spatial predictions from model II based on augmented dataset (sample II)

For stratified random sampling along transects (sample II), the effort was measured as time taken to walk each transect multiplied by the number of times each transect was walked. We tried to maintain the same speed for each transect walk so that the effort was not confounded with terrain difficulty. Total time spent in sampling each stratum was estimated as sum of the time spent in all transects in the strata.

Spatial information

All the spatial data was handled in ArcGIS version 9 (ESRI) with the spatial analyst extension. Digital elevation models (DEMs) were obtained from Shuttle Radar Topography Mission (SRTM, 90 m) data. We downscaled the DEM to 30 m as finer resolution DEM was not available for the study area. To ensure that the spatial information was not lost by downscaling, we compared the values of the variables extracted using 90 m DEM to the 30 m downscaled DEM and evaluated the differences and used them only when they were not significant.

Environmental variables

We developed a set of five environmental variables (Table 1) related to terrain and vegetation. Altitude, slope, and aspect were extracted from DEM. Aspect was converted into a continuous variable as northness—cosine(aspect) representing a north–south exposure gradient (Guisan et al. 1999), whose values varied from −1 (south) to 1 (north).

Ruggedness was estimated as Slope and Aspect ruggedness index (SARI) (Nellemann and Fry 1995; Danks and Klein 2002; Jepsen et al. 2005). SARI combines the attributes of slope and terrain heterogeneity to give high index values where the terrain is simultaneously rugged (heterogeneous) and sloping. We used slope and aspect raster grids extracted from the DEM. We calculated slope heterogeneity as the standard deviation (SD) of slope (SLSD) in a circular moving window with radius $s = 900$ m (radius selected based on minimum distance of flight by argali). The aspect grid was binned into groups of 45 (0–45°, 45–90° etc.) yielding a total of 8 aspect bins. Aspect heterogeneity was calculated as the number of different aspect bins found within the circular moving window (ASPVAR). We calculated SARI as $(SLSD_s * ASPVAR_s)/(SLSD_s + ASPVAR_s)$ following Nellemann and Fry (1995). We used normalized difference vegetation index (NDVI; Pettorelli et al. 2005) as the index of greenness due to non availability of high quality vegetation maps. NDVI was calculated as the difference between the red and infra red channels divided by the sum of the same two channels. NDVI was calculated from a Landsat image from August 20, 2002.

Statistical analysis

We used logistic regression models, assuming a Bernoulli distribution for the response variable and a logit link (Venables and Ripley 1999) to derive resource selection functions (RSFs). RSFs provide estimates of relative probability of use of a given unit, which can be interpreted as a measure of habitat suitability (Boyce et al. 2002). Different designs exist for estimating RSFs, and we used use-availability design because true absence is difficult to establish (Boyce et al. 2002). Note that availability in RSFs corresponds to pseudo-absence in niche models (e.g. Engler et al. 2004). Johnson et al. (2006) have shown that in the case of use-available data, logistic regression provides correct estimates of the regression coefficients defining the habitat suitability/RSF, but that the intercept and therefore unconditional habitat use are not estimable without additional information on sampling intensity. We generated random points for availability (160 in 2005–2006 and 259 in 2007). In 2005–2006, as the vantage points did not cover the low altitude areas and argali generally do not use low flat areas, available points were restricted to slopes >5° and altitude <5,400 m. In 2007, the whole area (except >5,400 m) was considered as available and we did not stratify the random sampling as the number of transects sampled was approximately proportional to area of strata. Presence was coded as 1 and availability as 0.

Logistic regression models were fitted in R (R Development Core Team 2008) and all combinations of variables were assessed as no subset could be chosen a priori. Model selection was performed using the Akaike's Information Criterion corrected for small sample size (AIC_c), and AIC_c weights (Burnham and Anderson 2004). We assessed the influence of the number of availability points on model selection by sub sampling the availability points. We used a random subsample (without replacement) made of half of the availability points, and calculated the AICc for the five top models and 1,000 sub samples. We used the average differences in AICc as a measure of the reliability of model selection. Non-linear relationships were assessed using partial residuals and were found to be adequately described using second-order polynomials. Models containing quadratic terms without linear terms were discarded. Note that because of constraints in the sampling design (repeated visits of the same points or transects), we could expect some dependency in the observations. However, extensive movements of argali and the large distances between points of observation and argali locations made this dependency undetectable in the datasets (strong dependency would result in overdispersion). We nevertheless used a bootstrap procedure for clustered data, the cluster boostrap (see Field and Welsh 2007 for discussion). We implemented the bootstrap in two parts: (1) vantage points (2005–2006) or transects (2007) were resampled with replacement (i.e. not individual observations but the observation clusters were resampled), and (2) availability points were sampled with replacement. We generated 2,000 datasets and used 2.5–97.5% percentiles to estimate 95% confidence intervals. We assessed the goodness of fit (GOF) of full models for data set 1 and 2 using the deviance and the le Cessie–van Houwelingen normal test z-statistic (le Cessie and van Houwelingen 1991) implemented in Design library for R (Harrell 2001).

Spatial predictions were prepared using the model coefficients generated for the best models selected using AIC_c. The variable importance was assessed using the sum of AIC_c weights for the model including this variable (Anderson et al. 2001). The predictive performance of models was evaluated using Boyce's index (Boyce et al. 2002). We allocated the data randomly into cross validation groups (4 for the initial dataset, 5 for the augmented dataset; the difference is due to the different number of observations). Each model was then developed iteratively by using 3 (resp. 4) of the 4 (resp. 5) cross-validation groups and the predictions were evaluated using the remaining group. We used logistic regression models to train our model iteratively on four of the five data sets. We estimated the parameters of variables in full models. Model performance was evaluated by dividing map predictions into RSF bins and comparing the frequencies of argali use sites (adjusted for bin area) within each bin with bin rank. A Spearman-rank correlation (r_s) between area-adjusted frequency of cross-validation points within individual RSF bins and the bin rank was calculated for each cross-validated model r_s (Boyce et al. 2002). Area-adjusted frequencies in this case were simply the frequency of cross-validated use locations with a bin adjusted (divided) by the area of that range of RSF scores available across the landscape. Due to limited sample size in the cross-validated groups ($n = 26$), the total number of bins was 6. A good predictive model is expected to have an increasing area adjusted frequency of argali use sites in higher bins and hence a significant ($P < 0.05$) and positive r_s.

Results

The logistic regression models selected according to their AIC_c weights are presented in Table 2 and parameter estimates for the full and best models are presented in Table 3. Both full models fitted the data (model I, Le Cessie's test: $z = 0.17$, $P = 0.86$, Deviance = 233.38, d.f = 231; model II—Le Cessie's test: $z = -2.25$, $P = 0.02$, Deviance = 293.7, d.f. = 379; the negative z value and the low deviance indicate no overdispersion). The number and sub sampling of availability points did not influence the order of the selected models as the average difference in AICc based on sub sampled datasets had a similar order to the one obtained from the full datasets (sample I: observed differences/average differences in AICc between best model and next 4 best models: 0.77/0.52; 0.79/0.34; 1.23/0.37; 1.54/0.30; augmented dataset/sample II: 0.32/−0.17; 1.60/0.33; 1.64/0.85; 2.09/1.85). The best model I based on the sample I (Table 2), indicated that altitude, slope, NDVI and ruggedness (SARI) were the most important variables determining habitat selection by argali. However, when making spatial predictions we dropped slope as it had a lower importance than other variables and slope and ruggedness might have indicated similar predation risk factors. Results from the model II based on sample II indicate that altitude, slope, northness and NDVI were the most important variables, but ruggedness was not important (Table 2). According to best model I, argali selected areas in a range of altitude, neither very high nor very low (second order term for altitude; note that available elevations were already fairly

Table 2 Comparisons of ΔAIC_c and Akaike weights for the best 10 models (bold italics) after AICc selection from models I (sample I) and model II (sample II)

Alt	Alt^2	ndvi	$ndvi^2$	nness	sari	$sari^2$	slp	slp^2	Model 1 ΔAICc	Akaike Weights 1	Model II ΔAICc	Akaike Weights 2
Best models for sample I												
−1.652	−1.672	−0.390			0.455		0.266	−0.434	***0.000***	***0.2897***	*3.660*	*0.010*
−1.557	−1.757	−0.653	0.126		0.458		0.269	−0.440	***0.767***	***0.1346***	*3.720*	*0.009*
−1.705	−1.991	−0.348			0.414				***0.794***	***0.1309***	*73.994*	*0.000*
−1.793	−1.652				0.424		0.228	−0.417	***1.228***	***0.0848***	*10.360*	*0.000*
−1.838	−1.959				0.386				***1.542***	***0.0620***	*79.288*	*0.000*
−1.619	−2.088	−0.597	0.118		0.418				***1.652***	***0.0555***	*75.267*	*0.000*
−1.807	−1.965	−0.367			0.421		0.230		***1.724***	***0.0517***	*75.010*	*0.000*
−1.656	−1.656	−0.397			0.459	−0.046	0.265	−0.432	***2.045***	***0.0375***	*4.315*	*0.005*
−1.652	−1.671	−0.390		−0.011	0.454		0.265	−0.434	***2.139***	***0.0341***	*2.393*	*0.035*
−1.727	−2.065	−0.617	0.119		0.426		0.232		***2.583***	***0.0219***	*76.257*	*0.000*
1.00	**1.00**	**0.82**	**0.53**	**0.06**	**1.00**	**0.08**	**0.71**	**0.62**				
Best models for sample II												
−0.734	−2.243	−0.728	0.124	−0.108			−0.016	−2.002	*8.347*	*0.000*	***0.000***	***0.365***
−0.729	−2.219	−0.730		−0.372			−0.050	−2.003	*7.556*	*0.000*	***0.320***	***0.281***
−0.728	−2.255	−0.731					−0.006	−1.999	*5.491*	*0.001*	***1.598***	***0.078***
−0.719	−2.376	−0.898	0.123				0.004	−2.028	*6.286*	*0.001*	***1.642***	***0.075***
−0.719	−2.346	−0.922	0.132	−0.395	0.019		−0.043	−2.035	*2.921*	*0.014*	***2.086***	***0.048***
−0.777	−2.146	−1.013	0.151	−0.404	−0.009	−0.233	−0.063	−1.964	*4.980*	*0.002*	***2.156***	***0.045***
−0.728	−2.229	−0.726		−0.371	−0.020		−0.047	−2.002	*2.139*	*0.031*	***2.393***	***0.035***
−0.779	−2.045	−0.772		−0.376	−0.047	−0.195	−0.064	−1.939	*4.201*	*0.004*	***3.012***	***0.019***
−0.727	−2.265	−0.727			−0.021		−0.004	−1.997	*0.000*	*0.260*	***3.660***	***0.010***

Table 2 continued

Alt	Alt^2	ndvi	$ndvi^2$	nness	sari	$sari^2$	slp	slp^2	Model 1 ΔAICc	Akaike Weights 1	Model II ΔAICc	Akaike Weights 2
−0.720	−2.370	−0.903	0.124		0.016		0.002	−2.029	*0.767*	*0.121*	***3.720***	***0.009***
1.00	**1.00**	**1.00**	**0.57**	**0.81**	**0.18**	**0.08**	**1.00**	**1.00**				

Values are parameter estimates for the variables included in the model, empty cell indicating that the variable is not included. Support for the best models for sample I but estimated using sample II (resp. sample II and I) are given in italics. Variable importance is indicated in bold roman in the last row of model sets

(int.), Intercept; alt, altitude; alt2, $altitude^2$; ndvi, normalised difference vegetation index; ndvi2, $ndvi^2$; sari, slope aspect ruggedness index; sari2, $sari^2$; slp, slope; nness, northness

Table 3 A Parameter estimates with SE for the full models estimated using initial dataset (sample I) and augmented dataset (sample II). Full models are given as they are directly comparable. B Parameter estimates, SE and 95% CI for the best model according to AICc for the initial data set. Median estimate and 95% CI based on blocked bootstrap are also given. C Same as B but for the augmented dataset

A	Sample I (2005–2006)		Sample II (2005–2006 + 2007)	
Variables	Estimate	SE	Estimate	SE
Intercept	0.065	0.259	1.48	0.245
Altitude	−1.56	0.445	−0.784	0.321
Altitude^2	−1.73	0.494	−2.17	0.431
NDVI	−0.664	0.318	−0.987	0.290
NDVI^2	0.127	0.106	0.141	0.064
Northness	−0.018	0.232	−0.112	0.216
Ruggedness	0.461	0.175	−0.012	0.197
Ruggedness^2	−0.051	0.149	−0.232	0.167
Slope	0.265	0.224	−0.023	0.254
Slope^2	−0.436	0.230	−1.96	0.331

B	GLM estimate (SE)	GLM 95% CI	Bootstrap median estimate	Bootstrap 95% CI
Intercept	0.059 (0.238)	[− 0.406; 0.530]	0.052	[-0.611; 0.637]
Altitude	−1.65 (0.441)	[− 2.60; 0.863]	−1.73	[-3.07; −0.776]
Altitude^2	−1.67 (0.480)	[− 2.69; −0.808]	−1.77	[-3.29; −0.867]
NDVI	−0.390 (0.228)	[− 0.886; 0.026]	−0.398	[− 1.32; 0.102]
Ruggedness	0.455 (0.172)	[0.128; 0.806]	0.479	[− 0.002; 0.948]
Slope	0.266 (0.222)	[− 0.169; 0.709]	0.284	[− 0.438; 0.946]
Slope^2	−0.434 (0.230)	[− 0.906; −0.0003]	−0.478	[− 1.06; −0.037]

C	GLM estimate (SE)	GLM 95% CI	Bootstrap median estimate	Bootstrap 95% CI
Intercept	1.46 (0.239)	[1.01; 1.95]	1.51	[0.894; 2.19]
Altitude	−0.717 (0.330)	[− 1.42; −0.112]	−0.792	[− 2.12; −0.095]
Altitude^2	−2.35 (0.418)	[− 3.24; −1.60]	−2.49	[− 4.40; −1.30]]
NDVI	−0.916 (0.275)	[− 1.46; −0.376]	−0.969	[− 1.84; −0.269}
NDVI^2	0.131 (0.061)	[− 0.051; 0.234]	0.130	[− 0.404; 0.312]
Northness	−0.395 (0.205)	[− 0.801; 0.006]	−0.414	[− 0.861; 0.005]
Slope	−0.041 (0.255)	[− 0.559; 0.448]	−0.063	[− 0.814; 0.529]
Slope^2	−2.03 (0.334)	[− 2.73; −1.43]	−2.10	[− 2.94; −1.54]

high (> 4,200 m) and below the vegetation limit at about 5,200 m). Similar was the case with slope, argali seemed to select intermediate slopes (second order term for slope). NDVI and ruggedness (SARI) also contribute to the habitat selection. Non linear affect of NDVI indicated again a preference for a restricted range of NDVI values.

The spatial predictions (RSF, 0–0.99) from the best model I based on variables altitude, altitude^2, NDVI and ruggedness (SARI) predicted about 12.8% of the study area as high suitability (RSF, 0.4–0.99), 18.5% area of medium suitability (RSF, 0.2–0.4) and 68.6% area of low suitability (RSF, 0–0.2) (Fig. 3b). The K fold cross validation of model I (Fig. 4) gave an average Spearman rank correlation r_s = 0.94 between RSF bin ranks and

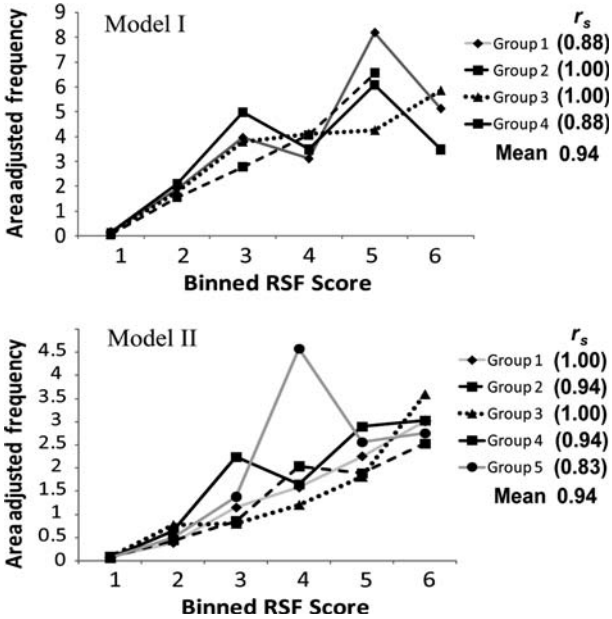

Fig. 4 Area adjusted frequency of categories (bins) of RSF scores for observed locations for argali for models I and II. Spearman rank correlations between the area adjusted frequencies and bin ranks are presented in *parenthesis*. Frequency values for individual cross validation sets (model 1: groups = 4 and model 2: groups = 5) are depicted with *unique symbols*

area adjusted frequencies for individual and average model sets. An average p value was 0.021 (Group 1 = 0.03, 2 = 0.017, 3 = 0.003, 4 = 0.033). The cross validation of the model II (Fig. 4) gave an average $r_s = 0.94$ between RSF bin ranks and area adjusted frequencies for individual and average model sets and average P values equal to 0.019 (Group 1 = 0.0028, 2 = 0.017, 3 = 0.0028, 4 = 0.017, 5 = 0.058) (Fig. 4).

Area adjusted frequencies displayed significant positive rank values across RSF bins for model I and II. model II was somewhat more significant on average (Fig. 4). All individual model sets demonstrated significant Spearman's correlations, indicating good model performance. model II appears to be more consistent across all RSF bins than model I (Fig. 4), with both low frequency in lower bins and high frequency in higher bins. However, the main result is that both models have similar predictive power.

A total of 13 vantage points were used for sampling during the initial surveys (sample I), the observations which formed the basis of model I. The total sampling output or efficiency (Table 4) could be quantified as hours per observation or observations per hour. For the former the initial sampling was estimated to be 1.97 ± 0.41 h/obs and 1.92 ± 0.39 obs/h for the years 2005 and 2006 respectively. The stratified random prospective sampling significantly enhanced sampling efficiency in the best stratum. The time spent sampling the highest probability stratum was 0.74 ± 0.059 h/obs relative to 4.51 ± 0.89 h/obs in the lowest probability stratum (Table 4). Even the medium stratum resulted in lower number of hours needed per observation (1.52 ± 0.35). An area-weighted average would have resulted in 3.47 h/obs, a higher value than in 2005/6, most likely because vantage points were located in better habitats. Therefore, compared to the use of random or a priori vantage points, stratified random sampling with a higher sampling intensity in the

Table 4 Total survey effort with characteristics of the sampled strata

Period	RSF class	Area of class (%)	Observations (%)	Effort	
				Obs/h	h/obs*
2007					
High	0.4–0.99	12.8	50	1.33 ± 0.125	0.74 ± 0.059
Medium	0.2–0.4	18.5	38	0.807 ± 0.166	1.52 ± 0.35
Low	0–0.2	68.6	12	0.0798 ± 0.101	4.51 ± 0.89
2006				0.740 ± 0.123	1.92 ± 0.39
2005				0.730 ± 0.130	1.97 ± 0.41

Sampling in years 2005 and 2006 was conducted from 13 vantage points and 2007 survey was based on stratified random sampling

Values represent mean ± SE

* Note that E[h/obs] $\neq$ 1/E[obs/h]

highest probability stratum would result in an effort between 2 and 3 times lower per observation. In terms of number of observations per hour, the difference is a factor of 16 (1.33 vs. 0.08 obs/h). Given that all strata need to be sampled, the difference in overall effort would not be as large, but still quite substantial.

Discussion

We present the first predictive fine scale model of habitat suitability for Tibetan argali in a Transhimalayan area using an iterative process based on sampling strata defined by an initial RSF model (model I) and fitting a new model based on an augmented data set (model II) (Fig. 3c). Survey sites added after stratification of the area according to habitat suitability captured the habitat variation in the survey region better (Edwards et al. 2005; Guisan et al. 2006).

Altitude, slope, NDVI, and northness emerged as main effects in the best model II. Argali prefer a range of intermediate altitudes. The lower flat steppes contain higher biomass but also, high disturbance from people and livestock. On the other hand, at the upper limit of vegetation disturbance is low but forage is absent. Thus argali may be forced to use the narrow belt of sparse vegetation in the mountains. The non linear effect of NDVI was somewhat surprising since one could expect argali to select areas with higher biomass (at a given altitude). One possible explanation is that vegetation is very sparse in the narrow elevation band used and vegetation cover does not exceed more than 25% on average (Rawat and Adhikari 2005). Also, argali may be forced to use higher slopes that do not contain the highest biomass, but provide a better overview of the area to look for wolves (Schaller 1998). For such habitats, NDVI may not reflect biomass but rather other aspects such as soil water content or vegetation stress (Kogan et al. 2004; Hassan et al. 2007). Another explanation could be disturbance due to livestock use of high biomass areas (Namgail et al. 2006). There is clearly a need for a better understanding of vegetation indices in high altitude habitats with very low primary productivity (Karnieli et al. 2006).

The best model II obtained with augmented data also presented similar results as model I but the ruggedness variable was dropped and NDVI^2 and northness were included. First, this could result from removal of sampling bias during the second sampling stage by adoption of a stratified random sampling. This concerns both availability (flat areas were not considered as available in 2005–2006) and use. The results of the second model confirmed which one among slope and ruggedness was more important. The model change could also be due to different habitat selection between years. Habitat selection is scale sensitive (Manly et al. 1993) on both spatial and temporal scales (Johnson 1980; Boyce 2006). Boyce (2006) argued that different habitats might be selected in different years, and different RSFs might apply for different years. Even if the years sampled here were not characterized by large climatic differences, this cannot be excluded. To capture such patterns spanning multiple years, sampling must be on a temporal scale of sufficient duration to understand the significance of habitat selection that varies through time (Boyce 2006).

The model based prospective sampling resulted in large variation in sampling efficiency. Although, only 16.6% of the total sampling effort was spent sampling the highest probability stratum, it still contained more than 50% of the observations. The number of hours spent per observation was seven times less in the high suitability stratum relative to the lower suitability strata, indicating that if the sampling effort is concentrated on highest suitability areas, it would save survey effort significantly. It also presents a robust approach for stratification as it is based on resource selection functions.

An important caveat here is that habitat selection does not necessarily reflect quality of habitat (van Horne 1983; Johnson and Seip 2008). It describes the current species distribution or the realised niche of the species, which results from competition with livestock, disturbance, predation and several other biotic and abiotic factors. For example, argali abandoned their preferred plant communities and moved closer to cliffs with lower vegetation cover following the introduction of livestock in the Gya Miru wildlife sanctuary adjacent to the study area (Namgail et al. 2006).

The purpose of the study was to initiate the iterative approach proposed by Guisan et al. (2006) with continued improvements over time. Lack of surveys in some habitats and absence of proper methodology for sparse, mobile, clustered or extremely seasonal populations were some of the major reasons identified for failures to successfully survey rare populations (MacDonald 2004). The model-based sampling approach addresses most of these issues by providing a standardized method which not only saves survey time but is also accurate and efficient in identifying the survey areas and thus increasing precision.

When global model precision is the intent of use of predictive models, then sampling intensity should probably be similar over all habitats (see Hand and Vinciotti 2003). In a conservation perspective we might focus on the best habitats. Directing the sampling to ensure inclusion of areas with a higher probability of finding the species is thus a desirable approach for increasing survey efficiency and reducing sampling costs when dealing with rare species (Guisan et al. 2006).

The study has potential implications for surveying and monitoring argali populations in the study area as well as in the other areas where argali are known to occur. Application of this methodology to the rest of eastern Ladakh would also test the transferability (Randin et al. 2006) of the RSF-based models of animal species distribution, which is an important issue to be considered in such approaches. Finally, focusing sampling effort in the higher probability strata will also be important for obtaining an accurate population size estimate of this rare species.

Acknowledgments The University of Tromsø and Rufford Foundation for Nature conservation generously supported the fieldwork. Thanks to Dr. Charudutt Mishra for initial and valuable advice during the preparation of manuscript. Department of Wildlife protection, Jammu and Kashmir provided necessary permits for undertaking the study, thanks to them. Thanks to Nature Conservation Foundation for the support during the preparation of the manuscript. Rinchen, Jigmet and *Ajaan* Stanzin Dorjey provided essential and inspiring field assistance during field work. We are grateful to all the people and organizations involved in the study. An anonymous referee provided helpful comments on a previous version.

References

Anderson DR, Link WA, Johnson DH et al (2001) Suggestions for presenting results of data analyses. J Wildl Manage 65:373–378. doi:10.2307/3803088

Bhatnagar YV, Wangchuk R (2001) Status survey of large mammals in eastern Ladakh and Nubra. In: Anonymous (ed) Conserving biodiversity in the Trans-Himalaya, technical report (1999–2000). Wildlife Institute of India, International Snow Leopard Trust and USFWS, Dehradun, pp 108–135

Boyce MS (2006) Scale for resource selection functions. Divers Distrib 12:269–276. doi:10.1111/j.1366-9516.2006.00243.x

Boyce MS, Vernier PR, Nielsen SE et al (2002) Evaluating resource selection functions. Ecol Model 157:281–300. doi:10.1016/S0304-3800(02)00200-4

Burnham KP, Anderson DR (2004) Multimodel inference—understanding AIC and BIC in model selection. Sociol Methods Res 33:261–304. doi:10.1177/0049124104268644

Chundawat RS, Qureshi Q (1999) Planning wildlife conservation in Leh and Kargil districts of Ladakh, Jammu and Kashmir. Wildlife Institute of India, Dehradun

Danks FS, Klein DR (2002) Using GIS to predict potential wildlife habitat: a case study of muskoxen in northern Alaska. Int J Remote Sens 23:4611–4632. doi:10.1080/01431160110113890

Edwards TC, Cutler DR, Geiser L et al (2004) Assessing rarity of species with low detectability: lichens in Pacific northwest forests. Ecol Appl 14:414–424. doi:10.1890/02-5236

Edwards TC, Cutler DR, Zimmermann NE et al (2005) Model-based stratifications for enhancing the detection of rare ecological events. Ecology 86:1081–1090. doi:10.1890/04-0608

Engler R, Guisan A, Rechsteiner L (2004) An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. J Appl Ecol 41:263–274. doi:10.1111/j.0021-8901.2004.00881.x

Field CA, Welsh AH (2007) Bootstrapping clustered data. J R Stat Soc Ser B Stat Methodol 69:369–390. doi:10.1111/j.1467-9868.2007.00593.x

Green RH, Young RC (1993) Sampling to detect rare species. Ecol Appl 3:351–356. doi:10.2307/1941837

Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. Ecol Model 135:147–186. doi:10.1016/S0304-3800(00)00354-9

Guisan A, Weiss SB, Weiss AD (1999) GLM versus CCA spatial modelling of plant species distribution. Plant Ecol 143:107–122. doi:10.1023/A:1009841519580

Guisan A, Broennimann O, Engler R et al (2006) Using niche-based models to improve the sampling of rare species. Conserv Biol 20:501–511. doi:10.1111/j.1523-1739.2006.00354.x

Hand DJ, Vinciotti V (2003) Local versus global models for classification problems: fitting models where it matters. Am Stat 57:124–131. doi:10.1198/0003130031423

Harrell FE (2001) Regression modelling strategies with applications to linear models, logistic regression and survival analysis. Springer, New York

Harris RB, Loggers CO (2004) Status of Tibetan plateau mammals in Yeniugou, China. Wildl Biol 10: 91–99

Harris RB, Pletscher DH, Loggers CO et al (1999) Status and trends of Tibetan plateau mammalian fauna, Yeniugou, China. Biol Conserv 87:13–19. doi:10.1016/S0006-3207(98)00046-9

Harris RB, Ali A, Loggers CO (2005) Trend monitoring of large mammals: two case studies. Acta Theriol Sin 25:319–325

Hassan QK, Bourque CPA, Meng FR et al (2007) A wetness index using terrain-corrected surface temperature and normalized difference vegetation index derived from standard MODIS products: an evaluation of its use in a humid forest-dominated region of eastern Canada. Sensors 7:2028–2048. doi:10.3390/s7102028

Hill NM, Keddy PA (1992) Prediction of rarities from habitat variables: coastal plain plants on Nova Scotian lakeshores. Ecology 73:1852–1859. doi:10.2307/1940036

Jackson RM, Hunter DO (1996) Snow leopard information management handbook, Seattle. International Snow Leopard Trust, Seattle

Jepsen JU, Madsen AB, Karlsson M et al (2005) Predicting distribution and density of European badger (*Meles meles*) setts in denmark. Biodivers Conserv 14:3235–3253. doi:10.1007/s10531-004-0444-2

Johnson DH (1980) The comparison of use and availability measurements for evaluating resource preference. Ecology 61:65–71. doi:10.2307/1937156
Johnson CJ, Seip DR (2008) Relationship between resource selection, distribution, and abundance: a test with implications to theory and conservation. Popul Ecol 50:145–157. doi:10.1007/s10144-008-0078-4
Johnson CJ, Nielsen SE, Merrill EH et al (2006) Resource selection functions based on use-availability data: theoretical motivation and evaluation methods. J Wildl Manage 70:347–357. doi:10.2193/0022-541X(2006)70[347:RSFBOU]2.0.CO;2
Kadmon R, Heller J (1998) Modelling faunal responses to climatic gradients with GIS: land snails as a case study. J Biogeogr 25:527–539. doi:10.1046/j.1365-2699.1998.2530527.x
Karnieli A, Bayasgalan M, Bayarjargal Y et al (2006) Comments on the use of the vegetation health index over Mongolia. Int J Remote Sens 27:2017–2024. doi:10.1080/01431160500121727
Kogan F, Stark R, Gitelson A et al (2004) Derivation of pasture biomass in Mongolia from AVHRR-based vegetation health indices. Int J Remote Sens 25:2889–2896. doi:10.1080/01431160410001697619
le Cessie S, van Houwelingen JC (1991) A goodness of fit test for binary regression models, based on smoothing methods. Biometrics 47:1267–1282. doi:10.2307/2532385
Macdonald LL (2004) Sampling rare populations. In: Thompson WL (ed) Sampling rare or elusive species concepts designs and techniques for estimating population parameters. Island Press, Washington, DC, pp 11–42
Mackenzie DI, Royle JA (2005) Designing occupancy studies: general advice and allocating survey effort. J Appl Ecol 42:1105–1114. doi:10.1111/j.1365-2664.2005.01098.x
Manly BFJ, Mcdonald LL, Thomas DL (1993) Resource selection by animals: statistical design and analysis for field studies. Chapman and Hall, London
Namgail T, Fox JL, Bhatnagar YV (2006) Habitat shift and time budget of the Tibetan argali: the influence of livestock grazing. Ecol Res 22:22–31
Nellemann C, Fry G (1995) Quantitative analysis of terrain ruggedness in reindeer winter grounds. Arctic 48:172–176
Peterson AT (2001) Predicting species' geographic distributions based on ecological niche modelling. Condor 103:599–605. doi:10.1650/0010-5422(2001)103[0599:PSGDBO]2.0.CO;2
Peterson AT, Soberon J, Sanchez-Cordero V (1999) Conservatism of ecological niches in evolutionary time. Science 285:1265–1267. doi:10.1126/science.285.5431.1265
Pettorelli N, Vik JO, Mysterud A et al (2005) Using the satellite-derived NDVI to assess ecological responses to environmental change. Trends Ecol Evol 20:503–510. doi:10.1016/j.tree.2005.05.011
R Development Core Team (2008) A language and environment for statistical computing. R Foundation for statistical computing, Vienna
Randin CF, Dirnböck T, Dullinger S et al (2006) Are niche-based species distribution models transferable in space? J Biogeogr 33:1689–1703. doi:10.1111/j.1365-2699.2006.01466.x
Rawat GS, Adhikari BS (2005) Floristics and distribution of plant communities across moisture and topographic gradients in Tso Kar basin, Changthang plateau, eastern Ladakh. Arct Antarct Alp Res 37:539–544. doi:10.1657/1523-0430(2005)037[0539:FADOPC]2.0.CO;2
Schaefer JA, Messier F (1995) Habitat selection as a hierarchy: the spatial scales of winter foraging by muskoxen. Ecography 18:333–344. doi:10.1111/j.1600-0587.1995.tb00136.x
Schaller GB (1977) Mountain monarchs. University of Chicago Press, Chicago and London
Schaller GB (1998) Wildlife of the Tibetan steppe. University of Chicago Press, Chicago and London
Schaller GB, Gu BY (1994) Ungulates in northwest Tibet. Res Explor 10:266–293
Shackleton DM (1997) Wild sheep and goats and their relatives: status survey and conservation action plan. IUCN Gland, Switzerland, Cambridge
Thompson WL (2004) Future directions in estimating abundance of rare or elusive species. In: Thompson WL (ed) Sampling rare or elusive species. Island Press, Washington DC, pp 389–399
van Horne B (1983) Density as a misleading indicator of habitat quality. Journal of Wildlife Management 47:893–901
Venables WN, Ripley BD (1999) Modern applied statistics with S-plus. Springer, New York
Walker ABD, Parker KL, Gillingham MP et al (2007) Habitat selection by female stone's sheep in relation to vegetation, topography, and risk of predation. Ecoscience 14:55–70. doi:10.2980/1195-6860(2007)14[55:HSBFSS]2.0.CO;2
Williams PH, Margules CR, Hilbert DW (2002) Data requirements and data sources for biodiversity priority area selection. J Biosci 27:327–338. doi:10.1007/BF02704963
Wiser S, Peet RK, White PS (1998) Prediction of rare-plant occurrence: a southern Appalachian example. Ecol Appl 8:909–920. doi:10.1890/1051-0761(1998)008[0909:PORPOA]2.0.CO;2
Yoccoz NG, Nichols JD, Boulinier T (2001) Monitoring of biological diversity in space and time. Trends Ecol Evol 16:446–453. doi:10.1016/S0169-5347(01)02205-4