A Journal of Conservation Biogeography

Diversity and Distributions

BIODIVERSITY RESEARCH

Modelling invasive alien species distributions from digital biodiversity atlases. Model upscaling as a means of reconciling data at different scales

Arnald Marcer[1*], Joan Pino[1], Xavier Pons[2] and Lluís Brotons[1,3]

[1]CREAF (Centre de Recerca Ecològica i Aplicacions Forestals), Universitat Autònoma de Barcelona, E-08193, Bellaterra, Catalonia, Spain, [2]Departament de Geografia, Universitat Autònoma de Barcelona, E-08193, Bellaterra, Catalonia, Spain, [3]Grup d'Ecologia del Paisatge, Àrea de Biodiversitat, Centre Tecnològic Forestal de Catalunya, E-25280, Solsona, Catalonia, Spain

ABSTRACT

Aim There is a wealth of information on species occurrences in biodiversity data banks, albeit presence-only, biased and scarce at fine resolutions. Moreover, fine-resolution species maps are required in biodiversity conservation. New techniques for dealing with this kind of data have been reported to perform well. These fine-resolution maps would be more robust if they could explain data at coarser resolutions at which species distributions are well represented. We present a new methodology for testing this hypothesis and apply it to invasive alien species (IAS).

Location Catalonia, Spain.

Methods We used species presence records from the Biodiversity data bank of Catalonia to model the distribution of ten IAS which, according to some recent studies, achieve their maximum distribution in the study area. To overcome problems inherent with the data, we prepared different correction treatments: three for dealing with bias and five for autocorrelation. We used the MaxEnt algorithm to generate models at 1-km resolution for each species and treatment. Acceptable models were upscaled to 10 km and validated against independent 10 km occurrence data.

Results Of a total of 150 models, 20 gave acceptable results at 1-km resolution and 12 passed the cross-scale validation test. No apparent pattern emerged, which could serve as a guide on modelling. Only four species gave models that also explained the distribution at the coarser scale.

Main conclusions Although some techniques may apparently deliver good distribution maps for species with scarce and biased data, they need to be taken with caution. When good independent data at a coarser scale are available, cross-scale validation can help to produce more reliable and robust maps. When no independent data are available for validation, however, new data gathering field surveys may be the only option if reliable fine-scale resolution maps are needed.

*Correspondence: Arnald Marcer, CREAF, Centre for Ecological Research and Forestry Applications, Autonomous University of Barcelona, Bellaterra E–08193, Spain. E-mail: arnald.marcer@uab.cat

INTRODUCTION

For centuries, species occurrences have been recorded in an *ad hoc* way by natural historians, museums, scientists and the like in the form of museum specimens, site inventories, citations in technical and scientific literature, etc. (Chapman & Busby, 1994; Chapman, 2005). In the last three decades, both governments and non-governmental organizations have invested considerable financial resources on the digitizing of these data into digital species distribution atlases and making them publicly available. Ideally, they should offer reliable high-quality digital data, which withstand public, scientific

and legal scrutiny (Robertson *et al.*, 2010). Such databases represent a wealth of information on species distribution and an indispensable asset for science and conservation (Funk & Richardson, 2002; Graham *et al.*, 2004; Suarez & Tsutsui, 2004; Franklin, 2009; Robertson *et al.*, 2010). However, as these data usually come from opportunistic or *ad hoc* sources rather than well-planned surveys, they present some important drawbacks: they are presence-only in nature, are highly biased (not representing a random sample) and may show spatial aggregation derived from sampling biases. Also, because they are laboriously compiled from analogue sources, they are difficult and costly to georeference, and hence, coarser resolutions tend to dominate (Margules & Pressey, 2000; Pressey, 2004). As a result, most data in these biodiversity databases are often too coarse for use in conservation planning and management, where fine-grained maps (i.e. 1 km or better) are needed (McPherson *et al.*, 2006; Guralnick & Hill, 2009; Niamir *et al.*, 2011).

Planned systematic surveys of species presence, absence and abundance could provide the most precise, accurate and unbiased information on the spatial distribution of biodiversity. However, such surveys are expensive to conduct for large regions, even for a single species (Robertson *et al.*, 2010). Given the current accelerated trend in world-wide biodiversity loss and the urge for addressing conservation problems, it becomes of utmost importance to find ways and methodologies to make the best use of this existing information (Newbold, 2010; Venette *et al.*, 2010).

Such databases or atlases of species occurrences still represent a largely untapped potential of information that can play an important role in conservation biogeography (Elith & Leathwick, 2009; Richardson & Whittaker, 2010; Elith *et al.*, 2011). Methods for dealing with biases and evaluating results need more development. Fortunately, in the last few years, the community of species distribution modelling has witnessed the appearance of new tools and methodologies from the fields of statistics and artificial intelligence which have the potential to address the problems inherent in these data. Some of these tools, if used properly (Segurado *et al.*, 2006; Pearson *et al.*, 2007; Phillips *et al.*, 2008; Elith *et al.*, 2011), can handle presence-only data even when only few occurrences are available, and provided attention is given to problems such as bias and autocorrelation (Elith *et al.*, 2011). Despite that, in presence-only models, the lack of absence data compromises the model reliability (Aranda & Lobo, 2011) and that, some issues have been raised concerning the maximum entropy technique [e.g. (Peterson *et al.*, 2007, 2008)], it has been judged among the best performers in distribution modelling for such kind of data (Elith *et al.*, 2006) and is still widely used by the research community.

In many cases, applying these novel techniques to the scarce fine-resolution data can yield distribution maps with high validation scores. The question is whether we can take advantage of coarser data, which are relatively abundant and for some regions close to the species true distribution, to further validate these maps and assess their reliability.

Having models that explain data at fine resolutions while being consistent with coarser resolution data is important (McPherson *et al.*, 2006; Vallecillo *et al.*, 2009; Niamir *et al.*, 2011) as it can yield more robust and reliable distribution maps for conservation. The existence of databases with such information provides an opportunity to test this hypothesis. A specially relevant piece of information for conservation, which can potentially be derived from these costly data banks, is current and historical distributions of invasive alien species (IAS).

The spread of IAS, driven mainly by human activities, is increasing worldwide (Butchart *et al.*, 2010) and poses potential problems not only to native biodiversity but also to economic development and human well-being (Vitousek *et al.*, 1997; Taylor & Irwin, 2004; Pimentel *et al.*, 2005; Chytrý *et al.*, 2009; Pejchar & Mooney, 2009; Pyšek *et al.*, 2010; Vilà *et al.*, 2011). Having risk maps representing the potential distribution of IAS is a necessary step towards effective management (Richardson & Whittaker, 2010; Jiménez-Valverde *et al.*, 2011). Using all information in species inventories and atlases, coarse and fine-resolution records, is essential for maximizing limited financial resources (Nielsen *et al.*, 2008).

We test whether fine-resolution maps of IAS can be obtained from existing biodiversity databases that are robust enough to explain occurrences at different scales held in the database. By using a probabilistic upscaling methodology, we translate 1-km resolution maps into 10-km resolution maps that are then used to further validate the models against well-known distributions at the coarser scale. With this approach, each model is evaluated with data at its own resolution. We use different bias and autocorrelation treatments to deal with problems inherent in the finer resolution data. Results suggest that, despite current modelling methodologies can deliver good fine-resolution models, these models cannot always explain well-known distributions at coarser scales.

METHODS

Area of study

Catalonia is a region of around 32,000 km^2 located in the northeastern part of the Iberian Peninsula (Fig. 1). It ranges in elevation from 0 to over 3000 m.a.s.l, from the Mediterranean coast up to the Pyrenees. Its environmental conditions are highly variable because of its location and complex topography. Although dominated by the Mediterranean climate, it also has continental and Atlantic influences. Mediterranean and Eurosiberian biogeographic regions dominate, while Subalpine and Alpine types appear in the upper zones of the Pyrenees. There is a trend of decreasing precipitation and increasing temperature towards the south (Ninyerola *et al.*, 2000). It is a highly populated region, particularly around the Barcelona metropolitan area. The rest of the region is dominated by forests and agroforestry mosaics with
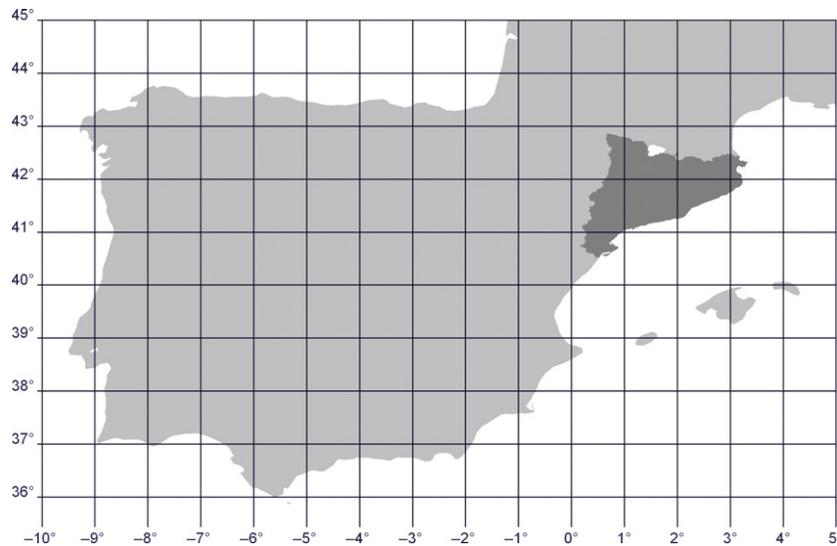
Figure 1 Study area.

relatively large human influence although with lower intensity.

Database

As a case study, we used the Biodiversity Databank of Catalonia (BDBC, http://biodiver.bio.ub.es/biocat/) (Font *et al.*, 2009). Catalonia is a region with a rich history in Botany which is reflected in the holdings of the database. For an area of around 32,000 km^2, the BDBC contains more than 1.5 million plant species citations from sources such as scientific articles, PhD theses and local floras. Most of its data are at a coarse resolution of 10 km for historical reasons – the recording tradition among botanists and the fact that, before the use of GPS, manually recording occurrences at finer resolutions was a time-consuming process. Catalonia is a well-surveyed region for vascular plants at 10-km resolution. The BDBC contains also more than 180,000 plant occurrence records at 1-km resolution. The geographical distribution of IAS as reflected by 10-km resolution occurrence records in the BDBC can be considered to approximate its true distribution for our study area (Pino *et al.*, 2005). Therefore, when developing models at finer resolutions, their geographic distribution should be coherent with the distribution obtained from mapping the coarser data.

Species data

IAS may violate the assumption made in species distribution modelling that species have had the opportunity to realize their full potential range (Austin, 2002; Araújo & Pearson, 2005; Peterson, 2005; De Marco *et al.*, 2008; Zimmermann *et al.*, 2010; Vàclavík & Meentemeyer, 2012). According to some studies (Williamson *et al.*, 2009; Gassó *et al.*, 2010), neophytes reach their maximum range around 150 years after their introduction into the Iberian Peninsula. The 150-

year period was hence taken as the main criterion for species selection, together with the availability of data at 1-km resolution. The 10 species selected are presented in Table 1. Four of them are considered to be highly invasive in Europe (DAISIE European Invasive Alien Species Gateway, 2011): *Ailanthus altissima*, *Opuntia ficus-indica*, *Oxalis pes-caprae* and *Robinia pseudoacacia*. See Appendix S1 in Supporting Information for maps of occurrences for each species.

Data independence across scales

Occurrence data at different resolutions in biodiversity atlases may not be independent; that is, occurrence records at coarser resolutions may have their origin in records at finer resolutions. To overcome this difficulty, we only

Table 1 List of invasive alien species selected for modelling

Species	Abbr	1 km	10 km	Intr	Yrs
Agave Americana L.	aga	20	124	XVIth	411
Ailanthus altissima (P.Mill) Swingle	aia	43	213	1818	192
Amaranthus albus L.	ama	29	194	1861	149
Conyza Canadensis (L.) Cronquist	coc	73	307	1784	226
Datura stramonium L.	das	31	230	XVIth	411
Oenothera biennis L.	oeb	55	80	1848	162
Opuntia ficus-indica (L.) Mill.	opf	13	102	XVIth	411
Oxalis pes-caprae L.	oxp	12	41	1850	160
Robinia pseudoacacia L.	rop	66	257	XVIIIth	211
Xanthium spinosum L.	xas	56	252	XVIIIth	211

Abbr, species abbreviation; 1 km, number of 1 km occurrences; 10 km, number of 10 km occurrences; Intr, introduction date; Yrs, number of years since introduction (conservative estimate).

accepted 10 km squares which had at least one citation more per species than the sum of citations for the one hundred 1 km squares contained within the 10 km square; that is, there is at least one 10 km occurrence record which is independent from 1 km data. This procedure also allowed us to use all occurrences records at 1-km resolution.

Environmental data

We used 19 bioclimatic variables (Nix, 1986) (Table 2) that represent a combination of annual trends, seasonality and extreme conditions relevant to species physiological tolerances. We added two more variables regarding radiation (mean radiation of the least radiated quarter and mean radiation of the most radiated quarter) and three more variables that may partially explain the distribution of IAS (distance to main harbours, distance to the coast and degree of anthropization) (Brooks, 2007; Vicente *et al.*, 2010) (See Table 2). We calculated the bioclimatic variables using the Digital Climatic Atlas of Catalonia (DCAC) (Ninyerola *et al.*, 2000) that holds monthly data on temperature, precipitation and radiation for the whole of Catalonia. We calculated the

Table 2 Set of environmental predictors used in modelling

Bioclimatic variables
  Annual mean temperature
  Minimum temperature of the coldest month
  Mean temperature of the coldest year quarter
  Mean temperature of the warmest year quarter
  Mean temperature of the wettest year quarter
  Mean temperature of the driest year quarter
  Maximum temperature of the warmest month
  Annual mean precipitation
  Precipitation of the coldest year quarter
  Precipitation of the driest month
  Precipitation of the driest year quarter
  Precipitation of the warmest year quarter
  Precipitation of the wettest month
  Precipitation of the wettest year quarter
  Annual temperature range
  Mean temperature diurnal range
  Isothermality
  Temperature seasonality
  Precipitation seasonality
  Mean solar radiation of the least radiated quarter
  Mean solar radiation of the most radiated quarter
Landscape and physical variables
  Anthropization degree
    1 – Natural forests, shrublands, wetlands, grasslands, rock
      outcrops and screens, bare soil, beaches, glaciers and snow
      cover and continental waters, 2 – recently burnt areas
      and reforestations, 3 – crops and tree plantations,
      4 – agricultural water bodies and quarrying areas,
      5 – dense and sparse urban areas and roads
  Distance to coast
  Distance to closest harbour

degree of anthropization using the Land Cover Map of Catalonia (CREAF – Centre for Ecological Research & Forestry Applications, 2009; Ibáñez & Burriel, 2010). Each land cover category was assigned a value ranging from one (least anthropization) to five (most anthropization) (Table 2). Then, to represent the degree of anthropization, we calculated a weighted average scaled between 0 and 100 for each 1 km square grid.

As our goal is to predict species distributions rather than to understand which factors affect their distribution, all predictors were used for modelling each species. Extracting collinearity from the model was not necessary. Although collinearity can hinder the explanation of which variables affect species distribution, it does not affect MaxEnt predictive performance (Kuemmerle *et al.*, 2010).

Species distribution modelling

Modelling involved a five-step process as shown in Fig. 2. In the first step, we modelled the species distribution at the finer resolution of 1 km following the methodology described in Case 1 of Elith *et al.*, (2011). These authors use different alternative background scenarios to account for bias, and cross-validation techniques to validate models developed with presence-only data of *Banksia prionotes* from an atlas database. Accounting for bias and autocorrelation is an important issue in species distribution modelling, especially in presence-only models (Legendre, 1993; Legendre *et al.*, 2002; Segurado *et al.*, 2006; Phillips *et al.*, 2009; Newbold, 2010; Merckx *et al.*, 2011). As we expect fine-grained casually collected data to show a number of biases, we included three bias correction treatments and five spatial autocorrelation (SAC) correction treatments (see below) to evaluate the potential of these data to derive ecologically sound species distribution models.

This resulted in a total of 15 models per species. In a second step, only those models with an AUC (area under the curve) $\geq$ 0.7 not showing residual SAC were selected. In a third step, these selected models were upscaled to a coarser resolution of 10 km using a probabilistic model (see equation later). In step four, upscaled models were validated against the independent 10 km data set and only those with an AUC $\geq$ 0.7 at 10-km resolution (AUC_{10K}) were selected. Therefore, the 1-km resolution models which, once upscaled, resulted in these selected 10 km models are the only ones which showed acceptable predictions at both scales. Finally, in step five, if more than one fine-scale model per species had been selected, we determined the best one by selecting that with the highest AUC at 1-km resolution (AUC_{1K}). Despite concerns about the use of AUC to compare species distribution models, this metric can safely be applied when evaluating model performance within species (Lobo *et al.*, 2008; Blach-Overgaard *et al.*, 2010) and when interpreting it as a measure of discrimination between presence and background rather than presence and absence (Phillips *et al.*, 2006; Phillips & Dudík, 2008).
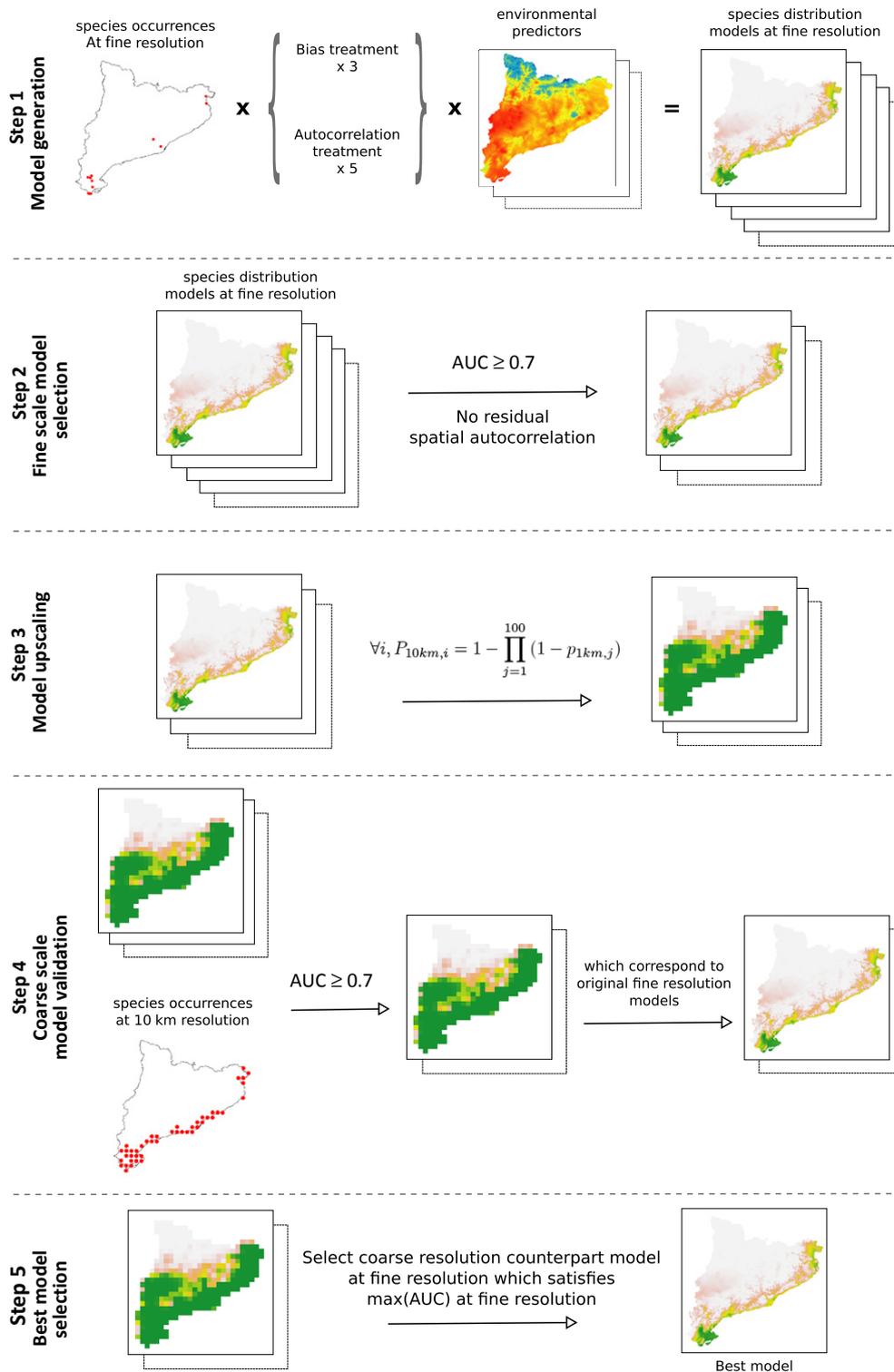
Figure 2 Outline for the proposed modelling workflow. Step one corresponds to the modelling of each species at 1-km resolution with three different bias treatments and five different autocorrelation treatments, which gives a total of 15 models per species. In step two, we check for residual autocorrelation, calculate the AUC and select only those models with no residual spatial autocorrelation and with an AUC $\geq$ 0.7. In step three, previously selected models are upscaled to 10-km resolution by probabilistic calculations. In step four, a ROC analysis is performed using independent data at 10-km resolution. Models with an AUC $\geq$ 0.7 tell us which models at 1-km resolution are accepted. Finally, in step five, if more than one model per species at 1-km resolution has been accepted, we define the best model as the one which has the maximum AUC_{1K}.

Bias correction treatments

Background samples should be chosen to reflect the spatial bias and thus to minimize the effects of bias in the data (Phillips *et al.*, 2009; Veloz, 2009; Elith *et al.*, 2011). We prepared three different background scenarios: (a) the entire study area (coded as 'whole_area'), (b) 1 km squares with the presence of vascular plants citations (about 14.0% of the whole area, coded as 'vasculars1k') and (c) 1 km squares with the presence of IAS citations (about 2.3% of the whole area, coded as 'invasive1k'). See Table 3.

Autocorrelation correction treatments

SAC may falsely inflate AUC measures for species distribution models with presence-only data (Segurado *et al.*, 2006; Veloz, 2009) and environmental autocorrelation may have the same effects. There is no established methodology for accounting for SAC when dealing with presence-only data (Dormann *et al.*, 2007; Elith & Leathwick, 2009). Autoregressive models are not applicable because both presence and absence data would be needed (Allouche *et al.*, 2008). We took an a priori approach similar to (Segurado *et al.*, 2006; Pearson *et al.*, 2007) which consisted in filtering occurrences by setting a minimum spatial and environmental distance between them and then checking for residual autocorrelation.

We prepared five treatments for modelling each species. The first involved including all available presences without filtering them. The second and third involved randomly filtering and selecting occurrences so that any occurrence was at least at a spatial distance of 2830 m (two 1 km squares) and 4250 m (three 1 km squares) from each other, respectively. For the fourth and fifth treatment, we used a minimum multivariate environmental distance based on the Gower's distance index with values 0.05 and 0.1, respectively (higher values resulted in an excessive reduction in occurrences). Models were then checked for significant residual autocorrelation [observed occurrence minus probability of occurrence as in (De Marco *et al.*, 2008; Nuñez & Medley, 2011; Vàclavík & Meentemeyer, 2012)] by using Monte-Carlo simulation of Moran's I autocorrelation coefficient using package spdep in R (Bivand, 2011). Only those models with a P-value ≥ 0.05 were accepted (as shown in Table 3).

Modelling and validation at 1-km resolution

We used MaxEnt software, version 3.3.3e, (Phillips *et al.*, 2006; Phillips & Dudík, 2008). MaxEnt is a presence-background modelling tool based on the maximum entropy principle. There is wide agreement in the species distribution modelling community that it is the best available tool for presence-only data, even when only a limited number of

Table 3 Models with $AUC_{1K} \geq 0.7$ and no residual spatial autocorrelation at 1 km. Finally accepted models (AUC_{10K} at 10 km with ≥ 0.7) at 10-km resolution are indicated with a Y in column 'Accepted'. In column 'Best', those with the highest AUC_{1K} at 1 km from the accepted models are marked with an asterisk.

Sp	Bias tr.	Aut. type	Min. dist.	AUC_{1k}	M P-value	AUC_{10k}	Accepted	Best
aga	whole_area	Spatial	4250 m	0.79	0.064	0.86	Y	*
aga	vasculars1k	Spatial	4250 m	0.75	0.066	0.78	Y	
aia	invasive1k	Spatial	2830 m	0.82	0.052	0.50	N	
aia	vasculars1k	Spatial	2830 m	0.78	0.076	0.65	N	
aia	invasive1k	Spatial	4250 m	0.72	0.164	0.45	N	
aia	vasculars1k	Spatial	4250 m	0.72	0.124	0.67	N	
oeb	invasive1k	Environmental	0.10	0.75	0.086	0.57	N	
opf	whole_area	Environmental	0.10	0.87	0.074	0.85	Y	
opf	whole_area	Spatial	2830 m	0.87	0.072	0.85	Y	*
opf	vasculars1k	Spatial	2830 m	0.80	0.054	0.82	Y	
opf	whole_area	Spatial	4250 m	0.86	0.076	0.86	Y	
oxp	whole_area	Environmental	0.10	0.92	0.172	0.92	Y	
oxp	invasive1k	Environmental	0.10	0.70	0.078	0.78	Y	
oxp	vasculars1k	Environmental	0.10	0.87	0.136	0.89	Y	
oxp	whole_area	Spatial	0.00	0.94	0.054	0.93	Y	*
oxp	vasculars1k	Spatial	2830 m	0.86	0.056	0.90	Y	
rop	invasive1k	Environmental	0.05	0.83	0.252	0.52	N	
rop	invasive1k	Environmental	0.10	0.81	0.405	0.53	N	
xas	vasculars1k	Environmental	0.05	0.75	0.150	0.69	N	
xas	vasculars1k	Environmental	0.10	0.76	0.577	0.70	Y	*

Sp, species abbreviation; Bias tr., bias treatment (*whole_area*, whole study area as background; *vasculars1k*, UTM squares with citations of vascular plants as background; *invasive1k*, UTM squares with citations of invasive plants as background), Aut. Type, autocorrelation treatment type (*spatial*, based on spatial distance; *environmental*, based on environmental distance), Min. dist., autocorrelation minimum distance value; AUC_{1K}, AUC value for 1 km models; M P-value, Moran's I P-value from Monte-Carlo simulation; AUC_{10K}, AUC value for 10 km models; Accepted, models accepted; Best, overall best models.

occurrence records are available (Elith *et al.*, 2006; Hernandez *et al.*, 2006; Phillips & Dudík, 2008; Wisz *et al.*, 2008; Elith & Graham, 2009; Thorn *et al.*, 2009; Costa *et al.*, 2010) and with bias present (Rebelo & Jones, 2010). MaxEnt estimates the distribution of maximum entropy constrained in a way that expected values for predictor variables match their empirical average (Phillips *et al.*, 2006). We used the logistic output of the model that indicates the relative environmental suitability of each pixel in relation to background for the study area (Phillips *et al.*, 2006; Phillips & Dudík, 2008).

We ran the model for each species with default options using the whole set of environmental predictors (Table 2) and following the methodology explained in Case 1 of Elith *et al.*, 2011. A total of 150 models were generated, which correspond to ten species times three bias scenarios times five autocorrelation correction treatments. When dealing with data from atlas databases, randomly partitioning occurrence data into training and test sets and using cross-validation techniques is often the only solution available to calibrate and test a model. We used 10-fold cross-validation and then used the average of all models as the final one. As a goodness-of-fit measure, we used the test AUC. As it is usually the norm in species distribution modelling, we accepted only models with an AUC $\geq$ 0.7. Models with an AUC $\geq$ 0.9 are considered excellent (Swets, 1988). As mentioned earlier, we only accepted models with no residual autocorrelation as tested by Moran's I autocorrelation coefficient.

Upscaling and validation at a coarser scale

We assumed that habitat quality is related to probability of presence and upscaled each accepted model at 1-km resolution (AUC$_{1K}$ $\geq$ 0.7 and no residual SAC) to 10-km resolution by a basic calculation of probabilities (see equation below). We computed the probability of presence for each i_{th} 10 km square of the study area ($P_{10km,\ i}$), given the predicted probability of presence for each 1 km square contained within it ($P_{1km,\ j}$). If we subtract this probability from 1, we obtain the probability of absence for this j_{th} 1 km square. For a given i_{th} 10 km square to have an absence, all of its one hundred 1 km squares need also to be absences. Therefore, by multiplying the probabilities of absence for each j_{th} 1 km square, we get the probability of absence for the i_{th} 10 km square. Finally, by subtracting the probability of absence for an i_{th} 10 km square from 1, we get its probability of presence ($P_{10km,i}$).

$$\forall_i, P_{10km,i} = 1 - \prod_{j=1}^{100}(1 - p_{1km,j})$$

We then performed a receiver operating characteristic (ROC) analysis [ROCR package in R (Sing *et al.*, 2009)] and computed the AUC$_{10K}$ value for each upscaled 10 km model using the independent data set at 10-km resolution. To ensure accurate prediction assessment, independent test sets

should be available (Loiselle *et al.*, 2008; Veloz, 2009). Again, those models with an AUC$_{10K}$ value $\geq$ 0.7 were accepted. Finally, of all models accepted for each species, we selected the one with the highest AUC$_{1K}$ value at 1-km resolution as the best one. In summary, we obtained a set of distribution maps that perform well at the finer resolution and that also acceptably predict independent records at the coarser resolution. We think these models can be considered robust and reliable given the data available.

RESULTS

Overall, AUC test values at 1-km resolution (AUC$_{1K}$) ranged from as low as 0.37 to as high as 0.96 (including models with residual autocorrelation), while their corresponding upscaled models at 10-km resolution ranged from 0.45 to 0.93 (Table 3). Of 150 models, 101 (67%) had an AUC$_{1K}$ $\geq$ 0.7. Of these, only 20 showed no significant residual SAC (Moran's *P*-value from Monte-Carlo simulation $\geq$ 0.05). The 20 that performed well at 1-km resolution are shown in Table 3. AUC$_{1K}$ test values for the accepted 20 models ranged from 0.7 to 0.94 and correspond to seven of the ten modelled species. The other three, *Amaranthus albus*, *Conyza canadensis* and *Datura stramonium*, did not perform well when modelling at 1-km resolution. *Oxalis pes-caprae* had the highest number of acceptable models at 1-km resolution but, nevertheless, unacceptable models predominated (10 of 15). The rest had between 11 and 14 unacceptable models. The worst models, those with an AUC$_{1K}$ $\leq$ 0.5, were four models of *Amaranthus albus* and one of *Datura stramonium*. All of these models used the *invasive1k* bias treatment.

When evaluating the performance at 10-km resolution, 12 of these final 20 models (60%) had an AUC$_{10K}$ $\geq$ 0.7 and were considered acceptable distribution models given the data available (see Table 3). Models marked with an asterisk correspond to our best models (see Table 3 and Fig. 3); *that is*, those with the maximum AUC$_{1K}$ value, when more than one model per species was accepted.

Half of the 12 models finally accepted required no bias treatment, while the other half performed better when a bias treatment was applied, although only one of them showed preference for the background offered by IAS citation areas. With respect to autocorrelation treatment, six performed better with some sort of SAC correction, while five did so with environmental autocorrelation correction. One model needed no autocorrelation correction, while none seemed to prefer the environmental correction with the shortest distance, and finally, only one model did not need either bias or autocorrelation treatment, which corresponded to *Oxalis pes-caprae*. This model also coincides with the best one of all, although care should be taken when comparing AUC values between species (Lobo *et al.*, 2008; Blach-Overgaard *et al.*, 2010). See Table 4 for a summary.

Three species, *Ailanthus altissima*, *Oenothera biennis* and *Robinia pseudoacacia*, did not pass the cross-scale validation cut (see Table 3). They had models that were acceptable at
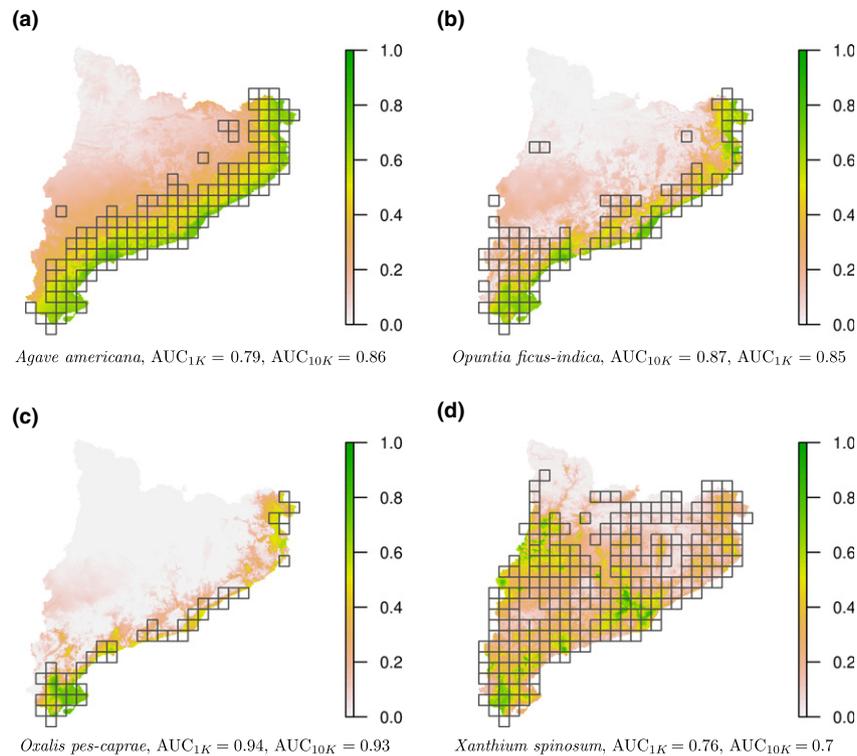
Figure 3 Best models per species among all the accepted models. Only four species resulted in finally valid models at 1-km resolution. For these species, those shown in the figure are the ones with max (AUC$_{1K}$). Legend scale ranges from 1.0 (maximum suitability) to 0.0 (no suitability). Black empty squares represent records of presence at 10-km resolution.

1-km resolution but which, once scaled, did not offer acceptable predictive power at 10 km. Thus, their finer resolution models were discarded as not robust enough: *that is*, they could not explain the independent data set at 10-km resolution. As an example, Fig. 4 shows two models that, while having passed the cut at 1-km resolution modelling, show an AUC$_{10K}$ around 0.5 that is not better than random.

On a per species basis, *Agave americana* performed well under the *whole_area* and *vasculars1k* bias treatments and for SAC correction with a minimum distance of 4250 m. Its best model was the one with the bias treatment *whole_area*. *Opuntia ficus-indica* performed well under *whole_area* and

vasculars1k bias treatments and under both environmental and spatial occurrence filtering, its best model being the one with the *whole_area* bias treatment and a SAC correction with a minimum distance of 2830 m. *Oxalis pes-caprae* performed well under all three bias treatments and under both spatial and environmental autocorrelation correction. Its best model required no bias or autocorrelation treatment at all. Finally, for *Xanthium spinosum*, the only successful treatment was the *vasculars1k* bias treatment and the environmental autocorrelation correction with a minimum distance of 0.1. Reductions in the number of available occurrences after autocorrelation correction for the final four best models were
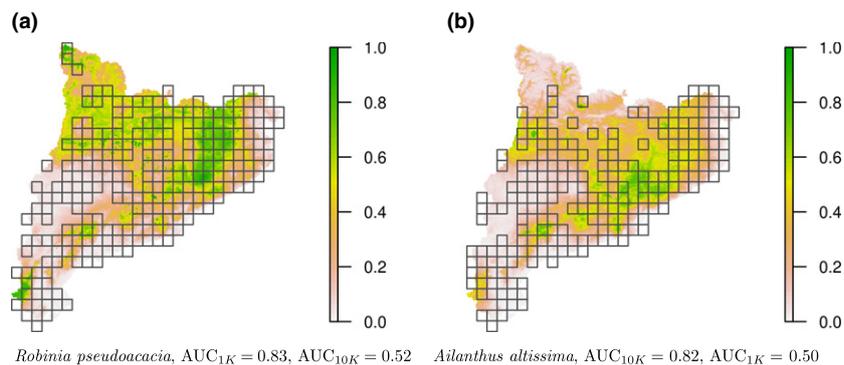

Figure 4 Examples of models that did not work. Even though these two models had a high AUC$_{1K}$ value and showed no residual autocorrelation, they had an AUC$_{10K}$ close to 0.5 and are thus not better than random. Legend scale ranges from 1.0 (maximum suitability) to 0.0 (no suitability). Black empty squares represent records of presence at 10-km resolution.

Table 4 1-km models with residual spatial autocorrelation and models finally accepted

Treatments			Bias										Total		
			None			invasive1k			vasculars1k						
			n	%	a	n	%	a	n	%	a	n	%	a
SAC	None	0	9	90.0	1	9	90.0	0	10	100.0	0	28	93.3	1
	Env. dist.	0.05	10	100.0	0	7	70.0	0	7	70.0	0	24	80.0	0
		0.01	7	70.0	2	3	30.0	1	3	30.0	2	13	43.3	5
	Sp. dist.	2830 m	9	90.0	1	8	80.0	0	5	50.0	2	22	73.3	3
		4250 m	8	80.0	2	7	70.0	0	7	70.0	1	22	73.3	3
		Total	43	86.0	6	34	68.0	1	32	64.0	5	109	72.7	12

This table offers a summary of the 109 models at 1-km resolution which were discarded because of residual spatial autocorrelation. Bias treatments are shown in columns and spatial autocorrelation treatments (SAC) are shown in rows. For each treatment combination, three numbers are shown: n is the number of discarded models, % is the percentage of discarded models, and a is the number of finally accepted models for the rest of the models. Note that for n and %, the lower the number, the better the combination treatment performed with respect to residual spatial autocorrelation.

as follows: *Agave americana* from 20 to 12, *Opuntia ficus-indica* from 13 to 10, *Xanthium spinosum* from 56 to 22 and no reduction for *Oxalis pes-caprae* because its best model was the one without autocorrelation correction.

Table 4 presents a summary of accepted and discarded models according to treatments. For SAC treatments, the option of no treatment was the worst (93.3% of these models showed residual SAC), while the best was the treatment corresponding to an environmental distance (Gower's index) of 0.1 (less than half (43.3%) of the models showed residual SAC). The rest of SAC treatments had similar results: only a quarter to a fifth of the models showed no residual SAC. For bias treatments, using some kind of treatment worked better (68% for *invasive1k* and 64% for *vasculars1k*) than no treatment (86%). The best treatments for removing residual SAC were the combination of an environmental distance of 0.1 with some bias treatment (*invasive1k* or *vasculars1k*). As expected, the number of models with residual SAC is inversely proportional to the intensity of the SAC treatment applied.

DISCUSSION

Our results show that species distribution maps derived from presence-only records held in biodiversity databases or atlases should be used with caution. Apparently, high scores in predictive power from species distributions can be obtained from scarce, biased and autocorrelated presence records using modern tools such as MaxEnt. However, our work shows that these results can be misleading when confronted with independent data at different scales. Other authors have reached similar conclusions (Wisz *et al.*, 2008). If the distribution of a species was well-known at two different scales, these should necessarily be coherent with one another. To generate reliable fine-resolution distribution maps, these need to be in accordance across scales (Niamir *et al.*, 2011). For a given species, its real distribution map at a fine scale should match its real distribution map at a coarser scale once upscaled. This seems not to be the case for some species, indicating that either the modelled distributions at fine resolution are wrong or that the *known* distributions at the coarser scale are, in fact, incomplete (unlikely for a well-surveyed region for vascular plants such as our study area). Therefore, if one accepts this assumption, our results suggest that distribution maps at the finer scales are not as good as they appear to be. However, this could also be due to the fact that our models have been built without explanatory variables which can account for other environmental factors and biotic interactions (e.g. interspecific competition), thus not reflecting the realized niche of the species. Although at macroecological scales climate is the main factor affecting species distributions, biotic interactions may also play a role (Araújo & Luoto, 2007; Heikkinen *et al.*, 2007; Kissling *et al.*, 2010). Such variables, if available, could positively affect our models and make them more in accordance with well-known distributions at coarser scales. Unless this problem can be solved, if these models are used for decision-making in conservation, they may not always accomplish the objectives for which they are meant.

Of the 20 models that performed well at 1-km resolution, only 12 were coherent with data at 10-km resolution. Specieswise, it might seem that good fine-scale predictive maps could be derived from the biodiversity database for seven species. However, fine-scale distribution maps were in accordance with their coarser scale data for only four of them (Table 3). We can thus consider the fine-resolution maps for these four species to be sufficiently reliable for biodiversity conservation. Coarse resolution data do not often match the requirements of conservation planning (Araújo *et al.*, 2005), but, when these data are assumed to reflect the distribution of the species at the coarse scale, they can be used to make a cross-scale validation of modelled fine-scaled distribution maps, even if high predictive scores had been obtained. The resulting maps will be much more reliable and robust and will help decision-makers to better meet their conservation goals.

Atlas data commonly suffer from bias and autocorrelation problems (Robertson *et al.*, 2010). Treating both hindrances is paramount to developing robust and reliable species distribution models (Segurado *et al.*, 2006; Merckx *et al.*, 2011). We tried several bias and autocorrelation correction scenarios but did not find any particular pattern in our results that can help in establishing protocols for distribution modelling. Although the overall outcome of the modelling process can only be known on a case-by-case basis, with respect only to the correction of residual SAC, our results suggest that the use of data filtering (either by setting minimum environmental or spatial distances) helps to remove residual SAC: the larger the minimum distance, the better. Also, to a lesser extent, bias treatments help to remove it. Except for *Oxalis pes-caprae*, for which the environmental predictors used were able to remove residual SAC without applying any treatment to the occurrence data, the rest of the species needed some sort of occurrence filtering to remove it. This suggests that, for the other species in this study, some missing explanatory variables (environmental or biotic), if available, may have resulted in models with a better fit and no residual SAC without having to discard occurrences to build it.

The lesson that can be learnt from this study is to be sceptical of fine-resolution maps obtained when modelling species distributions from scarce and biased data, even though they score high when measuring their predictive power. Given that current modelling techniques and computer power allow us to run many models per species taking into account different scenarios of bias and autocorrelation correction treatments, it is always sensible to do so and check, on a case-by-case basis, which one works best. Having a set of independent data to validate the model seems indispensable. Cross-scale validation, when possible, is a good solution to produce reliable and robust maps which can then be used to make better conservation decisions.

CONCLUSIONS

Casual observations at fine resolution in biodiversity atlases or databases have the potential to generate continuous species distribution maps through species distribution modelling, providing powerful tools for conservation management and planning. However, it may be impossible to obtain them for many species because of scarcity and strong biases in the data. Although high validation scores can be obtained when modelling such data, there is a risk that the distribution maps reflect the data distribution rather than the true species distribution. Cross-scale validation of the data with species distribution information at a coarser scale appears to be a consistent protocol to test the validity and robustness of fine-resolution models and thus to make them much more reliable for decision-makers in conservation.

Fine-resolution maps can be derived from biodiversity atlases with data scarcity, bias and autocorrelation. Bias, and specially autocorrelation, treatments can help to remove residual SAC in models. However, if no independent set of data is available to further validate them, results should be viewed with caution. When good coarser scale data are available, cross-scale validation appears to be effective in assessing the robustness of the models. When these options are not available, new field surveys may be the only option if reliable fine-scale maps are needed.

REFERENCES

Allouche, O., Steinitz, O., Rotem, D., Rosenfeld, A. & Kadmon, R. (2008) Incorporating distance constraints into species distribution models. *Journal of Applied Ecology*, **45**, 599–609.

Aranda, S .C. & Lobo, J.M. (2011) How well does presence-only-based species distribution modelling predict assemblage diversity? A case study of the Tenerife flora. *Ecography*, **34**, 31–38.

Araújo, M.B. & Luoto, M. (2007) The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography*, **16**, 743–753.

Araújo, M.B. & Pearson, R.G. (2005) Equilibrium of species' distributions with climate. *Ecography*, **28**, 693–695.

Araújo, M., Thuiller, W., Williams, P. & Reginster, I. (2005) Downscaling European species atlas distributions to a finer resolution: implications for conservation planning. *Global Ecology and Biogeography*, **14**, 17–30.

Austin, M.P. (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, **157**, 101–118.

Bivand, R. (2011). spdep: spatial dependence: weighting schemes, statistics and models. R package version 0.5-40.

Blach-Overgaard, A., Svenning, J.C., Dransfield, J., Greve, M. & Balslev, H. (2010) Determinants of palm species distributions across africa: the relative roles of climate, non-climatic environmental factors, and spatial constraints. *Ecography*, **33**, 380–391.

Brooks, M.L. (2007). Effects of land management practices on plant invasions in wildland areas. *Biological invasions*

(ed. by W. Nentwig), pp. 147–162. Springer, Berlin, Heidelberg.

Butchart, S.H.M., Walpole, M., Collen, B. *et al.* (2010) Global biodiversity: indicators of recent declines. *Science*, **328**, 1164–1168.

Chapman, A.D. (2005). *Uses of primary species-occurrence data, version 1.0.* Global Biodiversity Information Facility, Copenhagen.

Chapman, A.D. & Busby, J.R. (1994). *Mapping the Diversity of Nature,* chapter Linking plant species information to continental biodiversity, inventory, climate modeling and environmental monitoring. pp. 179–195. Chapman & Hall, London.

Chytrý, M., Pyšek, P., Wild, J., Pino, J., Maskell, L.C. & Vilà, M. (2009) European map of alien plant invasions based on the quantitative assessment across habitats. *Diversity and Distributions*, **15**, 98–107.

Costa, G., Nogueira, C., Machado, R. & Colli, G. (2010) Sampling bias and the use of ecological niche modeling in conservation planning: a field evaluation in a biodiversity hotspot. *Biodiversity and Conservation*, **19**, 883–899.

CREAF – Centre for Ecological Research and Forestry Applications (2009). *Land cover map of Catalonia*, 3rd edn. http://www.creaf.uab.cat/mcsc/.

DAISIE European Invasive Alien Species Gateway (2011). DAISIE European Invasive Alien Species Gateway. Web. http://http://www.europe-aliens.org, accessed on January 31st, 2011.

De Marco, P. Jr, Diniz-Filho, J.A.F. & Bini, L.M. (2008) Spatial analysis improves species distribution modelling during range expansion. *Biology Letters*, **4**, 577–580.

Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A.H., Jetz, W., Kissling, W.D., Kühn, I., Ohlemüller, R., Peres-Neto, P.R., Reineking, B., Schröder, B., Schurr, F.M. & Wilson, R. (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, **30**, 609–628.

Elith, J. & Graham, C.H. (2009) Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography*, **32**, 66–77.

Elith, J. & Leathwick, J.R. (2009) Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology Evolution and Systematics*, **40**, 677–697.

Elith, J., Graham, C.H., Anderson, R.P. *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E. & Yates, C.J. (2011) A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, **17**, 43–57.

Font, X., de Cáceres, M., Quadrada, R. & Navarro, A. (2009). Biodiversity Data Bank of Catalonia. http://biodiver.bio.ub.es/biocat/homepage.html (Accessed on April 30th 2009).

Franklin, J. (2009). *Mapping species distributions: spatial inference and prediction*. Cambridge University Press, Cambridge, UK.

Funk, V.A. & Richardson, K.S. (2002) Systematic data in biodiversity studies: use it or lose it. *Systematic Biology*, **51**, 303–316.

Gassó, N., Pyšek, P., Vilà, M. & Williamson, M. (2010) Spreading to a limit: the time required for a neophyte to reach its maximum range. *Diversity and Distributions*, **16**, 310–311.

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

Guralnick, R. & Hill, A. (2009) Biodiversity informatics: automated approaches for documenting global biodiversity patterns and processes. *Bioinformatics*, **25**, 421–428.

Heikkinen, R.K., Luoto, M., Virkkala, R., Pearson, R.G. & Körber, J.H. (2007) Biotic interactions improve prediction of boreal bird distributions at macro-scales. *Global Ecology and Biogeography*, **16**, 754–763.

Hernandez, P.A., Graham, C.H., Master, L.L. & Albert, D.L. (2006) The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, **29**, 773–785.

Ibáñez, J.J. & Burriel, J.A. (2010). Mapa de cubiertas del suelo de Cataluña: características de la tercera edición y relación con SIOSE. In Actas del XIV Congreso Nacional de Tecnologías de la Información Geográfica., Sevilla.

Jiménez-Valverde, A., Peterson, A.T., Soberón, J., Overton, J., Aragón, P. & Lobo, J. (2011) Use of niche models in invasive species risk assessments. *Biological Invasions*, **13**, 2785–2797.

Kissling, W.D., Field, R., Korntheuer, H., Heyder, U. & Böhning-Gaese, K. (2010) Woody plants and the prediction of climate-change impacts on bird diversity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**, 2035–2045.

Kuemmerle, T., Perzanowski, K., Chaskovskyy, O., Ostapowicz, K., Halada, L., Bashta, A.T., Kruhlov, I., Hostert, P., Waller, D.M. & Radeloff, V.C. (2010) European bison habitat in the Carpathian Mountains. *Biological Conservation*, **143**, 908–916.

Legendre, P. (1993) Spatial autocorrelation: trouble or new paradigm? *Ecology*, **74**, 1659–1673.

Legendre, P., Dale, M.R.T., Fortin, M.J., Gurevitch, J., Hohn, M. & Myers, D. (2002) The consequences of spatial structure for the design and analysis of ecological field surveys. *Ecography*, **25**, 601–615.

Lobo, J.M., Jiménez-Valverde, A. & Real, R. (2008) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, **17**, 145–151.

Loiselle, B.A., Jorgensen, P.M., Consiglio, T., Jiménez, I., Blake, J.G., Lohmann, L.G. & Montiel, O.M. (2008) Predicting species distributions from herbarium collections:

does climate bias in collection sampling influence model outcomes? *Journal of Biogeography*, **35**, 105–116.

Margules, C.R. & Pressey, R.L. (2000) Systematic conservation planning. *Nature*, **405**, 243–253.

McPherson, J.M., Jetz, W. & Rogers, D.J. (2006) Using coarse-grained occurrence data to predict species distributions at finer spatial resolutions–possibilities and limitations. *Ecological Modelling*, **192**, 499–522.

Merckx, B., Steyaert, M., Vanreusel, A., Vincx, M. & Vanaverbeke, J. (2011) Null models reveal preferential sampling, spatial autocorrelation and overfitting in habitat suitability modelling. *Ecological Modelling*, **222**, 588–597.

Newbold, T. (2010) Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. *Progress in Physical Geography*, **34**, 3–22.

Niamir, A., Skidmore, A.K., Toxopeus, A.G., Muñoz, A.R.M. & Real, R. (2011) Finessing atlas data for species distribution models. *Diversity and Distributions*, **17**, 1173–1185.

Nielsen, C., Hartvig, P. & Kollmann, J. (2008) Predicting the distribution of the invasive alien *Heracleum mantegazzianum* at two different spatial scales. *Diversity and Distributions*, **14**, 307–317.

Ninyerola, M., Pons, X. & Roure, J.M. (2000) A methodological approach of climatological modelling of air temperature and precipitation through GIS techniques. *International Journal of Climatology*, **20**, 1823–1841.

Nix, H. (1986). A biogeographic analysis of Australian Elapid snakes. *Atlas of elapid snakes of Australia* (ed. by R. Longmore), pp. 4–15. Australian Flora and Fauna Series 7. Australian Government Publishing Service, Canberra.

Nuñez, M.A. & Medley, K.A. (2011) Pine invasions: climate predicts invasion success; something else predicts failure. *Diversity and Distributions*, **17**, 703–713.

Pearson, R.G., Raxworthy, C.J., Nakamura, M. & Peterson, A.T. (2007) Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography*, **34**, 102–117.

Pejchar, L. & Mooney, H.A. (2009) Invasive species, ecosystem services and human well-being. *Trends in Ecology & Evolution*, **24**, 497–504.

Peterson, A.T. (2005) Predicting potential geographic distributions of invading species. *Current Science*, **89**, 9.

Peterson, A.T., Papes, M. & Eaton, M. (2007) Transferability and model evaluation in ecological niche modeling: a comparison of GARP and MaxEnt. *Ecography*, **30**, 550–560.

Peterson, A.T., Papes, M. & Soberón, J. (2008) Rethinking receiver operating characteristic analysis applications in ecological niche modelling. *Ecological Modelling*, **213**, 63–72.

Phillips, S.J. & Dudík, M. (2008) Modeling of species distributions with MaxEnt: new extensions and a comprehensive evaluation. *Ecography*, **31**, 161–175.

Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

Phillips, S.J. (2008) Transferability, sample selection bias and background data in presence-only modelling: a response to Peterson et al. (2007). *Ecography*, **31**, 272–278.

Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J. & Ferrier, S. (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, **19**, 181–197.

Pimentel, D., Zuniga, R. & Morrison, D. (2005) Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics*, **52**, 273–288.

Pino, J., Font, X., Carbó, J., Jové, M. & Pallarès, L. (2005) Large-scale correlates of alien plant invasion in Catalonia (NE of Spain). *Biological Conservation*, **122**, 339–350.

Pressey, R.L. (2004) Conservation planning and biodiversity: assembling the best data for the job. *Conservation Biology*, **18**, 1677–1681.

Pyšek, P., Jarošík, V., Hulme, P.E. *et al.* (2010) Disentangling the role of environmental and human pressures on biological invasions across Europe. *Proceedings of the National Academy of Sciences USA*, **107**, 12157–12162.

Rebelo, H. & Jones, G. (2010) Ground validation of presence-only modelling with rare species: a case study on barbastelles *Barbastella barbastellus* (Chiroptera: Vespertilionidae). *Journal of Applied Ecology*, **47**, 410–420.

Richardson, D.M. & Whittaker, R.J. (2010) Conservation biogeography – foundations, concepts and challenges. *Diversity and Distributions*, **16**, 313–320.

Robertson, M.P., Cumming, G.S. & Erasmus, B.F.N. (2010) Getting the most out of atlas data. *Diversity and Distributions*, **16**, 363–375.

Segurado, P., Araújo, M.B. & Kunin, W.E. (2006) Consequences of spatial autocorrelation for niche-based models. *Journal of Applied Ecology*, **43**, 433–444.

Sing, T., Sander, O., Beerenwinkel, N. & Lengauer, T. (2009). ROCR: visualizing the performance of scoring classifiers. R package version 1.0-4.

Suarez, A.V. & Tsutsui, N.D. (2004) The value of museum collections for research and society. *BioScience*, **54**, 66–74.

Swets, J. (1988) Measuring the accuracy of diagnostic systems. *Science*, **240**, 1285–1293.

Taylor, B.W. & Irwin, R.E. (2004) Linking economic activities to the distribution of exotic plants. *Proceedings of the National Academy of Sciences USA*, **101**, 17725–17730.

Thorn, J.S., Nijman, V., Smith, D. & Nekaris, K.A.I. (2009) Ecological niche modelling as a technique for assessing threats and setting conservation priorities for Asian slow lorises (Primates: Nycticebus). *Diversity and Distributions*, **15**, 289–298.

Vàclavík, T. & Meentemeyer, R.K. (2012) Equilibrium or not? Modelling potential distribution of invasive species in

different stages of invasion. *Diversity and Distributions*, **18**, 73–83.

Vallecillo, S., Brotons, L. & Thuiller, W. (2009) Dangers of predicting bird species distributions in response to land-cover changes. *Ecological Applications*, **19**, 538–549.

Veloz, S.D. (2009) Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *Journal of Biogeography*, **36**, 2290–2299.

Venette, R.C., Kriticos, D.J., Magarey, R.D., Koch, F.H., Baker, R.H., Worner, S.P., Gómez Raboteaux, N.N., McKenney, D.W., Dobesberger, E.J., Yemshanov, D., De Barro, P.J., Hutchison, W.D., Fowler, G., Kalaris, T.M. & Pedlar, J. (2010) Pest risk maps for invasive alien species: a roadmap for improvement. *BioScience*, **60**, 349–362.

Vicente, J., Alves, P., Randin, C., Guisan, A. & Honrado, J. (2010) What drives invasibility? A multi-model inference test and spatial modelling of alien plant species richness patterns in Northern Portugal. *Ecography*, **33**, 1081–1092.

Vilà, M., Espinar, J.L., Hejda, M., Hulme, P.E., Jarošík, V., Maron, J.L., Pergl, J., Schaffner, U., Sun, Y. & Pyšek, P. (2011) Ecological impacts of invasive alien plants: a meta-analysis of their effects on species, communities and ecosystems. *Ecology Letters*, **14**, 702–708.

Vitousek, P.M., Mooney, H.A., Lubchenco, J. & Melillo, J.M. (1997) Human domination of Earth's Ecosystems. *Science*, **277**, 494–499.

Williamson, M., Dehnen-Schmutz, K., Kühn, I., Hill, M., Klotz, S., Milbau, A., Stout, J. & Pyšek, P. (2009) The distribution of range sizes of native and alien plants in four European countries and the effects of residence time. *Diversity and Distributions*, **15**, 158–166.

Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A. & NCEAS Predicting Species Distributions Working Group (2008) Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, **14**, 763–773.

Zimmermann, N.E., Edwards, T.C., Graham, C.H., Pearman, P.B. & Svenning, J.C. (2010) New trends in species distribution modelling. *Ecography*, **33**, 985–989.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Maps of occurrences.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

BIOSKETCHES

Arnald Marcer is a researcher at the Center for Ecological Research and Forestry Applications (CREAF), Autonomous University of Barcelona. His research interests are in the fields of biodiversity conservation and modelling, and environmental information handling and management through the development of information systems.

Author contributions: A.M. conceived the ideas, analysed the data and wrote the manuscript. L.B., J.P. and X.P helped to refine the ideas and methodology and to revise the final manuscript. The authors' research groups and interests can be found at the following hyperlinks:

http://www.grumets.uab.cat/index_en.htm

http://biodiversitylandscapeecologylab.blogspot.com/

Editor: Mark Robertson