Global Ecology and Biogeography, (Global Ecol. Biogeogr.) (2009) **18**, 50–63

RESEARCH PAPER

Characterizing and predicting species distributions across environments and scales: Argentine ant occurrences in the eye of the beholder

S. B. Menke[1]*, D. A. Holway[1], R. N. Fisher[2] and W. Jetz[1]

[1]*Division of Biological Sciences, University of California at San Diego,USA,* [2]*US Geological Survey, Western Ecological Research Center, San Diego Field Station, USA*

ABSTRACT

Aim Species distribution models (SDMs) or, more specifically, ecological niche models (ENMs) are a useful and rapidly proliferating tool in ecology and global change biology. ENMs attempt to capture associations between a species and its environment and are often used to draw biological inferences, to predict potential occurrences in unoccupied regions and to forecast future distributions under environmental change. The accuracy of ENMs, however, hinges critically on the quality of occurrence data. ENMs often use haphazardly collected data rather than data collected across the full spectrum of existing environmental conditions. Moreover, it remains unclear how processes affecting ENM predictions operate at different spatial scales. The scale (i.e. grain size) of analysis may be dictated more by the sampling regime than by biologically meaningful processes. The aim of our study is to jointly quantify how issues relating to region and scale affect ENM predictions using an economically important and ecologically damaging invasive species, the Argentine ant (*Linepithema humile*).

Location California, USA.

Methods We analysed the relationship between sampling sufficiency, regional differences in environmental parameter space and cell size of analysis and resampling environmental layers using two independently collected sets of presence/absence data. Differences in variable importance were determined using model averaging and logistic regression. Model accuracy was measured with area under the curve (AUC) and Cohen's kappa.

Results We first demonstrate that insufficient sampling of environmental parameter space can cause large errors in predicted distributions and biological interpretation. Models performed best when they were parametrized with data that sufficiently sampled environmental parameter space. Second, we show that altering the spatial grain of analysis changes the relative importance of different environmental variables. These changes apparently result from how environmental constraints and the sampling distributions of environmental variables change with spatial grain.

Conclusions These findings have clear relevance for biological inference. Taken together, our results illustrate potentially general limitations for ENMs, especially when such models are used to predict species occurrences in novel environments. We offer basic methodological and conceptual guidelines for appropriate sampling and scale matching.

Keywords

Argentine ant, California, ecological niche models, GIS, invasion, *Linepithema humile*, model averaging, spatial grain, species distribution models.

*Correspondence: Sean Menke, Department of Entomology, North Carolina State University, 3324 Gardner Hall, Box 7613, Raleigh, North Carolina 27695-7613, USA.
E-mail: sean_menke@ncsu.edu

DOI: 10.1111/j.1466-8238.2008.00420.x © 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd www.blackwellpublishing.com/geb

INTRODUCTION

Species distribution models (SDMs) or ecological niche models (ENMs) are increasingly used to draw inferences about the determinants of species occurrences, to generate predictions about species distributions in novel regions and to forecast how distributions may shift as a result of global change (Guisan & Zimmermann, 2000; Peterson, 2003, 2006; Guisan & Thuiller, 2005). A common goal of these modelling efforts involves the production of a statistical model that is general not only across the present global distribution of a species but also across its future distribution. The development of universally general models may be hindered in a number of ways (Fielding & Bell, 1997; Pearson & Dawson, 2003). Mechanistically, ENMs are static models in that they assume no changes in the fundamental niche of a species (Peterson, 2003). Methodologically, analyses must often contend with data that cover only portions of the geographic range of a species: incomplete samples may fail to encompass the full range of environmental conditions present within a region (Fig. 1). When such occurrence data are then used to parametrize ENMs, the models may not be transferable to other regions or scales (Guisan & Zimmermann, 2000; Pearson & Dawson, 2003; Thuiller *et al.*, 2004; Randin *et al.*, 2006). Moreover, predictive models cannot account for novel climates (Williams *et al.*, 2007). These issues are exacerbated when independent data from a different sampling regime are unavailable to robustly validate model predictions (Araújo *et al.*, 2005). Each of these limitations can confound inferences concerning the relative importance of different abiotic constraints and result in inaccurate predictions about occurrences. Given that ENM predictions will continue to inform management decisions about the potential ranges of introduced species and range shifts caused by global change, it is imperative to develop a more quantitative understanding of biases that result from sampling limitations.

Accurate predictions of species distributions based on environmental conditions hinge on adequate sampling of environmental variation (Gillison & Brewer, 1985). Because any two geographical regions will differ both in the distribution

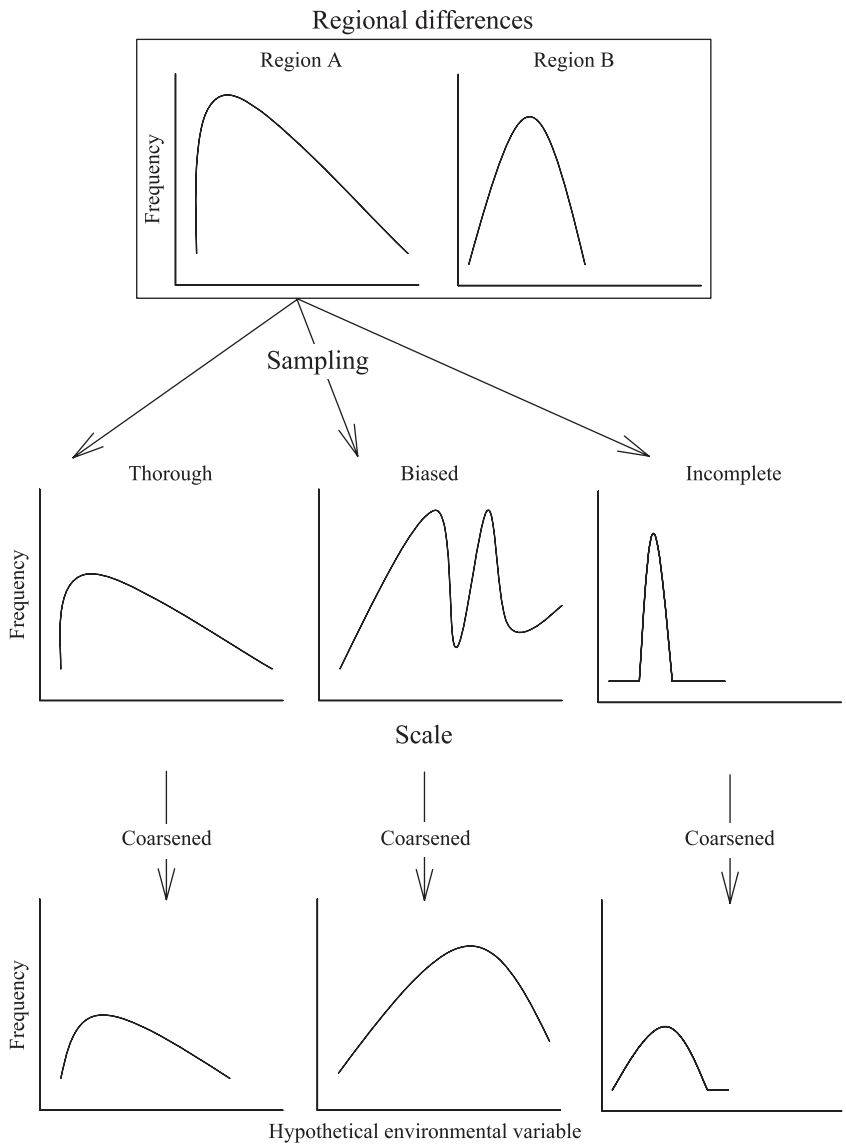

Figure 1 Conceptual diagram of issues that result from inter-regional disparities in environmental variation, insufficient sampling of environmental variation and coarsening of data across spatial grain. Biased sampling covers the full range of data but misrepresents parts of the gradient. Incomplete sampling fails to cover the range of the gradient.

© 2008 The Authors

and range of their environmental variation, unrealistic model extrapolations may result: (1) when the environmental profile of the region into which predictions are being made does not match that of the sampling region, or (2) when biased or incomplete environmental sampling occurs (Fig. 1). Additional sampling problems can result when environmental data collected at one spatial grain are used to predict species occurrences at a different grain (McPherson *et al.*, 2006). This circumstance may result from scale-dependent changes in the sampling distribution (Fig. 1) or mechanistically from changes in the importance of environmental variables at different grains (Luoto *et al.*, 2007; Whittingham *et al.*, 2007).

Because introduced species disrupt ecosystems and drain financial resources, modelling their distributions now represents an important part of ENM research. Predicting the potential ranges of species actively undergoing range expansion presents unique challenges and opportunities (Peterson, 2003). First, biotic interactions can influence where species successfully invade (Levine & D'Antonio, 1999; Bruno *et al.*, 2003), and these interactions may alter the effects of environmental change on patterns of species occurrence (Suttle *et al.*, 2007). Second, obtaining accurate absence data for species that are actively expanding their ranges is often impossible (Guisan & Thuiller, 2005). The lack of reliable absence data complicates interpretations of the factors controlling species occurrence. Unlike the case for ENMs that focus on species restricted to a single biogeographical realm, however, modelling efforts that focus on widespread introduced species can exploit cases where the native range and multiple invaded regions can all be used to create spatially independent tests of model predictions (Roura-Pascual *et al.*, 2006; Fitzpatrick *et al.*, 2007). Distribution models for introduced species now encompass a range of organisms and spatial scales (Korzukhin *et al.*, 2001; Morrison *et al.*, 2004; Rew *et al.*, 2005; Muñoz & Real, 2006).

Here we use two extensive, independently collected presence/absence data sets for the Argentine ant (*Linepithema humile*) in southern California to quantify the effects of incomplete sampling and scale dependency on ENM accuracy. Other studies have addressed questions relating to the performance of different modelling techniques (Elith *et al.*, 2006; Pearson *et al.*, 2006), the role of sample size (Stockwell & Peterson, 2002; McPherson *et al.*, 2004) and the effects of spatial grain (McPherson *et al.*, 2006; Guisan *et al.*, 2007). However, with one exception (McPherson *et al.*, 2004), these studies rely either on presence-only data (Kadmon *et al.*, 2003) or pseudo-absence data (Roura-Pascual *et al.*, 2004, 2006; Fitzpatrick *et al.*, 2007). Our study is distinct in that we use real presence/absence data to simultaneously and comprehensively analyse how sampling, scale and regional disparities affect ENM predictions for a widespread invader. Specifically, we address the following questions. (1) How does insufficient sampling of environmental variation affect model predictions? (2) How transferable are landscape-scale models created in different regions? (3) What is the effect of spatial resolution (i.e. cell size of analysis and resampling of environmental data) on the performance of predictive variables?

METHODS

Study system

The Argentine ant is a widespread, abundant and ecologically damaging invasive species (Holway *et al.*, 2002a). Native to central South America (Tsutsui *et al.*, 2001; Wild, 2004), *L. humile* now occurs worldwide, especially in areas with Mediterranean-type climates (Suarez *et al.*, 2001). Argentine ants have been present in California for approximately a century, but they only invade areas with suitable levels of soil moisture (Ward, 1987; Holway, 2005; Menke & Holway, 2006). Biotic resistance from native ants is at most a weak influence in determining where Argentine ants invade in this region (Holway, 1998; Menke *et al.*, 2007). Because of the close and well-understood correspondence between the physical environment and Argentine ant occurrence, the lack of biotic resistance from native ants and time since establishment, this system is ideally suited to the goals of this study.

Several recent studies have predicted the potential distribution of Argentine ants. Using presence-only data collected from the native range, Roura-Pascual *et al.* (2004) used the Genetic Algorithm for Rule-set Prediction (GARP) to predict the global distribution of Argentine ants under different climate change scenarios. Roura-Pascual *et al.* (2006) again used GARP with presence-only data collected from different invaded regions to predict *L. humile* distributions. Adopting a more mechanistic approach, Hartley *et al.* (2006) used the physiological tolerances of the Argentine ant to develop a bioclimatic envelope model. Although the resulting global distribution overlapped that predicted by Roura-Pascual *et al.* (2004), the model produced by Hartley and colleagues yielded a broader potential distribution on every continent considered. At a regional scale, Hartley & Lester (2003) and Krushelnycky *et al.* (2005) created predictive models based on temperature-dependent colony growth in New Zealand and Hawaii, respectively. Our study extends this body of previous work in that we use two independently collected, high-resolution presence/absence data sets to test limitations of ENMs across both region and scale.

Sampling regions

To assess how differences in sampling environmental variation affects model generality, we collected presence/absence data from two adjacent and climatically similar regions of southern California (Fig. 2), where Argentine ants are widespread. Our definition of region is arbitrarily based on political boundaries; in fact both sampling regions encompass the same climatological realms, therefore *a priori* we expect models based on data collected in either region to perform comparably. The first data set was collected in the northern region (area 25,550 km^2). This sampling effort was initiated by researchers at UC San Diego (R. N. Fisher & T. J. Case) and the US Geological Survey (USGS) to inventory reptiles and amphibians in natural areas (Fisher *et al.*, 2002); an ant sampling protocol was later added (Laakkonen *et al.*, 2001). The northern region included 348 sites, 69 of which were invaded (Fig. 2). The second data set was collected by S.B.M. in the southern

© 2008 The Authors

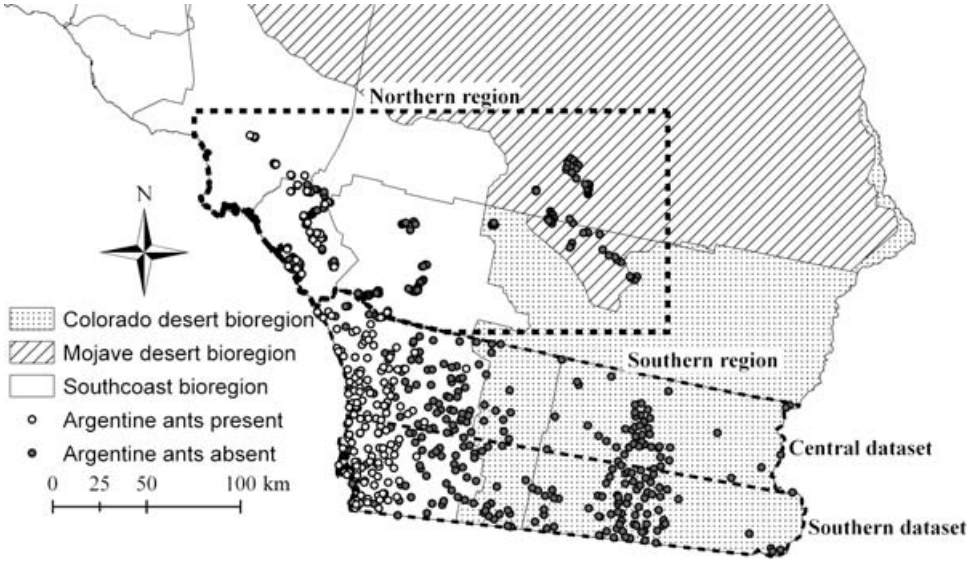

Figure 2 Map of southern California showing the distribution of Argentine ants in the northern and southern survey regions.

region (22,584 km^2). Sampling in this region was specifically tailored to the environmental heterogeneity in the region and to known patterns of Argentine ant occurrence. The southern region included 399 sites, 139 of which were invaded by Argentine ants. We split the data from the southern region into two roughly equal-area data sets (south and central) to produce two separate, well-sampled data sets, one for model training and one for testing the models created by both the haphazardly and well-sampled data sets (Fig. 2).

Sampling protocols differed between the two regions, but they both yielded accurate information about Argentine ant presences and absences. Because Argentine ants forage throughout the year, displace nearly all above-ground foraging native ant species and are strongly constrained by physical limitations in California's seasonally dry environment, it is possible to unambiguously determine whether *L. humile* is present or absent (Ward, 1987; Holway, 1995; Menke *et al.*, 2007). At each northern site, ants were sampled using five pitfall traps configured like the five on a die, with corner traps separated by 40 m. At all northern sites classified as absences, no Argentine ants were captured in any pitfall trap at that site. All northern sites were sampled four or more times, including summer and winter, in each of 2 years between 1999 and 2005.

To sample the full habitat and climate heterogeneity of the southern region, we allocated sampling points to different land-cover types based on their spatial extent and presumed probability of *L. humile* occurrence given pitfall trap results from the northern region. Land-cover types included agriculture, barren ground, conifer, desert, hardwood, herbaceous, shrub and urban; these land-cover types were further subdivided into 30 vegetation categories. We used Hawth's analysis tools for ArcGIS® to randomly place sampling points in each vegetation category based on the proportion of the land-cover type occupied by that category. To determine whether or not Argentine ants were present at sites in the southern region, we used visual surveys and bait transects (tuna baits placed every 5 m along two 50-m transects in the shape of a cross). We surveyed each site for 45 min or until we detected Argentine ants. Surveys took place in the spring and summer of 2006. A random subset of survey sites ($n = 40$) in which Argentine ants were recorded as absent were re-sampled at least 1 month later and no Argentine ants were detected.

Environmental predictor layers

Potential predictor variables of Argentine ant occurrence included a set of nine environmental layers selected from a range of possible variables because of their common usage in bioclimatic and ENMs and their putative importance in influencing Argentine ant occurrence (Holway, 1998; Hartley *et al.*, 2006; Menke & Holway, 2006; Roura-Pascual *et al.*, 2006; Menke *et al.*, 2007). Climate variables included maximum July temperature (*maxT*), minimum January temperature (*minT*), average annual rainfall (*ppt*) and normalized difference vegetation index (*NDVI*). *MaxT*, *minT* and *ppt* are averages from 1966–95 at 1-km resolution (Franklin *et al.*, 2001). *NDVI* was averaged from 16-day composites (28 July–12 August) recorded between 2000 and 2002 at 250-m resolution by the National Oceanic and Atmospheric Administration's (NOAA) advanced high-resolution radiometer satellite. Habitat variables included land-cover type (*veg*, eight levels), distance to intermittent water (*Iwater*) and distance to perennial water (*Pwater*). *Veg* is based on the life-form category in the Multi-source Land Cover Data (version 02_2) (resolution 100 m) compiled by the California Department of Forestry and Fire Protection. *Iwater* and *Pwater* were derived from high-resolution data sets in the National Hydrography Database. Human-impact variables included distance to agricultural area and distance to nearest highway or urban area (*human*). These distances were determined using the Multi-source Land Cover Data (version 02_2) and the US Census Bureau Tiger 2k (version 7 June 2002).

© 2008 The Authors

All variables were re-sampled from their native resolution to 100-m resolution. We attempted to account for potential nonlinear combinations of predictors by log $(x + 1)$ transforming all continuous variables (except for *minT*, which was log $(x + 10)$ transformed). The 100-m spatial resolution was selected to match that of the Argentine ant sampling points (all of which were separated by at least 500 m) and the highest resolution habitat data layers.

Occurrence models

We first identified the important predictors of Argentine ant occurrence in the northern and southern regions separately. For each of the nine environmental predictors, we performed logistic regression (GLM, binary with logit link). For every predictor we randomly selected two-thirds of the presences and absences as training data. We then used three measurements of model accuracy: area under the curve (AUC) of receiver operating characteristic (ROC) plots, Cohen's kappa and sensitivity and specificity to quantify the predictive ability of every model in each region with the remaining third of the data (Cumming, 2000; McPherson *et al.*, 2004). ROC plots were obtained by plotting all sensitivity values on the *y*-axis against their equivalent (1 – specificity) values on the *x*-axis. The resulting AUCs (which range from 0 to 1 with '1' indicating a perfect fit of the model, and '0.5' indicating randomness) provide a measure of model accuracy that is independent of a particular probability cut-off (Fielding & Bell, 1997). Cohen's kappa records overall agreement between predictions and observations, corrected for agreement expected to occur by chance. Values range between –1 and 1 with '1' indicating perfect agreement and values ≤ 0 suggesting a performance no different from random (Fielding & Bell, 1997). The calculation of sensitivity/specificity and Cohen's kappa requires the transformation of probabilistic predictions of occurrence into binary predictions of presence/absence; we used a threshold of 0.5 for this transformation. Training and testing were iterated 30 times and the average results are reported. Variables with AUC values above 0.75 were considered to be strong predictors (Swets, 1988; Hosmer & Lemeshow, 2000) of Argentine ant occurrence, and those environmental predictors were used in further analyses.

Model generality and ecological inference

Using simple logistic regressions, it remained unclear whether the predictors that explain *L. humile* occurrence are the same between regions and also how they should be combined to create an overall multi-predictor model. This uncertainty arose from high information overlap among certain predictors (i.e. collinearity) (see Table S1 in Supporting Information). We therefore used a model averaging technique described by Burnham & Anderson (2002) to test how the relative fits of models containing all possible combinations of predictors varied between regions. This approach uses Akaike's Information Criterion (AIC) to compare the relative fits of a suite of candidate models (Stephens *et al.*, 2005). The absolute size of the AIC is unimportant, instead differences in AIC values among models indicate the relative support for different models. For each model, we calculated an 'Akaike weight,' w_i. For a set of models, the w_i sum to 1 and have a probabilistic interpretation: of these models, w_i is the probability that model i would be selected as the best-fitting model if the data were collected again under identical circumstances. This approach yields 95% confidence sets (i.e. the smallest subset of candidate models for which the w_i sum to 0.95).

Using this model averaging approach we created 95% confidence sets to determine how well data collected in one region predict occurrence in the adjacent region. We calculated separate confidence sets for the northern and southern regions with the five environmental predictors with AUCs ≥ 0.75 from the simple logistic regression analyses. We also included *NDVI* because of its demonstrated importance in previous distribution models of Argentine ants (Roura-Pascual *et al.*, 2004; Roura-Pascual *et al.*, 2006) and *ppt* because of the known importance of soil moisture in determining *L. humile* distribution. The performances of the best two models for each region were then tested with data from the opposite region.

Model generality across spatial grain

We used data from the thoroughly sampled southern region to determine how the spatial resolution of the analysis affects the importance of different environmental predictors. First, we re-sampled Argentine ant occurrence data and the seven environmental layers used in the regional analyses (all data layers were re-sampled from their native resolution) to three coarser resolutions: 1 km, 5 km and 10 km. For each environmental predictor, we then used Pearson's chi-square goodness of fit tests to compare the sampling distribution at each resolution to its native distribution. At spatial grains coarser than 100 m, cells with multiple samples were assigned a 'presence' value if any of the subsamples included a presence. To assess the relative importance of individual variables at each spatial grain, we used AUC values. Confidence sets were then created for each spatial grain using all combinations of the seven environmental predictors.

RESULTS

Sampling of environmental variation

Region

We first evaluate how regional disparities in environmental conditions and the thoroughness of sampling may bias ENM predictions. Even though the north and south regions are adjacent, span a north–south distance of 200 km and encompass similar climatological realms (Fig. 2 & Fig. S1), the south and central data sets from the southern region differ strikingly from the northern region with respect to the frequency distributions of particular environmental variables, e.g. *maxT* (central vs north: $\chi^2 = 630.35$, d.f. = 21, $P < 0.001$; southern vs north: $\chi^2 = 694.79$, d.f. = 21, $P < 0.001$) (Fig. 3a). While the ranges of *maxT* are similar among the three data sets, the distributions of

© 2008 The Authors

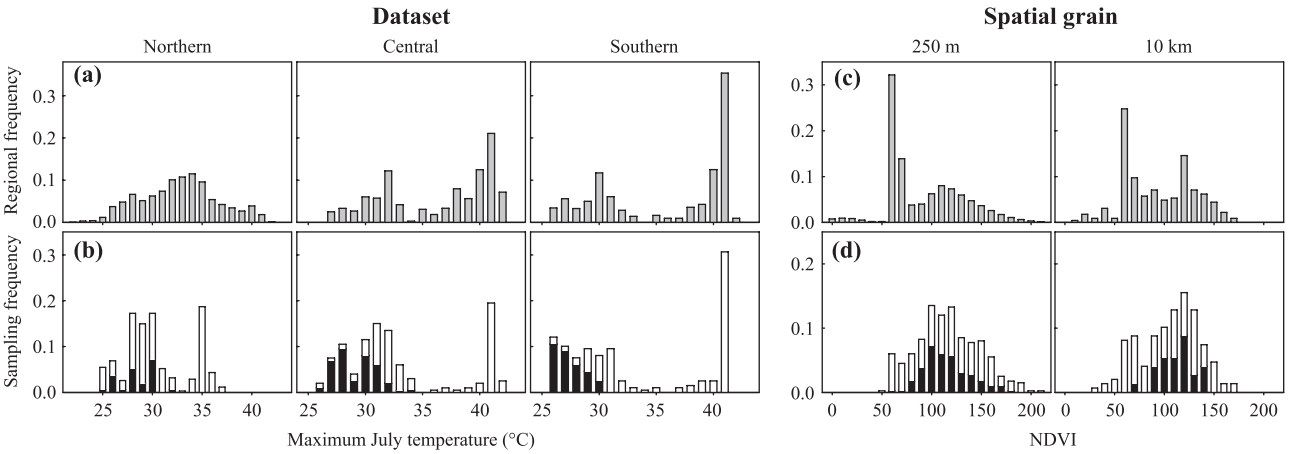

Figure 3 Empirical distributions of maximum July temperature (°C) for (a) 1-km cells in each data set and (b) at every sampled location in each data set. Empirical distributions of normalized difference vegetation index (NDVI) in the southern region at (c) 250-m and 10-km resolution and (d) at every sampled location. Gray bars represent empirical distribution; black bars are presences; white bars are absences.

maxT in the southern and central data sets more closely resemble one another; both distributions exhibit strong bimodality with the desert portions frequently experiencing temperatures exceeding 40 °C. The distribution of *maxT* for the northern region, in contrast, is unimodal with temperatures averaging 32.7 °C. Insufficient sampling of environmental conditions exacerbates these existing regional disparities. Based on the sampled environmental variation in our three data sets, insufficient sampling appears to be of greatest concern in the northern region where sampled temperatures did not exceed 37 °C, which was a modal temperature in the southern region (Fig. 3b). Sampled localities for Argentine ants occurred in areas beyond their physiological tolerances (Holway *et al.*, 2002b; Schilman *et al.*, 2005) in each region (Fig. 3b).

Spatial grain

The ranges and frequency distributions of environmental conditions vary with grain; this problem may further bias ENM predictions. The distribution of *NDVI* in the southern region, for example, becomes truncated and changes in shape when the spatial resolution of analysis is coarsened from its native resolution of 250 m to 10 km ($\chi^2 = 202.17$, d.f. = 16, $P < 0.0001$) (Fig. 3c). These changes are reflected in shifts in the distributions of sampled data; the distribution at 10 km resembles a mirror image of the distribution at 250 m (Fig. 3d). Also, the range of NDVI values associated with *L. humile* presences at 10 km (NDVI 70–140) is compressed compared with the distribution of values at 250 m (NDVI 60–180) ($\chi^2 = 44.21$, d.f. = 16, $P = 0.0002$) (Fig. 3d).

Model performance

Region

The regional disparities in environmental variation summarized in Fig. 3 result in diverging and erroneous model predictions. To illustrate this problem, we examine model predictions for *L. humile* occurrence that are based on the bioclimatic model ($y \sim maxT + minT + ppt + NDVI$; see Hartley *et al.*, 2006, and Roura-Pascual *et al.*, 2004 & 2006). When this model is parametrized with data from the northern region and tested with a subset of data from the northern region that was excluded from model creation, the model performs well over all (AUC 0.909) but is weak in predicting Argentine ant presences (sensitivity 0.587). Model predictions result in a disjointed pattern of occurrence along the coast and in the interior deserts (Fig. 4a). When the same model is transferred to the southern region and tested with the well-sampled central data set, performance is poor (AUC 0.650, sensitivity 0.311, specificity 0.746): this model predicts widespread Argentine ant occurrence in many desert locations where *L. humile* does not occur, and it fails to predict their presence along much of the coast where this species is widespread and common (Fig. 4b). In contrast, when the same bioclimatic model was parametrized with only the southern data set, it performs well throughout the southern region, both when tested with the excluds central data set (AUC 0.956) and when tested with a subset of the southern data set (AUC 0.972) (Fig. 4c).

Spatial grain

As with the regional differences summarized in Fig. 3, disparities in the frequency distributions of occurrence data and environmental data also arise across changing spatial resolution. The best AIC model created from the well-sampled southern region predicts Argentine ant presence in urban environments at low elevations irrespective of whether or not those locations are in the desert (where *L. humile* has not been found) at both the 100-m scale (the native resolution of habitat variables) and the 1-km scale (the native resolution of the climate variables) (Fig. 5a). As the spatial resolution is coarsened to 5 and 10 km, the best AIC models increasingly predict more widespread Argentine ant occurrence along the coastal plain (e.g. extending farther inland

© 2008 The Authors

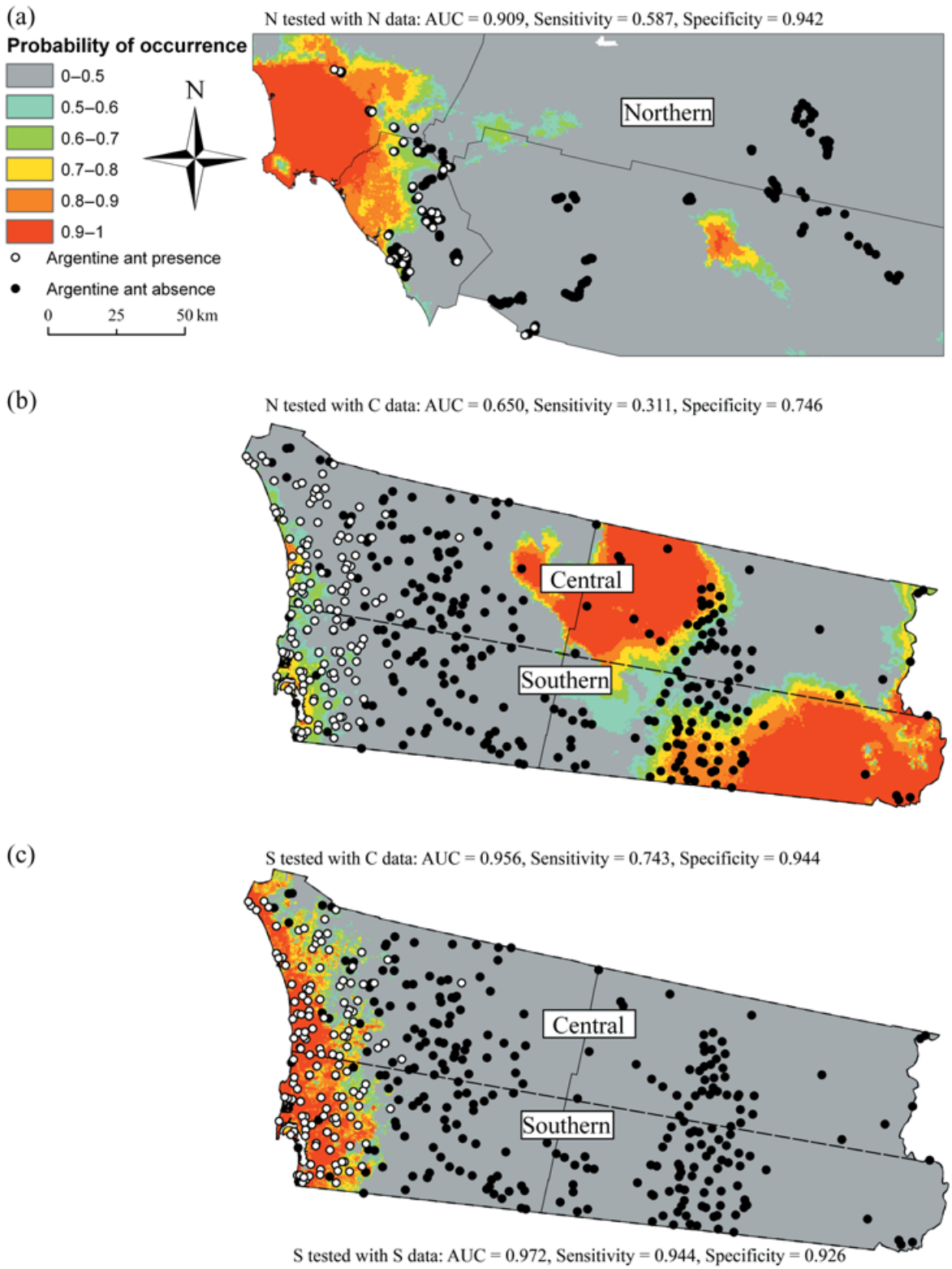

Figure 4 Effects of region on predicted Argentine ant occurrence. Regional differences are based on a climate envelope GLM ($y \sim maxT + minT + NDVI + ppt$) that used occurrence data from (a) the northern region (N) to predict Argentine ant distribution in N, (b) N to predict Argentine ant distribution with the central data set (C), and (c) the southern data set (S) to predict Argentine ant distribution with C and S. Model performance within region (N to N and S to S) was evaluated with the remaining third of the data that were not included in model creation and results are the average of 50 iterations. All data from C were used to evaluate the northern and southern models. A threshold of 0.5 was used to calculate sensitivity and specificity. *maxT*, maximum July temperature; *minT*, minimum January temperature; *ppt*, average annual precipitation; *NDVI*, normalized difference vegetation index.

© 2008 The Authors

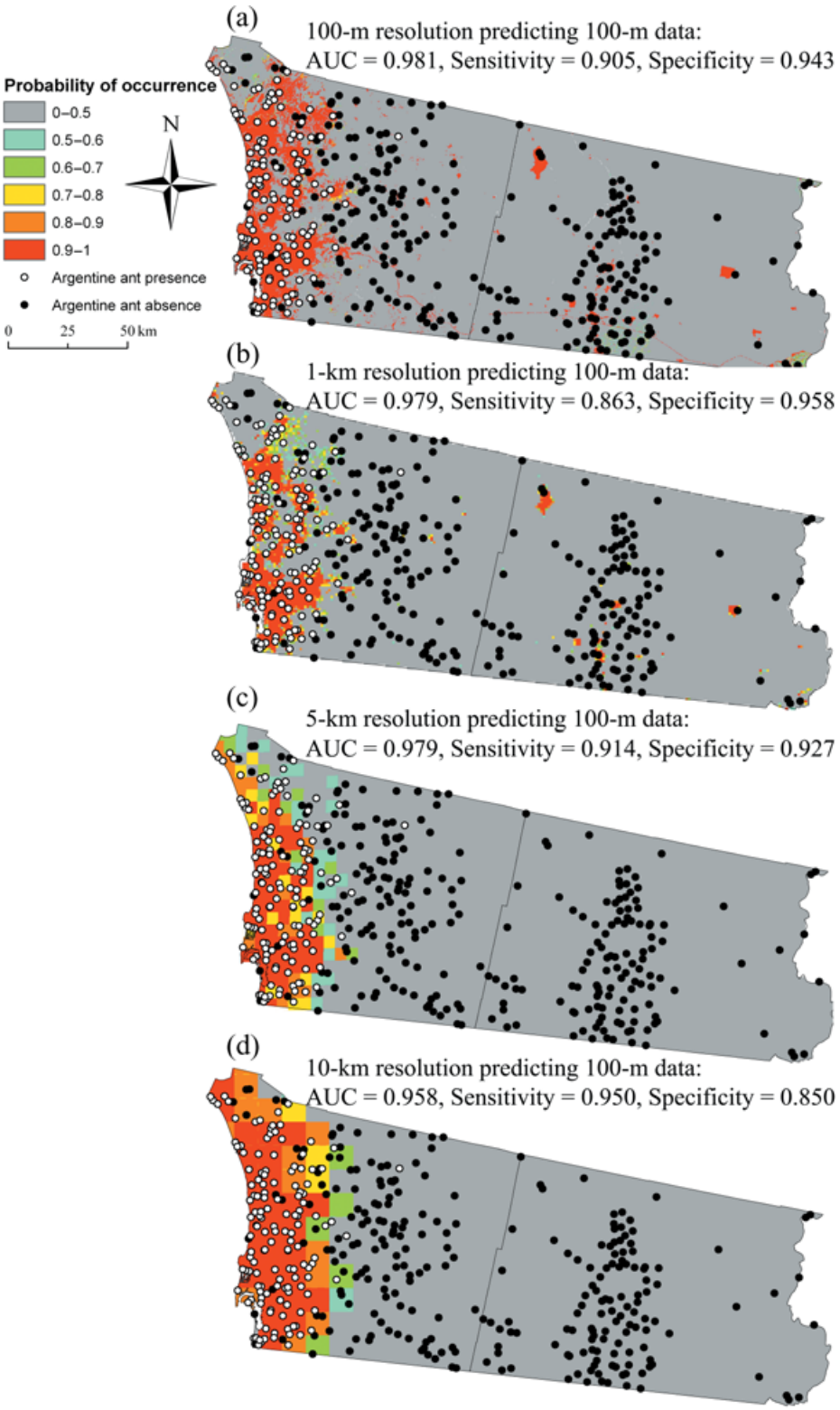

Figure 5 Effects of spatial resolution on predicted Argentine ant occurrence. Differences in spatial resolution are based on the best Akaike Information Criterion (AIC) model parametrized with data from the southern region at (a) 100-m resolution ($y \sim minT + NDVI + ppt + veg + Pwater$), (b) 1-km resolution ($y \sim minT + ppt + veg + Pwater + human$), (c) 5-km resolution ($y \sim maxT + minT + NDVI + ppt + human$) and (d) 10-km resolution ($y \sim minT + NDVI + ppt$). The performance of each model was evaluated with occurrence data at the 100-m resolution. *maxT*, maximum July temperature; *minT*, minimum January temperature; *ppt*, average annual precipitation; *NDVI*, normalized difference vegetation index; *veg*, land-cover type; *Pwater*, distance to perennial water; *human*, distance to nearest highway or urban area.

© 2008 The Authors

Table 1 Single-predictor logistic regressions for nine environmental variables used to predict Argentine ant occurrence. The best explanatory variables (AUC > 0.75) are highlighted in bold; parameter estimates that switched signs between regions are italicized.

	Region							
	Northern				Southern			
Variable	AUC	*k*/sens/spec	Slope	Intercept	AUC	*k*/sens/spec	Slope	Intercept
Maximum summer temperature (°C)	0.69	–0.01/0.00/0.99	–8.15	26.60	**0.91**	**0.62/0.72/0.89**	–22.28	76.61
Minimum winter temperature (°C)	**0.91**	**0.47/0.48/0.94**	26.62	–74.48	**0.90**	**0.69/0.76/0.92**	22.46	–61.43
NDVI	0.67	0.02/0.02/0.99	3.93	–20.28	0.46	0/0/1	0.09	–1.04
Precipitation (cm)	0.64	–0.01/0/0.99	3.76	–14.48	0.57	–0.22/0/0.81	0.97	–3.74
Land cover	**0.80**	**0.54/0.47/0.98**	Cat.	Cat.	**0.85**	**0.64/0.68/0.93**	Cat.	Cat.
Distance to intermittent water (m)	0.51	0/0/1	–0.19	1.40	0.70	0.05/0.11/0.94	–0.45	3.14
Distance to perennial water (m)	**0.77**	**0.21/.018/0.97**	*–0.91*	*5.03*	0.58	–0.05/0/0.96	*0.22*	*–2.10*
Distance to agricultural zone (m)	0.56	0.01/0.01/1	*–0.26*	*0.63*	0.55	0/0/1	*0.10*	*–1.21*
Distance to human modified area (m)	**0.82**	**0.31/0.30/0.95**	–0.91	4.20	**0.88**	**0.54/0.62/0.90**	–0.51	1.74

AUC, area under the curve; NDVI, normalized difference vegetation index; *k* = Cohen's kappa; sens, sensitivity; spec, specificity; Cat., categorical. A threshold of 0.5 was used to calculate *k*, sens and spec.

from the coast, consolidating in non-urbanized environments where they were previously predicted to be absent and disappearing from interior deserts (Fig. 5a–d)). Model performance declines at coarser resolutions, mostly as a result of a decreased ability to correctly predict Argentine ant absences.

Model generality and ecological inference

Region

To quantify how model generality changes among regions, we compared the biological interpretations of models parametrized in the insufficiently sampled northern region with those from the well-sampled southern region. Despite the large number of samples and the geographical proximity of the two survey regions (Fig. 2 & Fig. S1), environmental predictors of Argentine ant occurrence differed between the two regions (Table 1). While *minT*, *veg* and *human* were strong predictors in both regions, other predictor variables were unique to one region or the other. *MaxT* was the best predictor in the southern region (AUC = 0.91), but it only had a weak effect on occurrence in the northern region (AUC = 0.69). Conversely, *Pwater* was a strong predictor in the northern region (AUC = 0.77) but performed no better than random in the southern region (AUC = 0.58). Even though *veg* was important in both regions, parameter estimates for two land-cover types shifted in sign from north to south: hardwood tree (–1.82, 0.63) and herbaceous (–1.28, 0.74).

The importance of individual environmental predictors (Table 1, Fig. S2) did not necessarily correspond to how often those predictors appeared in model confidence sets from the model averaging analyses (Fig. 6a). Though there was inter-regional overlap in model confidence sets, models that were shared between regions had a 62% probability of being selected as the best model according to AIC in the southern region but only a 23% probability of being the best model in the northern region (Fig. 6c). Further, individual models parametrized in the southern region performed somewhat better in the northern region than did models created in the northern region and tested in the southern region (Table S2).

Spatial grain

Model averaging analyses revealed that with coarsening spatial grain all habitat predictors became less common in model confidence sets, whereas climate predictors tended to become more prevalent or to not change in frequency (Fig. 6b). For example, *NDVI* appeared in a higher percentage of models at the coarsest spatial scale, whereas *veg* and *Pwater* became less common in model confidence sets (Fig. 6b). *Human* remained in at least 99% of all models until the 10-km spatial grain at which point it occurred in only 36% of the models and also became an unimportant univariate predictor (Fig. 6b, Table S4). Unlike the model averaging results, no other predictors aside from *human* changed in importance (0.75 AUC cut-off) at coarser spatial resolutions (Table S4). The composition of confidence sets changed dramatically with changes in spatial grain (Fig. 6d). The extent to which model confidence sets overlapped with the confidence set at the 100-m scale diminished with decreasing spatial resolution to the point that by 10 km no models were shared in common (Fig. 6d); a similar pattern held when the analysis was re-run using the 1-km scale (i.e. the original scale of the climate variables) as the smallest spatial resolution. This loss in model overlap resulted in large part from habitat variables decreasing in importance at coarser spatial grains (Fig. 6b). For example, at the 100-m grain, the confidence set comprised three models, each with at least six predictor variables, while at the 10-km grain, the confidence set included 15 models, each with only three important predictor variables (Table S3).

© 2008 The Authors

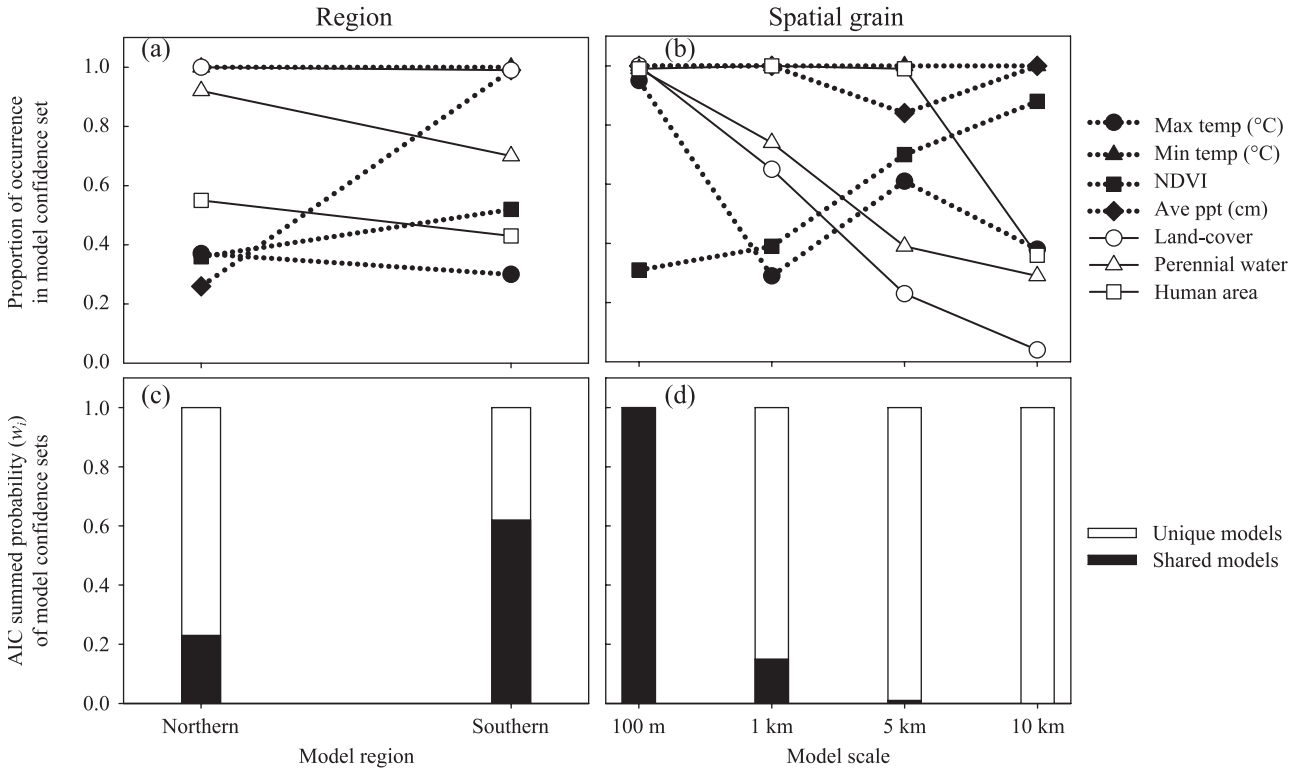

Figure 6 Importance of model parameters across region and spatial grain. The proportion of models in the model averaging confidence sets that contained each predictor (a) between the northern and southern regions and (b) across spatial grains (southern region only). Amount of the Akaike information criterion (AIC) summed probability (w_i) explained by model confidence sets comparing (c) northern and southern regions and (d) each spatial grain with 100-m spatial grain (southern region only).

Changes in the predictive ability of different environmental variables and in the frequency of their inclusion in model confidence sets across spatial resolution may be related to how the sampling distributions of different variables change with coarsening spatial resolution (Fig. 3d). We performed a simple analysis to test this idea. A significantly negative rank correlation exists between the difference in predictive strength of a variable [measured as ΔAUC = (the performance of a variable parameterized at 100 m resolution and tested at 10 km) – (its performance at 100 m)] and the difference in its frequency distribution between 10 km and 100 m (measured as the Pearson's chi-square goodness-of-fit values: Spearman rank correlation: $\rho = -0.79$, $P = 0.023$). The same trend was confirmed for 100 m vs 5 km ($\rho = -0.82$, $P = 0.0362$), but not for 100 m vs 1 km ($\rho = -0.43$, $P = 0.337$). This association suggests that as the spatial resolution is coarsened, the sampling distributions of habitat variables become more compressed relative to those of climate variables. The relatively reduced variability of habitat variables at coarser spatial resolutions may restrict their predictive ability.

DISCUSSION

Efforts to model species distributions attempt to identify factors that determine current range limits, to predict potential ranges of actively spreading introduced species and to forecast future range shifts resulting from climate change. The findings of our study illustrate that caution is warranted when transferring ENMs to new regions and in making biological inferences from ENMs. We add nuance to such general caution by identifying two factors that mediate the relative performance of models, namely space and scale.

The two regions in our study are in close proximity, and we obtained sizeable and accurate presence/absence data sets for both regions (Fig. 2). We therefore expected that particular environmental variables and patterns of Argentine ant occurrence would exhibit similar relationships throughout the sampling area. In contrast we found divergent environmental associations in the two regions. While models performed reasonably well in predicting occurrences within the region in which the models were parametrized (e.g. AUC north = 0.909), we found that the ability of those same models to accurately predict into adjacent regions was poor (e.g. AUC north predicting south = 0.650). The observer perspective, in terms of environments sampled, determined the environmental associations identified: occurrences and their correlates were in the eye of the beholder, and the generality of models beyond their training regions was poor. Simply sampling with different levels of thoroughness (Figs 1 & 3b) yielded multivariable models that differed in their composition, ecological interpretation and ability to predict occurrence (Fig. 6) (Table S1). These findings illustrate the importance of

© 2008 The Authors

sufficiently sampling the multivariate environmental space not only of the region in which the model is created but also in the region into which predictions are being made.

Prediction errors resulting from disparities in sampling as well as regional variation in environmental conditions are cause for concern given that a key goal of ENMs involves accurately predicting distributions outside the original study area and into regions that may differ strongly in their environmental conditions (Thuiller *et al.*, 2004; Randin *et al.*, 2006; Peterson *et al.*, 2007). Unfortunately, the accuracy of ENMs is usually assessed with data collected from the same region or with a random subset of the data that are withheld from model creation (Kadmon *et al.*, 2003; Elith *et al.*, 2006; Bahn & McGill, 2007). The resulting proximity of test data points to those used for training causes a high spatial autocorrelation of both response and predictor variables (Bahn & McGill, 2007). These associations compromise statistical inference (i.e. model fit) and limit the interpretation of model generality outside the environments sampled and, often with that, outside the region studied. Our finding that incomplete sampling led to poor model accuracy contrasts with results from presence-only models, for which weaker model performance can occur when the full extent of species distributions are sampled (Kadmon *et al.*, 2003). This discrepancy may be caused by differences in model prediction error from omission to commission that result from the increasing ratio of presence-to-absence points in a sample (McPherson *et al.*, 2004). These findings underscore the necessity of using independently collected data to gauge model accuracy (Fielding & Bell, 1997). Recent modelling efforts have begun to use occurrence data from widely disparate regions (often different continents) to study model transferability (Roura-Pascual *et al.*, 2006; Fitzpatrick *et al.*, 2007; Peterson *et al.*, 2007). Our results demonstrate that one potential risk with this approach concerns the error associated from building a model based on sampling of the environmental gradient in one region and then transferring that model to another region in which the environmental gradient is not represented.

We find that when models are not parametrized for the environmental gradient in a particular region, large errors in the predicted species distribution can result (Fig. 4). Studies that attempt to predict the spread of introduced species must confront such problems when they consider independent regions with distinct and often non-overlapping distributions of environmental conditions, or when they use data sets that differ in the resolution of environmental data layers or in the resolution at which occurrence data are recorded (Roura-Pascual *et al.*, 2006; Fitzpatrick *et al.*, 2007). These problems may be compounded in studies that attempt to predict range shifts caused by global change because climate predictions call for the appearance of environments with no present analogue (Pearson & Dawson, 2003; Williams *et al.*, 2007). These issues may be further complicated when ENMs are used to predict range shifts of introduced species that have altered their niche requirements following establishment in a novel environment (Urban *et al.*, 2007).

In addition to issues related to sufficient sampling and independently collected data, model accuracy can also be affected by the scale of analysis. While model performance did not decline with spatial scale as dramatically as it did between regions, the predicted spatial extent of Argentine ant occurrence did change dramatically (Fig. 5). Factors constraining the occurrence of taxa change with scale (McPherson *et al.*, 2006; Luoto *et al.*, 2007), but with such changes so does a researcher's ability to isolate these key factors. The representation of environmental conditions in data used for model parametrization is scale dependent (Figs 1 & 3d). In general, coarsening the spatial grain of data tends to decrease model performance (McPherson *et al.*, 2006; Guisan *et al.*, 2007). In our analyses, for example, a loss of model accuracy apparently resulted from an increase in the over-prediction of occurrences (Fig. 5), which were caused in large part by the diminishing importance of habitat variables with increasing grain (Fig. 6b). Recent studies have also noted that habitat variables, as well as variables influenced by anthropogenic activity, often decrease in explanatory power at coarser spatial grains (Luoto *et al.*, 2007; Pautasso, 2007; Whittingham *et al.*, 2007). This form of scale dependence may be related to the size of species ranges: narrowly ranging or specialist taxa may be more strongly associated with fine-grain habitat variables, compared to how more wide-ranging species or generalists respond to environmental variation (McPherson *et al.*, 2004; Menéndez *et al.*, 2007). With respect to the Argentine ant, a species strongly limited by its environmental tolerances, patterns of occurrence at the local scale in southern California strongly depend on fine-scale differences in soil moisture (Holway, 2005; Menke & Holway, 2006; Menke *et al.*, 2007). This association weakens at coarser spatial grains when climatic factors assume greater importance (Fig. 5b). Changes in the relative importance of environmental variables with spatial grain result from a combination of changes in the sampling distribution of a variable and its scale-dependent performance. Taken together, these findings illustrate that care should be exercised to match the spatial resolution of predictions to the environmental correlates used to create the models (Karl *et al.*, 2000; McPherson *et al.*, 2006).

CONCLUSIONS

In this study, we systematically analysed how the accuracy of model predictions and the inferences drawn from them hinge on sufficient sampling, independently collected data from different regions and the spatial grain of environmental variables. Model predictions were strongly influenced by the thoroughness of sampling (Figs 3 & 4). Moreover, we found that variables changed in importance at different spatial resolutions: climatic factors became more important at coarser resolutions, while habitat variables became less important. These results are due in part to changes in the sampling distribution and in the performance of variables at different spatial resolutions. In the light of these findings, we recommend that efforts to model species distributions: (1) use both presence and absence data when appropriate, (2) sample across the environmental tolerance of a species, (3) sufficiently sample the environmental parameter space of the region into which predictions will be made, (4) test model predictions in a distinct region with independently collected data, (5) use variables at the appropriate spatial grain, and (6) make model predictions

© 2008 The Authors

at the same spatial resolution as model parametrization. Exercising appropriate levels of caution in efforts to model species distributions will increase the likelihood that the interpretations and predictions of SDMs are biologically meaningful.

ACKNOWLEDGEMENTS

This research was supported through a NOAA-GRF to S.B.M., USDA-NRI award (2006-35302-17255) to D.A.H and NSF grant BCS-0648733 to W.J. For technical support, we thank C. Brown, S. Daeschner, R. Freckleton, E. Goldberg, S. Hathaway, D. Martin, T. Matsuda, J. McPherson, G. Nichols, K. Pease, C. Rochester, A. Vandergast and P. Zarnetske. The BLM, Nature Reserve of Orange County, The Nature Conservancy, California Department of Fish & Game, City of San Diego, USGS, Bureau of Reclamation, Cabrillo National Monument Foundation, Chino Hills State Park and Mountain Recreation Conservation Authority funded ant collections. For comments on the manuscript, we thank R. Dunn, M. Fitzpatrick, J. Franklin, A. T. Peterson, A. Syphard, P. Ward and two anonymous referees. The use of trade, product or firm names in this publication is for descriptive purposes only and does not imply endorsement by the US government.

REFERENCES

Araújo, M.B., Whittaker, R.J., Ladle, R.J. & Erhard, M. (2005) Reducing uncertainty in projections of extinction risk from climate change. *Global Ecology and Biogeography*, **14**, 529–538.

Bahn, V. & McGill, B.J. (2007) Can niche-based distribution models outperform spatial interpolation? *Global Ecology and Biogeography*, **16**, 733–742.

Bruno, J.F., Stachowicz, J.J. & Bertness, M.D. (2003) Inclusion of facilitation into ecological theory. *Trends in Ecology and Evolution*, **18**, 119–125.

Burnham, K.P. & Anderson, D.R. (2002) *Model selection and multimodel inference: a practical information – theoretic approach*, 2nd edn. Springer – Verlag, New York, NY.

Cumming, G.S. (2000) Using between-model comparisons to fine-tune linear models of species ranges. *Journal of Biogeography*, **27**, 441–455.

Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Townsend Peterson, A., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S. & Zimmermann, N.E. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38–49.

Fisher, R.N., Suarez, A.V. & Case, T.J. (2002) Spatial patterns in the abundance of the coastal horned lizard. *Conservation Biology*, **16**, 205–215.

Fitzpatrick, M.C., Weltzin, J.F., Sanders, N.J. & Dunn, R.R. (2007) The biogeography of prediction error: why does the introduced range of the fire ant over-predict its native range? *Global Ecology and Biogeography*, **16**, 24–33.

Franklin, J., Keeler-Wolf, T., Thomas, K.A., Shaari, D.A., Stine, P.A., Michaelsen, J. & Miller, J. (2001) Stratified sampling for field survey of environmental gradients in the Mojave desert ecoregion. *GIS and remote sensing applications in biogeography and ecology* (ed. by A.C. Millington, S.J. Walsh & P.E. Osborne), pp. 229–253. Kluwer Academic Press, Boston.

Gillison, A.N. & Brewer, K.R.W. (1985) The use of gradient directed transects or gradsects in natural resource surveys. *Journal of Environmental Management*, **20**, 103–127.

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993–1009.

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

Guisan, A., Graham, C.H., Elith, J. & Huettmann, F. (2007) Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions*, **13**, 332–340.

Hartley, S. & Lester, P.J. (2003) Temperature-dependent development of the Argentine ant, *Linepithema humile* (Mayr) (Hymenoptera: Formicidae): a degree-day model with implications for range limits in New Zealand. *New Zealand Entomologist*, **26**, 91–100.

Hartley, S., Harris, R. & Lester, P.J. (2006) Quantifying uncertainty in the potential distribution of an invasive species: climate and the Argentine ant. *Ecology Letters*, **9**, 1068–1079.

Holway, D.A. (1995) Distribution of the Argentine ant (*Linepithema humile*) in northern California. *Conservation Biology*, **9**, 1634–1637.

Holway, D.A. (1998) Factors governing rate of invasion: a natural experiment using Argentine ants. *Oecologia*, **115**, 206–212.

Holway, D.A. (2005) Edge effects of an invasive species across a natural ecological boundary. *Biological Conservation*, **121**, 561–567.

Holway, D.A., Lach, L., Suarez, A.V., Tsutsui, N.D. & Case, T.J. (2002a) The causes and consequences of ant invasions. *Annual Review of Ecology and Systematics*, **33**, 181–233.

Holway, D.A., Suarez, A.V. & Case, T.J. (2002b) Role of abiotic factors in governing susceptibility to invasion: a test with Argentine Ants. *Ecology*, **83**, 1610–1619.

Hosmer, D.W. & Lemeshow, S. (2000) *Applied logistic regression*, 2nd edn. John Wiley & Sons, Inc., New York, NY.

Kadmon, R., Farber, O. & Danin, A. (2003) A systematic analysis of factors affecting the performance of climatic envelope models. *Ecological Applications*, **13**, 853–867.

Karl, J.W., Heglund, P.J., Garton, E.O., Scott, J.M., Wright, N.M. & Hutto, R.L. (2000) Sensitivity of species habitat-relationship model performance to factors of scale. *Ecological Applications*, **10**, 1690–1705.

Korzukhin, M.D., Porter, S.D., Thompson, L.C. & Wiley, S. (2001) Modeling temperature-dependent range limits for the fire ant *Solenopsis invicta* (Hymenoptera : Formicidae)

© 2008 The Authors

in the United States. *Environmental Entomology*, **30**, 645–655.

Krushelnycky, P.D., Joe, S.M., Medeiros, A.C., Daehler, C.C. & Loope, L.L. (2005) The role of abiotic conditions in shaping the long-term patterns of a high-elevation Argentine ant invasion. *Diversity and Distributions*, **11**, 319–331.

Laakkonen, J., Fisher, R.N. & Case, T.J. (2001) Effect of land cover, habitat fragmentation and ant colonies on the distribution and abundance of shrews in southern California. *Journal of Animal Ecology*, **70**, 776–788.

Levine, J.M. & D'Antonio, C.M. (1999) Elton revisited: a review of evidence linking diversity and invasibility. *Oikos*, **87**, 15–26.

Luoto, M., Virkkala, R. & Heikkinen, R.K. (2007) The role of land cover in bioclimatic models depends on spatial resolution. *Global Ecology and Biogeography*, **16**, 34–42.

McPherson, J.M., Jetz, W. & Rogers, D.J. (2004) The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *Journal of Applied Ecology*, **41**, 811–823.

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.

Menéndez, R., González-Megías, A., Collingham, Y., Fox, R., Roy, D.B., Ohlemüller, R. & Thomas, C.D. (2007) Direct and indirect effects of climate and habitat factors on butterfly diversity. *Ecology*, **88**, 605–611.

Menke, S.B. & Holway, D.A. (2006) Abiotic factors control invasion by Argentine ants at the community scale. *Journal of Animal Ecology*, **75**, 368–376.

Menke, S.B., Fisher, R.N., Jetz, W. & Holway, D.A. (2007) Biotic and abiotic controls of Argentine ant invasion success at local and landscape scales. *Ecology*, **88**, 3164–3173.

Morrison, L.W., Porter, S.D., Daniels, E. & Korzukhin, M.D. (2004) Potential global range expansion of the invasive fire ant, *Solenopsis invicta*. *Biological Invasions*, **6**, 183–191.

Muñoz, A.-R. & Real, R. (2006) Assessing the potential range expansion of the exotic monk parakeet in Spain. *Diversity and Distributions*, **12**, 656–665.

Pautasso, M. (2007) Scale dependence of the correlation between human population presence and vertebrate and plant species richness. *Ecology Letters*, **10**, 16–24.

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

Pearson, R.G., Thuiller, W., Araújo, M.B., Martinez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, T.P. & Lees, D.C. (2006) Model-based uncertainty in species range prediction. *Journal of Biogeography*, **33**, 1704–1711.

Peterson, A.T. (2003) Predicting the geography of species' invasions via ecological niche modeling. *Quarterly Review of Biology*, **78**, 419–433.

Peterson, A.T. (2006) Uses and requirements of ecological niche models and related distributional models. *Biodiversity Informatics*, **3**, 59–72.

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

Randin, C.F., Dirnböck, T., Dullinger, S., Zimmermann, N.E., Zappa, M. & Guisan, A. (2006) Are niche-based species distribution models transferable in space? *Journal of Biogeography*, **33**, 1689–1703.

Rew, L.J., Maxwell, B.D. & Aspinall, R. (2005) Predicting the occurrence of nonindigenous species using environmental and remotely sensed data. *Weed Science*, **53**, 236–241.

Roura-Pascual, N., Suarez, A.V., Gómez, C., Pons, P., Touyama, Y., Wild, A.L. & Peterson, A.T. (2004) Geographical potential of Argentine ants (*Linepithema humile* Mayr) in the face of global climate change. *Proceedings of the Royal Society B: Biological Sciences*, **271**, 2527–2534.

Roura-Pascual, N., Suarez, A.V., McNyset, K., Gómez, C., Pons, P., Touyama, Y., Wild, A.L., Gascon, F. & Peterson, A.T. (2006) Niche differentiation and fine-scale projections for Argentine ants based on remotely sensed data. *Ecological Applications*, **16**, 1832–1841.

Schilman, P.E., Lighton, J.R.B. & Holway, D.A. (2005) Respiratory and cuticular water loss in insects with continuous gas exchange: comparison across five ant species. *Journal of Insect Physiology*, **51**, 1295–1305.

Stephens, P.A., Buskirk, S.W., Hayward, G.D. & Martinez del Rio, C. (2005) Information theory and hypothesis testing: a call for pluralism. *Journal of Applied Ecology*, **42**, 4–12.

Stockwell, D.R.B. & Peterson, A.T. (2002) Effects of sample size on accuracy of species distribution models. *Ecological Modelling*, **148**, 1–13.

Suarez, A.V., Holway, D.A. & Case, T.J. (2001) Patterns of spread in biological invasions dominated by long-distance jump dispersal: insights from Argentine ants. *Proceedings of the National Academy of Sciences USA*, **98**, 1095–1100.

Suttle, K.B., Thomsen, M.A. & Power, M.E. (2007) Species interactions reverse grassland responses to changing climate. *Science*, **315**, 640–642.

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

Thuiller, W., Brotons, L., Araújo, M.B. & Lavorel, S. (2004) Effects of restricting environmental range of data to project current and future species distributions. *Ecography*, **27**, 165–172.

Tsutsui, N.D., Suarez, A.V., Holway, D.A. & Case, T.J. (2001) Relationships among native and introduced populations of the Argentine ant (*Linepithema humile*) and the source of introduced populations. *Molecular Ecology*, **10**, 2151–2161.

Urban, M.C., Phillips, B.L., Skelly, D.K. & Shine, R. (2007) The cane toad's (*Chaunus* [*Bufo*] *marinus*) increasing ability to invade Australia is revealed by a dynamically updated range model. *Proceedings of the Royal Society B: Biological Sciences*, **274**, 1413–1419.

Ward, P.S. (1987) Distribution of the introduced Argentine ant (*Iridomyrmex humilis*) in natural habitats of the Lower Sacramento Valley and its effects on the indigenous ant fauna. *Hilgardia*, **55**, 1–16.

© 2008 The Authors

Whittingham, M.J., Krebs, J.R., Swetnam, R.D., Vickery, J.A., Wilson, J.D. & Freckleton, R.P. (2007) Should conservation strategies consider spatial generality? Farmland birds show regional not national patterns of habitat association. *Ecology Letters*, **10**, 25–35.

Wild, A.L. (2004) Taxonomy and distribution of the Argentine ant, *Linepithema humile* (Hymenoptera : Formicidae). *Annals of the Entomological Society of America*, **97**, 1204–1215.

Williams, J.W., Jackson, S.T. & Kutzbach, J.E. (2007) Projected distributions of novel and disappearing climates by 2100 AD. *Proceedings of the National Academy of Sciences USA*, **104**, 5738–5742.

SUPPORTING INFORMATION

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

Table S1 Pearson correlation matrix for the seven environmental variables used to predict Argentine ant occurrence in the model averaging analysis.

Table S2 Model averaging analysis of the regional confidence sets.

Table S3 Model averaging analysis of the spatial grain confidence sets.

Table S4 Single predictor effects of seven environmental variables on Argentine ant occurrence at four different spatial grains using logistic regression.

Figure S1 Empirical distributions of the four dominant environmental variables.

Figure S2 Effects of the seven environmental predictors used in the model averaging analysis on the probability of Argentine ant occurrence using data from the southern region.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

BIOSKETCHES

Sean B. Menke integrates the study of ecology, behaviour and evolution into the fields of biogeography, community ecology and invasion biology by using modelling and experimental techniques to link patterns of species occurrence with processes that determine species persistence, coexistence and movement across spatial scales.

David A. Holway studies the causes and consequences of social insect invasions.

Robert N. Fisher is a research biologist for the USGS and focuses primarily on the conservation of southern California's biota.

Walter Jetz is interested in how environment, ecology, evolutionary history and chance affect ecological phenomena at the level of the individual, population and community and how these combine to form patterns at continental and global scales.

Editor: Tim Blackburn

© 2008 The Authors