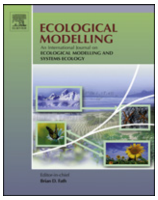
Ecological Modelling 230 (2012) 22–33

Contents lists available at SciVerse ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel



Occurrence and abundance models of threatened plant species: Applications to mitigate the impact of hydroelectric power dams

Ernestino de Souza Gomes Guarino[a,*], Ana Márcia Barbosa[b,c], Jorge Luiz Waechter[a]

[a] Programa de Pós-Graduação em Botânica, Universidade Federal do Rio Grande do Sul. Avenida Bento Gonçalves, 9500, Prédio 43432.2, Bloco 4, Sala 203, Campus do Vale Bairro Agronomia, CEP 91501-970, Porto Alegre, Brazil
[b] Cátedra "Rui Nabeiro" Biodiversidade, CIBIO – Universidade de Évora, 7004-516, Évora, Portugal
[c] Departament of Life Sciences, Imperial College London, Silwood Park Campus, Ascot (Berkshire) SL5 7PY, United Kingdom

ARTICLE INFO

Article history:
Received 14 June 2011
Received in revised form
25 December 2011
Accepted 6 January 2012
Available online 17 February 2012

Keywords:
Species reintroduction
In situ and ex situ plant conservation
Environmental assessment
Model validation

ABSTRACT

Species occurrence and abundance models are important tools that can be used in biodiversity conservation, and can be applied to predict or plan actions needed to mitigate the environmental impacts of hydropower dams. In this study our objectives were: (i) to model the occurrence and abundance of threatened plant species, (ii) to verify the relationship between predicted occurrence and true abundance, and (iii) to assess whether models based on abundance are more effective in predicting species occurrence than those based on presence–absence data. Individual representatives of nine species were counted within 388 randomly georeferenced plots (10 m × 50 m) around the Barra Grande hydropower dam reservoir in southern Brazil. We modelled their relationship with 15 environmental variables using both occurrence (Generalised Linear Models) and abundance data (Hurdle and Zero-Inflated models). Overall, occurrence models were more accurate than abundance models. For all species, observed abundance was significantly, although not strongly, correlated with the probability of occurrence. This correlation lost significance when zero-abundance (absence) sites were excluded from analysis, but only when this entailed a substantial drop in sample size. The same occurred when analysing relationships between abundance and probability of occurrence from previously published studies on a range of different species, suggesting that future studies could potentially use probability of occurrence as an approximate indicator of abundance when the latter is not possible to obtain. This possibility might, however, depend on life history traits of the species in question, with some traits favouring a relationship between occurrence and abundance. Reconstructing species abundance patterns from occurrence could be an important tool for conservation planning and the management of threatened species, allowing scientists to indicate the best areas for collection and reintroduction of plant germplasm or choose conservation areas most likely to maintain viable populations.

1. Introduction

According to statistics from the Brazilian Electricity Regulatory Agency (ANEEL, 2008), more than 150 hydropower plants were in operation in Brazil as of 2008, amounting to an operating capacity in excess of 74 GW. Hydroelectric energy accounts for more than 70% of the electricity consumed in Brazil, and this percentage is likely to increase because Brazil possesses the world's largest hydroelectric potential, with more than 260 GW available. Some estimates indicate that by the end of the twentieth century, more than 30,000 km^2 of land were flooded by hydropower dam reservoirs in Brazil. However, few initiatives have been implemented to improve procedures for the mitigation and management of landscapes affected by dam reservoirs.

Due to Brazil's high biodiversity, it is almost impossible to define sampling procedures that take into consideration every organism affected by the development of an artificial reservoir. An insufficient number of field experts, time and money constraints, and the large areas covered by hydropower dam reservoirs add to this challenge. Thus, to overcome these obstacles, species distribution models (SDMs) constitute a viable alternative that is effective in predicting the occurrence of different organisms affected by dam construction. The use of these models has greatly increased over the last three decades due to the development and dissemination of Geographic Information Systems (Guisan and Zimmermann, 2000). Today, these models are essential tools within conservation biology (Elith et al., 2006). Using different algorithms, species distribution

---

[*] Corresponding author. Current address: Embrapa Acre, Rodovia BR 364, km 14, P.O. Box 321, CEP 69908-970, Rio Branco, Brazil. Tel.: +55 5192273184.
E-mail address: esguarino@cpafac.embrapa.br (E.d.S.G. Guarino).

0304-3800/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2012.01.007

models correlate species occurrence with environmental data (e.g. climate, soil, topography) to predict species presence on a map (Soberón and Peterson, 2005).

Due to the need to find the highest quality areas for the conservation of biodiversity (Pearce and Ferrier, 2001) and to improve the statistical techniques used to predict species abundance, abundance models are undergoing an expansion similar to that observed in the beginning of the last decade. A search for papers in indexed journals on the ISI Web of Science (performed on March 5, 2010; http://www.isiknowledge.com), using the terms *species abundance models* and *species distribution models*, identified over 3700 papers on this topic published in the last 30 years (1979–2009). More than 50% of these articles were published in the last six years (2004–2009), indicating an increasing interest in understanding patterns of species abundance. Information about the abundance of a species provides the following intuitive idea: the greater the number of individuals of a certain species in one area, the greater the probability of maintaining a viable population for that species (Araújo and Williams, 2000). However, this intuitive assumption is a target of a broad discussion (i.e., Skagen and Yackel Adams, 2011; Van Horne, 1983). According to Van Horne (1983), population density is not a good direct measure of habitat quality. This author suggests that habitat-quality assessment is based only on simple estimations of total density and forgets to take into account the demography of the species and of the factors influencing population levels through their influence on survival and production. Although important, this information is expensive to collect and time demanding, undesirable factors when it comes to planning conservation actions. In some circumstances there is no choice but to assume that habitat quality is correlated with abundance (Pearce and Ferrier, 2001).

The positive relationship between occupancy and abundance is one of the most interesting topics within ecology (Brown, 1984; Gaston et al., 2000; He and Gaston, 2007) and is the subject of constant discussion (Blackburn and Gaston, 2009; Komonen et al., 2009; Kotiaho et al., 2009). This macroecological pattern suggests that locally abundant species tend to be more widely distributed than locally rare species, which tend to be of restricted occurrence (He and Gaston, 2000). This pattern can be attributed to the relationship between the population density of a species with the spatial distribution of environmental factors which determine its distribution. Both density (or abundance) and probability of occurrence are intuitive indicators of habitat quality, so we can expect a positive relationship between them (e.g. Pearce and Ferrier, 2001). However, this relationship is not always observed. There is a large number of non-environmental controls of abundance, including biotic interactions such as predation or interspecific competition (Holt et al., 2002; Van Horne, 1983), dispersal limitation (Holt et al., 2002; Pulliam, 2006; Verbek et al., 2010) and different species detectability among habitats, seasons, weather conditions (Gu and Swihart, 2004; Pearce and Ferrier, 2001) and observers (Chen et al., 2009). This can lead to the population of a species reaching high values of abundance within low probability of occurrence areas.

Few studies have attempted to reproduce abundance patterns using probability of occurrence data generated by species distribution models, with different statistical approaches and inconsistent results (Jiménez-Valverde et al., 2009; Nielsen et al., 2005; Pearce and Ferrier, 2001; Real et al., 2009; VanDerWal et al., 2009). Working with a large number of species from different groups (arboreal marsupials, small reptiles, diurnal birds, vascular plants and arthropods), Pearce and Ferrier (2001) and Jiménez-Valverde et al. (2009) showed that for a small number of species the probability of occurrence may also serve as a proxy of abundance. Jiménez-Valverde et al. (2009) suggest that this relationship tends to be more common for species with high dispersal ability. VanDerWal et al. (2009) showed that probability of occurrence generated by presence-only species distribution models could predict upper limits of local abundance for rain forest vertebrates in the Australian wet tropics. Real et al. (2009) showed that abundance of Iberian lynx and European rabbit correlates positively, although not strongly, with probability of occurrence. Contrasting patterns were described by Nielsen et al. (2005) for bracken fern and moose.

Regardless of geographical range or organism, species distribution models have been successfully used in a large variety of conservation biology studies (Cayuela et al., 2009; Rodríguez et al., 2007). Good examples of application of species distribution models in conservation biology problems are given by Alves and Fontoura (2009) with fish in a hydrographic basin in southern Brazil; Barbosa et al. (2003, 2010) and Real et al. (2009) with otters, lynx and rabbits, and desmans, respectively, in the Iberian Peninsula; Willis et al. (2009) with birds in sub-Saharan Africa; Zhu et al. (2007) with invasive plants in China; and Parolo et al. (2008) with plant reintroduction in Alps. However, examples of species abundance models applied to solve conservation problems are rare, probably due to the difficulty to obtain and to analyse abundance or density data.

In this study, we modelled both the occurrence and abundance of a set of plant species in the surroundings of a power dam and tested the following hypotheses: (i) there is a positive relationship between predicted probability of occurrence and true abundance; (ii) there is a positive relationship between a model's discrimination capacity and its relationship with true abundance; and (iii) models based on abundance are more effective in predicting species occurrence than those based on presence–absence data. We discuss the application of our methods and the implications of our results for the conservation of species affected by the construction of hydroelectric power dams.

2. Materials and methods

2.1. Study area

The field survey was conducted around the Barra Grande dam reservoir in southern Brazil (Fig. 1). Located in the Pelotas River Basin between the states of Santa Catarina and Rio Grande do Sul, this reservoir encompasses approximately 90 km^2, and its surroundings, delimited by watershed boundaries, cover an area of roughly 4600 km^2. Elevation ranges between 500 and 1200 m above sea level, and the climate types range from Cfa (humid subtropical) to Cfb (oceanic or marine temperate), depending on the elevation quota (Köppen climate classification). Annual precipitation is 1.412 mm, and the mean temperature is 15.2 °C (Maluf, 2000). Topography varies from rolling highlands (Southern Brazilian Plateau), where cambisol and alfisol are the most common soil classes, to steep slopes near the Pelotas River, where lithosol is the most frequent class (Potter et al., 2004; sensu Brazilian soil classification system – SiBCS).

The vegetation is characterised by continuous areas of semideciduous forests predominantly located close to the Pelotas River. In the highlands ($\geq$800 m), Araucaria forests cover a large area that is naturally fragmented by grasslands (Joly et al., 1999; Klein, 1975).

2.2. Target species

The time and resources available allowed the field sampling of nine selected species: *Araucaria angustifolia* (Bertol.) Kuntze (Araucariaceae), *Butia eriospatha* (Mart. ex Drude) Becc. (Arecaceae), *Clethra scabra* Pers. (Clethraceae), *Dicksonia sellowiana* (Presl.) Hook. (Dicksoniaceae), *Erythrina falcata* Benth. (Fabaceae), *Maytenus ilicifolia* (Schrad.) Planch. (Celastraceae), *Myrocarpus frondosus* Allemão (Fabaceae), *Podocarpus lambertii* Klotzsch ex

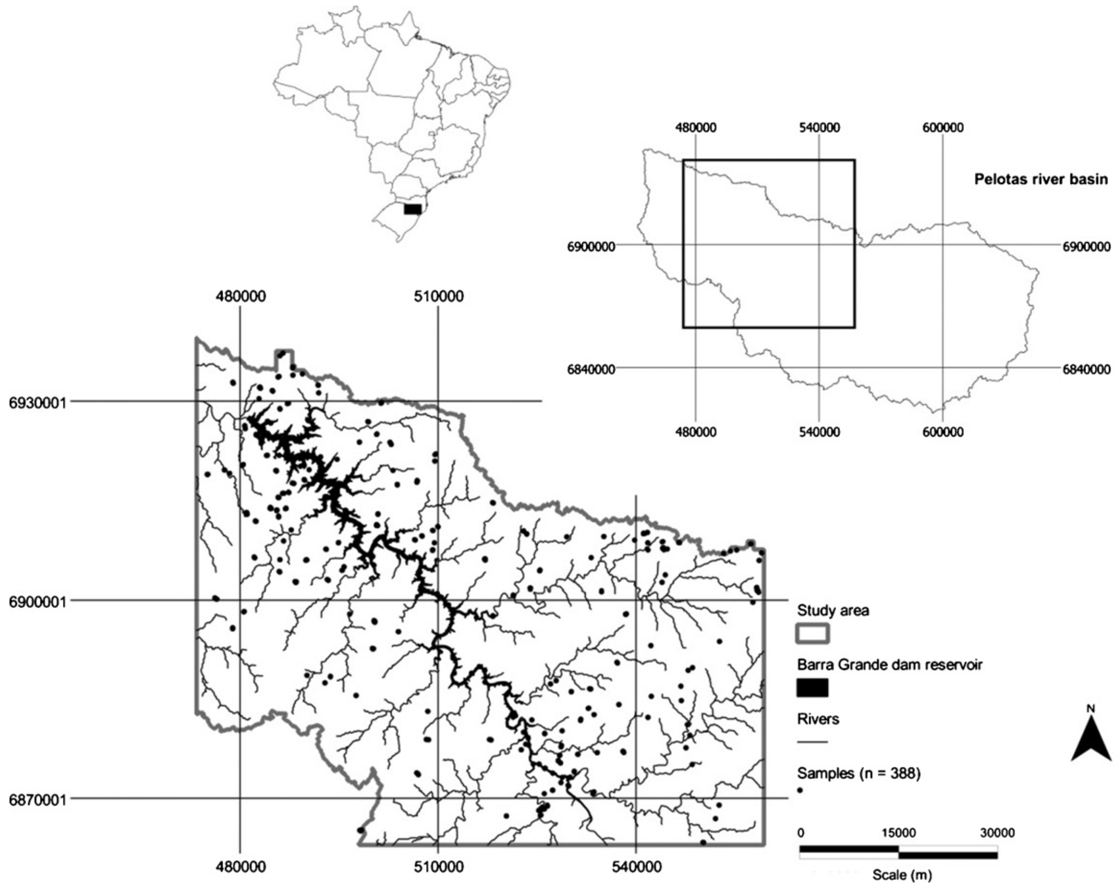


Fig. 1. Location map of the study area and sampling sites (Universal Transverse Mercator coordinate system, Zone 22J, southern Brazil).

Endl. (Podocarpaceae) and *Trithrinax brasiliensis* Mart (Arecaceae). Except for *E. falcata*, all of these species possess threat categories, both globally (IUCN Red List) and locally (Rio Grande do Sul State list of threatened species). Due to the intense commercial exploitation of *D. sellowiana*, this fern had its market regulated by the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES, Appendix II). All of these species are easily identified in the field, minimising sampling errors (false absences).

2.3. Sampling

Past studies have used standardised abundance estimates obtained from different surveys or estimates of abundance from indirect methods (F-igueiredo and Grelle, 2009; He and Gaston, 2007). These methods, according to Austin and Meyers (1996), can create an unwanted bias. In contrast, our work was based on reliable occurrence and abundance data. These data were obtained by conducting a detailed field survey and germplasm collection expeditions for the purpose of ex situ conservation of the target species. We randomly sampled 388 georeferenced plots (10 m × 50 m) around the Barra Grande hydropower dam reservoir. In each plot we counted the number of individuals of each target species with height greater than 1.5 m. To prevent or reduce the influence of topography in sampling, plots were allocated along different topographic contour levels, similar to those described by Magnusson et al. (2005) for rapid surveys of biodiversity.

To avoid any effects of spatial autocorrelation, the minimum distance between plots was 50 m. The value of this precaution was later validated by Moran's *I* correlograms. All correlograms were calculated using the software SAM v3.1 (Rangel et al., 2010). For all target species, the variation follows a random pattern with a small oscillation around the zero value, which represents the absence of significant autocorrelation (Appendix A; Fortin and Dale, 2005; Legendre and Legendre, 1998).

2.4. Environmental variables

We used 15 environmental variables (direct and indirect, sensu Austin and Meyers, 1996) to predict the spatial distribution of the target species (Table 1). All environmental maps had a spatial resolution of 30 m. All spatial data were stored and analysed using the

Table 1
Environmental variables used to predict the spatial distribution of the target species.

Group	Variable
Soil	pH H_2O
	$Ca^{+2} + Mg^{2+}$[*]
	$H^{+} + Al^{3-}$[*]
	K^{+}
	S^{2-}
	P^{+}
	Total nitrogen
	Silt:clay ratio[*]
	Bulk density
Topography	Elevation (m)[*]
	Northness
	Eastness
	Topographical wetness index (TWI)
	Slope (°)
Current vegetation	Normalised Difference Vegetation Index (NDVI)

[*] Variables excluded from analysis due to high Pearson correlation coefficient (r) with one or more environmental variable ($r \geq 0.7$; $P < 0.01$).

software Quantum GIS v.1.5.0 (Quantum GIS Development Team, 2009) and its interface to GRASS (GRASS Development Team, 2010).

2.4.1. Topographic variables

Elevation values were obtained from a Digital Elevation Model (DEM) generated by the ASTER sensor (ASTER GDEM). From this DEM, four new topographic variables were generated (northness, eastness, topographic wetness index and slope). (i) Northness and eastness (Roberts, 1986): the occurrence of different vegetation physiognomies is intimately connected to the amount of available solar radiation. In the southern hemisphere, places with a northeast solar orientation have greater Sun exposure and consequently a higher rate of evapotranspiration, resulting in the occurrence of vegetation with xerophytic characteristics. Conversely, slopes facing south–southeast, especially during the winter, are exposed to less sunlight (Kirkpatrick et al., 1988; Kirkpatrick and Nunez, 1980). To explore this relationship, aspect-modified maps were created in this study. The maps indicate a trend to the north (northness = *cos* [aspect]) and east (eastness = *sin* [aspect]). After this transformation, both variables reached values between 1 and −1, indicating a gradient north to south and west to east. (ii) Topographic wetness index (TWI) describes the spatial pattern of soil wetness and is defined as a function of slope. TWI is obtained using the function $\ln(A/\tan\beta)$, where A is the upslope draining through a determined point x grid cell size and β is the point slope (Neteler and Mitasova, 2008; Sørensen et al., 2005). (iii) Slope was used to indicate soil depth. A higher degree of slope corresponds to a shallower soil depth (Penížek and Borůvka, 2006; Tsai et al., 2001).

2.4.2. Soil variables

Due to the lack of soil maps at the spatial resolution adopted, we collected soil samples at 381 of our sampling plots (depth: 0–20 cm) and analysed their chemical and physical properties (Table 1). Owing to logistical problems we were not able to analyse soil samples from the remaining seven sites. To overcome this problem, we performed a spatial interpolation of the properties of our 381 soil samples, using the regularised spline with tension algorithm (RST; Mitasova and Hofierka, 1993; Mitasova and Mitas, 1993), to cover all the sites studied. RST is a robust and flexible method used to select parameters that control the properties of the interpolation (tension and smoothing). In addition, estimates generated by RST have accuracy similar to traditional geostatistical methods (e.g. ordinary and universal kriging; see Chaplot et al., 2006 for more details). The best interpolation control parameter combination was selected iteratively using cross-validation (*leave-one-out* method; Tomczak, 1998). To ensure accuracy, we examined the decrease in root mean square error (RMSE) and mean differences between the observed and predicted values (the closer to zero, the better the estimate). In addition, our soil maps were validated using other, coarser-scale soil maps available for this region.

2.4.3. Current vegetation cover

To estimate current vegetation cover, we used the Normalised Difference Vegetation Index (NDVI), obtained by dividing bands 3 (visible red) and 4 (near infrared) of Landsat 5 TM. NDVI values vary between −1.0 and +1.0, and high pixel values represent plentiful vegetation. To avoid any effects of leaf phenology during the year (forest deciduousness), we used Landsat 5 TM images collected during summer in the southern hemisphere.

2.5. Data analysis

Except for the Moran's I correlograms described above, all other statistical analyses were carried out with R software (v. 2.10.0; R Development Core Team, 2009).

2.5.1. Presence–absence models

Generalised Linear Models (GLMs) are the most common regression method used to predict species' spatial distributions. According to McCarthy and Elith (2002), GLMs provide a rigorous and statistically robust method to predict the occurrence (or abundance) of species. We modelled the observed presence–absence of each target species with GLM (binomial distribution, logistic link function) using the 15 environmental variables described in Table 1 as predictors. The predictors were selected through a forward–backward stepwise procedure based on small-sample-size-corrected Akaike's Information Criterion (AIC_c; Burnham and Anderson, 2002), using a modified version of the *stepAIC* function of the R MASS package. Variables were thus added to or removed from the model according to how they changed its AIC_c; the best (or minimal adequate) model for each species is achieved when no variable can be added or removed without an increase in AIC_c. Stepwise selection is a useful and effective tool to infer distribution patterns inductively from observed data, when no theory or previous hypotheses exist about the importance of each variable (Guisan and Zimmermann, 2000; Real et al., 2009).

We evaluated the models' discrimination capacity (i.e., their ability to distinguish presence from absence cases) using the Area under the Receiver Operating Characteristic (ROC) curve (AUC; Fielding and Bell, 1997). AUC values may vary between 0 and 1. Values close to 0.5 indicate that model predictions are no better than random, and AUC values equal to 1 indicate a 100% chance for the model to correctly classify an event (in our case, species presence or absence). AUC values lower than 0.5 indicate that the model is discriminating presences from absences, but using the information in a reversed way (Fawcett, 2006). The null hypothesis that the area under ROC curve is ≤ 0.5 was tested using a Mann–Whitney U-test (Mason and Graham, 2002).

Model predictions were also compared with the abundance data collected in the same sites. Using Spearman's rank correlation (*rho*), we tested the relationship between observed species abundance and the predicted probability of occurrence for each species. We also tested the hypothesis that if a model correctly predicts abundance, it has a high discrimination capacity (Jiménez-Valverde et al., 2007). We repeated this analysis using only locations with observed abundance >0 as suggested by Pearce and Ferrier (2001). We also verified the relationship between mean number of occupied sites per species (which is the sample size for abundance >0) and abundance using a Wilcoxon rank sum test (Zar, 1999).

2.5.2. Abundance models

Regression models based on Poisson distributions are often used to analyse count data (such as abundance). However, these data are often Zero-inflated, i.e., the incidence of zeros is larger than expected by chance (Ridout et al., 1998; Welsh et al., 1996). Zuur et al. (2009) described five sources of zeros, two related to species occupancy patterns (i.e., the habitat is not suitable and the species is not present, or the habitat is suitable but is not used by the species) and three related to sampling errors (i.e., design error, low species detectability, and sampling outside species' habitat range – "naughty naughts" sensu Austin and Meyers (1996)). The rigorousness of our field survey and the high detectability of our target species warrant that our samples do not contain false negatives (sampling errors), so the zeros that do occur are only related to patterns of species occupancy.

To model abundance we used the same 15 environmental predictors described in Table 1 and tested three different algorithms: (i) GLM with Poisson (P) and negative binomial (NB) distributions (GLM_P and GLM_{NB}), (ii) Hurdle models (H_P and H_{NB}, Zeileis et al., 2008) and (iii) Zero-Inflated Count Data Regression (ZI_P and ZI_{NB}, respectively). GLM_P (standard Poisson) is the simplest method used to model count data and assumes equi-dispersion (variance = mean), but this assumption is not always met in practice due to zero-inflation (variance > mean). One alternative to deal with overdispersion is the use of negative-binomial regressions, where variance is estimated as a quadratic function of the mean (Ver Hoef and Boveng, 2007; Lindén and Mäntyniemi, 2011).

Hurdle models are two-component class models capable of accounting for overdispersion, or underdispersion, using Poisson (Hurdle_{P}) or negative binomial ($\text{Hurdle}_{\text{NB}}$) distributions. Hurdle models slack the assumption that zeros and values >0 come from the same process (Cameron and Trivedi, 1998). As a first step, Hurdle models use a truncated count component for values >0, assuming that these values arise from the effect of conditions that result in passing a probability threshold or zero-hurdle (Cameron and Trivedi, 1998; Gray, 2005; Potts and Elith, 2006). As a second step, a Hurdle component models zeros vs. non-zero values using a binomial GLM (Zeileis et al., 2008). Zero-inflated models are two-component mixture models in which zeros are modelled as originating from two stochastic processes, the binomial process and the count process. Similar to Hurdle models, Zero-Inflated models use binomial GLM to model the probabilities of measuring zeros, and the count process is modelled by a Poisson (ZI_{P}) or negative binomial (ZI_{NB}) GLM (Zuur et al., 2009). The predictors of all abundance models (GLM, Hurdle and ZI_{P} models) were selected using the same AIC_{c}-based approach applied to presence–absence models (see Section 2.5.1).

We selected the best abundance model for each target species using AIC_{c} and evaluated it with the following four criteria: (i) Pearson correlation coefficient (*r*), which varies from −1 to +1 and provides an indication of agreement between observed and predicted abundance values (note that a perfect adjustment ($r = 1$) does not imply an exact prediction); (ii) Spearman's rank correlation (*rho*), which also varies from −1 to +1 and provides an indication of similarity in rank between observed and predicted abundance values; (iii) linear regression coefficients, which are obtained by fitting a simple linear regression (observed values = *m*(predicted values) + *b*; in a perfectly calibrated model, *m* should equal 1 and *b* should equal 0); and (iv) root mean square error (RMSE) and average error ($\text{AVE}_{\text{error}}$), both of which are dependent on sample size and measure divergences between observed and predicted abundance values (Potts and Elith, 2006). Confidence intervals (95%) for each evaluated parameter were calculated using a bootstrap procedure (1000 replicates).

3. Results

3.1. Presence–absence models

AUC values ranged from 0.71 to 0.96 and in all cases were significantly different from 0.5 (Mann–Whitney *U*-test, $P < 0.01$; for details about environmental variables included in the GLM models see Appendix B). Overall, the relationship between observed abundance and occurrence probability was positive and statistically significant (Table 2 and Fig. 2). When we truncated the abundance data to eliminate zeros, the results were different. Only *A. angustifolia* (which had the highest prevalence) and *C. scabra* (with a relatively low prevalence) retained significant correlations between observed abundance and probability of occurrence values (Table 2 and Fig. 2). For the remaining seven species, the relationship between abundance and probability of occurrence was not significant when unoccupied sites were excluded. Note, however, that this exclusion of unoccupied sites implied that the mean sample size for these species was significantly smaller (Mann–Whitney *U*-test, $P < 0.001$).

The correlation between AUC (i.e., model discrimination capacity) and *rho* (i.e., the rank correlation between occurrence probability and abundance) was 0.18 (Spearman coefficient, $P = 0.63$). Therefore, we found no proof that models with better discrimination capacity were better at predicting species abundance.

Table 2
Presence–absence model evaluation. Abundance: median and range, *P* (%): (Prevalence = [true positives + false negatives]/sample size) × 100; *rho*: Spearman's rank correlation; AUC: area under the receiver operator characteristic curve (Mann–Whitney *U*-test).

Target species	Abundance	*P* (%)	*rho*	*rho* (Abd > 0)	AUC
A. angustifolia	4 (0–49)	72.42	0.42*	0.17*	0.77**
B. eriospatha	0 (0–24)	1.55	0.21*	0.39	0.95**
C. scabra	0 (0–27)	9.79	0.30*	0.32*	0.79**
D. sellowiana	0 (0–153)	13.40	0.25*	0.02	0.71**
E. falcata	0 (0–4)	4.38	0.23*	0.16	0.83**
M. ilicifolia	0 (0–14)	21.13	0.47*	0.08	0.83**
M. frondosus	0 (0–29)	5.15	0.32*	0.12	0.92**
P. lambertii	0 (0–24)	5.67	0.18*	0.00	0.73**
T. brasiliensis	0 (0–21)	4.12	0.32*	0.21	0.96**

* $P < 0.05$.
** $P < 0.001$.

3.2. Abundance models

Due to the high prevalence of zero counts (98%), we were not able to model the abundance of *B. eriospatha*. Based on AIC_{c} values, Hurdle and Zero-Inflated models (Poisson or negative binomial distributions) were better than GLM with Poisson or negative binomial distributions (Table 3).

We could clearly distinguish three groups of species using abundance evaluation parameters. (i) *C. scabra*, *D. sellowiana*, *M. frondosus* and *P. lambertii* showed the worst-fit models with the highest values of RMSE and $\text{AVE}_{\text{error}}$ (Fig. 3). *D. sellowiana* had a strong and inconsistent bias ($b = 1.461$ and $m = 0.151$), while *C. scabra*, *M. frondosus* and *P. lambertii* showed similar values of *b* (0.230, 0.203 and 0.304, respectively) and *m* (0.01, 0.05 and 0.00, respectively). (ii) *E. falcata*, *M. ilicifolia* and *T. brasiliensis* showed accurate abundance estimates, with *b* and *m* equal or closer to zero and one, respectively (Fig. 3). These species had relatively high values of *r* (≥0.55) and *rho* (≥0.33), indicating that both observed and predicted abundance measures were similar in magnitude, but not similarly ordered (Potts and Elith, 2006). Also, these species gave the smallest RMSE and $\text{AVE}_{\text{error}}$ (Fig. 3). (iii) Finally, a group composed by only *A. angustifolia* which had consistent bias ($b = 1.06$ and $m = 0.83$; Fig. 3).

3.3. Environmental correlates of occurrence and abundance

Soil pH was negatively related to the abundance of *A. angustifolia*, *C. scabra*, *D. sellowiana* and *M. frondosus*, and positively related to the abundance of *T. brasiliensis* (Table 4). Soil pH provides an indirect nutrient gradient in the soil, and its values directly affect the uptake of K^+, S^{2-} and P^+ (low or negative β values,

Table 3
Selected abundance models based on AIC_{c}. ZI_{P}, Zero-Inflated Count Data Regression with Poisson distribution; ZI_{NB}, Zero-inflated Count Data Regression with Negative Binomial distribution; $\text{Hurdle}_{\text{NB}}$, Hurdle Model Regression with Negative Binomial distribution.

Target species	Model	AIC_{c}
A. angustifolia	ZI_{NB}	2148.84
C. scabra	ZI_{NB}	347.28
D. sellowiana	ZI_{NB}	617.37
E. falcata	ZI_{NB}	144.35
M. ilicifolia	$\text{Hurdle}_{\text{NB}}$	569.96
M. frondosus	ZI_{P}	173.67
P. lambertii	ZI_{P}	241.48
T. brasiliensis	ZI_{NB}	158.12

Table 4
Variables included in the abundance models (count and zero components) obtained for the studied species. β, coefficient; S.E., standard error; $P(z)$, significance of the z-statistic test.

Target species	Model component	Variable	β	S.E.	$P(z)$
A. angustifolia	Count	pH H_2O	−0.72	0.14	<0.001
		S^{2-}	−0.07	0.03	0.03
		Eastness	1.82	0.40	<0.001
		Slope	0.03	0.01	<0.001
	Zero	pH H_2O	1.33	0.58	0.02
		S^{2-}	−0.18	0.05	<0.001
		N total	−11.44	3.75	0.002
		Northness	10.44	3.22	<0.001
		Eastness	5.95	2.39	0.01
		NDVI	− 5.13	1.76	0.005
		Slope	−0.18	0.07	0.01
C. scabra	Count	Constant	58.82	12.63	<0.001
		pH H_2O	−1.27	0.61	0.03
		S^{2-}	−0.07	0.02	0.003
		P^+	−0.56	0.16	<0.001
		Bulk density	−17.54	4.23	<0.001
		NDVI	−20.18	4.00	<0.001
		Slope	0.27	0.05	<0.001
		S^{2-}	−0.13	0.05	0.01
		P^+	−1.03	0.51	0.04
		NDVI	−20.63	8.55	0.01
D. sellowiana	Count	Constant	7.94	3.38	0.02
		pH H_2O	−1.30	0.61	0.03
		K^+	1.68	0.61	0.006
		N total	6.22	2.81	0.02
		TWI	−0.05	0.19	0.009
		Slope	0.14	0.06	0.01
E. falcata	Count	Constant	−31.89	3.59	<0.001
		pH H_2O	2.67	0.73	<0.001
		K^+	−3.46	0.69	<0.001
		N total	19.75	4.36	<0.001
		Bulk density	6.27	1.52	<0.001
		Eastness	−8.38	2.13	<0.001
	Zero	K^+	−55.17	27.09	0.04
M. ilicifolia	Count	P^+	0.69	0.26	0.007
		Northness	2.03	0.88	0.02
	Zero	Constant	−17.60	2.07	<0.001
		pH H_2O	2.66	0.35	<0.001
		S^{2-}	0.08	0.01	<0.001
M. frondosus	Count	Constant	76.65	13.24	<0.001
		K^+	1.09	0.23	<0.001
		S^{2-}	−0.36	0.06	<0.001
		P^+	−1.02	0.17	<0.001
		Bulk density	−23.20	3.54	<0.001
		Eastness	8.09	0.91	<0.001
		Slope	0.13	0.02	<0.001
	Zero	Constant	673.77	288.92	0.02
		pH H_2O	−42.25	18.68	0.02
		K^+	−6.83	3.35	0.04
		S^{2-}	−3.38	1.47	0.02
		P^+	−2.82	1.30	0.03
		N total	89.08	43.27	0.04
		Bulk density	−137.98	59.40	0.02
		Eastness	38.07	16.31	0.02
		Slope	−1.25	0.57	0.03
P. lambertii	Count	Constant	−3.75	1.46	0.01
		K^+	−2.16	0.47	<0.001
		P^+	0.53	0.11	<0.001
		N total	4.68	2.31	0.04
		Northness	8.88	2.36	<0.001
		Eastness	−6.21	1.65	<0.001
		TWI	0.52	0.08	<0.001
		NDVI	9.48	2.00	<0.001
		Slope	−0.56	0.08	<0.001
	Zero	P^+	0.91	0.30	<0.001
		Northness	12.57	5.23	0.01
		Slope	−0.55	0.21	0.09
T. brasiliensis	Count	pH H_2O	7.86	2.56	0.002
		K+	5.70	1.20	<0.001
		S^{2-}	−0.25	0.09	0.06
		TWI	0.92	0.19	<0.001
		NDVI	6.13	1.60	<0.001
		Slope	0.52	0.10	<0.001

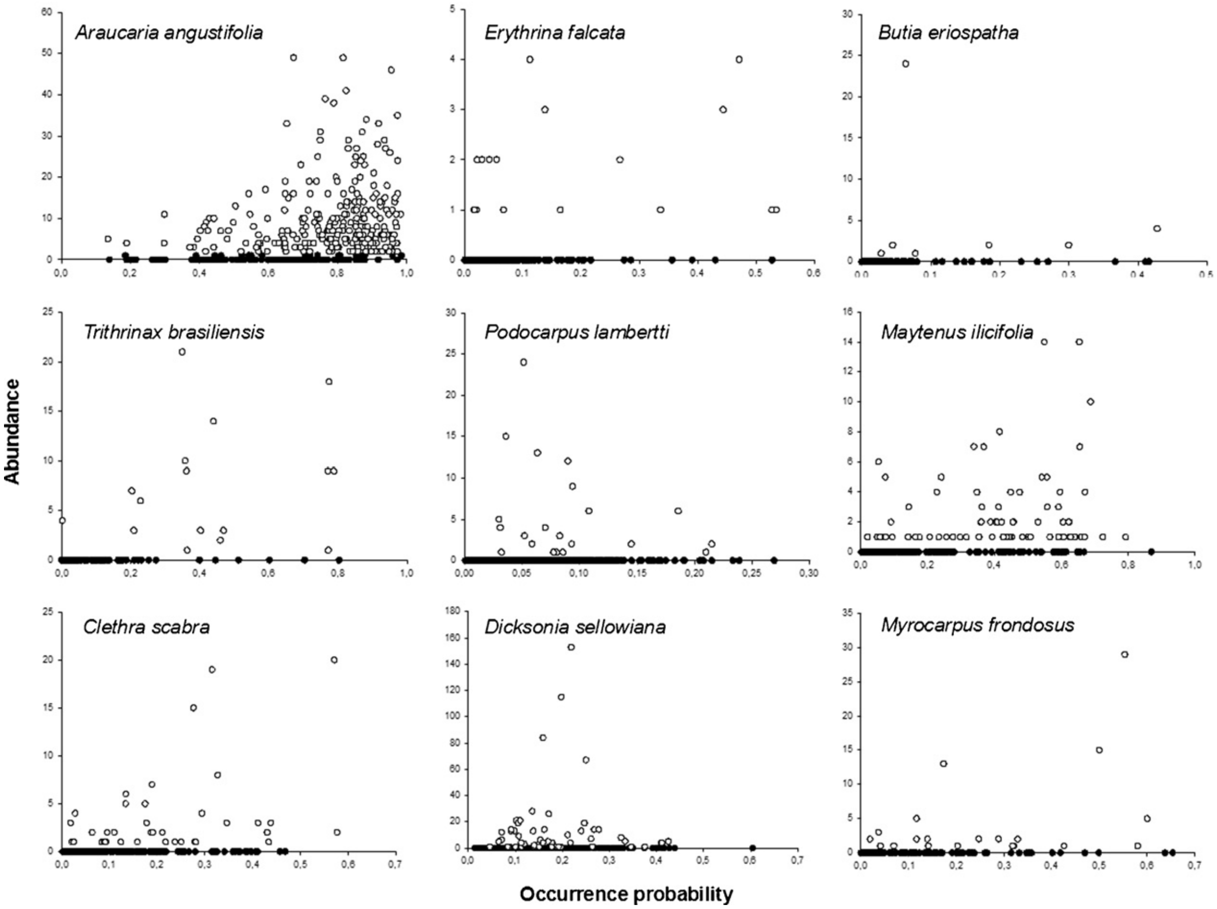


Fig. 2. Relationship between occurrence probability and true abundance values (abundance ≥ 0) for each studied species. Black dots, species absence; white dots, species presence.

Table 4) for each species. The occurrence of *M. ilicifolia* correlated with soil pH. However, neither abundance nor occurrence of *P. lambertii* was affected by soil pH. The abundance of *E. falcata*, was positively related to the amount of nitrogen in soil, while the abundance of *P. lambertii* was positively related to nitrogen (Table 4). Northness, eastness, slope and TWI each had a different relationship with the values of abundance and occurrence, occasionally affecting both simultaneously (Table 4). The values of NDVI, an indirect indicator of the stage of vegetation succession (current vegetation biomass), were negatively related to the abundance of *C. scabra*. Occurrence of *C. scabra* was also associated with low values of NDVI, while *P. lambertii* and *T. brasiliensis* were positively related to NDVI (Table 4).

4. Discussion

4.1. Presence–absence models and their relationship with abundance

Because of the need for reliable species distribution models to aid in developing conservation strategies, methods used to assess model accuracy are one of the most important issues in species distribution modelling (see e.g. Elith et al., 2006; Fielding and Bell, 1997; Manel et al., 2001). Lobo et al. (2008) conducted a detailed description of the different issues involved with the misuse of the AUC, which is widely used as a measure of accuracy of species' distribution models. One alternative proposed by the authors is the use of abundance data to validate these models. This idea is based on an inductive relationship, where probabilities of occurrence are functionally related to species abundance (Nielsen et al., 2005).

Pearce and Ferrier (2001) attempted to use predicted probabilities of occurrence as a surrogate method to predict species abundance. However, according to these authors, the relationship between the probabilities generated by linear models and observed abundance is weak and restricted to a few species. Similar results have been reported by Nielsen et al. (2005) and Jiménez-Valverde et al. (2009). In both studies, the probability of occurrence is not correlated with abundance when points with zero abundance are excluded from analysis. Pearce and Ferrier (2001) suggest that the overall correlation between abundance and probability of occurrence is due mostly to the difference in mean predicted probability between occupied and unoccupied sites. However, the fact that this correlation often disappears when unoccupied sites are excluded might also be due, at least in part, to a reduction in sample size and the consequent loss of analytical power. Indeed, in our data, sample size was highly significantly reduced when unoccupied sites were excluded (mean occupied sample size = 59.33 ± 86.32; see "prevalence" in Table 2), which may have impeded the detection of significant relationships between abundance and probability of occurrence in occupied sites. Also in Pearce and Ferrier's (2001) study, for 59 species, the relationship between abundance and probability of occurrence (mean sample size = 55.49, standard deviation = 47.50) was no longer significant when unoccupied sites were excluded (mean sample size = 42.98, standard deviation = 38.43). We found that the mean sample size considering only occupied sites was also significantly lower in their study (Wilcoxon rank sum test, $P < 0.001$) than for 13 species for which this

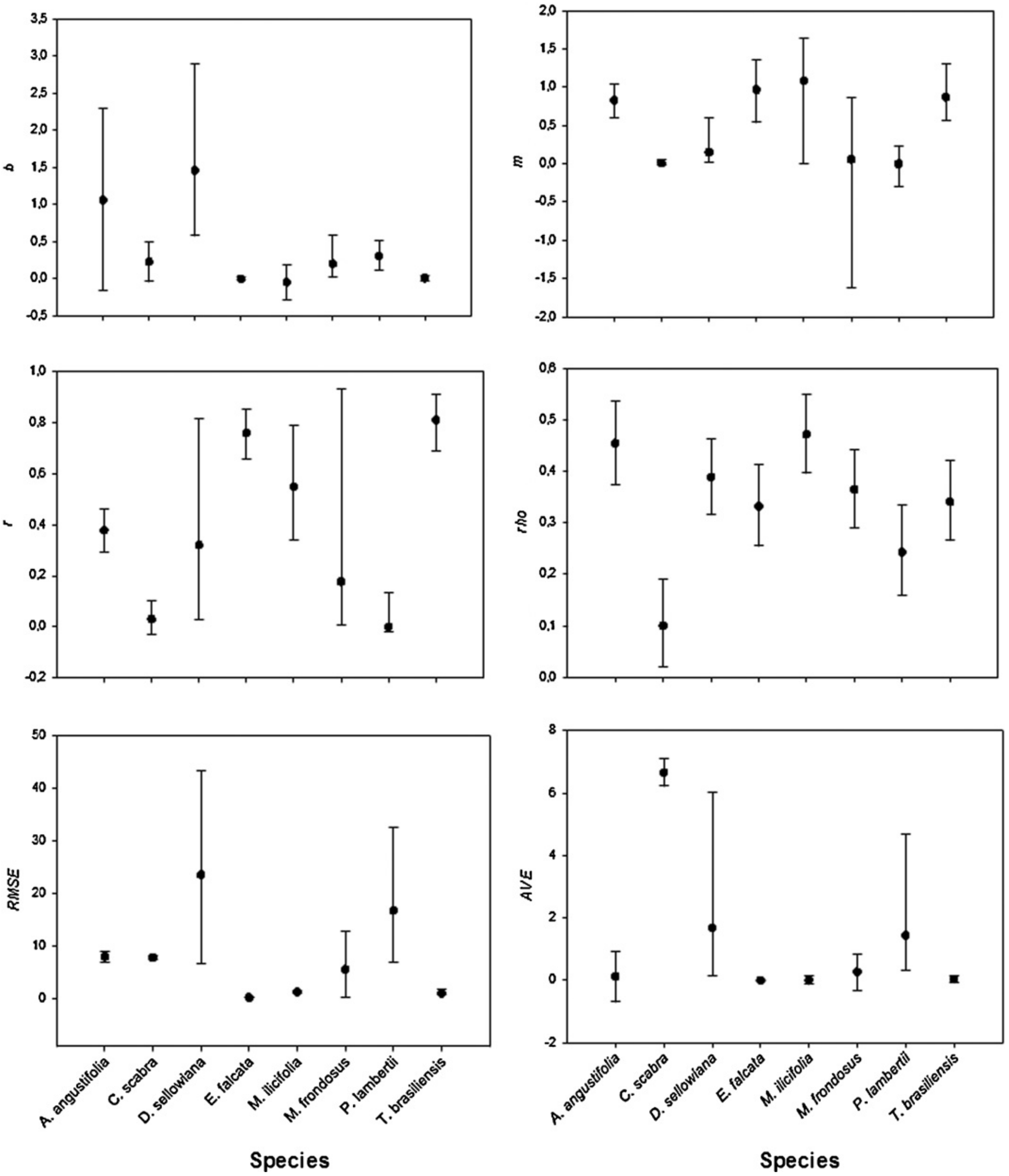


Fig. 3. Abundance evaluation parameters (95% CI). *r*, Pearson correlation coefficient; *rho*, Spearman's rank correlation; *b*, intercept; and *m*, gradient of the fitted line (observed = *b* + *m*(predicted)); RMSE, root mean square error; AVE_{error}, average error.

correlation remained significant (mean sample size = 111.38, standard deviation = 47.31). These results suggest that sample size might be influencing the detection of an occupancy–abundance relationship.

Real et al. (2009) also showed that predictions of GLMs based on the presence–absence data of the Iberian lynx (*Lynx pardinus*) and wild rabbit (*Oryctolagus cuniculus*) in Spain are significantly correlated with independent abundance data. Even when we repeated their analysis excluding unoccupied sites, the correlations remained significant for the abundance of rabbit (Kendall's *Tau-b* = 0.100, *P* = 0.036, *n* = 300) and lynx in 1950 (there were no zero-abundance data for this year), 1965 (*Tau-b* = 0.108, *P* = 0.005, *n* = 355), and 1975 (*Tau-b* = 0.086, *P* = 0.035, *n* = 315). Only in one case (lynx data from 1985) was the correlation no longer significant between their predicted favourability and observed abundance, possibly due to the lower sample size (*n* = 215) when zero-abundance cells were excluded for this year. This provides further support to the idea that, given sufficient sample size, predicted probability of occurrence may be a rough indicator of actual species abundance.

Reconstructing or inferring abundance from species occurrence data can potentially be an important tool for conservation planning efforts and species management, especially given the difficulty of obtaining and analysing abundance data. Current techniques are often not reliable because they generate unstable, low-quality results (spurious estimates; Joseph et al., 2009). The occupancy–abundance relationship may additionally be conditioned by ecological or life-history traits of the modelled species, as we expose in the next section. Further research may provide useful insights into this matter. In the meanwhile, and given that urgent measures are needed to mitigate the impacts of hydroelectric power dams on threatened species, we advocate the use of occurrence models as surrogates for species abundance when the latter cannot be obtained. We defend that, given sufficient sample sizes, probability of occurrence measures can not only be an indicator of model accuracy, as suggested by Albert and Thuiller (2008) and Lobo et al. (2008), but also provide a simple and inexpensive alternative for the abundance measure for most species. This alternative method also has great potential to be used in environmental assessment (see also Araújo and Williams, 2000; Real et al., 2009).

4.2. Ecological constraints on the occupancy–abundance relationship

Jiménez-Valverde et al. (2009) suggest that generalist species and those with a high degree of dispersal show positive and significant relationships between abundance (>0) and the probability of occurrence. Nielsen et al. (2005) recommend an approach based on organism life history to explain the relationship between occurrence and abundance patterns, which is different from the traditional approach based on extensive and exploratory fitting exercises. In our study, *A. angustifolia* and *C. scabra* were the only species that exhibited a positive and significant correlation between probabilities of occurrence and observed abundance when zero-abundance plots were excluded from the analysis. Both species are tied to early successional stages in the studied region (Duarte et al., 2006; Sampaio and Guarino, 2007), and several studies indicate that *A. angustifolia* is able to advance forest expansion over natural grasslands and colonise new areas of grasslands in the highlands of southern Brazil (Duarte et al., 2006). Furthermore, the seeds of *A. angustifolia* are dispersed by birds and small mammals (Anjos, 1991; Iob and Vieira, 2008), whereas *C. scabra*'s seeds are wind-dispersed and can thus travel greater distances.

Another interesting case is that of subtropical palms, *B. eriospatha* and *T. brasiliensis*, which are gregarious species with dense but sporadic populations in the study area. Because of the importance of fibre and the high nutritional value of the fruits, both species were influenced by ancient human inhabitants of the region. The population of these species was purposefully dispersed by natives and European settlers, who grazed livestock for centuries in the area we studied (Reitz et al., 1974). Past human-mediated dispersion can affect the current patterns of plant distribution, creating artificially clustered populations that are often unnoticed in extensive exploratory model-fitting exercises. Although this history of plant manipulation is important, it has rarely been included in studies of species distributions (Lutolf et al., 2009).

The relatively low strength of the correlation between predicted occurrence probabilities and observed abundance (<0.5 in all cases) could also be due to biotic interactions not included in the occurrence models. Austin et al. (1990) and Guisan and Thuiller (2005), for example, suggested that the response curve of a species along an environmental gradient can be seriously constrained by interaction with biotic factors. This hypothesis was recently tested by Heikkinen et al. (2007) and Ritchie et al. (2009), using different organisms, and they both found similar results. When data related to interspecific competitors were incorporated into models, species predictions were significantly improved. In addition, this effect likely plays a role in the relationship between occurrence and abundance, obscuring the true relationship between occupied area and abundance. Ritchie et al. (2009) confirmed this idea demonstrating that the predicted abundance of wallaroos and kangaroos was improved with the addition of the occurrence and abundance of interspecific competitors into the models. However, obtaining these data is difficult and model practitioners must continue to use mostly abiotic factors (Barbosa et al., 2009; Elith and Leathwick, 2009).

4.3. Abundance models

In general, Zero-Inflated models performed better than Hurdle models. This result is contrary to other empirical studies on both real populations (Potts and Elith, 2006) and theoretical studies with pseudo-populations (Miller and Miller, 2008), suggesting that Hurdle models generally perform better than Zero-Inflated models. Because the techniques used in this study to model abundance deal differently with zeros, the results can be interpreted in several ways (Potts and Elith, 2006; Zuur et al., 2009). Two-part modelling techniques, such as GLM Hurdle models, analyse abundance data in the following two steps: (i) zero vs. non-zero values are modelled with a logistic regression (binomial distribution), and (ii) non-zero observations are modelled with a truncated Poisson or negative-binomial regression. Mixed techniques, such as Zero-Inflated models, classify zeros as originating from two different processes, binomial and count processes (Zuur et al., 2009). Interpreting Hurdle models is simpler than interpreting Zero-Inflated models (Potts and Elith, 2006), but, according to Welsh et al. (2000), mixed models provide a better tool when there is overdispersion and a large number of true zeros, as was the case with our data.

4.4. Variables related to species occurrence and abundance

Our main objective with the fitted models was to predict species' distribution and abundance rather than to test the effects of different ecological drivers on species occurrence and abundance. However, we can draw some conclusions about the ecological factors that are associated to the analysed species' distributions. According to Barrows et al. (2005), environmental managers must understand the anthropogenic and environmental factors that influence the occurrence and abundance of species. With this information, managers can employ adaptive management strategies to maintain viable populations of desired species. Overall, occurrence and abundance of studied species are determined, to some extent, by different suites of environmental variables. Regardless of organism type, this pattern has been widely reported (Heinanen et al., 2008; Illán et al., 2010; Truscott et al., 2008), suggesting that biotic and abiotic events associated with plant establishment may be different than those influencing their abundance (Truscott et al., 2008).

Here we highlight the relationship between the abundance of *E. falcata* and total nitrogen availability. Abundance (count model component) of *E. falcata* was negatively related to low soil nitrogen values. This species, along with others of the same genus, are generally able to produce nitrogen-fixing root nodules (Faria et al., 1984, 1989; Schimann et al., 2008). However, its roots have a unique strategy to acquire soil nitrogen. The roots do not exploit available nitrogen in the soil, and instead use nitrogen in the leaf litter (Chesney and Vasquez, 2007; Payan et al., 2009 on *E. poeppigiana* in agroforestry systems), an environmental variable that was analysed in our study. Conversely, *M. frondosus*, another Fabaceae species, was not reported to produce nitrogen-fixing nodules (Allen and Allen, 1981; Faria et al., 1984).

5. Conclusions and future applications

Our study is the first to apply a macroecological approach to support in situ and ex situ plant conservation practices in a region affected by a hydropower dam reservoir. Currently in Brazil, Environmental Impact Assessment (EIA) and Strategic Environmental Assessment (SEA) methods require costly and time-consuming field surveys (floristic and phytosociological studies), often conducted by professionals with an incomplete understanding of the extreme diversity of Brazilian flora. As a result, these surveys often produce low-quality, inapplicable data that are susceptible to environmental error. Despite methodological limitations, i.e., low (although significant) correlation between predicted probability of occurrence and observed abundance, and a possible dependence on ecological and life-history traits, we believe that EIA/SEA projects in

Brazil could profit from using species distribution models to predict future impacts and plan landscape management strategies and mitigation actions. This approach is quicker and less expensive than the current approach, although this method does require staff training in data collection and analysis.

An effective way to organise data collection is to combine the *gradsect* method proposed by Austin and Heyligers (1989) and Wessels et al. (1998) with the rapid assessment survey method used in this study and described by Magnusson et al. (2005). For example, instead of measuring the height and diameter of each tree species within plots, we record only the presence of each species in georeferenced plots. Austin and Heyligers (1989) and Austin and Meyers (1996) support the view that for most environmental goals, including species niche modelling, representative samples are more important than accurate basal area estimations, which is the method currently used in the preparation of EIA/SEA in Brazil. Even in a situation where field surveys are not performed, presence-only data obtained from herbarium or museum collections are viable alternatives. Even with a coarse resolution, they can produce pertinent information for the development of fine-scale species conservation planning (Araújo, 2004; Barbosa et al., 2003, 2010).

Even weakly, the occupancy–abundance relationship can optimise ex situ conservation actions in similar situations. Based on species distribution models built with presence–absence field data obtained in rapid and systematic biodiversity assessments, we can be able to, a priori, define the rare species set using occupancy–abundance relationships (Flather and Sieg, 2006). Regardless of threat levels, we can concentrate germplasm collection efforts, at a first glance, on locally rare species with low abundance and narrow area of occurrence, such as *B. eriospatha* and *T. brasiliensis*. In a second step, species distribution models can be used as a guide to define the most suitable areas to collect germplasm samples of the selected species. This approach is less money and time-consuming and can accelerate conservation actions in landscapes submitted to strong anthropogenic impacts, like hydroelectric power dam reservoirs. Practical examples of this application of distribution models as a guide to define suitable areas to collect species germplasm samples were first described by Jones et al. (1997) and followed by Jarvis et al. (2003, 2005a,b), Villordon et al. (2006) and Ramírez-Villegas et al. (2010). However, in all these previous examples authors used presence-only models to guide germplasm collecting expeditions of crop wild relatives (i.e., peanut, potato, pepper) on a large scale.

The mere occurrence of one species is not sufficient to ensure the persistence of a viable population within a system of protected areas (Barrows et al., 2005). A system of protected areas should contain samples of the largest possible number of local ecosystems, thereby maintaining the ecological processes of these ecosystems (Australia, 1997). An alternative to accomplish this goal would be to integrate information about species occurrence (or abundance) and regional land use in an interactive decision-making system. This method would ensure the conservation of viable samples in different ecosystems occurring in the region of interest (Ferrier et al., 2002).

Acknowledgments

The field work was supported by Embrapa Recursos Genéticos e Biotecnologia. E.S.G.G. was supported by Brazilian National Council for Scientific and Technological Development (CNPq; proc. 140712/2007-0) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES; proc. 2298/09-0) scholarships. A.M.B. is supported by a Post-Doctoral Fellowship (SFRH/BPD/40387/2007) from Fundação para a Ciência e a Tecnologia (Portugal). The "Rui Nabeiro" Biodiversity Chair is funded by Delta Cafés and an FCT project (PTDC/AAC-AMB/098163/2008). We are grateful to A.A. Santos, M.B. Sampaio, J.A. Santos and others who helped with field work. M.B. Araújo, D.L.M. Vieira, A.B. Sampaio, F.S. Rocha, P.R. Vieira and R.L.L. Orihuela helped with important suggestions. We are also grateful to C. Scherber who kindly provided the script for stepwise model selection with AIC_c.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ecolmodel.2012.01.007.

References

Albert, C.H., Thuiller, W., 2008. Favourability functions versus probability of presence: advantages and misuses. Ecography 31, 417–422.

Allen, O.N., Allen, E.K., 1981. The Leguminosae: A Source Book of Characteristics, Uses, and Nodulation, 1st ed. The University of Wisconsin Press, Madison.

Alves, T.P., Fontoura, N.F., 2009. Statistical distribution models for migratory fish in Jacui basin, South Brazil. Neotrop. Ichthyol. 7, 647–658.

Anjos, L., 1991. O ciclo anual de *Cyanocorax caeruleus* em floresta de araucária (Passeriformes: Corvidae). Ararajuba 2, 19–23.

Araújo, M.B., 2004. Matching species with reserves – uncertainties from using data at different resolutions. Biol. Conserv. 118, 533–538.

Araújo, M.B., Williams, P.H., 2000. Selecting areas for species persistence using occurrence data. Biol. Conserv. 96, 331–345.

Austin, M.P., Heyligers, P.C., 1989. Measurement of the realized qualitative niche: environmental niches of five *Eucalyptus* species. Biol. Conserv. 50, 13–32.

Austin, M.P., Meyers, J.A., 1996. Current approaches to modelling the environmental niche of eucalypts: implication for management of forest biodiversity. Forest Ecol. Manag. 85, 95–106.

Austin, M.P., Nicholls, A.O., Margules, C.R., 1990. Measurement of the realized qualitative niche: environmental niches of five *Eucalyptus* species. Ecol. Monogr. 60, 161–177.

Australia Co., 1997 Nationally Agreed Criteria for the Establishment of a Comprehensive, Adequate and Representative Reserve System for Forest in Australia. A Report by the Joint ANZECC/MCFFA National Forest Policy Statement Implementation Sub-committee.

Barbosa, A.M., Real, R., Olivero, J., Vargas, J.M., 2003. Otter (*Lutra lutra*) distribution modelling at two resolution scales suited to conservation planning in the Iberian Peninsula. Biol. Conserv. 114, 377–387.

Barbosa, A.M., Real, R., Vargas, J.M., 2009. Transferability of environmental favourability models in geographic space: the case of the Iberian desman (*Galemys pyrenaicus*) in Portugal and Spain. Ecol. Model. 220, 747–754.

Barbosa, A.M., Real, R., Vargas, J.M., 2010. Use of coarse-resolution models of species' distributions to guide local conservation inferences. Conserv. Biol. 24, 1378–1387.

Barrows, C.W., Swartz, M.B., Hodges, W.L., Allen, M.F., Rotenberry, J.T., Li, B.L., Scott, T.A., Chen, X.W., 2005. A framework for monitoring multiple-species conservation plans. J. Wildlife Manage. 69, 1333–1345.

Blackburn, T.M., Gaston, K.J., 2009. Sometimes the obvious answer is the right one: a response to 'Missing the rarest: is the positive interspecific abundance-distribution relationship a truly general macroecological pattern?'. Biol. Lett. 5, 777–778.

Brown, J.H., 1984. On the relationship between abundance and distribution of species. Am. Nat. 124, 255–279.

Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference: A Practical Information–Theoretic Approach. Springer, New York.

Cameron, A.C., Trivedi, P.K., 1998. Regression Analysis of Count Data. Cambridge, New York, p. 432.

Cayuela, L., Golicher, D.J., Newton, A.C., Kolb, M., Albuquerque, F.Sd., Arets, E.J.M.M., Alkemade, J.R.M., Perez, A.M., 2009. Species distribution in the tropics: problems, potentialities, and the role of biological data for effective species conservation. Trop. Conserv. Sc. 2, 319–352.

Chaplot, V., Darboux, F., Bourennane, H., Leguedois, S., Silvera, N., Phachomphon, K., 2006. Accuracy of interpolation techniques for the derivation of digital elevation models in relation to landform types and data density. Geomorphology 77, 126–141.

Chen, G., Kéry, M., Zhang, J., Ma, K., 2009. Factors affecting detection probability in plant distribution studies. J. Ecol. 97, 1383–1389.

Chesney, P., Vasquez, N., 2007. Dynamics of non-structural carbohydrate reserves in pruned *Erythrina poeppigiana* and *Gliricidia sepium* trees. Agroforest. Syst. 69, 89–105.

Duarte, L.D.S., Machado, R.E., Hartz, S.M., Pillar, V.D., 2006. What saplings can tell us about forest expansion over natural grasslands. J. Veg. Sci. 17, 799–808.

Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S., Zimmermann, N.E., 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29, 129–151.

Elith, J., Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. Annu. Rev. Ecol. Evol. Syst. 40, 677–697.

Faria, B.S.M., Franco, A.A., Jesus, R.M., Menandro, M.S., Baitello, J.B., Mucci, E.S.F., Döbereiner, J., Sprent, J.I., 1984. New nodulating legume trees from south-east Brazil. New Phytol. 98, 317–328.

Faria, B.S.M., Lewis, G.P., Sprent, J.I., Sutherland, J.M., 1989. Occurrence of nodulation in the Leguminosae. New Phytol. 111, 607–619.

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recogn. Lett. 27, 861–874.

Ferrier, S., Watson, G., Pearce, J., Drielsma, M., 2002. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Species-level modelling. Biodivers. Conserv. 11, 2275–2307.

Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ. Conserv. 24, 38–49.

Figueiredo, M.S.L., Grelle, C.E.V., 2009. Predicting global abundance of a threatened species from its occurrence: implications for conservation planning. Divers. Distrib. 15, 117–121.

Flather, C.H., Sieg, C.H., 2006. Species rarity: definition, causes, and classification. In: Raphael, M.G., Molina, R. (Eds.), Conservation of Rare or Little-known Species: Biological, Social, and Economic Considerations. Island Press, Washington.

Fortin, M.J., Dale, M., 2005. Spatial Analysis: A Guide for Ecologists, 1st ed. Cambridge University Press, Cambridge.

Gaston, K.J., Blackburn, T.M., Greenwood, J.J.D., Gregory, R.D., Quinn, R.M., Lawton, J.H., 2000. Abundance–occupancy relationships. J. Appl. Ecol. 37, 39–59.

Gray, B.R., 2005. Selecting a distributional assumption for modelling relative densities of benthic macroinvertebrates. Ecol. Model., 1–12.

Gu, W., Swihart, R.K., 2004. Absent or undetected? Effects of non-detection of species occurrence on wildlife-habitat models. Biol. Conserv. 116, 195–203.

Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. Ecol. Lett. 8, 993–1009.

Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecol. Model. 135, 147–186.

He, F., Gaston, K.J., 2000. Occupancy–abundance relationships and sampling scale. Ecography 23, 503–511.

He, F., Gaston, K.J., 2007. Estimating abundance from occurrence: an underdetermined problem. Am. Nat. 170, 655–659.

Heikkinen, R.K., Luoto, M., Virkkala, R., Pearson, R.G., Korber, J.H., 2007. Biotic interactions improve prediction of boreal bird distributions at macro-scales. Global Ecol. Biogeogr. 16, 754–763.

Heinanen, S., Ronka, M., von Numers, M., 2008. Modelling the occurrence and abundance of a colonial species, the arctic tern Sterna paradisaea in the archipelago of SW Finland. Ecography 31, 601–611.

Holt, A.R., Gaston, K.J., He, F., 2002. Occupancy-abundance relationships and spatial distribution: a review. Basic Appl. Ecol. 3, 1–13.

Illán, J.G., Gutierrez, D., Wilson, R.J., 2010. The contributions of topoclimate and land cover to species distributions and abundance: fine-resolution tests for a mountain butterfly fauna. Global Ecol. Biogeogr. 19, 159–173.

Iob, G., Vieira, E.M., 2008. Seed predation of *Araucaria angustifolia* (Araucariaceae) in the Brazilian Araucaria Forest: influence of deposition site and comparative role of small and 'large' mammals. Plant Ecol. 198, 185–196.

Jarvis, A., Ferguson, M.E., Williams, D.E., Guarino, L., Jones, P.G., Stalker, H.T., Valls, J.F.M., Pittman, R.N., Simpson, C.E., Bramel, P., 2003. Biogeography of wild *Arachis*: assessing conservation status and setting future priorities. Crop Sci. 43, 1100–1108.

Jarvis, A., Williams, K., Williams, D., Guarino, L., Caballero, P.J., Mottram, G., 2005a. Use of GIS for optimizing a collecting mission for rare wild pepper (*Capsicum flexuosum* Sendtn.) in Paraguay. Genet. Resour. Crop Eval. 52, 671–682.

Jarvis, A., Yeaman, S., Guarino, L., Tohme, J., 2005b. The role of geographic analysis in locating, understanding, and using plant genetic diversity. Methods Enzymol. 395, 279–296.

Jiménez-Valverde, A., Diniz, F., de Azevedo, E.B., Borges, P.A.V., 2009. Species distribution models do not account for abundance: the case of arthropods on Terceira Island. Ann. Zool. Fennici 46, 451–464.

Jiménez-Valverde, A., Ortuno, V.M., Lobo, J.M., 2007. Exploring the distribution of *Sterocorax ortuno*, 1990 (Coleoptera, Carabidae) species in the Iberian peninsula. J. Biogeogr. 34, 1426–1438.

Joly, C.A., Aidar, M.P.M., Klink, C.A., McGrath, D.G., Moreira, A.G., Moutinho, P., Nepstad, D.C., Oliveira, A.A., Pott, A., Rodal, M.J.N., Sampaio, E.V.S.B., 1999. Evolution of the Brazilian phytogeography classification systems: implications for biodiversity conservation. Cienc. Cult. 51, 331–348.

Jones, P.G., Beebe, S.E., Tohme, J., Galwey, N.W., 1997. The use of geographical information systems in biodiversity exploration and conservation. Biodivers. Conserv. 6, 947–958.

Joseph, L.N., Elkin, C., Martin, T.G., Possingham, H.P., 2009. Modeling abundance using N-mixture models: the importance of considering ecological mechanisms. Ecol. Appl. 19, 631–642.

Kirkpatrick, J.B., Fensham, R.J., Nunez, M., Bowman, D.M.J.S., 1988. Vegetation–radiation relationships in the wet–dry tropics: granite hills in northern Australia. Vegetatio 76, 103–112.

Kirkpatrick, J.B., Nunez, M., 1980. Vegetation–radiation relationships in mountainous terrain: eucalypt-dominated vegetation in the Risdon Hills, Tasmania. J. Biogeogr. 7, 197–208.

Klein, R.M., 1975. Southern Brazilian phytogeographic features and the probable influence of upper quaternary climatic changes in the floristic distribution. Bol. Parana. Geocienc. 33, 67–88.

Komonen, A., Paivinen, J., Kotiaho, J.S., 2009. Missing the rarest: is the positive interspecific abundance–distribution relationship a truly general macroecological pattern. Biol. Lett. 5, 492–494.

Kotiaho, J.S., Komonen, A., Paivinen, J., 2009. On the obvious positive interspecific relationship between abundance and distribution: a reply to Blackburn and Gaston. Biol. Lett. 5, 779–780.

Legendre, P., Legendre, L., 1998. Numerical Ecology, 2nd ed. Elsevier Science B.V., Amsterdam.

Lindén, A., Mäntyniemi, S., 2011. Using the negative binomial distribution to model overdispersion in ecological data. Ecology, 1414–1421.

Lobo, J.M., Jiménez-Valverde, A., Real, R., 2008. AUC: a misleading measure of the performance of predictive distribution models. Global Ecol. Biogeogr. 17, 145–151.

Lutolf, M., Guisan, A., Kienast, F., 2009. History matters relating land-use change to butterfly species occurrence. Environ. Manage. 43, 436–446.

Magnusson, W.E., Lima, A.P., Luizão, R., Costa, F.R.C., Castilho, C.V., Knupp, V.F., 2005. RAPELD, a modification of the Gentry method for biodiversity surveys in long-term ecological research sites. Biot. Neotrop. 5, 19–24.

Maluf, J.R.T., 2000. Nova classificação climática do Estado do Rio Grande do Sul. Rev. Bras. Agrometeorol. 8, 141–150.

Manel, S., Williams, H.C., Ormerod, S.J., 2001. Evaluating presence–absence models in ecology: the need to account for prevalence. J. Appl. Ecol. 38, 921–931.

Mason, S.J., Graham, N.E., 2002. Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: statistical significance and interpretation. Quart. J. Roy. Meteorol. Soc. 128, 2145–2166.

McCarthy, M.A., Elith, J., 2002. Species mapping for conservation. Gap Anal. Bull. 2, 50–57.

Miller, J.M., Miller, M.D., 2008. No zero left behind: comparing the fit for zero-inflation models as a function of skew and proportion of zeros. InterStat 11, 35.

Mitasova, H., Hofierka, J., 1993. Interpolation by regularized spline with tension: II. Application to terrain modeling and surface geometry analysis. Math. Geol. 25, 656–669.

Mitasova, H., Mitas, L., 1993. Interpolation by regularized spline with tension: I. Theory and implementation. Math. Geol. 25, 641–655.

Neteler, M., Mitasova, H., 2008. Open Source GIS: A GRASS GIS Approach, 3rd ed. Springer, New York.

Nielsen, S.E., Johnson, C.J., Heard, D.C., Boyce, M.S., 2005. Can models of presence–absence be used to scale abundance? Two case studies considering extremes in life history. Ecography 28, 197–208.

Parolo, G., Rossi, G., Ferrarini, A., 2008. Toward improved species niche modelling: *Arnica montana* in the Alps as a case study. J. Appl. Ecol. 45, 1410–1418.

Payan, F., Jones, D.L., Beer, J., Harmand, J.M., 2009. Soil characteristics below *Erythrina poeppigiana* in organic and conventional Costa Rican coffee plantations. Agroforest. Syst. 76, 81–93.

Pearce, J., Ferrier, S., 2001. The practical value of modelling relative abundance of species for regional conservation planning: a case study. Biol. Conserv. 98, 33–43.

Penížek, V., Borůvka, L., 2006. Soil depth prediction supported by primary terrain attributes: a comparison of methods. Plant Soil Environ. 52, 424–430.

Potter, R.O., Carvalho, A.M., Flores, C.A., Bognola, I., 2004. Solos do Estado de Santa Catarina. Bol. Pesquisa, 46. Embrapa Solos.

Potts, J.M., Elith, J., 2006. Comparing species abundance models. Ecol. Model. 1999, 153–163.

Pulliam, H.R., 2000. On the relationship between niche and distribution. Ecol. Lett., 349–361.

Ramírez-Villegas, J., Khoury, C., Jarvis, A., Debouck, D.G., Guarino, L., 2010. A gap analysis methodology for collecting crop genepools: a case study with *Phaseolus* beans. PLoS One 10, 1–18.

Rangel, T.F., Diniz, J.A.F., Bini, L.M., 2010. SAM: a comprehensive application for spatial analysis in macroecology. Ecography 33, 46–50.

Real, R., Barbosa, A.M., Rodriguez, A., Garcia, F.J., Vargas, J.M., Palomo, L.J., Delibes, M., 2009. Conservation biogeography of ecologically interacting species: the case of the Iberian lynx and the European rabbit. Divers. Distrib. 15, 390–400.

Reitz, R., 1974. Palmeiras. In: Itajaí Herbário Barbosa Rodrigues.

Ridout, M., Demétrio, C.G.B., Hinde, J., 1998. Models for count data with many zeros. In: International Biometric Conference, Cape Town, pp. 1–13.

Ritchie, E.G., Martin, J.K., Johnson, C.N., Fox, B.J., 2009. Separating the influences of environment and species interactions on patterns of distribution and abundance: competition between large herbivores. J. Anim. Ecol. 78, 724–731.

Roberts, D.W., 1986. Ordination on the basis of fuzzy set theory. Plant Ecol. 66, 123–131.

Rodríguez, J.P., Brotons, L., Bustamante, J., Seoane, J., 2007. The application of predictive modelling of species distribution to biodiversity conservation. Divers. Distrib. 13, 243–251.

Sampaio, M.B., Guarino, E.S.G., 2007. Efeitos do pastoreio na estrutura populacional de plantas em fragmentos de Floresta Ombrófila Mista. Rev. Árvore 31, 1035–1046.

Schimann, H., Ponton, S., Hattenschwiler, S., Ferry, B., Lensi, R., Domenach, A.M., Roggy, J.C., 2008. Differing nitrogen use strategies of two tropical rainforest late successional tree species in French Guiana: evidence from N-15 natural abundance and microbial activities. Soil Biol. Biochem. 40, 487–494.

Skagen, S.K., Yackel Adams, A.A., 2011. Potential misuse of avian density as a conservation metric. Conserv. Biol. 25, 48–55.

Soberón, J., Peterson, A.T., 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. Biodivers. Inform. 2, 1–10.

Sørensen, R., Zinko, U., Seibert, J., 2005. On the calculation of the topographic wetness index: evaluation of different methods based on field observations. Hydrol. Earth Syst. Sci. 2, 1807–1834.

Tomczak, M., 1998. Spatial interpolation and its uncertainty using automated anisotropic inverse distance weighting (IDW) – cross-validation/Jackknife approach. J. Geogr. Inform. Decis. Anal. 2, 18–30.

Truscott, A.M., Palmer, S.C.F., Soulsby, C., Hulme, P.E., 2008. Assessing the vulnerability of riparian vegetation to invasion by *Mimulus guttatus*: relative importance of biotic and abiotic variables in determining species occurrence and abundance. Divers. Distrib. 14, 412–421.

Tsai, C., Chen, Z., Duh, C., Horng, F., 2001. Prediction of soil depth using a soil-landscape regression model: a case study on forest soils in Southern Taiwan. Proc. Natl. Sci. Counc. ROC [B] 25, 34–39.

VanDerWal, J., Shoo, L.P., Johnson, C.N., Williams, S.E., 2009. Abundance and the environmental niche: environmental suitability estimated from niche models predicts the upper limit of local abundance. Am. Nat. 174, 282–291.

Van Horne, B., 1983. Density as a misleading indicator of habitat quality. J. Wildlife Manage., 893–901.

Ver Hoef, J.M., Boveng, P.L., 2007. Quasi-Poisson vs. negative-binomial regression: how should we model overdispersed count data. Ecology 88, 2766–2772.

Verbek, W.C.E.P., Van der Velde, G., Esselink, H., 2010. Explaining abundance–occupancy relationships in specialists and generalists: a case study on aquatic macroinvertebrates in standing waters. J. Anim. Ecol. 79, 589–601.

Villordon, A., Njuguna, W., Gichuki, S., Ndolo, P., Kulembeka, H., Jeremiah, S.C., LaBonte, D., Yada, B., Tukamuhabwa, P., Mwanga, R.O.M., 2006. Using GIS-based tools and distribution modeling to determine sweetpotato documentation priorities in Sub-Saharan Africa. Hortscience 41, 1377–1381.

Welsh, A.H., Cunningham, R.B., Chambers, R.L., 2000. Methodology for estimating the abundance of rare animals: seabird nesting on North East Herald Cay. Biometrics 56, 22–30.

Welsh, A.H., Cunningham, R.B., Donnelly, C.F., Lindenmayer, D.B., 1996. Modelling the abundance of rare species: statistical models for counts with extra zeros. Ecol. Model. 88, 297–308.

Wessels, K.J., Van Jaarsveld, A.S., Grimbeek, J.D., Van der Linde, M.J., 1998. An evaluation of the gradsect biological survey method. Biodivers. Conserv. 7, 1093–1121.

Willis, S.G., Hole, D.G., Collingham, Y.C., Hilton, G., Rahbek, C., Huntley, B., 2009. Assessing the impacts of future climate change on protected area networks: a method to simulate individual species' responses. Environ. Manage. 43, 836–845.

Zar, J.H., 1999. Biostatistical Analysis, 4th ed. Prentice Hall.

Zeileis, A., Kleiber, C., Jackman, S., 2008. Regression models for count data in R. J. Stat. Softw. 27, 1–25.

Zhu, L., Sun, O.J., Sang, W.G., Li, Z.Y., Ma, K.P., 2007. Predicting the spatial distribution of an invasive plant species (*Eupatorium adenophorum*) in China. Landscape Ecol. 22, 1143–1154.

Zuur, A.F., Ieno, E.N., Walker, N., Saveliev, A.A., Smith, G.M., 2009. Mixed Effects Models and Extensions in Ecology with R, 1st ed. Springer.

Web references

ANEEL, 2008. Atlas de energia elétrica do Brasil. Brazilian Electricity Regulatory Agency, Brasília, Available from: http://www.aneel.gov.br/biblioteca/EdicaoLivros2009atlas.cfm (accessed September 2009).

ASTER GDEM, ASTER Global Digital Elevation Model. Available from: http://www.gdem.aster.ersdac.or.jp (accessed September 2009).

CITES, CITES Convention on International Trade in Endangered Species of Wild Fauna and Flora. Available from: http://www.cites.org (accessed November 2008).

GRASS Development Team, 2010. Geographic Resources Analysis Support System (GRASS). Open Source Geospatial Foundation, Vancouver, British Columbia, Available from: http://grass.osgeo.org (accessed July 2010).

IUCN, International Union for Conservation of Nature, International Red List. Available from: www.iucnredlist.org (accessed November 2008).

Quantum GIS Development Team, 2009. Quantum GIS. Geospatial Foundation, Vancouver, British Columbia, Available from: http://qgis.osgeo.org (accessed September 2009).

R Development Core Team, 2009. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, Available from: http://www.R-project.org (accessed November 2009).

Rio Grande do Sul State list of threatened species, Rio Grande do Sul State law n° 41.672, June 11, 2002. Available from: http://www.al.rs.gov.br (accessed September 2009).