Diversity and Distributions, (Diversity Distrib.) (2007) **13**, 289–298

BIODIVERSITY RESEARCH

An integrated framework to map animal distributions in large and remote regions

Alejandro Travaini[1]*, Javier Bustamante[2], Alejandro Rodríguez[2], Sonia Zapata[1], Diego Procopio[1], Julieta Pedrana[1] and Rolando Martínez Peck[1]

[1]*Centro de Investigaciones de Puerto Deseado, Universidad Nacional de la Patagonia Austral, CONICET, PO Box 238, Avenida Prefectura Naval S/N, 9050 Puerto Deseado, Santa Cruz, Argentina,* [2]*Department of Applied Biology, Estación Biológica de Doñana, Consejo Superior de Investigaciones Científicas, Avda. María Luisa S/N, 41013 Sevilla, Spain*

*Correspondence: Alejandro Travaini, Centro de Investigaciones de Puerto Deseado, Universidad Nacional de la Patagonia Austral, PO Box 238, Avenida Prefectura Naval S/N, 9050 Puerto Deseado, Santa Cruz, Argentina.
Tel./Fax: (54) 297 4870511;
E-mail: atravaini@pdeseado.com.ar

ABSTRACT

In this paper we show how new technologies can be incorporated from the gathering of field data on wildlife distribution to the final stage of producing distribution maps. We describe an integrated framework for conducting wildlife censuses to obtain data to build predictive models of species distribution that when integrated in a GIS will produce a distribution map. Field data can be obtained with greater accuracy and at lower costs using a combination of Global Positioning System, Personal Digital Assistant, and specific wildlife recording software. Sampling design benefits from previous knowledge of environmental variability that can be obtained from free remote sensing data. Environmental predictors derived from this remote sensing information alone, combined with automatic procedures for predictor selection and model fitting, can render cost-effective predictive distribution models for wildlife. We show an example with guanaco distribution in the Patagonian steppes of Santa Cruz province, Argentina.

Keywords

Gathering field information, GIS, Patagonia, PDA, wildlife censuses, wildlife distribution maps, CYBERTRACKER**.**

INTRODUCTION

There is growing interest in the study of ecological processes on large scales (Ormerod & Watkinson, 2000). Large-scale data of field distributions are useful to describe patterns (e.g. delineation of geographical ranges, Fortin *et al*., 2005; composition of communities, Nichols *et al*., 1998; determinants of distributions, Root & Schneider, 1993; Bailey *et al*., 1996; Calvert & Gauthier, 2005) as well as to guide management (exploitation of natural resources, Thompson *et al*., 2004; pest control, van der Werf *et al*., 2005; ecological restoration, Lamb, 1998; identification of areas with conservation value, Lenton *et al*., 2000; setting conservation priorities, Battersby & Greenwood, 2004). However, the current focus on large-scale processes is particularly outstanding in species distribution modelling (Guisan & Zimmermann, 2000) and monitoring of species and communities (Manley *et al*., 2004; La Sorte & Boecklen, 2005).

Large-scale modelling and monitoring of species distributions face a number of methodological difficulties related with design, data collection, data treatment, and analysis, such as imperfect species detectability, low spatial accuracy of records, and incomplete consideration of the sources of variation affecting animal observations (Link *et al*., 1994; Ottaviani *et al*., 2004; Royle *et al*., 2005). Moreover, gathering field data over large areas is expensive, time-consuming, and prone to errors (Waddle *et al*., 2003), usually within a context of limited funding. These constraints greatly influence sampling design. Carlson & Schmiegelow (2002) identify cost-effective sampling strategies and suggest that the effort should be best allocated extensively, covering an area as big as possible and spreading sampling sites over it, rather than adopting an intensive schedule in a small number of sites. In turn, increasing the area covered without reducing the density of sites sampled requires high efficiency in data gathering and processing. Methods designed for a fast and reliable recording of animal positions, saving time, and reducing errors during data processing (Logan & Smith, 1997; Stoleson *et al*., 2004) would harmonize funding limitation with the appropriate survey of large areas. Moreover, the increasing availability of remote sensing information provides the potential for suitable environmental predictors at large scales (Hepinstall & Sader, 1997). Remote sensing data in combination with simple automatic model-fitting techniques may result in improved estimates of species distributions (Bustamante & Seoane, 2004; Seoane *et al*., 2005).

In this paper we describe an integrated framework for modelling species distribution in large and remote areas. This protocol improves the efficiency of animal surveys in three ways: (1) it makes data gathering faster, (2) it increases the spatial accuracy of records, and (3) it minimizes the error propagation during data processing, thus reducing the number of steps necessary to store the data in a database. Moreover, with the combination of remote sensing data, statistical modelling, and geographical information systems (GIS), allocation of field sampling can be

DOI: 10.1111/j.1472-4642.2007.00338.x
www.blackwellpublishing.com/ddi

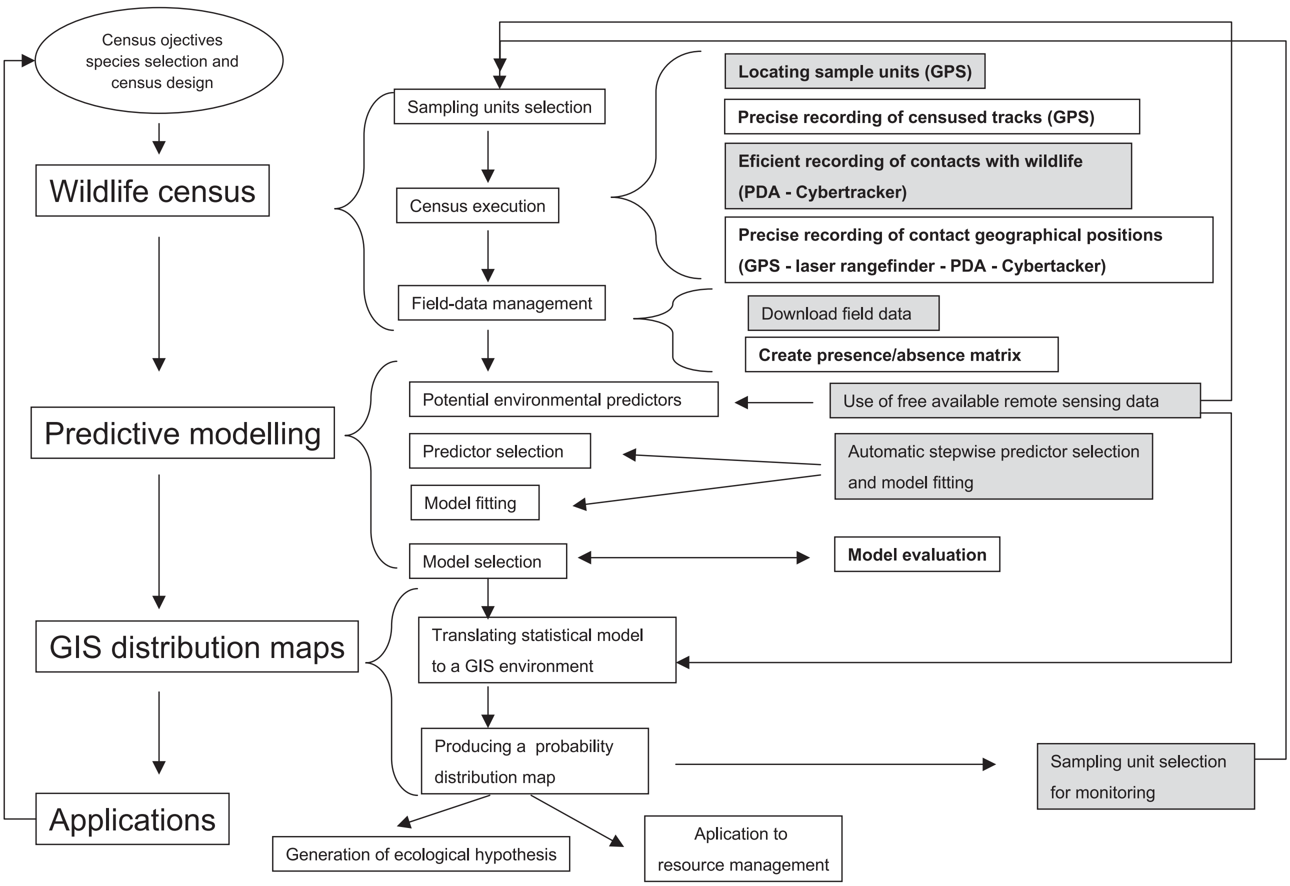

Figure 1 A flowchart of the integrated framework for species distribution mapping that we describe. Different steps in the process are subdivided in tasks. In bold we indicate the places in which our methods provide increased accuracy, boxes with grey background indicate the places in which the methods provide a reduction in costs (time or money).

optimized and species distributions can be extrapolated with known reliability to a larger area. We illustrate its application by developing a predictive model of guanaco (*Lama guanicoe*) distribution in southern Patagonia, a remote, sparsely inhabited and relatively inaccessible region of South America. This protocol of wildlife surveys is currently being used to build distribution maps of species susceptible of innovative sustainable uses in the region. The ultimate aim of such effort is to identify areas where the sustainable exploitation of wildlife resources could potentially alleviate the pressure from traditional sheep ranching on steppe ecosystems.

Integrated framework for species distribution mapping

Producing species distribution maps can be greatly improved with the integration of new technology at the different steps that go from animal sightings in the field to generating a product that reports how the species is distributed in space (usually a density or distribution map). We present an integrated framework that comprises three steps: (1) wildlife census, (2) predictive distribution modelling, and (3) generation of distribution maps in a GIS environment. At each of these steps we identify how new technology can help increase map accuracy and reduce costs (Fig. 1). None of the proposals in this paper are completely new, but we integrate technological advances coming from different fields. First, we suggest how to optimize field data recording using a PDA (Personal Digital Assistant) connected to a Global Positioning System (GPS) and employing specific wildlife recording software. We show how this combination can reduce data acquisition costs, reduce error propagation, and increase data accuracy. Second, we suggest fitting predictive models to contacts with wildlife using environmental predictors as a way of generating potential distribution models, which can be extrapolated from the sample to the whole region of interest. There is an everyday increasing availability of free remote sensing data that can be used as potential predictors in these models and that can also help optimize the initial sampling strategy. This is particularly relevant for countries where environmental cartography is scarce or no-existent. Finally, GIS provides a way of translating models to maps that can be used for planning the sustainable exploitation of animal populations, establishing ecological hypothesis about the factors determining species distribution, and even refining the sampling strategy for long-term population monitoring.

This integrated framework is of particular interest in large and remote areas where animal distributions remain poorly known, where the usual survey methods produce animal records that are not spatially accurate and the funding to obtain them is scarce.

Study area and model species

Patagonia is the southern-most portion of temperate South America. In Argentina it extends from the Colorado River (38° S) to the Horn Cape (56° S), encompassing, together with the Chilean portion, an approximate surface of 1,140,000 km^2.

We selected Santa Cruz province as a start point for this study. Santa Cruz is the biggest Patagonian province, with a surface of about 250 000 km^2 and several circumstances that makes it interesting for our study. Sheep ranging has been the prevalent land use since the early European settlers. Arid conditions preclude shifting this activity to the more profitable cattle ranging. For long, unsuitable livestock management, including keeping sheep numbers above carrying capacity and continuous grazing of mesic sites, has lead to steppe degradation (Golluscio *et al*., 1998). Unsustainable use of rangelands has been followed by progressive land abandonment in over 60% of Santa Cruz area. Currently, there is a socioeconomic trend to resume sheep ranging activity, unfortunately based on the same criteria responsible for previous failure.

The guanaco is the biggest herbivore in Patagonia. It was historically considered as a pest by rangers on the basis of an assumed competition with sheep for water and food (Baldi *et al*., 2001, 2004). The Patagonian steppe seems a relatively uniform environment at first glance. Consequently, guanacos are considered to be widely and uniformly distributed over this large area. Therefore we chose the guanaco as an adequate model for testing the potential of our integrated methodological approach in improving knowledge of species distributions in large and remote areas. Moreover, since this species could be potentially amenable to a number of sustainable uses (e.g. game, meat, and high quality wool), enhanced distribution maps may help to highlight areas where such uses could be implemented.

METHODS

Sampling unit selection

We performed a stratified random sampling by dividing the study area into 12 regions. We considered two environmental variables for the stratification: mean Normalized Difference Vegetation Index (NDVI) obtained from the VEGETATION project (see below) that was divided into three classes (because we felt that vegetation productivity could be an important driver for guanaco abundance) and slope obtained from SRTM (Shuttle Radar Topographic Mission, see below) that was divided into two classes (because guanaco abundance could increase with local heterogeneity and detection could also be affected by terrain irregularity). This generated six possible environmental combinations. To create 12 regions we selected the two biggest polygons from each class as a seed and assigned the remaining polygons to the seed that was geographically closer. Using a vectorial road cover of the study area we randomly selected road segments that added up to 4.500 linear km. To guarantee that every strata would be censused, 1500 km were equally distributed among census strata (125 km on each strata), and the remaining 3000 km were distributed proportionally to strata surface.

Census procedure

We selected a line transect method (Bibby *et al*., 1992) for our census because the open nature of the steppe environment, the low density of wildlife, and the possibility of covering more ground in a fixed time than with other methods. Although we present here results only for guanacos, a set of other nine species with some potential sustainable use was censused simultaneously (Table 1). Censuses were performed from a vehicle, at a maximum speed of 40 km h^{-1}, by the driver and one observer.

We drove to the start of a selected road segment with the help of a GPS (GARMIN MAP 76S). The GPS was used to record the actual trajectory of the census track — because in some instances maps were inaccurate in relation to road location — and registered the date, time, and actual speed for each track position. A PDA (Tungsten T3, using the Palm OS™, Palm Inc., Sunnyvale, CA, USA) was used to register information on census tracks characteristics and wildlife sightings.

Wildlife sightings were collected in the PDA using the free software CYBERTRACKER (CyberTracker Software (Pty) Ltd Reg. no. 97/01908/07, http://www.cybertracker.co.za). Using CYBERTRACKER we created our own data entry template to gather an almost unlimited amount of data. CYBERTRACKER interface is designed with a series of screens that follow a logical sequential order and then loop back to the start in order to enter a new observation. Smaller loops can also be made to enter observations that are very similar. Information that does not change often is entered only once, in the beginning, and linked into the sequence in such a way that it is stored in every sighting. For example census track number, driver, and observer names are entered only once for each census track, but automatically recorded on each sighting. Data for items which have a small set of appropriate values, such as observer, species, group category,

Table 1 Number of contacts with species susceptible of some sustainable use in 4500 km of road census in Santa Cruz (Argentinean Patagonia).

Species	Contacts
Least seedsnipe *Thinocorus rumicivorus*	307
Grey-breasted seedsnipe *Thinocorus orbignyianus*	1
Eared dove *Zenaida auriculata*	72
Ashy-headed goose *Chloephaga poliocephala*	5
Upland goose *Chloephaga picta*	507
Elegant crested tinamou *Eudromia elegans*	14
Patagonian tinamou *Tinamotis ingoufi*	16
Darwin's Rhea *Pterocnemia pennata*	341
Guanaco *Lama guanicoe*	526
European hare *Lepus europaeus*	73
Total	1862

etc., are entered via menus in a very easy way, reducing the possibility of recording errors. The PDA was connected to the GPS. When a particular observation had been completed it was saved together with a GPS reading (instantly georeferencing the data). This made the path loop back, making ready the screen for the next observation to be recorded. The combination of progressive refinement in screen sequences and looping back alternatives makes it possible to capture complex information very efficiently. We tested CYBERTRACKER against other database management software for PDAs, and CYBERTRACKER proved the most efficient for recording wildlife sightings.

Once a guanaco was observed, we stopped the vehicle and recorded group size, sex, and age of each individual; the distance to the animal or the group centre (with a laser range finder Leica LRF 1200 Rangemaster; Leica Camera AG, Solms, Germany); our bearing relative to north obtained from the inertial compass in the GPS; and the angle of the animal relative to our bearing. We also recorded all these data for Darwins' Rhea sightings, but smaller species detected at shorter distances were assumed to be at the vehicle position and could be recorded in the PDA without stopping.

The required equipment had a moderate cost. Laser Range Finder Leica LRF 1200 Rangemaster ($US599). PDA Tungsten T3 ($US350), plus an additional 128 Mb Secure Digital memory card ($US20), on which we performed backup of data at regular times, Garmin GPS MAP 76CS ($US420) with an external antenna (GA 27C, $US69). Both GPS and PDA were placed in waterproof transparent bags (AQUAPACK $US35 each), so we were able to read the GPS and use the PDA while they were protected from moisture, dust, and shock. Both PDA and GPS were powered by a 12-volt connection to the car cigarette lighter (cables $US38). The total cost of our equipment was around $US1566. These costs can be lower if based on cheaper, but also functional, components. For instance there are PDAs at half the cost of ours, which can be connected to a cheaper GPS (Garmin e-Trex, $US99). This can reduce as much as 50% of the equipment costs.

Field-data management

After each census, GPS and PDA data were downloaded to a laptop computer. In an absence of a laptop, a backup copy was performed in a Secure Digital card in the PDA. CYBERTRACKER data in the PDA were automatically downloaded to a CYBERTRACKER database when the PDA was synchronized with a computer. GPS track logs were downloaded using GPS TRACKMAKER version 12.3 for Windows and saved as a text file that was later imported to a database. The position of the animal was calculated using the georeferenced position of the observer and the distance and the bearing angle from the observer. The GPS track log was imported to a GIS (IDRISI Kilimanjaro; Eastman, 2003), keeping the date and time information. The GPS track was used to define the area effectively sampled by the census, and to assign date and time information to sampling units in case this had an effect on detection probabilities.

Potential environmental predictors

Both for census stratification and as potential environmental predictors, we used readily free available digital environmental information. We used the NDVI, which is an index of green vegetation vigour and photosynthetic activity obtained in digital format from the VEGETATION Programme. The VEGETATION Programme (a joint programme of the European Commission, French, Italian, and Swedish space agencies) offers free access to data of an earth observation sensor onboard of the SPOT satellites that monitor daily terrestrial vegetation cover at 1-km spatial resolution.

The instrument and associated ground services for processing and distribution are operational since 1998 (the first instrument is part of the SPOT 4 satellite and a second payload was decided to be onboard SPOT 5). Details of the program can be found at http://www.spot-vegetation.com. We used the VGT-S10 product, a synthesis image in which each ground value corresponds to the maximum NDVI recorded in a 10-day period. It can be obtained free from the VEGETATION distribution site http://free.vgt.vito.be. We estimated different environmental predictors related to mean productivity, annual variation, and vegetation growth seasonality using five consecutive years of data (April 1999 to March 2003; Table 2).

We obtained topographical data from the Shuttle Radar Topography Mission (SRTM). SRTM utilized dual Spaceborne

Table 2 Variables used as potential environmental predictors in the guanaco model.

Acronym	Variable description
AREA_SURVEYED	Fraction of the 1-km grid cell surface that is included in the 400-m buffer on both sides of the census track.
MEAN_NDVI	Mean VEGETATION NDVI VGT-S10 product, corrected for missing values in a 1-km pixel.
CV_NDVI	Coefficient of variation of NDVI excluding values < 28 to reduce noise.
GROWTH_PERIOD	Length of the vegetation growth period defined as the mean number of 10-day periods with NDVI values > 85.
MONTH_MAX_NDVI	Month at which the NDVI on average reaches its annual maximum value, being August month 1 and July month 12.
ALTITUDE	Mean altitude in m above sea level in a 1-km pixel. Obtained form the SRTM as the mean value of 11 × 11 3-arc-second cells after excluding missing values.
SLOPE	Mean slope in degrees in a 1-km pixel calculated in IDRISI Kilimanjaro. Obtained from the SRTM as the mean slope of 11 × 11 3-arc-second cells after excluding slope values of cells neighbours of those with missing altitude values.
LATITUDE	Latitude in m. UTM19S y coordinate (WGS84) of the 1-km grid cell centre.
LONGITUDE	Longitude in m. UTM19S x coordinate (WGS84) of the 1-km grid cell centre.

Imaging Radar (SIR-C) and dual X-band Synthetic Aperture Radar (X-SAR) configured as a baseline interferometer, acquiring two images at the same time. These images, when combined, produced a single 3-D image. Flown aboard the NASA Space Shuttle Endeavour on 11–22 February 2000, SRTM successfully collected data over 80% of the Earth's land surface, for all area between 60 degrees N and 56 degrees S latitude. Data were processed at NASA's Jet Propulsion Laboratory generating a 3-arc-second Digital Elevation Model (DEM) for South America (resolution of about 90 m at the equator). Details of this product can be obtained at http://seamless.usgs.gov. We estimated the mean altitude and mean slope at a 1-km pixel resolution from the DEM (Table 2).

Predictor selection and model fitting

The census tracks recorded with the GPS defined initially the area surveyed in our census. As 75% of all guanaco sightings were less that 400 m from the track line, we considered that a 400-m buffer on both sides of the track line defined the area effectively covered for guanacos by the observer. Presence/absence modelling requires defining units in which presence or absence is recorded. In this case we decided to use 1-km grid cells defined by the spatial distribution of VEGETATION NDVI data. We overlaid the surveyed tracks with 400-m buffers on top of this 1-km grid and selected all grid cells that partially or totally overlapped with buffers. This defined our sampling universe. The positions of guanacos ($N = 526$) were overlaid on top of selected cells. Within the sampling universe, 1-km grids with one or more guanaco sightings were considered presences ($N = 427$) and the remaining cells were considered absences ($N = 8268$).

We built generalized additive models with binomial error and logistic link using as response variable guanaco presence/absence, and as potential predictors variables indicative of vegetation productivity, topography, and spatial position, all derived from the digital sources already mentioned (Table 2). Predictors were selected in a stepwise manner using the step.gam procedure of S-Plus 2000 (MathSoft, 1999) that performs a stepwise forward–backward predictor selection based on Akaike's information criterion (AIC; Sakamoto *et al.*, 1986). We initially tested 3 d.f. smooth splines for all predictors. Once a predictor was included in the model we tested for a reduction in number of degrees of freedom of the spline, up to a linear term. The variable AREA_SURVEYED was included as a fixed term spline with 3 d.f. in the model to compensate for the effect on the probability of detection due to the variable proportion of the 1-km cell that had been effectively censused.

Cells with guanaco absences are 20 times more frequent than cells with presence. These zero inflated data create problems with making sound statistical inferences by violating basic assumptions of binomial distribution. There are different approaches that can be used to model this kind of data. The high prevalence of absences could be partially due to imperfect guanaco detection. Martin *et al.* (2005) suggest using a zero-inflated binomial (ZIB) mixture model when this is the case to model two processes: true presence–absence and probability of detection. We cannot use this approach here because it requires that at least some cells have been censused two or more times. Our approach was to resample absences to obtain a balanced presence–absence matrix, repeating the process several times. We built the model in two stages. First, we selected the environmental predictors and the degree of smoothing for each one. Sample cells were divided in presences and absences. We selected 75% of cells with presence and the same number of cells with absences (320 cells) in a random draw, fitted a model with the step-wise procedure previously described, and kept track of the predictors selected in the final model. This procedure was repeated 100 times. On average each absence was used in four, and each presence was used in 75 out of the 100 models. We selected the predictors that entered at least in 50% of the model with the degree of smoothing more frequently selected. This final model was cross-validated. Cells with presence were divided into two sets one-construction set with 75% of the cells and a test set with the remaining 25%. We drew at random without replacement a number of absences equal to the number of presences for the construction set (320 cells) and for the test set (107 cells). The final model was fitted to the construction set and used to predict the test set. The final model was cross-validated using a random sample of 75% of the observations to build the model and the remaining 25% to test it. The cross-validation procedure was repeated 100 times and we obtained mean and SE values for percentage of agreement, kappa, and AUC of the ROC plot (area under the curve of a receiver operating characteristic plot). To translate the probabilities given by each single model to predicted presences/absences we used a cut-point of 0.5, which is adequate for models with an equal number of presences and absences.

Translating models to maps

Generalized additive models have the disadvantage of not having an equation that can be easily incorporated in a GIS to make predictions and produce a map. To produce our map we used IDRISI (version 14.1 Kilimanjaro; Eastman, 2003) possibility of exporting predictors as a data matrix to S-PLUS, fitted the final model to the cells with presence and a random draw of 420 cells with absence, and applied the predict.gam procedure to predict on the new data matrix, and then the predicted probability values, at the scale of the response, were exported from S-PLUS to IDRISI to produce a probability map.

RESULTS

Guanaco distribution model

The models selected by the stepwise procedure included always the predictors ALTITUDE, LATITUDE, and LONGITUDE, in 98% of the instances MEAN_NDVI and in 55% CV_NDVI. Other variables, which we excluded from the final model, were selected only in 32% (MONTH_MAX_NDVI), 16% (GROWTH_PERIOD), and 12% (SLOPE) of the instances. The degree of smoothing for each predictor was also stable. A linear term was selected for LONGITUDE (77% of the instances), ALTITUDE (64%), and CV_NDVI (67%), while a smooth spline

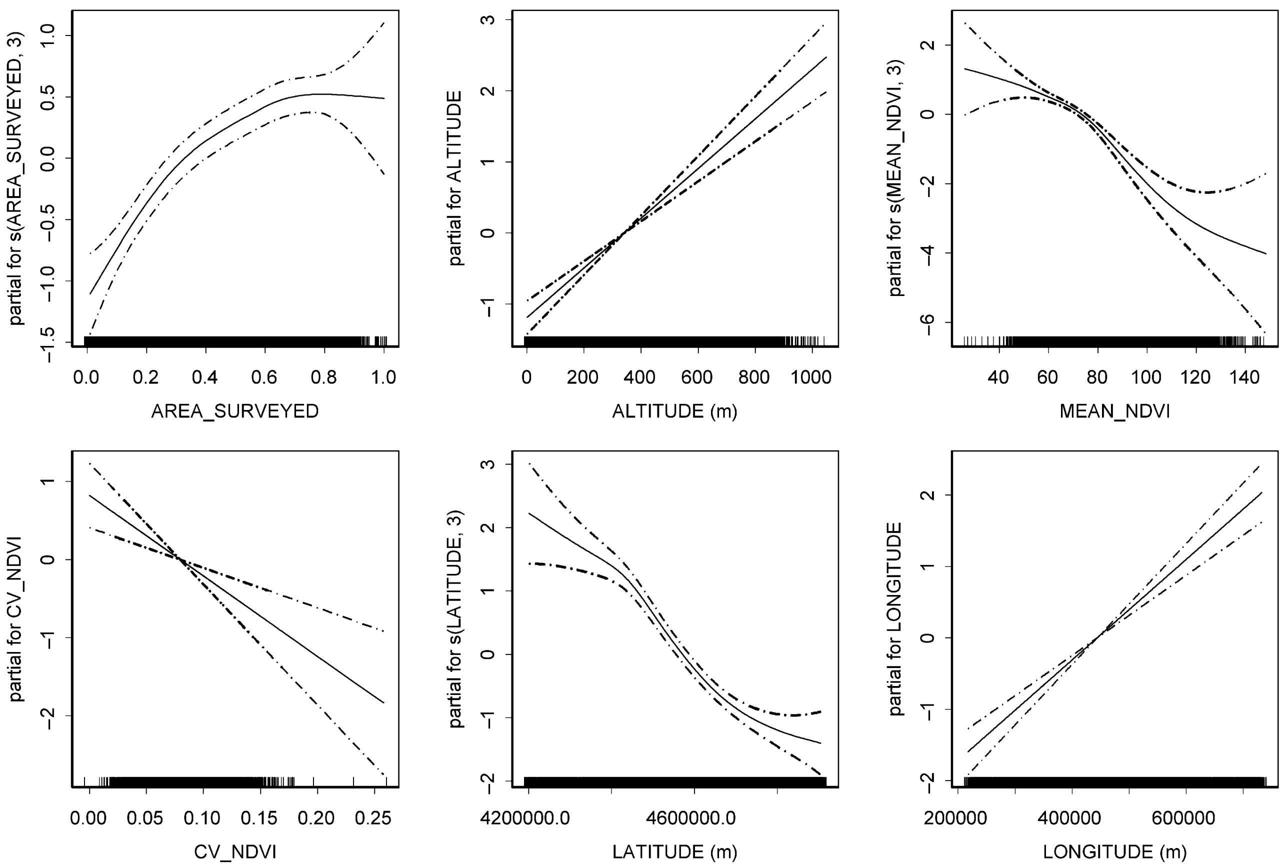

Figure 2 Partial effects of the predictors included in the Guanaco model.

with 3 d.f. was selected for LATITUDE (97%) and MEAN_NDVI (86%). The final model indicates, as expected, that the probability of contacting a guanaco group in a 1-km cell declines with the surface of the cell that has been surveyed. There is a strong linear increase in probability of sightings with altitude, and longitude, and a decline with mean NDVI, latitude, and NDVI variability (Fig. 2). So guanacos are more often found in high ground with low vegetation productivity and low variability in productivity. Corrected for those factors, the probability increases spatially to the south and to the east of Santa Cruz province. There is an east to west gradient of increasing altitude and a north-east to south-west gradient of increasing productivity across Santa Cruz, so it is difficult to visualize the spatial pattern of guanaco distribution without projecting the model in a GIS. The model cross-validation indicates a 67% mean correct classification rate, that is 34% better than a random classification (kappa = 0.341, SE = 0.062, Z = 5.489, $P < 0.001$) and a mean AUC = 0.732 (SE = 0.031). These parameters suggest that it is a relatively good predictive model (Elith, 2000; Harrel, 2001). The model was translated to a GIS assuming a constant 0.7 fraction of the area surveyed for all grid cells (Fig. 3). The model can be simplified into three probability classes to distinguish between areas of high, intermediate, and low guanaco density (Fig. 4) that can be used for conservation priorities or management planning.

We think that spatial heterogeneity in detection probabilities does not affect the model. All areas are steppe habitats with relatively low vegetation height in relation to guanaco size. Only rugged areas could have a reduced visibility for the observer, but SLOPE rarely entered the models, and there was no significant correlation between perpendicular distance to guanaco contacts and SLOPE (Pearson $r = 0.028$, d.f. = 402, $P = 0.574$).

Comparison with alternative censusing and mapping techniques

Traditional road census techniques rely on paper sheets or on a tape recorder to record wildlife sightings (Fuller & Mosher, 1987; Donázar *et al.*, 1993). Record sheets are bulky and require stopping the vehicle to fill the forms. The tape recorder is more convenient, because the information can be dictated without stopping, but is more prone to errors, especially for complex recording schemes, because there is no visual confirmation that all the relevant information has been recorded properly. The geographical position of the census has to be taken from road maps and the position of contacts with wildlife from the car odometer. The accuracy of estimates of the animal's true position depends on map quality, map scale, the precision of the car odometer assuming straight driving, and the ability of the observer to identify its own position in the map and to estimate the distance to the target.

The management of field data recorded in a traditional way to use them for predictive models is costly in time. Data on a tape

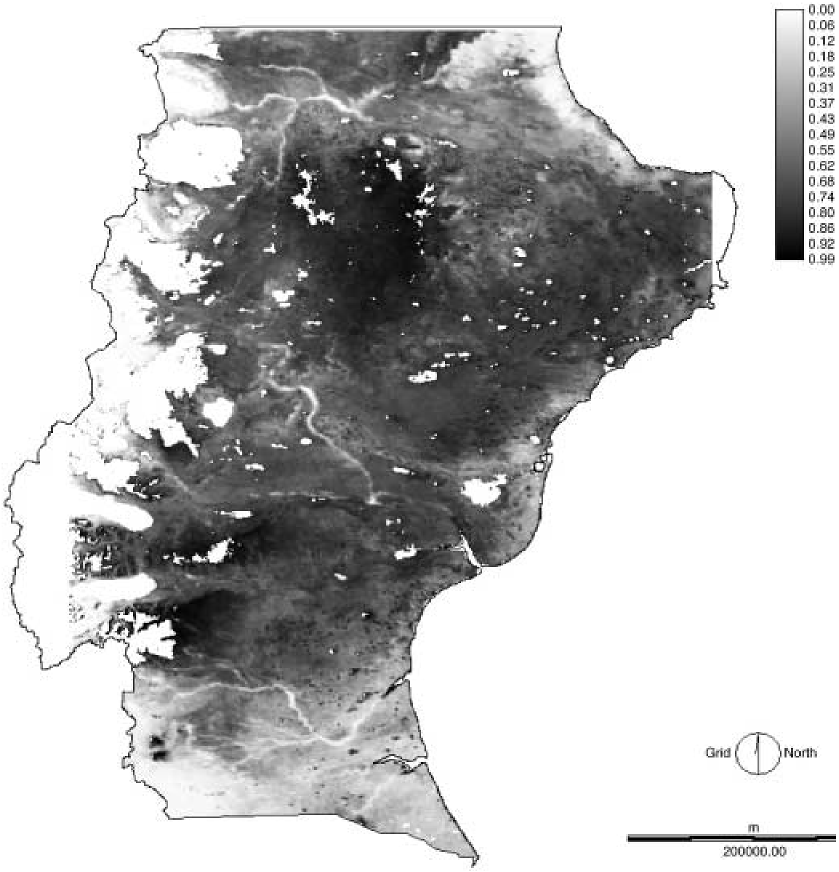

Figure 3 Relative probability of contact with a guanaco group. Darker areas have higher probability according to the model. Areas in white correspond to areas with no predictions, sea, lakes, forested areas, or out of the model environmental space.

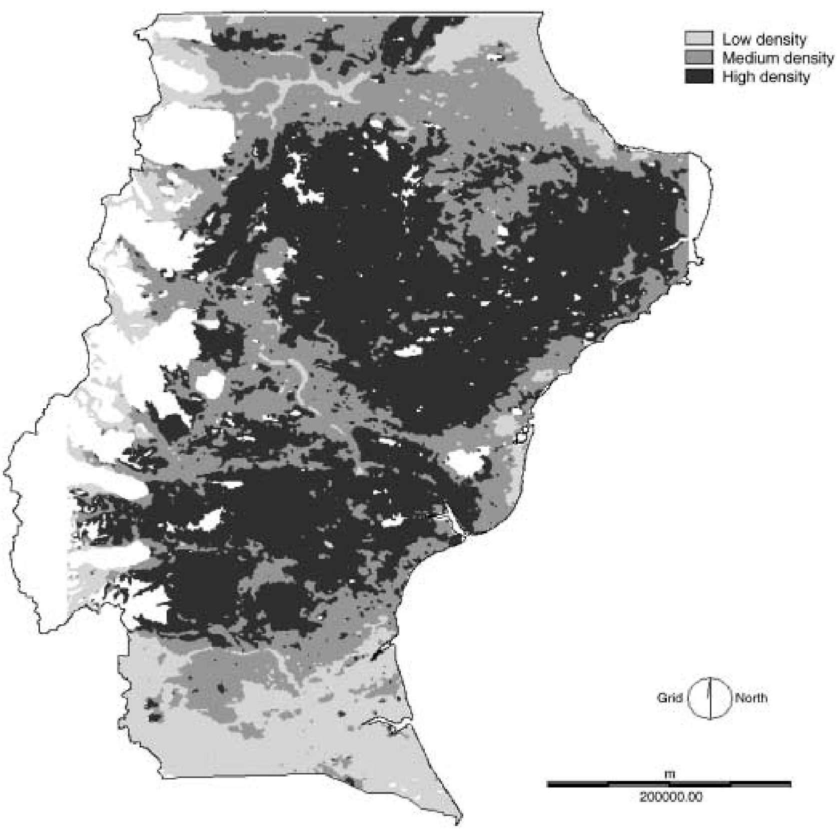

Figure 4 Guanaco density classes. The relative probability map is divided intro three classes (< 0.35, 0.35–0.65, and > 0.65) for a management purpose and filtered with a 3 × 3 mode filter.

Table 3 Time cost of translating field data taken on a tape recorder to a computer database.

	Traditional census*		
	Mean (s)	SD	*N*
Transcription of tapes (per contact)	146.4	96.9	1242
Entering records in the data base (per contact)	38.0	—	30
Georeferencing contacts (per contact)	207.3	71.8	30
Georeferencing census track (per km)	2.04	0.71	1179

*Taken from a similar road census in the same study area performed dictating contacts to a tape recorder and using the vehicle odometer to position the contact.

recorder have to be transcribed to paper, and these records have to be entered into a database. The census track has to be digitized from a map on a digitizing board or on a computer screen. The geographical position of contacts with wildlife has to be estimated, measuring the distance from the start of the census track on a map or on a digitized map on a computer screen. The average time per contact taken for each of these processes is indicated in Table 3 and, on average, adds up to 7 min per contact. Using the traditional approach, the data from our guanaco census would require 61 h after the census to get them ready, but the whole fieldwork with 10 species and 1862 contacts (Table 3) would require 217 h. Apart from the cost in time, the traditional method has two to three more steps to get the data from the field to the model than our procedure with a PDA. We can assume that errors accumulate in each of the processing steps so the traditional procedure would end up with two to three times more errors. Finally, our integrated framework is more accurate in the geographical positions of wildlife. The traditional methods would require locating the contact in a map. In our case in Santa Cruz we could use 1 : 250,000 scale maps. It is assumed that the maximum accuracy with which positions are indicated in a map is 0.5 mm and this corresponds to 125 m on the ground. This is approximately 10 times less accurate than the GPS (95% of the time below 15 m). The traditional method has to add to map errors, those due to the odometer accuracy, the digitizer's accuracy, and the difference between the vehicle position and the guanaco position. Guanacos were observed in our census up to 2000 m from the vehicle. This means that contacts with guanacos in a traditional road census could have minimum errors of 2–3 km which would preclude using 1-km grid cells.

Bustamante & Seoane (2004) have shown that by combining wildlife census with environmental predictive models, it is possible to produce distribution maps more accurate than the information available in field guides, reference works, and even recent atlas work, for forest raptors in southern Spain. In the case of the guanaco in Santa Cruz province the reference work available suggests a uniform distribution (Redford & Eisenberg, 1992), but our modelling results indicate that the distribution in the province is significantly different from a uniform distribution (kappa = 0.341, SE = 0.062, Z = 5.489, $P < 0.001$; see also Fig. 4). Although other modelling alternatives could be used (see Guisan & Zimmerman, 2000 for a review), automatic procedures for predictor selection and model fitting have been shown to be more cost-effective than methods involving expert criteria in the process of selecting and fitting variables when the aim is predictive accuracy (Seoane *et al.*, 2005).

The guanaco model used here as an example makes some untested assumptions. The first one is that 1 km is an adequate resolution to model guanaco spatial distribution. It is very likely that neighbouring cells of cells that contained guanacos also contain them. This can partially be explained by spatial correlation of environmental predictors, but can also be due to guanaco behaviour (e.g. tendency to aggregate). Our spatial terms (LATITUDE, LONGITUDE) partially model these effects. A better solution is to explore the spatial autocorrelation in guanaco contacts and include an autologistic term (Augustin *et al.*, 1996) or model guanaco contacts at a spatial resolution at which contacts can be considered independent. The second assumption is that roads sample the environment at random and do not affect guanaco distribution. Line transects on foot or horseback could be performed at random, but are not an adequate alternative to road transects because the observer would move too slow in relation to guanaco movements. Line transects from a plane (more expensive) could be used to test for any bias due to road transects.

CONCLUSIONS

We think that what we have defined as an integrated framework for species mapping, that is, implementing new technologies in wildlife census, the use of free remote sensing information for environmental predictors, and automatic methods of predictor selection and model fitting combined with GIS, is a promising way to improve wildlife distribution maps. The methods outlined are quite general, but will have to be adapted to each particular region, habitat, or species problem. For example, it may be difficult to use a GPS in a dense forest (although GPS receivers are constantly improving), and road transects are not the adequate survey method for all species. We have made particular emphasis on efficient data collection because it is an aspect, frequently disregarded, that makes up most cost in mapping species distributions and because other papers in this special volume emphasize modelling techniques. Implementing our methodology is actually cheaper than traditional methods because the relative higher cost of hardware is compensated by the reduction in man labour to process the data and by the increase in the accuracy of the information. These methods are especially applicable to large and remote areas of the Third World where qualified man labour and money to build or update wildlife distribution maps are usually scarce.

ACKNOWLEDGEMENTS

This work was primarily funded by the BBVA Foundation through a grant under the Conservation Biology Programme. Additional support was provided by Universidad Nacional de la Patagonia Austral, CONICET (PEI-6065), and the Consejo Agrario Provincial from Santa Cruz. Alejandro Rodríguez was supported by a research contract from the Junta de Andalucía (Spain) and Diego Procopio by a CONICET (Argentina) predoctoral fellowship. Emilio Daher, Miguel Santillán, Martín Yaya, Mara Brossman, and Juan Ignacio Zanón assisted during field work. We also thank the Universidad Internacional de Andalucía, sede Antonio Machado, Baeza, Spain, for organizing the workshop 'Predictive modelling of species distribution. New tools for the XXI century'.

REFERENCES

Augustin, N.H., Mugglestone, M.A. & Buckland, S.T. (1996) An autologistic model for the spatial distribution of wildlife. *Journal of Applied Ecology*, **33**, 339–347.

Bailey, D.W., Gross, J.E., Laca, E.A., Rittenhouse, L.R., Coughenour, M.B., Swift, D.M. & Sims, P.L. (1996) Mechanisms that result in large herbivore grazing distribution patterns. *Journal of Range Management*, **49**, 386–400.

Baldi, R., Elston, D. & Albon, S. (2001) Guanacos and sheep: evidence for continuing competition in arid Patagonia. *Oecologia*, **129**, 561–570.

Baldi, R., Pelliza-Sbriller, A., Elston, D. & Albon, S. (2004) High potential for competition between guanacos and sheep in Patagonia. *Journal of Wildlife Management*, **68**, 924–938.

Battersby, J.E. & Greenwood, J.J.D. (2004) Monitoring terrestrial mammals in the UK: past, present and future, using lessons from the bird world. *Mammal Review*, **34**, 3–29.

Bibby, C.J., Burgess, N.D. & Hill, D.A. (1992) *Bird census techniques*. Academic Press, London.

Bustamante, J. & Seoane, J. (2004) Predicting distribution of four species of raptors (Aves: Accipitridae) in southern Spain: statistical models work better than existing maps. *Journal of Biogeography*, **31**, 295–306.

Calvert, A.M. & Gauthier, G. (2005) Effects of exceptional conservation measures on survival and seasonal hunting mortality in greater snow geese. *Journal of Applied Ecology*, **42**, 442–452.

Carlson, M. & Schmiegelow, F. (2002) Cost-effective sampling design applied to large-scale monitoring of boreal birds. *Conservation Ecology*, **6**, 11 [online] URL: http://www.consecol.org/vol6/iss2/art11/.

Donázar, J.A., Ceballos, O., Travaini, A. & Hiraldo, F. (1993) Roadside raptor surveys in the Argentinian Patagonia. *The Journal of Raptor Research*, **27**, 106–110.

Eastman, J. (2003) IDRISI Kilimanjaro. *Guide to GIS and image processing*. Clark Labs, Clark University, Worcester, Massachusetts.

Elith, J. (2000) Quantitative methods for modeling species habitat: comparative performance and an application to Australian plants. *Quantitative methods for conservation biology* (ed. by M. Burgman), pp. 39–58. Springer, New York.

Fortin, M.J., Keitt, T.H., Maurer, B.A., Taper, M.L., Kaufman, D.M. & Blackburn, T.M. (2005) Species geographic ranges and distributional limits: pattern analysis and statistical issues. *Oikos*, **108**, 7–17.

Fuller, M.R. & Mosher, J.A. (1987) Raptor surveys techniques. *Raptor techniques manual* (ed. by B.A. Giron Pendeleton, B.A. Millsap, K.W. Cline and D.M. Bird), pp. 37–65. Science Technical Service No. 10. National Wildlife Federation, Washington DC.

Golluscio, R.A., Deregibus, V.A. & Paruelo, J.M. (1998) Sustainability and range management in the Patagonian steppes. *Ecología Austral*, **8**, 265–284.

Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.

Harrell, F.E. (2001) *Regression modelling strategies*. Springer, New York.

Hepinstall, J.A. & Sader, S.A. (1997) Using Bayesian statistics, Thematic Mapper satellite imagery, and breeding bird survey data to model bird species probability of occurrence in Maine. *Photogrammetric Engineering & Remote Sensing*, **63**, 1231–1237.

La Sorte, F.A. & Boecklen, W.J. (2005) Temporal turnover of common species in avian assemblages in North America. *Journal of Biogeography*, **32**, 1151–1160.

Lamb, D. (1998) Large-scale ecological restoration of degraded tropical forest lands: The potential role of timber plantations. *Restoration Ecology*, **6**, 271–279.

Lenton, S.M., Fa, J.E. & del Val, J.P. (2000) A simple non-parametric GIS model for predicting species distribution: endemic birds in Bioko Island, West Africa. *Biodiversity and Conservation*, **9**, 869–885.

Link, W.A., Barker, R.J., Sauer, J.R. & Droege, S. (1994) Within-site variability in surveys of wildlife populations. *Ecology*, **75**, 1097–1108.

Logan, N.G. & Smith, T. (1997) A hand-held, pen-based computer system for marine mammal sighting surveys. *Marine Mammal Science*, **13**, 694–700.

Manley, P.N., Zielinski, W.J., Schlesinger, M.D. & Mori, S.R. (2004) Evaluation of a multiple-species approach to monitoring species at the ecoregional scale. *Ecological Applications*, **14**, 296–310.

Martin, T.G., Wintle, B.A., Rhodes, J.R., Kuhnert, P.M., Field, S.A., Low-Chow, S.J., Tyre, A.J. & Possingham, H.P. (2005) Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecology Letters*, **8**, 1235–1246.

MathSoft, I. (1999) *S-Plus 2000 user's guide*. Data Analysis Products Division, Seattle, Washington.

Nichols, J.D., Boulinier, T., Hines, J.E., Pollock, K.H. & Sauer, J.R. (1998) Inference methods for spatial variation in species richness and community composition when not all species are detected. *Conservation Biology*, **12**, 1390–1398.

Ormerod, S.J. & Watkinson, A.R. (2000) Large-scale ecology and hydrology: an introductory perspective from the editors of the Journal of Applied Ecology. *Journal of Applied Ecology*, **37** (Suppl. 1), 1–5.

Ottaviani, D., Lasinio, G.J. & Boitani, L. (2004) Two statistical methods to validate habitat suitability models using presence-only data. *Ecological Modelling*, **179**, 417–443.

Redford, K.H. & Eisenberg, J.F. (1992) *Mammals of the Neotropics. The Southern Cone*, Vol. 2. The University of Chicago Press, Chicago.

Root, T.L. & Schneider, S.H. (1993) Can large-scale climatic models be linked with multiscale ecological-studies. *Conservation Biology*, **7**, 256–270.

Royle, J.A., Nichols, J.D. & Kery, M. (2005) Modelling occurrence and abundance of species when detection is imperfect. *Oikos*, **110**, 353–359.

Sakamoto, Y., Ishiguro, M. & Kitagawa, G. (1986) *Akaike information criterion statistics*. KTK Scientific Publishers, Tokyo.

Seoane, J., Bustamante, J. & Díaz-Delgado, R. (2005) Effect of expert opinion on the predictive ability of environmental models of bird distribution. *Conservation Biology*, **19**, 512–522.

Stoleson, S.H., Kirschbaum, K.J., Frank, J. & Atwood, C.J. (2004) From the field: Integrating GPS, GIS, and avian call-response surveys Using Pocket PCs. *Wildlife Society Bulletin*, **32**, 1309–1312.

Thompson, J.R., Anderson, M.D. & Johnson, K.N. (2004) Ecosystem management across ownerships: the potential for collision with antitrust laws. *Conservation Biology*, **18**, 1475–1481.

Waddle, J.H., Rice, K.G. & Percival, H.F. (2003) Using personal digital assistants for collection of wildlife field data. *Wildlife Society Bulletin*, **31**, 306–308.

van der Werf, W., Woldewahid, G., van Huis, A., Butrous, M. & Sykora, K. (2005) Plant communities can predict the distribution of solitarious desert locust *Schistocerca gregaria*. *Journal of Applied Ecology*, **42**, 989–997.