Process-based and correlative modeling of desert mistletoe distribution: a multiscalar approach

Andrés Lira-Noriega,[1,†] Jorge Soberón,[1] and Curtis P. Miller[1,2]

[1]Biodiversity Institute, University of Kansas, Lawrence, Kansas 66045 USA
[2]College of Pharmacy, MSC09 5360, 1 University of New Mexico, Albuquerque, New Mexico 87131-0001 USA

Abstract. Because factors affecting distributional areas of species change as scale (extent and grain) changes, different environmental and biological factors must be integrated across geographic ranges at different resolutions, to understand fully the patterns and processes underlying species' ranges. We expected climate factors to be more important at coarse resolutions and biotic factors at finer resolutions. We used data on occurrence of a parasitic plant (*Phoradendron californicum*), restricted to parts of the Sonoran and Mojave deserts, to analyze how climate and mobility factors explain its distributional area. We developed analyses at five spatial resolutions (1, 5, 10, 20, 50 km) within the distributional area of the disperser species, and compared ecological niche models from three commonly used correlative methods with a process-based model that estimates colonization and extinction rates in a metapopulation framework. Correlative models improved when layers associated with hosts and disperser were used as predictors, in comparison with models based on climate only; however, they tended to overfit to data as more layers were added. Dispersal-related parameters were more relevant at finer resolutions (1–5 km), but importance of extinction-related parameters did not change with scale. We observed greater coincidence between correlative and process-based models when based only on dimensions of the abiotic niche (i.e., climate), but a clearer and more comprehensive mechanistic understanding was derived from the process-based algorithm.

† **E-mail:** aliranoriega@gmail.com

INTRODUCTION

The relative roles of biotic and abiotic factors in determining distributions of species at specific spatial scales are a central organizing theme in ecology (Levin 1992). Understanding how patterns at one scale are manifestations of or influence processes operating at other scales is a particular challenge (Levin and Pacala 1997, Pearson and Dawson 2003, Hastings et al. 2010). The core of this challenge lies in disentangling how changes in scale affect the importance of different factors in shaping species' distributional patterns and processes.

When studying species' distributions, it is useful to distinguish interacting variables from those that are dynamically uncoupled from the presence or abundance of the species in question (scenopoetic variables; Hutchinson 1978). These variables may have contrasting effects at fine versus coarse spatial resolutions (Wiens 1989). The many ways in which scenopoetic variables operate at different extents and resolutions can be used to explain distributions of the species

and propose scale-dependent hypotheses regarding factors most relevant at different scales (Pulliam 2000, Gaston 2003, Soberón and Peterson 2005, Soberón 2007, 2010, Peterson et al. 2011).

Using data available documenting species' occurrence and climatic variation, researchers have been able to estimate environmental requirements of many species across broad regions (Guisan and Zimmermann 2000, Elith et al. 2006). Factors manifested at finer resolutions are less well studied at the scope of broad geographic ranges, leaving a gap in understanding as to how effects of these factors vary spatially and temporally. Also, knowledge of species' distributions at local scales may be sparse or biased spatially across geographic ranges, which makes analyzing the relative importance of variables even more difficult (MacArthur 1972, Levin 1992). These considerations explain why studies of spatial distributions of species that combine broad-scale and fine-scale views remain uncommon (e.g., Mackey and Lindenmayer 2001, Gaston et al. 2004, Heikkinen et al. 2007, Cunningham et al. 2009).

Correlative and process-based modeling are two approaches for exploring factors important in determining distributions of species at different spatial scales (Robertson et al. 2003, Kearney et al. 2010). The main difference between these approaches is that correlative methods simply seek associations between environments and occurrences from across broad geographic ranges, whereas process-based methods incorporate explicit hypotheses about biological processes (Robertson et al. 2003, Kearney and Porter 2009, Monahan 2009, Morin and Thuiller 2009, Cabral and Schurr 2010, Kearney et al. 2010). In particular, a set of factors that is emerging as crucial in process-based modeling studies is dispersal: incorporating dispersal factors improves model performance markedly (Allouche et al. 2008, Cabral and Schur 2010, Brotons et al. 2012).

In this paper, we explore how dispersal and scenopoetic factors act synergistically to explain the distribution of the Desert Mistletoe, *Phoradendron californicum*, at varying spatial resolutions. *P. californicum* is a hemi-parasitic plant associated with legume hosts in the Sonoran and Mojave deserts, and it is dispersed in largest part by the bird species *Phainopepla nitens* (Walsberg 1975, Kuijt 2003). This parasite-host-disperser system offers two advantages for understanding how scaling of biological processes govern distributions (Overton 1996, Aukema and Martınez del Rio 2002, Aukema 2003, 2004): (1) hosts are known, and their distributions can be studied at high spatial resolution using aerial photographs; and (2) dispersal is determined chiefly by a single disperser. We modeled the portion of the distribution of *P. californicum* in the U.S., where detailed data on *Phainopepla nitens* are available (almost 900,000 km^2; Fig. 1), encompassing approximately half of the parasite species' range. We compare commonly-used correlative niche modeling methods (GARP, Maxent, GAM) with a process-based model derived from a metapopulation-dynamic framework coupling rates of colonization and extinction. Following Pearson and Dawson (2003) and others, we hypothesize that, as resolution coarsens, variables associated with abiotic factors (e.g., climate) will become more important, but that the opposite will apply for biotic variables (e.g., dispersal).

Conceptual framework

To help understand causal factors underlying geographic distributions of species, a heuristic device called the BAM diagram is useful (Soberón and Peterson 2005, Peterson et al. 2011). The BAM diagram displays relationships between abiotic or uncoupled (**A**) and coupled biotic factors (**B**), and the movement or dispersal capacities of the species (**M**). This framework can be used to make explicit the possible arrangements of factors that determine distributions of species, and gives the opportunity to generate hypothetical scenarios, depending on the degree and geometry of their overlap (Fig. 2). Specifically, with this framework, we can analyze interactions of factors at fine and coarse spatial resolutions.

In the mistletoe-host-disperser system, the BAM diagram can be simplified significantly by making some reasonable assumptions. The conditions manifested across **A** (Soberón and Peterson 2005, Soberón 2007) represent those under which the species can have a positive growth rate (Hutchinson 1978); therefore, conditions in **A** are linked intimately to the environmental dimensions on which mechanisms of establishment and
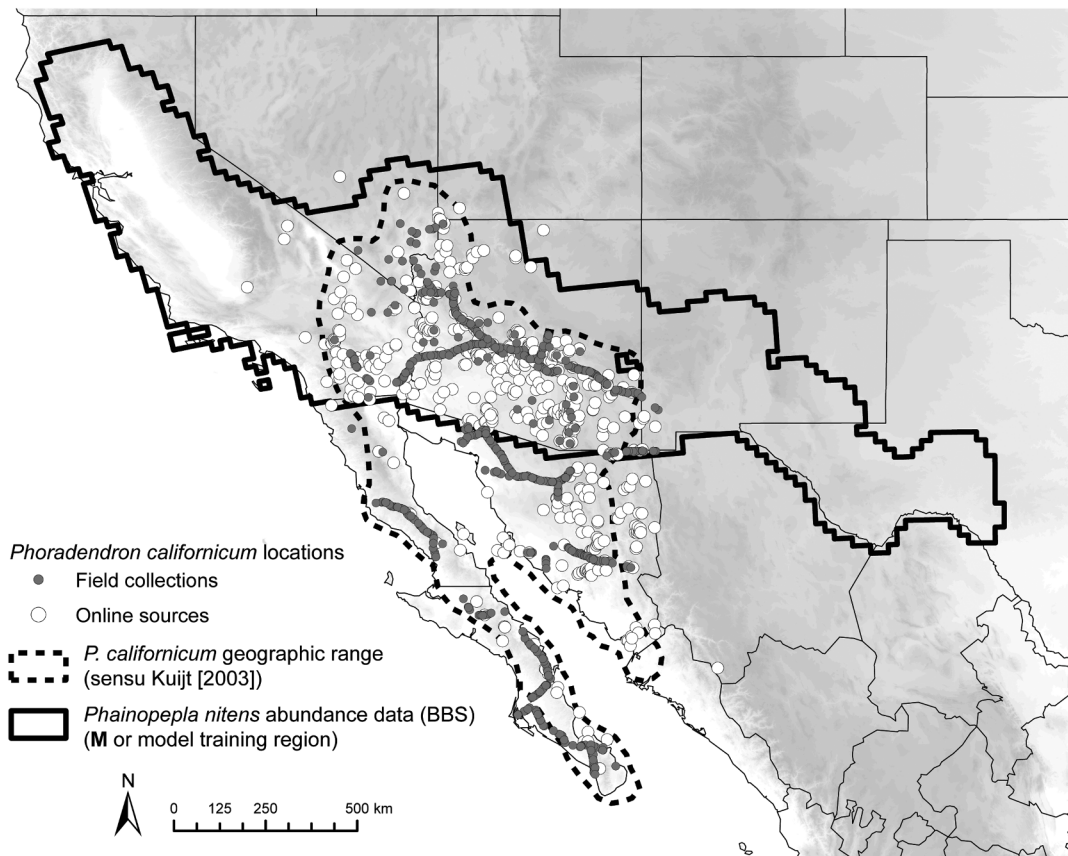
Fig. 1. Author-collected (A.L.-N.) and Internet-located occurrences of *P. californicum* and extent of study area. The study area corresponds to the distributional area of the principal disperser, assumed to represent accessible areas for the mistletoe (**M**), for which abundance data are available (thickest black line). Notice there are areas where the disperser or hosts can be present but not the mistletoe; these areas are relevant to understand the interaction of factors that limit the distribution of the species as well as challenging areas in which to test the models.

survivorship depend (Wiens 2000). Factors in **M** are related to movement: what happens across **M** determines how populations will be structured, demographically and genetically, although testing this particular idea will be accomplished in a later phase of this project (Lira-Noriega et al., *unpublished manuscript*).

We hypothesize that, to a first degree of approximation, the interaction with the host can be regarded as uncoupled, since hosts have expected lifetimes at least 10-fold longer than those of parasites (Overton 1996, 1997) and therefore their population dynamics can be regarded as approximately static relative to the faster demography of the parasites. The simplifying assumption of uncoupled dynamics of the

total host population is often made in epidemiological studies (Anderson 1981). This means that we regard climate and presence of hosts as part of the **A** circle. Second, we did not include the effect of competitors, which are not known to be present, nor of herbivores or pollinators, because information about them at the spatial extent of our study is simply unavailable.

Finally, in the mistletoe system, one major biotic factor, birds, act directly as dispersers. From a certain point of view the dispersers may be considered as part of **B**, but their effect is clearly felt in the dispersal circle, making **B** and **M** coincident. In consequence, we can simplify our system to a two-factor diagram: the abiotic climatic factors (**A**) and the hosts substrate, that
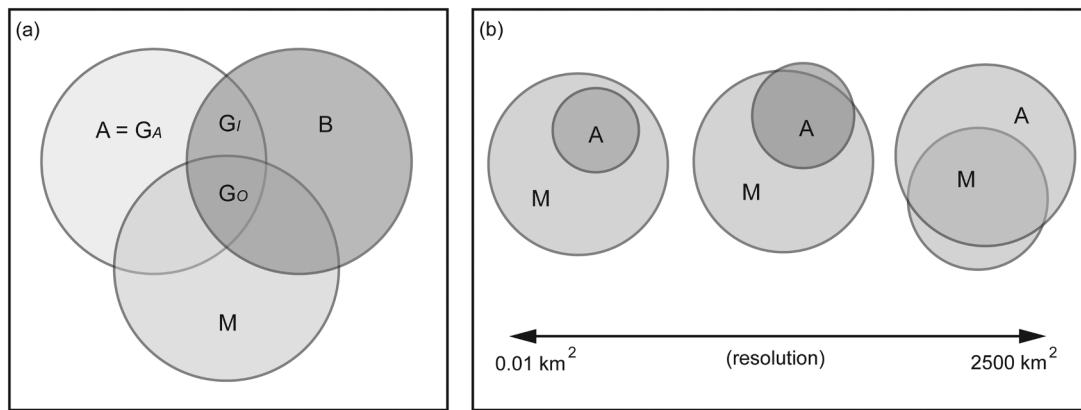
Fig. 2. Conceptual framework showing (a) the relationship between biotic-abiotic-mobility (BAM) conditions for a species to be present. (b) Interpretation of the BAM diagram in this modeling exercise restricts to the use of abiotic conditions (**A**) and accessible areas (**M**), assuming that the interaction between host and parasite is uncoupled, and that the positive interaction between disperser and parasite makes biotic conditions (**B**) and accessible area (**M**) coincident. As scale changes, limiting factors in **A** and **M** will also change. At higher resolutions, almost the entire area is accessible, and climate is not so limiting, whereas at coarser resolutions dispersal becomes more difficult and climate is more constant.

determine overall potential distributions, and the movement or dispersal capacity of the species (**M**), driven mainly by the bird that disperses seeds among trees. **M** represents a hypothesis of the regions that have been accessible for the species over relevant time periods; as such, **M** is the area across which models should be calibrated (Barve et al. 2011; see also Acevedo et al. 2012). Following this reasoning, we designated an **M** as the area of distribution of the mistletoe's main disperser, *P. nitens*, for which we have detailed abundance information (Fig. 1). Although other potential dispersers of the mistletoe species in question are known, none is as closely associated as this species (Walsberg 1975); adding the densities of other bird species to the model would not have changed our results, given that they overlap broadly in their distribution and behavior with that of *P. nitens*. This assumption can be relaxed in future studies, but for the present simplifies the process-based modeling exercise considerably.

METHODS

We designed a comparison of a process-based model with correlative models from three algorithms to understand the distribution of *P. californicum* at five spatial resolutions (1, 5, 10,

20, 50 km). We first describe the series of steps we followed in order to obtain data on the distribution of the parasite followed by the assemblage of three sets of environmental variables: spatial variation in disperser density, spatial variation in numbers of hosts, and climatic factors across the study area. We then describe the design and implementation of each model type and finally the comparisons among them.

Distribution of species presences and landscape characteristics

To estimate proportions of mistletoe-infected trees, we recorded geographic locations of infected host trees along roads. One of us (A.L.-N.) sampled occurrences of host trees infected by *P. californicum* across the southwestern United States and northwestern Mexico (Fig. 1). We used roads as sampling transects as an efficient means of surveying the species' range limits, allowing us to gather information regarding the broader set of climatic situations where the species occurs. We used a GPS unit to record geographic coordinates for each infected host tree located within 100 m of the road. In total, we sampled 16,000 km (4153 km within the modeling extent) of roads, and collected 17,371 unique geographic coordinates of infected trees (Fig. 1), 12,578 of which fell within the modeling extent (see

below).

Sampling along roads may potentially bias the data, since roads are usually areas of higher water accumulation, thus impacting quality of the hosts (Norton and Smith 1999). However, given that we are generalizing prevalence of the species from host trees at several spatial resolutions (cells of 1, 5, 10, 20 and 50 km by side), we assume that the effect of roads is constant and will not influence our general understanding of mistletoe distributions. Identification of host and mistletoe species was achieved via collection of herbarium specimens at sites every ~50 km across the study area or every time changes in host species composition were suspected. We used a second GPS unit to record landmarks by which to annotate general descriptions of species composition and vegetation physiognomy and landscape characteristics along the road; this information was used in processing aerial photographs (see below). All voucher specimens are deposited at the Ronald L. McGregor Herbarium, University of Kansas; copies of the collection were deposited at different herbaria in Mexico and the United States (see *Acknowledgments*). The two GPS units were Garmin 60CSx, with antennas to improve precision of coordinate estimates.

Host tree mapping

To derive proportions of infected trees for the process-based model, we extracted geographic coordinates of tree canopies along roads sampled. This step was achieved using National Agriculture Imagery Program (NAIP) aerial photographs, ideal for tree mapping because of their high spatial and spectral resolution (1 m and four bands—red, green, blue, and near infrared). We carried out object-oriented classification of these photographs with the software eCognition 3.0 (Baatz et al. 2003). We first identified all 3.75′ × 3.75′ aerial photographs that intersected the roads sampled. We further reduced the set of aerial photographs to those including the land use cover types in which mistletoe infections were found via comparisons with the 2001 National Land Cover Database (30 m resolution; Homer et al. 2004); land cover types considered were barren land, cultivated land, deciduous forest, development low density, development medium intensity, development

open space, emergent herbaceous wetlands, hay/pasture, herbaceous, open water, shrub/scrub, and woody wetlands. We selected at random 10% (67) of the aerial photographs for the states of Arizona (25; from 2007), New Mexico (6; 2009), California (13; 2009), and Nevada (23; 2006). Because of significant computational demands in the classification process, each 3.75′ × 3.75′ image was split into four equal portions for processing, generating a total of 268 images to classify.

In eCognition, we first segmented each image into two levels, with the following parameters: level 1 (scale = 10, color = 0.8, shape = 0.2, smoothness and compactness = 0.5) and level 2 (scale = 3, color = 0.8, shape = 0.2, smoothness and compactness = 0.5). Then, to classify trees, we selected polygons manually from the level 2 objects to be used as canopy samples, and classified each image using the standard nearest neighbor with the mean of each of the four bands (R, G, B, and NIR), the mean difference from neighbors, and the mean difference between the level 2 objects nested within the level 1 superobject. These procedures have been used in previous studies that have analyzed mesquite distributions in similar landscapes and regions (e.g., Laliberte et al. 2004). Using ArcGIS 9.3, we dissolved the raw polygon output from eCognition to obtain single-tree polygons. This step was necessary because the output sometimes contained multiple polygons subdividing single tree crowns. Although this procedure sometimes reduced numbers of trees estimated, it was not a common problem overall, and estimates of trees were close to true numbers of trees in each image ($R^2 = 0.617$, $P < 0.001$). This step allowed us to extract 5,786,586 polygons corresponding to tree crowns, and count actual numbers of trees within the 200 m strip along roads, to estimate proportions of trees infected (via the GPS coordinates of mistletoes described above).

Climate, host, and disperser summaries

These datasets were each derived at five spatial resolutions, as follows. Climatic layers used were annual mean temperature and annual precipitation, obtained from the WorldClim database (Hijmans et al. 2005), using the 30″, 2.5′, 5′, and 10′ resolution products to match 1 km, 5 km, 10 km, and 20 km resolutions, respectively; for the

50 km resolution, we scaled the values up with the mean of the 10′ raster layers using a bilinear interpolation from nearest neighbor cells. Because other climatic factors, particularly freezing temperatures and sunny days, can be important limiting factors for the mistletoe, we tested models with two other variables estimated directly from weather-station data (Easterling et al. 1999) for a 20-yr period (number of continuous days of freezing temperatures and number of continuous rainless days). Because models resulting showed similar or lower performance scores, these variables were not included in our final analyses.

The raster data layer summarizing abundances of *Phainopepla nitens* was obtained from the Breeding Bird Survey database (v.12.07.2011 [Sauer et al. 2011]), which comes as a vector-format data layer with cells at a resolution of 25 km, the product of interpolating point abundance information across the bird species' distribution in the United States. In contrast to the rest of the layers used for modeling, this data layer was used statically at its native resolution (i.e., not scale-dependent) for the first four resolutions of our analyses.

To estimate numbers of host trees across the region, we developed models to predict numbers of trees per cell at 1 km resolution using random forests (randomForest in R; Liaw and Wiener 2002). This estimate was based on 580 1-km cells sampled across the region from 85 images from the NAIP (55 images from Arizona, 7 from New Mexico, 7 from California, and 16 from Nevada) that were classified following the protocol described above. Specifically, we asked for a total of 10,000 trees developed under the random forest protocol, from which we estimated the logarithm of the number of host trees with the following statistical model: log 10 of host trees ~ GRS 2000 + NVG 2000 + annual mean temperature + annual precipitation + maximum temperature of the warmest month + precipitation of the wettest month + slope + elevation + aspect + latitude + longitude. The grass/scrub/woodland (GRS 2000) and barren/very sparsely vegetated land (NVG 2000) are raster data layers obtained from the Harmonized World Soil Database v 1.1 (Fischer et al. 2008). The final estimate of the number of host trees was more explanatory (67.0% of total variation) and showed better fit

between observed and predicted numbers of trees ($r = 0.91$, $P < 0.01$) and between observed and independent external data set aside from the classified images ($r = 0.60$, $P < 0.01$) than other random forests based on different combinations of variables, numbers of observations, and numbers of trees within the process. This procedure also gave better results than the two most explanatory (67.7–68.1% of total variation) generalized additive models (GAMs) of different combinations of same predictors as used in the random forest ($r = -0.22$, $P > 0.01$; $r = -0.16$, $P > 0.01$). Although the explained variances are similar, the shape of the relationship in the random forests is more regular and linear than with the GAMs.

Process-based model

To estimate probability of occurrence of *P. californicum* in grid cells via a process-based modeling approach, we developed a spatially explicit model incorporating the balancing rates of colonization and local extinction, as proposed in the theory of metapopulations (Hanski 1999, Vandermeer and Goldberg 2003, Roberts and Hamann 2012). The model was developed as follows.

After some assumptions related to using a mean-field approximation, we fit the following equation following Hanski's (1999) incidence function model:

$$\bar{p}_J = \frac{\bar{c}_J}{\bar{c}_J + \bar{e}_J},$$

where $\bar{p}_J$ represents the proportion of hosts in a composite cell J occupied by mistletoes. The parameter $\bar{c}_J$ is the mean colonization rate, assumed to be a function of the density of birds and host trees in the focal cell J and neighboring cells. The parameter $\bar{e}_J$ is the mean death rate of mistletoes, assumed to be a function of the distance between optimal climate, estimated as the difference of average climatic conditions across known occurrence points (for each resolution) and the climate in the cell J. This model was fit for each of the five spatial resolutions using a maximum likelihood routine implemented in R. Derivations of formulas for this model are provided in Appendix A.

To estimate variability in parameters, we resampled the data with replacement 115 times

to obtain a distribution of colonization and extinction-related parameter values with the following number of observations in each iteration: 200 (56% of the original data matrix) at 1 km resolution, 70 (63%) at 5 km, 50 (68%) at 10 km, 30 (54%) at 20 km, and 20 (64%) at 50 km. The distributions obtained were used to assess whether parameters differed from zero, and how different they were among scales, using t-tests. Specifically, we evaluated two colonization parameters, disperser abundance and disperser cost of movement across neighboring cells; in the extinction formula, we tested the deviation from a climate optimum as defined by Mahalanobis distance to the environments in the occurrence points for each resolution (see Appendix A).

Correlative models

We ran a total of 45 correlative models: three algorithms, three sets of environmental predictors, and five spatial resolutions. We used two presence-background correlative niche-modeling algorithms: DesktopGARP (best subsets version in OpenModeller v.1.1.0; de Souza Muñoz et al. 2011) and Maxent v.3.3.3 (Phillips et al. 2006), as well as the presence-absence algorithm generalized additive models (GAM; mgcv [Wood 2011]). We set parameters in GARP as follows: occurrence data for training 50%, soft omission threshold, 100 replicate runs, and best subsets option; the rest of the settings were as default. In Maxent, we set 50% of occurrence points for model calibration, and chose logistic output; the rest of settings were as default. To run models, we selected 5% ($N = 629$) of the total number of infected trees, after considering the spatial lag of spatial autocorrelations in environmental characteristics as follows. The spatial lag was estimated on the first principal component of the 19 bioclimatic variables as a surrogate for the environmental space in the study region. Principal components were calculated using a correlation matrix in R (stats; R Development Core Team 2011) and the spatial lag was estimated as the distance associated with the sill of a semi-variogram in the package Spatial Analysis in Macroecology (SAM 4.0; Rangel et al. 2010). We then sampled the 5% of the occurrence points subject to the constraint that they be separated by at least that distance in space.

We explored influences of different environ-mental factors on the distribution of *P. californi-cum* in correlative models using three sets of variables: (1) annual mean temperature and annual precipitation ("climate"); (2) abundance of *Phainopepla nitens* and numbers of host trees ("disperser + host"); and (3) a combination of all three ("climate + disperser + host"). All models were trained and projected within the extent that corresponds to the distribution of the disperser (**M**) at each spatial resolution. The importance of environmental predictors at each scale was estimated using the jackknife of regularized training gain in Maxent, recalculation of GARP accuracy using jacknife in openModeller, and the P-value for significance of each variable in GAMs. These procedures form part of the output of each algorithm.

Model evaluations and comparisons

All 50 models (45 correlative and 5 process-based) were evaluated using the partial AUC approaches implemented by Peterson et al. (2008), using 262 independent unique occurrence points for *P. californicum* from the Global Biodiversity Information Facility (www.gbif.org; search in September 2010). This procedure allowed us to evaluate performance of each model as compared to random expectations, as well as to compare performance across scales and modeling methods. Partial AUC approaches limit analysis to portions of the ROC curve that are relevant to the question at hand (i.e., within omission error tolerances); so one calculates the ratio between the area under the curve for observed values against the area under the line of random discrimination, AUC ratios are expected to depart upwards from one as model performance is better than random. The main advantage of this procedure is that the comparison covers only the range of values over which each algorithm predicts, thus avoiding problem caused by using equal scales of values when such is not the case in all comparisons (e.g., Maxent and GARP; Peterson et al. 2008).

To do this testing, we first multiplied grids by 1000, and converted each floating-point grid to an integer grid in ArcGIS 10. Using the modeled suitability values associated with each independent testing point, we implemented partial AUCs by running 1000 bootstrap simulations in a Visual Basic program (Barve, www.

biodiversity-informatics-training.org), with 50% of points resampled with replacement in each iteration of the bootstrap, and $E = 0.05$, given that these testing occurrences were obtained from GBIF and may be subject to some georeferencing or identification error. Distributions of the randomized ratios were compared with z-tests to see if ratios were consistently larger than 1 (1 corresponds to random discrimination). The ratios that resulted from this procedure were also used to compare models in a three-way ANOVA for type of algorithm, environmental predictors, resolution, and interactions of these three factors using R (stats; R Development Core Team 2011).

The degree of spatial agreement between model predictions was calculated using a fuzzy Kappa statistic in the software Map Comparison Kit (Visser and de Nijs 2006), and Pearson correlations were estimated between 3000 random points for resolutions 1–20 km and between the 527 pixels at the 50 km resolution following the Dutilleul method to correct for spatial autocorrelation in SAM 4.0 (Rangel et al. 2010), which corrects the number of degrees of freedom of estimated variances. Kappa statistics were calculated for models thresholded using the minimum training presence value (Liu et al. 2005, Wilson et al. 2005), given that we are confident that occurrence points used for model calibration correspond to a *P. californicum* with no error ($E = 0$). Finally, we compared the area predicted as suitable for each model using boxplots to examine effects of model type, environmental predictors, and spatial resolution visually.

RESULTS

All models, except for a marginally lower result in the 1 km process-based model, performed better than random expectations when tested against independent occurrence points (all $P < 0.05$; Appendix B: Table B1). Comparisons of model performance using AUC ratios indicated differences in performance of models depending on the algorithm, environmental predictors, and resolution, as well as interactions among these factors (Table 1). At all resolutions, the lowest AUC ratios were for models using climate only as environmental predictors; AUC ratios were in

Table 1. Three-way ANOVA on partial AUC ratios using as factors the different sets of variables ("climate", "disperser + host", "climate + disperser + host"), algorithms (process-based, GARP, Maxent, GAM), and resolutions (1, 5, 10, 20, 50 km), including interactions among factors. All comparisons were significantly different at $P < 10^{-15}$.

Factor	SS	df	F
Algorithm	184.031	3	16981.19
Resolution	11.253	4	778.75
Variables	110.021	2	15228.07
Algorithm : Resolution	55.616	11	1399.61
Algorithm : Variables	32.411	4	2243.01
Resolution : Variables	39.347	8	1361.5
Algorithm : Resolution : Variables	49.288	16	852.74
Residuals	177.01	49,000	

general higher for models with disperser + host and climate + disperser + host environmental predictors alone (Fig. 3). AUC ratios at resolutions of 1 and 5 km were highest when the environmental predictors were climate + disperser + host, followed by the AUC ratios when the environmental predictors were disperser + host, and comparatively lower AUC ratios when the environmental predictor was climate (Fig. 3). AUC ratios at resolutions of 10 and 20 km had lower mean ratios when the whole set of predictors were used. At the coarsest resolution (50 km), AUC ratios remained low for climate and disperser + host, but were considerably higher for climate + disperser + host (Fig. 3).

Coincidence among models, after converting them to binary maps, was highly variable in terms of both amount of area predicted as suitable and amount of overlap. Proportional areas identified as suitable differed depending on algorithm and environmental predictors, but not on resolution (Fig. 4; see also Appendix B: Fig. B1). On the other hand, amount of overlap between models at each resolution varied considerably when compared with the fuzzy Kappa statistic (Fig. 5). The spatial variation in the area identified as suitable can be divided in two broad groups: contiguous areas with high geographic coincidence and more disjunct areas with lower geographic coincidence (see an example in Appendix B: Fig. B2). In general, most map comparisons fell in the more disjunct category, meaning that completely overlapping areas are unusual (Fig. 5). Moreover, these broad groups
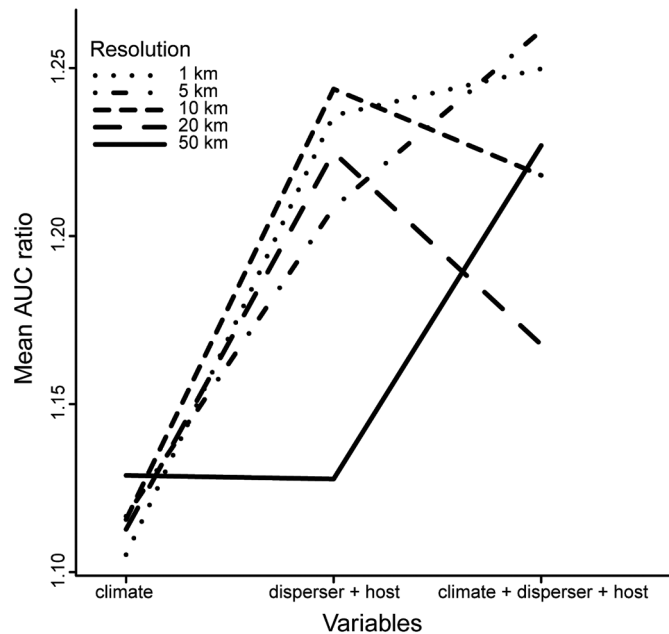
Fig. 3. Interaction plot of AUC ratios after 1000 iterations in the partial ROC based on independent occurrences of *Phoradendron californicum*. Differences in model performances are statistically significant ($P < 0.001$). Process-based model results were included in the set of variables "climate + disperser + host."

did not fall in the same categories when compared across scales. Correlations between models using raw output values instead of thresholded values coincided with results from fuzzy Kappa comparisons (Table 2).

In the process-based model, the colonization parameter associated with disperser cost of movement across the landscape showed significant departure from zero for the 1 and 5 km resolutions, and the parameter associated with disperser abundance showed significant departure from zero at all resolutions but was significantly higher at the three finest resolutions (Table 3; Fig. 6). Comparing across resolutions, the parameter associated with disperser cost of movement was significantly higher for the scales of 1 and 5 km. Also, when comparing across resolutions, the parameter associated with disperser abundance at 1 km resolution was significantly higher than those for 20 and 50 km; the value from 5 km was higher than that of the value of 20 km; and the value from 10 km was significantly higher than the values from 20 and 50 km (Table 4). Parameter values associated with the extinction function of the process-based model did not differ significantly from one

another across scales.

The importance of variables within each of the correlative models was generally consistent across algorithms and resolutions (Table 5). For the climate model, mean annual temperature was the most important variable, except in GAM at 5 km, where precipitation was more important. When models were trained with climate + disperser + host as environmental predictors, the most important variable was the disperser, except GARP at 20 km, where host was more important, and Maxent at 50 km, where annual mean temperature and annual precipitation were overall more relevant; GAMs also showed high importance of temperature at resolutions of 5 and 10 km, precipitation at 20 km, and host at 1 and 5 km. When trained with disperser + host as predictors, GARP showed host as most relevant, while Maxent and GAM showed disperser as most important (Table 5).

DISCUSSION

The two modeling approaches that we implemented are very different in nature (Morin and Thuiller 2009). In the process-based model, we
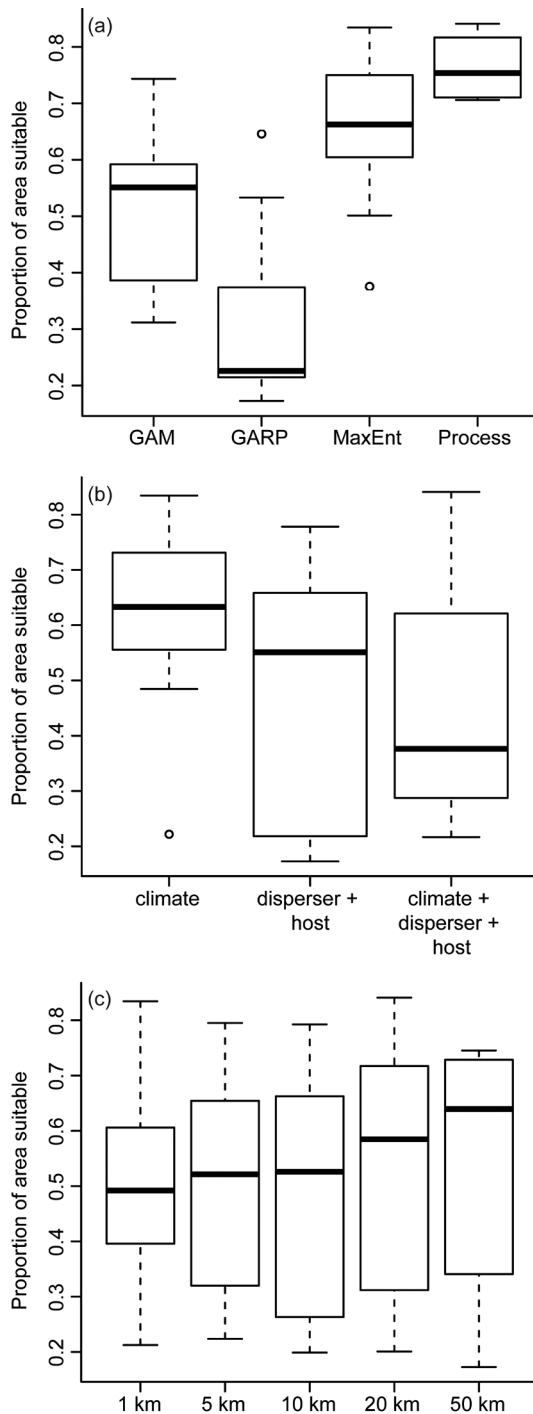
Fig. 4. Proportion of area predicted as suitable summarized by (a) model or algorithm type, (b) predictor variables, and (c) resolution. Suitable area was considered after converting models into binary predictions using a minimum training presence threshold approach.

made explicit assumptions about the way that the mistletoe could be dispersed and in this way "explore" the region, and reasons behind it successfully establishing or not given the availability of hosts and abiotic climatic conditions. On the other hand, in the correlative models, it is assumed that the effect of non-scenopoetic mechanisms covary with environmental factors. The differences and similarities between the results of the two approaches illuminate the roles of different factors across different spatial scales (Schurr et al. 2012). We could think of process-based models as a hypothetic-deductive approach by which to test hypothesis, whereas correlative models are an inductive way to propose hypotheses; the differences in results are explained by the ways in which the two methods make use of observations.

The results from the process-based model suggest that dispersal factors are most relevant at fine resolutions, but that they become practically irrelevant at coarser resolutions, where climatic factors dominate, as expected under the Eltonian Noise Hypothesis (Soberón and Nakamura 2009, Peterson et al. 2011). This result is consistent with previous general thinking about effects of scale on the relative importance of different factors on distributional ecology (Whittaker et al. 2001, Pearson and Dawson 2003), and also with previous results on spatial aggregation patterns of *P. californicum* (Aukema 2004). This later study found mistletoe aggregation within trees and at about 1.5–2 km, and again at a distance ≥4 km, that was correlated with elevation. This finding led the author to hypothesise that dispersal was key to the first two scales (within trees and ~1500 m) and abiotic factors to the >4 km scale. Although our simulations did not include these finer resolutions, our results are consistent where comparisons were possible. Coincident aggregation patterns have also been reported in studies of the Indian mistletoe *Taxillus tomentosus* (Lemieux et al. 2011).

The fact that parameters of the extinction rate did not show significant change across resolutions is consistent with the relative stability of scenopoetic variables across a broad set of scales (Soberón 2007). Scenopoetic variables are related to the abiotic fundamental niche, an adaptive feature related to physiology (James et al. 1984,
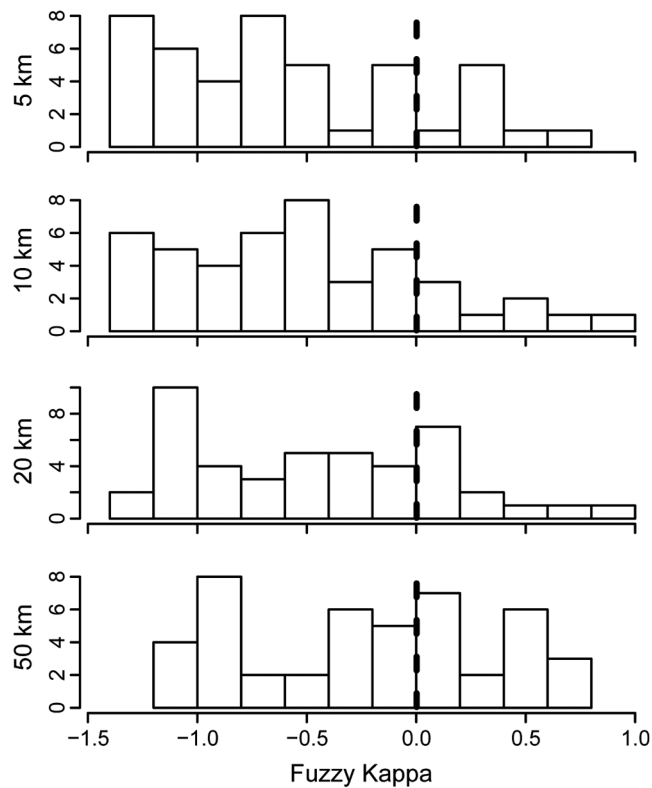
Fig. 5. Summary of degree of coincidence or overlap between pairs of maps (45 in total) using fuzzy Kappa statistics (Hagen 2003). Fuzzy Kappa shows values greater than 0 when maps have a large coincidence between areas predicted as suitable or non-suitable, and values less than 0 (either suitable or not) when maps show less contiguous more disjoint predictions.

Jackson and Overpeck 2000). In this case, climate is relevant because it adjusts species' survival limits in each cell of the grid. Overall, what is apparently changing with scale are factors associated with dispersal of the species, as seen in the colonization part of this model, whereas climate is a constant determinant of the species' range across scales. These results support the hypotheses proposed with the BAM diagram in the conceptual framework (Fig. 2).

Process-based models, which are much more time-consuming to implement than correlative models, are nevertheless potentially very informative. After comparing the models generated in this exercise, it is evident that incorporating colonization and extinction processes within a metapopulation framework sheds light on the importance that factors related to dispersal may have at finer scales for this species (Overton 1994, 1996, Aukema 2004, Lemieux et al. 2011), and

that more stable factors like climate either remain constant across resolutions or become more important in controlling the distribution of the species as the resolution becomes coarser.

Using correlative methods, we were able to perceive differences in contributions of variables as a function of changes in resolution, as other authors have shown (e.g., Buckley et al. 2010). However, interpretation of the roles of different variables, not being linked to mechanisms, was not straightforward, and potentially would lead to interpretations very different from what we observed using the process-based model (Cabral and Schurr 2010). Araújo and Luoto (2007) used GAMs to test whether biotic interactions played a role at macroscales for the butterfly *Parnassius mnemosyne*: they found that the type of interaction (contingent on species) and methods considered played significant roles in explaining distributions of species at coarse scales. Other

Table 2. Pairwise correlations between models after correcting for spatial autocorrelation with Dutilleul method in SAM (see *Methods* for details). g = GARP, gm = GAM, m = Maxent, process = metapopulation process-based model; c = climate, d = disperser, h = host.

Model comparison		1 km		5 km		10 km		20 km		50 km	
Mod 1	Mod 2	Pear.	Prob.	Pear.	Prob.	Pear.	Prob.	Pear.	Prob.	Pear.	Prob.
process	g_dh	0.38	0.011	0.31	0.082	0.31	0.086	0.18	0.241	0.24	0.182
	g_c	0.28	0.012	0.66	<0.001	0.66	<0.001	0.74	<0.001	0.66	<0.001
	g_cdh	0.34	0.017	0.34	0.052	0.37	0.04	0.2	0.188	0.3	0.112
	gm_dh	0.41	0.012	0.37	0.061	0.37	0.063	0.2	0.214	0.32	0.128
	gm_c	0.21	0.108	0.65	<0.001	0.68	<0.001	0.49	<0.001	0.42	0.008
	gm_cdh	0.37	0.023	0.42	0.036	0.44	0.019	0.26	0.095	0.31	0.14
	m_dh	0.39	0.009	0.37	0.036	0.39	0.03	0.19	0.208	0.32	0.107
	m_c	0.21	0.056	0.69	<0.001	0.78	<0.001	0.73	<0.001	0.61	<0.001
	m_cdh	0.28	0.023	0.43	0.005	0.48	0.002	0.27	0.063	0.38	0.056
g_dh	g_c	0.27	0.1	0.28	0.063	0.24	0.126	0.23	0.12	0.18	0.197
	g_cdh	0.91	<0.001	0.9	<0.001	0.89	<0.001	0.97	<0.001	0.79	<0.001
	gm_dh	0.92	<0.001	0.91	<0.001	0.93	<0.001	0.94	<0.001	0.86	<0.001
	gm_c	0.37	0.06	0.39	0.04	0.38	0.053	0.4	0.032	0.39	0.022
	gm_cdh	0.88	<0.001	0.86	<0.001	0.82	0.001	0.83	<0.001	0.76	0.001
	m_dh	0.91	<0.001	0.91	<0.001	0.94	<0.001	0.91	<0.001	0.9	<0.001
	m_c	0.26	<0.001	0.31	0.03	0.35	0.049	0.35	0.045	0.34	0.059
	m_cdh	0.75	<0.001	0.75	<0.001	0.8	<0.001	0.93	<0.001	0.76	0.003
g_c	g_cdh	0.29	0.064	0.34	0.023	0.34	0.023	0.27	0.063	0.19	0.181
	gm_dh	0.31	0.077	0.33	0.047	0.27	0.1	0.26	0.101	0.23	0.158
	gm_c	0.81	<0.001	0.78	<0.001	0.73	<0.001	0.67	<0.001	0.58	<0.001
	gm_cdh	0.4	0.023	0.42	0.015	0.4	0.012	0.38	0.012	0.3	0.064
	m_dh	0.29	0.072	0.29	0.057	0.26	0.085	0.24	0.101	0.22	0.158
	m_c	0.83	<0.001	0.88	<0.001	0.81	<0.001	0.85	<0.001	0.78	<0.001
	m_cdh	0.44	<0.001	0.46	<0.001	0.47	<0.001	0.34	0.018	0.32	0.036
g_cdh	gm_dh	0.9	<0.001	0.89	<0.001	0.91	<0.001	0.92	<0.001	0.86	<0.001
	gm_c	0.34	0.076	0.43	0.021	0.42	0.027	0.4	0.031	0.41	0.021
	gm_cdh	0.85	<0.001	0.87	<0.001	0.85	<0.001	0.83	<0.001	0.74	0.005
	m_dh	0.9	<0.001	0.91	<0.001	0.92	<0.001	0.9	<0.001	0.87	<0.001
	m_c	0.28	0.029	0.37	0.01	0.4	0.022	0.36	0.036	0.36	0.05
	m_cdh	0.75	<0.001	0.8	<0.001	0.87	<0.001	0.93	<0.001	0.81	0.002
gm_dh	gm_c	0.45	0.037	0.49	0.02	0.45	0.032	0.46	0.022	0.5	0.01
	gm_cdh	0.94	<0.001	0.95	<0.001	0.89	<0.001	0.89	<0.001	0.9	<0.001
	m_dh	0.89	<0.001	0.92	<0.001	0.96	<0.001	0.92	<0.001	0.96	<0.001
	m_c	0.26	0.074	0.34	0.032	0.39	0.039	0.39	0.034	0.43	0.038
	m_cdh	0.74	<0.001	0.77	<0.001	0.82	<0.001	0.92	<0.001	0.89	0.001
gm_c	gm_cdh	0.52	0.013	0.58	0.005	0.58	0.003	0.59	0.002	0.56	0.004
	m_dh	0.36	0.069	0.87	<0.001	0.41	0.033	0.41	0.026	0.46	0.018
	m_c	0.79	<0.001	0.42	0.01	0.89	<0.001	0.75	<0.001	0.82	<0.001
	m_cdh	0.48	0.002	0.8	<0.001	0.58	<0.001	0.46	0.01	0.6	<0.001
gm_cdh	m_dh	0.86	<0.001	0.41	0.031	0.83	0.002	0.8	0.002	0.87	<0.001
	m_c	0.33	0.022	0.83	<0.001	0.52	0.004	0.48	0.006	0.51	0.012
	m_cdh	0.76	<0.001	0.54	<0.001	0.86	<0.001	0.96	<0.001	0.86	0.002
m_dh	m_c	0.31	0.02	0.33	0.019	0.38	0.026	0.38	0.025	0.4	0.048
	m_cdh	0.81	<0.001	0.78	<0.001	0.86	<0.001	0.96	<0.001	0.89	0.001
m_c	m_cdh	0.5	<0.001	0.53	<0.001	0.56	<0.001	0.44	0.007	0.58	0.002

studies using correlative approaches have shown improvement in predictions of species' distributions when resource-related variables are incorporated (Heikkinen et al. 2007, Real et al. 2009, Wisz et al. 2012). In our exercise, when we used correlative models, we also found that biotic-related predictors play a role at coarse scales, but were most relevant at scales finer than the macroscale of Araújo and Luoto (2007). These authors worked only at a scale similar to the coarsest of our analysis (50 km); they might have

detected additional patterns if their analysis had extended to finer resolutions. Their results match ours at the coarsest resolution if we consider both disperser and host layers as part of the mistletoe's biotic interactions: in our case, the disperser layer showed some explanatory power, although not very significant, whereas, in Araújo and Luoto's example, the host plant distribution of the butterfly did not improve results drastically, but rather echoed climate signals (Figs. 3 and 4). Although these results from correlative models

Table 3. Departure from zero (*t*-test) of colonization parameters for disperser movement cost across the landscape given topographic heterogeneity and disperser abundance.

Parameter	Observed	Mean	SE	t	Prob.
Movement cost					
1 km	1.000	0.995	0.016	62.295	0.000***
5 km	0.135	0.130	0.008	16.681	0.000***
10 km	0.002	0.006	0.008	0.266	0.791
20 km	0.009	0.009	0.006	1.348	0.180
50 km	0.003	0.007	0.007	0.409	0.683
Disperser abundance					
1 km	1.649	1.434	0.514	3.206	0.002***
5 km	0.607	0.582	0.121	4.999	0.000***
10 km	1.649	0.981	0.382	4.318	0.000***
20 km	0.174	0.150	0.082	2.109	0.037
50 km	0.365	0.195	0.103	3.549	0.001

*P < 0.1, ** P < 0.05, *** P < 0.01.

Table 4. Comparisons of colonization parameters (disperser movement cost and abundance) across spatial resolutions (*t*-test) in the process-based model.

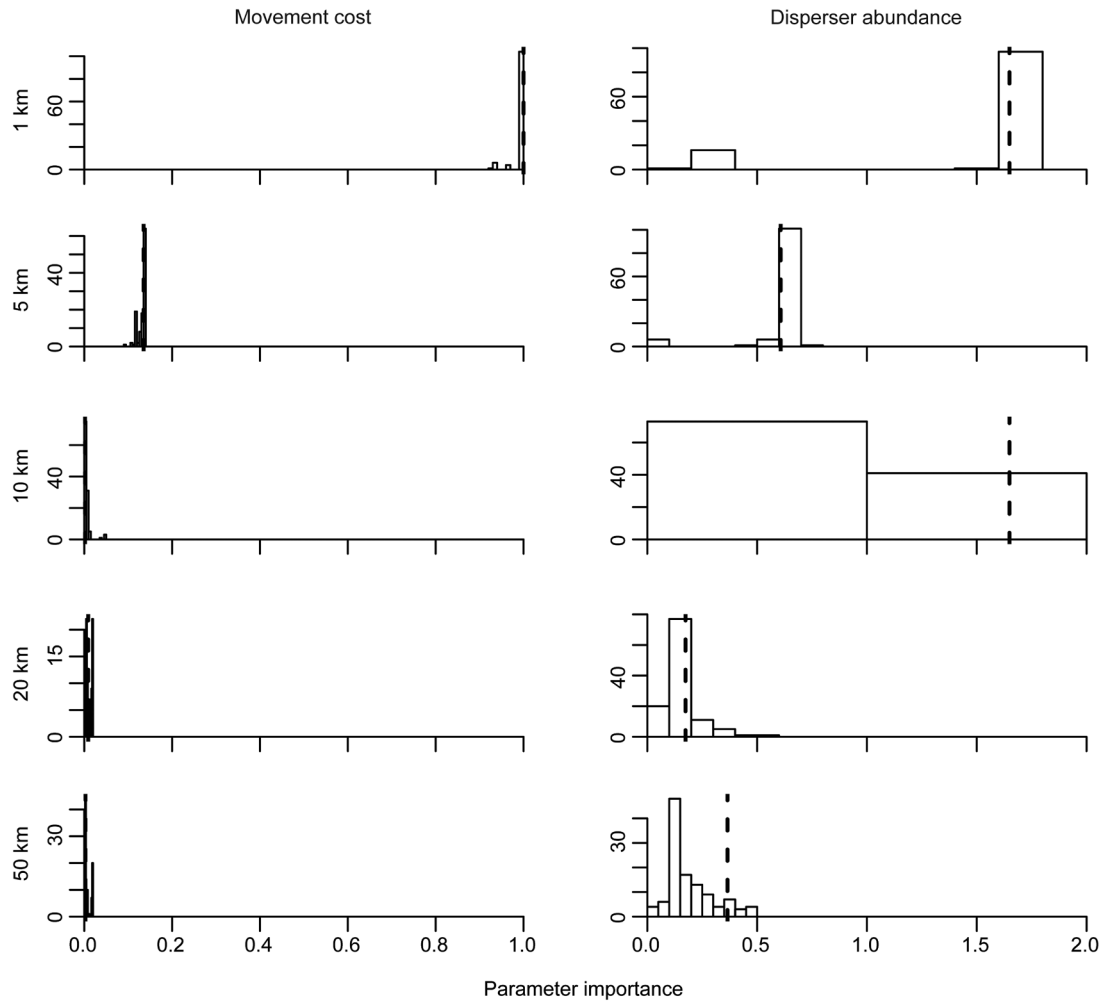| Resolution	t 1 km	t 5 km	t 10 km	t 20 km
Movement cost				
5 km	48.073***			
10 km	55.331***	11.530***		
20 km	57.345***	12.247***	−0.620	
50 km	57.386***	12.636***	−0.051	0.641
Disperser abundance				
5 km	1.972*			
10 km	0.000	−2.602**		
20 km	2.832***	2.951***	3.776***	
50 km	2.448**	1.518	3.247***	−1.451

*P < 0.1, ** P < 0.05, *** P < 0.01.

Fig. 6. Importance of parameters from the process-based model associated with (1) cost of movement for the disperser across the landscape given by topographic heterogeneity and (2) disperser abundance.

Table 5. Variable importance in correlative models calculated from jackknifing permutations in GARP and Maxent, and from significance values in GAM.

Algorithm	Model	Variable	1 km	5 km	10 km	20 km	50 km
GARP	climate	temperature	86.09	90.52	85.84	98.97	100.00
		precipitation	13.91	9.48	14.16	1.03	0.00
	climate + disperser + host	temperature	13.64	16.33	16.54	25.00	25.84
		precipitation	19.70	23.47	21.05	21.43	24.72
		disperser	45.45	25.51	41.35	14.29	35.96
		host	21.21	34.69	21.05	39.29	13.48
	disperser + host	disperser	30.30	29.31	12.00	34.38	45.45
		host	69.70	70.69	88.00	65.62	54.55
Maxent	climate	temperature	77.12	65.83	56.69	87.01	78.74
		precipitation	22.88	34.17	43.31	12.99	21.26
	climate + disperser + host	temperature	17.16	17.11	28.38	2.65	57.97
		precipitation	10.20	10.79	20.32	2.04	35.77
		disperser	70.75	66.48	51.29	93.86	2.38
		host	1.88	5.62	0.00	1.44	3.88
	disperser + host	disperser	94.50	89.88	93.34	95.47	100.00
		host	5.50	10.12	6.66	4.53	0.00
GAM	climate	temperature	0.00***	0.29	0.00***	0.00***	0.00***
		precipitation	0.67	0.005**	0.22	0.00***	0.00***
		(deviance explained)	48.80%	47.50%	47.20%	40.10%	39.60%
	climate + disperser + host	temperature	0.21	0.002**	0.00***	0.60	0.46
		precipitation	0.022*	0.65	0.014*	0.006**	0.0215*
		disperser	0.00***	0.00***	0.00***	0.00***	0.00***
		host	0.007**	0.004**	0.07	0.57	0.21
		(deviance explained)	91.60%	91.00%	91.80%	93.80%	94.20%
	disperser + host	disperser	0.00***	0.00***	0.00***	0.00***	0.00***
		host	0.80	0.85	0.70	0.38	0.002**
		(deviance explained)	87.30%	87.60%	85.60%	88.40%	87.10%

*$P < 0.1$, ** $P < 0.05$, *** $P < 0.01$.

suggest that non-scenopoetic variables are important at coarse scales, the process-based model helped us to understand interactions among predictors, and revealed contrasting importance of factors at different resolutions.

Our analysis suggests that different combinations of environmental predictors, whether biotic or abiotic, may affect interpretation of model predictions drastically, offering a step towards understanding model overfitting. Addition of the disperser layer resulted in concentration of the mistletoe's predicted distribution in the area of highest density of the disperser. Although model overfitting is known to complicate estimates of species' ecological niches (Peterson et al. 2007), in this case, it also shows how combinations of abiotic and dispersal-related predictors restrict the area predicted as suitable. An improvement on this exploration would be to test for significance of explanatory power of other (less important) dispersers and perhaps different suites of hosts in the model.

It is interesting to notice that the different modeling approaches produced predictions with only fair to good agreement. One possible explanation for such lack of coincidence is that the climate-only predictions agree closely with the distribution of the disperser, which is the element of the process-based model that covers the entire extent moving the parasite across cells in search for suitable conditions for its establishment (**M**; Barve et al. 2011). However, in the metapopulation framework of the process-based model, the disperser is just one part of the formula, and how the mistletoe reaches its distributional limits depends on the balance between colonization and extinction. Assuming that the disperser actually carries mistletoe seeds all across the study extent and that hosts are present, our study suggests that the mistletoe is able to reach and potentially establish in areas more distant than the known geographic range edges; however, its populations are restricted to an area smaller than its "accessible area," suggesting that other factors (including climate) constrain its survival. In a more realistic process-based modeling approach, actual seed processing times and dispersal distances could be incorporated (e.g., Overton 1996, 1997, Liu et al. 2011).

The fact that the species is not present in the northwestern and southeastern parts of the disperser's range suggests that the species is

responding to some combination of biotic and abiotic factors that reduces suitability in those regions (Fig. 1; see also Appendix B: Fig. B1). The two areas are consistently predicted as climatically suitable, but densities of disperser and hosts are much lower towards these two areas, very different from the central part of the study area, where we observed highest mistletoe abundances; areas similar to these areas may be worth exploring in detail for sink population dynamics. From observations during field work, the transition between absence and presence, moving from east to west, could also be explained by a combination of biogeographic barriers for the species and dramatic environmental gradients, which represent limits to dispersal (although other generalist and specialist species are present in these other regions) and successful establishment. Our observations do not suggest interference by other species (i.e., other biotic factors, such as competition) in this part of the range; rather, it is more likely that the species has not been able to adapt to abiotic conditions there, preventing it from being more widespread. This may explain why the species is restricted to the Sonoran Desert and parts of the Mojave Desert, and does not occur in other, adjacent desert areas in North America, such as California's Central Valley, the Chihuahuan Desert, or the Valle de Tehuacán in Mexico, that present different combinations of either hotter and wetter or colder and dryer climates, in comparison to areas where the species is present (Dimmit 2000).

Although our exercise is realistic according to what we know about the species' distribution and how biotic and abiotic factors operate across scales, it carries some significant assumptions and limitations. First, the environmental predictors we used and the way they were developed is far from ideal. We chose predictors that are likely important to the species, but generating them and matching their resolutions was not always possible. In particular, this concern applies to the disperser layer: it was originally created at a spatial resolution of 461 km^2 (Sauer et al. 2011), but data availability considerations prevented us from scaling it directly to other resolutions. Scaling it to high resolutions requires a larger density of survey routes than is actually available; as a consequence, we used this single resolution for the first four resolutions, and generalized it still more for the coarsest.

Other bioclimatic variables could have been chosen, such as those associated with extremes of temperature and precipitation; however, in tests using other sets of climate variables and the entire set of 19 bioclimatic variables of Hijmans et al. (2005), neither the process-based model nor the correlative approach appeared to benefit from such additional information (data not shown). To check this point more thoroughly, we tested the performance and geographic predictions of correlative models with two other environmental layers that were calculated using daily weather stations information for a period of 20 years: number of continuous days of freezing temperatures and number of continuous rainless days. However, these models did not produce results markedly better than annual mean temperature and annual precipitation visually, nor did models increase in performance, which suggests that different trends would not result in either the model prediction or in model performance.

For the process-based model, it would be possible to create and explore other colonization and extinction functions. In particular, the extinction function used was a simple exponential of the difference between the estimated optimal niche and the environments of a given cell: clearly, more complex may prove more appropriate (Martínez-Meyer et al. 2013). However, based on extensive initial exploration and simulations (not shown), in which our basic results were supported, it is not likely that major differences and conclusions from our current findings would emerge.

Other types of information could be incorporated in new models, as well as in experiments that could analyze details of mistletoe population dynamics in different parts of the range. For example, numbers of viable seeds dispersed by *Phainopepla nitens* and other bird species may indicate regions with unfavorable niche conditions and sink population dynamics, or perhaps even other physiological mechanisms of host-parasite recognition. Other studies of such dynamics in *P. californicum* have been carried out at much finer resolutions than ours, and are very illuminating when compared with our macro-geographic approximations (e.g., Overton 1994, 1996, Aukema and Martínez del Rio 2002,

Liu et al. 2011). From an ecological point of view, for example, these analyses would allow exploring whether abundance, genetic variation, population differentiation, or gene flow vary as functions of centrality within the geographic range (Thuiller et al. 2010, Torres et al. 2012, Martínez-Meyer et al. 2013) or the ecological niche of the species, thus presenting different opportunities for local optima and associated adaptation (VanDerWal et al. 2009, Martínez-Meyer et al. 2013).

ACKNOWLEDGMENTS

LITERATURE CITED

Acevedo, P., A. Jiménez-Valverde, J. M. Lobo, and R. Real. 2012. Delimiting the geographical background in species distribution modelling. Journal of Biogeography 39:1383–1390.

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

Anderson, R. 1981. Population ecology of infectious disease agents. Pages 318–355 in R. May, editor. Theoretical ecology: principles and applications. Blackwell, Oxford, UK.

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

Aukema, J. E. 2003. Vectors, viscin, and Viscaceae: mistletoes as parasites, mutualists, and resources. Frontiers in Ecology 1:212–219.

Aukema, J. E. 2004. Distribution and dispersal of desert mistletoes is scale-dependent, hierarchically nested. Ecography 27:137–144.

Aukema, J. E. and C. Martínez del Rio. 2002. Variation in mistletoe seed deposition: effects of intra- and interespecific host characteristics. Ecography 25:139–144.

Baatz, M., U. Benz, S. Dehghani, M. Heynen, A. Höltje, P. Hofmann, I. Lingenfelder, M. Mimler, M. Sohlbach, M. Weber, and G. Willhauck. 2003. eCognition 3 user guide. Definiens Imaging, Munich, Germany.

Barve, N., V. Barve, A. Jimenez-Valverde, A. Lira-Noriega, S. P. Maher, A. T. Peterson, J. Soberon, and F. Villalobos. 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. Ecological Modelling 222:1810–1819.

Brotons, L., M. De Caceres, A. Fall, and M. J. Fortin. 2012. Modeling bird species distribution change in fire prone Mediterranean landscapes: incorporating species dispersal and landscape dynamics. Ecography 35:458–467.

Buckley, L. B., M. C. Urban, M. J. Angilletta, L. G. Crozier, L. J. Rissler, and M. W. Sears. 2010. Can mechanism inform species' distribution models? Ecology Letters 13:1041–1054.

Cabral, J. S. and F. M. Schurr. 2010. Estimating demographic models for the range dynamics of plant species. Global Ecology and Biogeography 19:85–97.

Cunningham, H. R., L. J. Rissler, and J. J. Apodaca. 2009. Competition at the range boundary in the slimy salamander: using reciprocal transplants for studies on the role of biotic interactions in spatial

distributions. Journal of Animal Ecology 78:52–62.

de Souza Muñoz, M., R. de Giovanni, M. de Siqueira, T. Sutton, P. Brewer, R. Pereira, D. Canhos, and V. Canhos. 2011. openModeller: a generic approach to species' potential distribution modelling. GeoInformatica 15:111–135.

Dimmit, M. A. 2000. Biomes and communities of the Sonoran Desert region. *In* S. J. Phillips and P. Wentworth Comus, editors. A natural history of the Sonoran Desert. Arizona-Sonora Desert Museum Press and University of California Press, Berkeley and Los Angeles, California, USA.

Easterling, D. R., T. R. Karl, J. H. Lawrimore, and S. A. Del Greco. 1999. United States historical climatology network daily temperature, precipitation, and snow data for 1871-1997. ORNL/CDIAC-118, NDP-070. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee, USA.

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

Fischer, G., F. Nachtergaele, S. Prieler, H. T. van Velthuizen, L. Verelst, and D. Wiberg. 2008. Global agro-ecological zones assessment for agriculture. International Institute for Applied Systems Analysis, Laxenburg, Austria and FAO, Rome, Italy.

Gaston, K. J. 2003. The structure and dynamics of geographic ranges. Oxford University Press, Oxford, UK.

Gaston, K. J., D. R. Genney, M. Thurlow, and S. E. Hartley. 2004. The geographical range structure of the holly leaf-miner. IV. Effects of variation in host-plant quality. Journal of Animal Ecology 73:911–924.

Guisan, A., and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135:147–186.

Hagen, A. 2003. Fuzzy set approach to assessing similarity of categorical maps. International Journal of Geographical Information Science 17:235–249.

Hanski, I. 1999. Metapopulation ecology. Oxford University Press, Oxford, UK.

Hastings, A., S. Petrovskii, and A. Morozov. 2010. Spatial ecology across scales. Biology Letters 7:163–165.

Heikkinen, R. K., M. Luoto, R. Virkkala, R. G. Pearson, and J. H. Korber. 2007. Biotic interactions improve prediction of boreal bird distributions at macroscales. Global Ecology and Biogeography 16:754–763.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25:1965–1978.

Homer, C. C. H., L. Yang, B. Wylie, and M. Coan. 2004. Development of a 2001 National Landcover Database for the United States. Photogrammetric Engineering and Remote Sensing 70:829–840.

Hutchinson, G. E. 1978. An introduction to population ecology. Yale University Press, New Haven, USA.

Jackson, S. T., and J. T. Overpeck. 2000. Responses of plant populations and communities to environmental changes of the late Quaternary. Paleobiology 26:194–220.

James, F. C., R. F. Johnston, N. O. Wamer, G. J. Niemi, and W. J. Boecklen. 1984. The Grinnellian niche of the Wood Thrush. American Naturalist 124:17–30.

Kearney, M., and W. Porter. 2009. Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. Ecology Letters 12:334–350.

Kearney, M., B. A. Wintle, and W. Porter. 2010. Correlative and mechanistic models of species distribution provide congruent forecasts under climate change. Conservation Letters 3:203–213.

Kuijt, J. 2003. Monograph of *Phoradendron* (Viscaceae). Systematic Botany Monographs 66:1–643.

Laliberte, A. S., A. Rango, K. M. Havstad, J. F. Paris, R. F. Beck, R. McNeely, and A. L. Gonzalez. 2004. Object-oriented image analysis for mapping shrub encroachment from 1937 to 2003 in southern New Mexico. Remote Sensing of Environment 93:198–210.

Lemieux, C. J., T. J. Beechey, D. J. Scott, and P. A. Gray. 2011. The state of climate change adaptation in Canada's protected areas sector. Canadian Geographer 55:301–317.

Levin, S. A. 1992. The problem of pattern and scale in ecology. Ecology 73:1943–1967.

Levin, S. A., and S. W. Pacala. 1997. Theories of simplification and scaling of spatially distributed species. Pages 271–295 *in* D. Tilman and P. Kareiva, editors. Spatial ecology: the role of space in population dynamics and intersepecific interactions. Princeton University Press, Princeton, New Jersey, USA.

Liaw, A., and M. Wiener. 2002. Classification and regression by randomForest. R News 2:18–22.

Liu, R. S., C. Martínez del Rio, and J. H. Wu. 2011. Spatiotemporal variation of mistletoes: a dynamic modeling approach. Bulletin of Mathematical Biology 73:1794–1811.

Liu, C. R., P. M. Berry, T. P. Dawson, and R. G. Pearson. 2005. Selecting thresholds of occurrence in the prediction of species distributions. Ecography 28:385–393.

MacArthur, R. H. 1972. Geographical ecology: patterns in the distribution of species. Harper and Row, New York, New York, USA.

Mackey, B. G. and D. B. Lindenmayer. 2001. Towards a hierarchical framework for modelling the spatial distribution of animals. Journal of Biogeography 28:1147–1166.

Martínez-Meyer, E., D. Díaz-Porras, A. T. Peterson, and C. Yáñez-Arenas. 2013. Ecological niche structure determines rangewide abundance patterns of species. Biology Letters 9:20120637.

Monahan, W. B. 2009. A mechanistic niche model for measuring species' distributional responses to seasonal temperature gradients. PloS ONE 4:e7921.

Morin, X., and W. Thuiller. 2009. Comparing niche- and process-based models to reduce prediction uncertainty in species range shifts under climate change. Ecology 90:1301–1313.

Norton, D. A., and M. S. Smith. 1999. Why might roadside mulgas be better mistletoe hosts? Australian Journal of Ecology 24:193–98.

Overton, J. M. 1994. Dispersal and infection in mistletoe metapopulations. Journal of Ecology 82:711–723.

Overton, J. M. 1996. Spatial autocorrelation and dispersal in mistletoes: field and simulation results. Vegetatio 125:83–98.

Overton, J. M. 1997. Host specialization and partial reproductive isolation in desert mistletoe (*Phoradendron californicum*). Southwestern Naturalist 42:201–209.

Pearson, R. G., and T. P. Dawson. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global Ecology and Biogeography 12:361–371.

Peterson, A. T., M. Papeş, and M. Eaton. 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. Ecography 30:550–560.

Peterson, A. T., M. Papeş, and J. Soberón. 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. Ecological Modelling 213:63–72.

Peterson, A. T., J. Soberoìn, R. G. Pearson, R. P. Anderson, E. Martiìnez-Meyer, M. Nakamura, and M. B. Arauìjo. 2011. Ecological niches and geographic distributions. Princeton University Press, Princeton, New Jersey, USA.

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

Pulliam, R. 2000. On the relationship between niche and distribution. Ecology Letters 3:349–361.

R Development Core Team. 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rangel, T. F., J. A. F. Diniz-Filho, and L. M. Bini. 2010. SAM: a comprehensive application for Spatial Analysis in Macroecology. Ecography 33:46–50.

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

Roberts, D. R., and A. Hamann. 2012. Predicting potential climate change impacts with bioclimate envelope models: a palaeoecological perspective. Global Ecology and Biogeography 21:121–133.

Robertson, M. P., C. I. Peter, M. H. Villet, and B. S. Ripley. 2003. Comparing models for predicting species' potential distributions: a case study using correlative and mechanistic predictive modelling techniques. Ecological Modelling 164:153–167.

Sauer, J. R., J. E. Hines, J. Fallon, K. L. Pardieck, D. J. Ziolkowski Jr., and W. A. Link. 2011. The North American Breeding Bird Survey, Results and Analysis 1966-2010. Version 12.07.2011. USGS Patuxent Wildlife Research Center, Laurel, Maryland, USA.

Schurr, F. M., J. Pagel, J. S. Cabral, J. Groeneveld, O. Bykova, R. B. O'Hara, F. Hartig, W. D. Kissling, H. P. Linder, G. F. Midgley, B. Schröder, A. Singer, and N. E. Zimmermann. 2012. How to understand species' niches and range dynamics: a demographic research agenda for biogeography. Journal of Biogeography 39:2146–2162.

Soberón, J. 2007. Grinnellian and Eltonian niches and geographic distributions of species. Ecology Letters 10:1115–1123.

Soberón, J. 2010. Niche and area of distribution modeling: a population ecology perspective. Ecography 33:159–167.

Soberón, J. and M. Nakamura. 2009. Niches and distributional areas: concepts, methods, and assumptions. Proceedings of the National Academy of Sciences 106:19644–19650.

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

Thuiller, W., C. H. Albert, A. Dubuis, C. Randin, and A. Guisan. 2010. Variation in habitat suitability does not always relate to variation in species' plant functional traits. Biology Letters 6:120–123.

Torres, N. M., P. de Marco Júnior, T. Santos, L. Silveira, A. T. de Almeida Jácomo, and J. A. F. Diniz-Filho. 2012. Can species distribution modelling provide estimates of population densities? A case study with jaguars in the Neotropics. Diversity and Distributions 18:615–627.

Vandermeer, J. H., and D. E. Goldberg. 2003. Population ecology: first principles. Princeton University Press, Princeton, New Jersey, USA.

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

Visser, H., and T. de Nijs. 2006. The map comparison kit. Environmental Modelling & Software 21:346–358.

Walsberg, G. E. 1975. Digestive adaptations of *Phainopepla nitens* associated with the eating of mistletoe berries. Condor 77:169–174.

Whittaker, R. J., K. J. Willis, and R. Field. 2001. Scale and species richness: towards a general, hierarchical theory of species diversity. Journal of Biogeography 28:453–470.

Wiens, J. A. 1989. Spatial scaling in ecology. Functional Ecology 3:385–397.

Wiens, J. A. 2000. Ecological heterogeneity: an ontogeny of concepts and approaches. Pages 9–31 *in* M. J. Hutchings, E. A. John, and A. J. A. Stewart, editors. The ecological consequences of environmental heterogeneity: the 40th symposium of the British Ecological Society, held at the University of Sussex, 23-25 March 1999. Blackwell Science, Sussex, UK.

Wilson, K. A., M. I. Westphal, H. P. Possingham, and J. Elith. 2005. Sensitivity of conservation planning to different approaches to using predicted species distribution data. Biological Conservation 122:99–112.

Wisz, M. S., J. Pottier, W. D. Kissling, L. Pellissier, J. Lenoir, C. F. Damgaard, C. F. Dormann, M. C. Forchhammer, J.-A. Grytnes, A. Guisan, R. K. Heikkinen, T. T. Høye, I. Kühn, M. Luoto, L. Maiorano, M.-C. Nilsson, S. Normand, E. Öckinger, N. M. Schmidt, M. Termansen, A. Timmermann, D. A. Wardle, P. Aastrup, and J.-C. Svenning. 2012. The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. Biological Reviews 88:15–30.

Wood, S. N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society B 73:3–36.

Supplemental Material

Appendix A

Derivation of the formulas for the process-based model

We modeled population dynamics of mistletoes using formalism borrowed from metapopulation theory. We consider a grid of cells of maximum resolution, corresponding to the typical area under an average-sized legume tree, on the order of 20 m^2. We now construct coarser-resolution grids that contain maximum-resolution cells, asking for the proportion of the coarse-resolution cells that is occupied by infected trees (number of high-resolution infected cells/total number of cells in the coarse-resolution unit), that is, by trees with at least one individual of mistletoe. Let $i = 1, 2, 3 \ldots n$ denote the coarse-resolution cells in the grid. Then the following equation is proposed (Hanski 1999):

$$\frac{dp_i}{dt} = c_i(1 - p_i) - e_i p_i,$$

where p_i is the probability of species presence in cell i, dp_i/dt is its time derivative; c_i is the colonization rate in cell i and e_i is the extinction rate in cell i. In words, the growth rate of the proportion of space occupied by the mistletoe in the i-th cell increases proportionally to a colonization rate c_i and in inverse proportion to an extinction rate e_i.

Now, imagine that the grid is coarsened, such that inside a new larger cell J (we use a capital letter to denote an aggregate of smaller cells) is a collection of the higher-resolution cells i. This can be achieved by using square cells and doubling the side of the cells in the grid: one doubling contains 4 smaller cells, another doubling contains 16 smaller cells, and so on.

We are interested in the mean value of the occupied proportion of a cell:

$$\bar{p}_J = \frac{1}{|J|} \sum_{i \in J} p_i,$$

and therefore:

$$\frac{d\bar{p}_J}{dt} = \frac{d}{dt} \frac{1}{|J|} \sum_{i \in J} p_i = \frac{1}{|J|} \left(\sum_{i \in J} c_i(1 - p_i) - \sum_{i \in J} e_i p_i \right)$$

$$= \frac{1}{|J|} \left(\sum_{i \in J} c_i - \sum_{i \in J} c_i p_i - \sum_{i \in J} e_i p_i \right).$$

Now, we make the reasonable assumption that, inside every larger cell J, the values of colonization rate and extinction rates can be approximated as their averages over the smaller cells:

$$\frac{d\bar{p}_J}{dt} = \frac{1}{|J|} \left(\sum_{i \in J} c_i - \sum_{i \in J} c_i p_i - \sum_{i \in J} e_i p_i \right)$$
$$= \bar{c}_J - \bar{c}_J \bar{p}_J - \bar{e}_J \bar{p}_J,$$

which, in the steady state, implies:

$$\bar{p}_J* = \frac{\bar{c}_J}{\bar{c}_J + \bar{e}_J}.$$

The equations for the colonization and extinction rates are postulated after consideration of the BAM diagram: colonization is proportional to the abundance of birds and trees in cell J, and inversely related to the distance to other occupied cells (Hanski 1999):

$$\bar{c}_J = \beta_2 \sum_{K=1}^{n} L_K \exp\left(\frac{-d_{J,K}\beta_0}{1 + \beta_1 b_K} \right),$$

where β_0 and β_1 represent parameters to be fitted, $d_{J,K}$ is a measure of distance (Euclidean, corrected by topography) between cells J and K; b_K is the density of birds present in cell K, and L_K is the number of trees in cell K. These three last quantities are obtained from data (see *Methods* for more detail).

The extinction rate in cell J is assumed to be a function of the distance from the centroid of the fundamental niche of the species to the environmental data in the cell, as represented in an ellipsoid:

$$\bar{e}_J = f\left[(\mathbf{x}_J - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_J - \boldsymbol{\mu}) \right],$$

where the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ represent fitted values that define a multi-normal distribution of optimal, ideal environmental preferences. The values of x_J represent the observed environmental values in the cell J.

Appendix B

Table B1. Summary of model performance. For each model it is shown the area predicted as suitable or unsuitable after using a minimum training presence threshold and the partial AUC ratio statistics.

| Resolution	Model	Area 0	Area 1	AUC min	AUC max	AUC mean	AUC SD
1 km	process	178579	1181774	0.973	1.156	1.030	0.026
	g_dh	1076124	290319	1.122	1.550	1.321	0.069
	g_c	704875	662533	1.045	1.299	1.132	0.051
	g_cdh	1051953	314490	1.142	1.601	1.418	0.082
	gm_dh	694012	672431	1.035	1.270	1.089	0.026
	gm_c	561353	806055	1.029	1.328	1.073	0.027
	gm_cdh	825443	541000	1.041	1.265	1.089	0.026
	m_dh	538460	827983	1.110	1.530	1.297	0.087
	m_c	226380	1141028	1.048	1.270	1.110	0.042
	m_cdh	492828	873615	1.092	1.556	1.242	0.063
5 km	process	15471	38806	1.096	1.247	1.180	0.019
	g_dh	42135	12142	1.049	1.428	1.111	0.034
	g_c	25470	28807	1.032	1.281	1.091	0.035
	g_cdh	41897	12380	1.040	1.348	1.107	0.035
	gm_dh	26332	27651	1.036	1.570	1.239	0.089
	gm_c	22009	32268	1.038	1.276	1.130	0.049
	gm_cdh	36712	17271	1.184	1.599	1.397	0.088
	m_dh	18661	35322	1.114	1.617	1.276	0.081
	m_c	11120	43157	1.068	1.332	1.129	0.031
	m_cdh	33715	20268	1.099	1.510	1.360	0.084
10 km	process	2817	10764	1.106	1.272	1.203	0.024
	g_dh	10880	2701	1.043	1.309	1.099	0.030
	g_c	10567	3014	1.025	1.240	1.073	0.026
	g_cdh	10008	3573	1.071	1.552	1.165	0.062
	gm_dh	6099	7482	1.125	1.604	1.344	0.095
	gm_c	5731	7850	1.044	1.296	1.125	0.051
	gm_cdh	8471	5110	1.073	1.635	1.259	0.129
	m_dh	4583	8998	1.050	1.575	1.288	0.083
	m_c	3331	10250	1.094	1.267	1.149	0.036
	m_cdh	6775	6806	1.068	1.532	1.245	0.054
20 km	process	542	2866	1.082	1.278	1.188	0.030
	g_dh	2724	684	1.053	1.357	1.124	0.040
	g_c	1591	1817	1.028	1.279	1.087	0.035
	g_cdh	2670	738	1.052	1.614	1.122	0.042
	gm_dh	1472	1935	1.089	1.558	1.280	0.105
	gm_c	1070	2337	1.022	1.254	1.070	0.028
	gm_cdh	2345	1062	1.030	1.454	1.089	0.037
	m_dh	746	2617	1.112	1.546	1.270	0.065
	m_c	951	2412	1.081	1.330	1.182	0.031
	m_cdh	1340	2023	1.111	1.624	1.272	0.074
50 km	process	157	377	1.124	1.285	1.227	0.027
	g_dh	436	91	1.018	1.074	1.038	0.008
	g_c	189	345	1.019	1.214	1.075	0.033
	g_cdh	408	119	1.039	1.256	1.094	0.028
	gm_dh	137	397	1.002	1.420	1.056	0.053
	gm_c	196	338	1.033	1.241	1.133	0.055
	gm_cdh	352	182	1.077	1.563	1.249	0.098
	m_dh	143	384	1.105	1.571	1.289	0.101
	m_c	136	398	1.103	1.257	1.178	0.024
	m_cdh	212	322	1.167	1.595	1.337	0.085
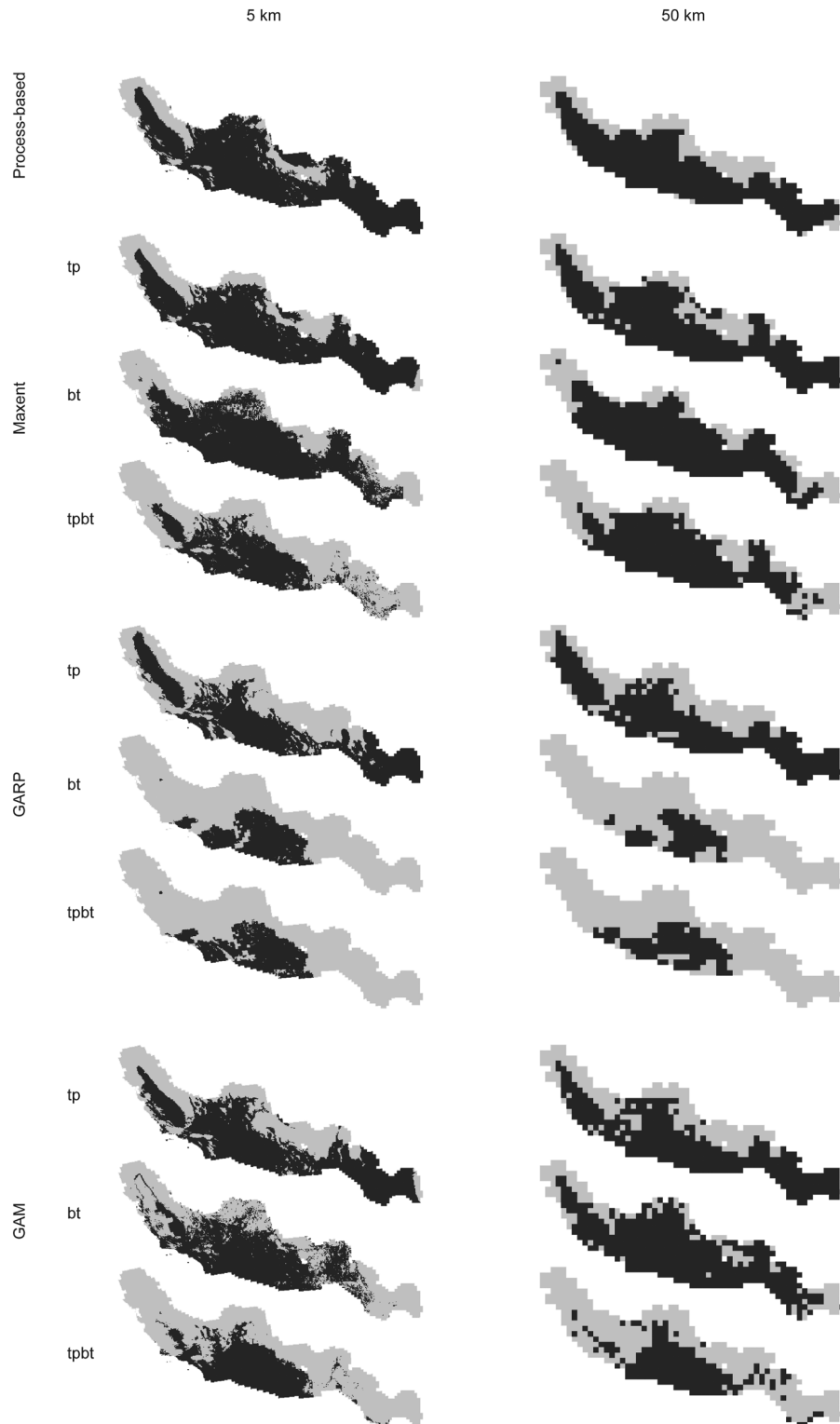
Fig. B1. Example binary maps for different models and sets of environmental predictors at the resolutions of 5 and 50 km. Binary predictions were generated using a minimum training presence threshold.
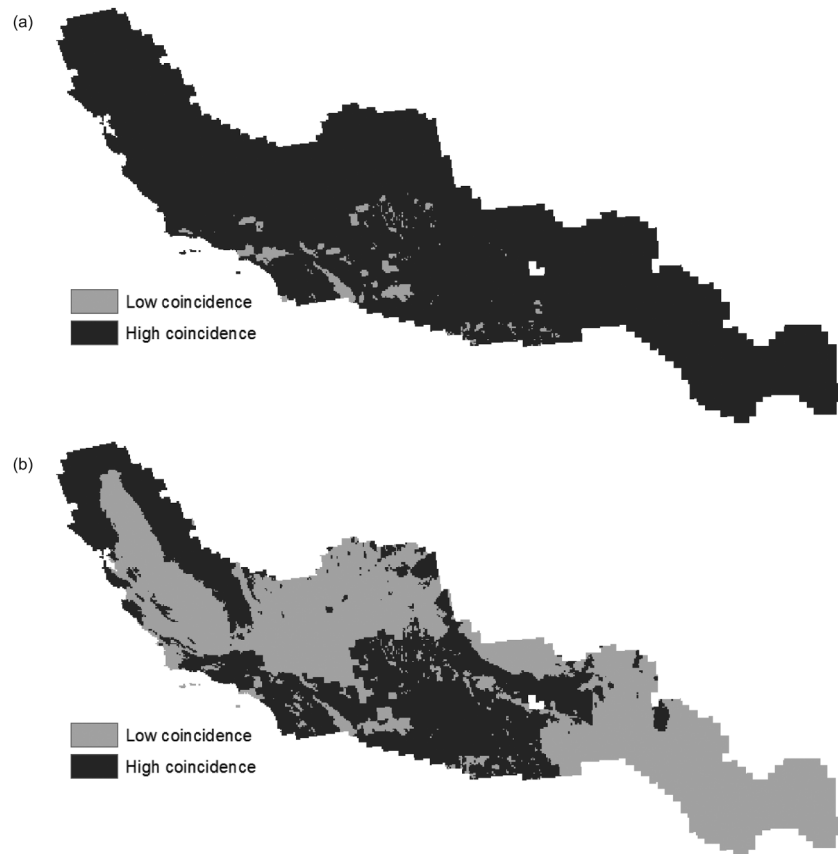
Fig. B2. Example of map comparison with fuzzy Kappa statistic. In (a), note an overall high coincidence between two maps (in this case GARP disperser + host versus GARP climate + disperser + host), and in (b), note an example of low coincidence (GARP climate + disperser + host versus Maxent climate) at a resolution of 5 km.