

LETTER

Quantifying uncertainty in the potential distribution of an invasive species: climate and the Argentine ant

Stephen Hartley^{1*}, Richard Harris² and Philip J. Lester¹

¹Centre for Biodiversity and Restoration Ecology, School of Biological Sciences, Victoria University of Wellington, PO Box 600, Wellington, New Zealand

²Landcare Research, Nelson, New Zealand

*Correspondence: E-mail:

stephen.hartley@vuw.ac.nz

Present address: Richard Harris, Department of Environmental Biology, Curtin University of Technology, Perth, Australia

Abstract

Maps of a species' potential range make an important contribution to conservation and invasive species risk analysis. Spatial predictions, however, should be accompanied by an assessment of their uncertainty. Here, we use multimodel inference to generate confidence intervals that incorporate both the uncertainty involved in model selection as well as the error associated with model fitting. In the case of the invasive Argentine ant, we found that it was most likely to occur where the mean daily temperature in mid-winter was 7–14 °C and maximum daily temperatures during the hottest month averaged 19–30 °C. Uninvaded regions vulnerable to future establishment include: southern China, Taiwan, Zimbabwe, central Madagascar, Morocco, high-elevation Ethiopia, Yemen and a number of oceanic islands. Greatest uncertainty exists over predictions for China, north-east India, Angola, Bolivia, Lord Howe Island and New Caledonia. Quantifying the costs of different errors (false negatives vs. false positives) was considered central for connecting modelling to decision-making and management processes.

Keywords

Bioclimatic model, concordance probability, confidence intervals, cross-validation, *Linepithema humile*, logistic regression, misclassification errors, multimodel inference, risk analysis, receiver operator's characteristic statistics.

Ecology Letters (2006) 9: 1068–1079

INTRODUCTION

Homogenization of the world's biota via human-assisted invasions is increasingly recognized as one of the greatest threats to global biodiversity (Mack *et al.* 2000). The ability to predict the potential for establishment and spread of an invasive species is, therefore, highly desirable. A common approach is to characterize a multidimensional 'climate envelope' of a species, based upon its existing range, and then to project this set of environmental requirements on to the geographical area to be assessed (Lehmann *et al.* 2002). Statistical methods for these analyses are continually being refined (Stockwell & Peters 1999; Pearce & Ferrier 2000; Guisan *et al.* 2002; Lehmann *et al.* 2002; Miller & Franklin 2002), and although limitations are acknowledged (Guisan & Thuiller 2005), relatively little attention has focussed on furnishing confidence intervals for these predictions. In this article, we illustrate a method for estimating and communicating uncertainty in predictions of an invasive species' potential range, using the Argentine ant, *Linepithema humile* (Mayr), as our case study.

When undertaking a modelling exercise there are four important processes to consider: (i) defining the problem and the aim of the exercise; (ii) model specification, i.e. deciding on the mathematical form of the model; (iii) model fitting or parameterization, i.e. estimating the values of the coefficients; and (iv) model validation, i.e. estimating the validity of the predictions. Each of these steps is subject to uncertainty or error (Chatfield 1995). Most attempts to quantify uncertainty have traditionally focussed on the statistical process of parameterization (step iii). In recent years, increasing attention has been paid to model validation (step iv), as it is realized that self-validation (testing model performance on the same data that are used for parameterization) will yield over-optimistic estimates of predictive ability (Fielding & Bell 1997; Harrell 2001). Potentially much more important, however, is the uncertainty in model specification (step ii) (Chatfield 1995). Only recently has this topic begun to be addressed in ecological models using multimodel inference and/or Bayesian methods (e.g. Burnham & Anderson 2002; Elith *et al.* 2002; Wintle *et al.* 2003; Johnson & Omland 2004; Clark 2005). The aim of this study

was to illustrate how defining and assessing the performance of a set of plausible models can help to characterize the uncertainty inherent in attempts to predict the potential distribution of a species. Recognizing and quantifying this uncertainty, as well as the cost of false predictions, should improve the basis for rational decision making in the management of invasive species and for introduction programmes with a conservation focus.

METHODS

The study system

The Argentine ant, *Linepithema humile*, is considered native to the Paraná drainage basin of South America (Wild 2004). In the past century, it has successfully invaded all continents apart from Antarctica, and its distribution continues to expand (Suarez *et al.* 2001). The arrival of Argentine ants is often associated with significant changes in the structure of local invertebrate and plant communities (Cole *et al.* 1992; Christian 2001), and in horticultural areas it can be a significant economic pest (Vega & Rust 2001).

The Argentine ant is typically found in Mediterranean-style climates, although the current distribution also extends partway into temperate and subtropical zones (Suarez *et al.* 2001). The influence of temperature on the autecology of Argentine ants is expressed through a variety of mechanisms. At a physiological level, queens do not lay eggs if the nest temperature falls below 18 °C (Newell & Barber 1913), while development of workers requires approximately 445 degree-days above 15.9 °C (Hartley & Lester 2003). Foraging outside the nest typically occurs when air temperatures are between 10 and 32 °C, with the range 15–30 °C considered optimal (Markin 1970). In the hotter, drier parts of its range, populations are restricted to areas close to water courses (Holway *et al.* 2002a), most likely reflecting the humidity requirements of Argentine ants (Walters & Mackay 2003). In colder areas (e.g. high-elevation Hawaii) excessive rainfall may limit population spread by depressing soil temperatures in the nests (Krushelnycky *et al.* 2005). In short, the available evidence suggests that both temperature and rainfall are important in determining the abiotic range limits of the Argentine ant.

Data sources

Global climate data for temperature and precipitation were obtained from the International Panel for Climate Change as an array of half-degree latitude × longitude grid cells (New *et al.* 1999). Monthly averages for the period 1961–1990 were used to calculate total annual precipitation (PPT), mean annual temperature (MAT), two measures of annual temperature minima (MINMIN and MINAVG) and two

measures of annual temperature maxima (MAXMAX and MAXAVG). For details see Table S1.

Records of Argentine ant presence were collected from the literature (Table S2). For the purposes of model-fitting we used all terrestrial grid cells between latitudes 60° N and 60° S, resulting in a data set of 49 966 presumed ‘absences’ and 220 confirmed presences. Areas beyond these latitudinal limits were deemed trivially unsuitable for Argentine ants and superfluous to the model-fitting procedure. Inevitably, the data will have contained some false absences. If they are more or less randomly distributed with respect to climate then the model-fitting procedure should be robust to their inclusion (but see below for their effect on the intercept of logistic regression models).

Model specification and the set of plausible models

All the models were of the logistic form (eqns 1 and 2), in which the logit of probability of occurrence at site i (p_i) is a linear function of the climate variables (x_k) and their beta-coefficients (β_k), and unexplained error (ε) is assumed to follow a binomial distribution (Hosmer & Lemeshow 2000)

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i = g(x_i) \quad (1)$$

$$\rightarrow p_i = \frac{1}{1 + e^{-g(x_i)}}. \quad (2)$$

Logistic regression has been widely used for modelling species presence and absence data, and the statistical properties of the model are well characterized (Harrell 2001; Hosmer & Lemeshow 2000). Other modelling approaches, such as generalized additive models, neural networks and classification trees are also possible. Comparative studies suggest that predictions from these different approaches do differ, but not greatly (e.g. Franklin 1998); however, this is a further source of between-model variation that we did not quantify here.

Based on our knowledge of Argentine ant ecophysiology (reviewed above) we constructed 11 plausible models involving different combinations of climate variables. Model *a* included all variables, but with no interactions, models *b–f* were reduced models relative to model *a*, and models *g–k* mirrored *b–f* but included interactions (Table S3).

In order to represent climatic optima in the models (unimodal responses), each climate variable was entered or removed as a pair of variables: a linear term and a transformed term. A squared term was used for the temperature variables, as their marginal distributions were approximately normal, and a logarithmic term was added for

precipitation, because its marginal distributions were left-skewed (Kay & Little 1987).

Model parameterization

The models were parameterized (i.e. the values of the beta-coefficients (eqn 1) were estimated) using the 'glm' command of the R2.2.0 statistical package (R Development Core Team 2005). This method uses maximum likelihood to minimize deviance between observed and predicted values.

It is worth pointing out that models are fitted so that the values of all the predicted probabilities of occurrences (p_i) sum to the number of observed positive values in the input data. This suggests that the absolute values of p_i are biased downward if there are false negatives in the input data. False negatives in the input could arise due to either: (i) under-recording of true occurrences; or (ii) the existence of suitable but unoccupied grid cells in the input data. To allow for either of these likely events the values of $g(x)$ should be adjusted upward by adding a constant to the intercept: $-\ln(\tau_1/\tau_0)$, where, τ_1 is the factor by which the true or potential prevalence is under-represented in the input data and τ_0 is the factor by which absences are over-represented. For further details see the case : control method described by Hosmer & Lemeshow (2000, p. 207). In this study, we are uncertain as to the degree of false negatives in the input data or the degree of future expansion to be expected and therefore we leave any such adjustment to the reader. In any event, the rank suitability of sites will be unaffected by such an adjustment, as it is simply a constant in terms of $g(x)$.

This complication in interpreting the absolute values of the predicted probabilities of occurrence reveals one of the conceptual difficulties in using climate envelope models for extrapolation when they are parameterized from imperfectly recorded and non-equilibrium distributions.

Model validation: assessing predictive ability

A common approach for assessing the performance of logistic regression models is to threshold the predicted probability of occurrence (a continuous measure) into a binary prediction of presence-absence; the predictions can then be cross-classified with the original observations in the form of a 2×2 confusion matrix (Table S4). A number of inter-related measures of predictive ability can be derived from a confusion matrix including the proportion of observed negatives correctly predicted (specificity), the proportion of observed positives correctly predicted (sensitivity), the false-positive rate and the false-negative rate (Fielding & Bell 1997; Pearce & Ferrier 2000).

A potential problem with all measures derived from a confusion matrix is that they are highly sensitive to the value of the threshold. This is evident from a receiver operator's

characteristic (ROC) curve, in which sensitivity is plotted against the proportion of false positives ($1 - \text{specificity}$) across the full range of threshold values (Zweig & Campbell 1993; Fielding & Bell 1997).

If the cost of false positives relative to false negatives (R) is made explicit, then a logical measure of performance would be the total cost of misclassification, assessed at whatever threshold minimizes this cost (Zweig & Campbell 1993). The cost of eradicating invasive species from a region is generally recognized to be much greater than the cost of surveillance and containment (Mack *et al.* 2000). Therefore, a false positive (a false alarm of invasion threat) is less costly than a false negative [a false sense of (bio)security]. For most species, however, it is difficult to obtain accurate estimates of R ; the Argentine ant is no different in this respect, hence, we assumed a constant value of $R = 0.1$. In terms of sensitivity (S_n) and specificity (S_p) the total cost of misclassifications is $(1 - S_n)(1 - P) + R(1 - S_p)(1 - P)$, where P is prevalence, the proportion of observed cases that are positive.

One performance measure that does not require the definition of a classification threshold is concordance probability (CP). It compares all possible pairs of observations involving one presence and one absence, and is calculated as the proportion of comparisons in which the predicted probability is higher for the observed presence than for the observed absence (Hanley & McNeil 1982). A purely random model would be expected to be correct half of the time ($CP = 0.5$), whereas a perfect model would be correct all of the time ($CP = 1.0$). CP is equivalent to the area under the curve (AUC) of a ROC plot (Harrell 2001) or it can be estimated by a Mann-Whitney statistic (Pearce & Ferrier 2000; Pearce *et al.* 2001). In general, models with a $CP > 0.8$ are considered to have 'useful' predictive ability (Fielding & Bell 1997; Harrell 2001). All the performance measures and statistics described above were calculated using the ROCR package in the 'R' statistical and programming environment (Sing *et al.* 2004).

So which measure of predictive ability is to be preferred? If biosecurity managers wish to know which sites justify a specified level of surveillance, and assuming the relative cost of false positives to false negatives is known, then the most relevant measure is surely the total cost of misclassification errors. However, if managers have an insufficient budget they may simply need to know the rank order importance of sites in terms of decreasing probability of establishment. In this case, the most appropriate measure of performance is CP, as the absolute value of the predicted probabilities is less important than their correct ordering (Pearce & Ferrier 2000).

Measuring the performance of a model with the same data that were used for parameterization, will yield an over-optimistic view of predictive ability, particularly, if the

model has been over-specified (Fielding & Bell 1997). To address this point, each model specification was cross-validated by a geographical k -fold partition or jack-knife procedure (Augustin *et al.* 2001; Hastie *et al.* 2001). Data from four continents were used as a 'training set' (for parameterization) while data from the two remaining continents were withheld as an independent 'test set' of predictive ability. Permuting all combinations of four continents from six allowed for 15 semi-independent tests between the 11-model specifications. We used large, spatially contiguous areas (i.e. continents) to partition the data, in preference to a random subset of grid cells, as we wished to maximize the climatic and biogeographical independence between the training and test data sets.

Model weights

Without doubt, all the models considered are simplifications of reality, therefore, model weights were calculated to represent our relative belief in the primacy of each model specification relative to the set of plausible models being considered (Burnham & Anderson 2002). Four methods of calculating weights were compared. The first two utilized the empirical performance measures: total cost and CP, calculated from the test sets of the cross-validation. For each geographical partitioning the 11 models were ranked according to the chosen performance measure and the top three performing models were awarded 3, 2 or 1 point(s) respectively. The scores were summed across partitions and standardized to produce a vector of weights whose elements sum to unity.

We also calculated two sets of Akaike weights using an information-theoretic approach (Burnham & Anderson 2002). The first set used Akaike's information criterion (AIC) and the second a quasi-AIC (QAIC) which includes a parameter to adjust for overdispersion in the data. These criteria assess a model's log-likelihood goodness-of-fit, relative to its structural complexity and the degrees of freedom in the data. Although the AIC is gaining widespread usage, in our case the effective degrees of freedom were overestimated due to the spatially autocorrelated nature of the data (Fig S1), therefore, the Akaike weights are presented here for comparison only.

Multimodel consensus prediction

Even though cross-validation is a useful method for assessing the relative predictive ability of different model specifications and for assigning weights, the best predictions are obtained when a model is parameterized from the complete data set (Fielding & Bell 1997). Hence, for each model, the data from all six continents were used to estimate $g(x_i)$ and its standard error (SE _{i}), for all grid cells between

90° N and 90° S. A weighted average of the $g(x_i)$ values was calculated across the 11-candidate models, prior to back-transformation into a single consensus estimate for the probability of occurrence (\hat{p}_i). There is some debate as to whether the best estimate is obtained from the single best performing model or from the weighted model average (Wintle *et al.* 2003). In this example, we favoured the weighted model average, although individual models were useful for generating a parsimonious description of the Argentine ant's climate envelope.

Uncertainty within and between models

Uncertainty in multimodel predictions was quantified by adding the weighted average of within-model variation (eqn 3) to the between-model variation (eqn 4) to obtain a measure of total variance (s_T^2) (S. Pledger, personal communication).

$$\text{Average within-model variation} : \sum w_m (s_m^2) \quad (3)$$

$$\text{Between-model variation} : \sum w_m (\bar{x}_m)^2 - \left(\sum w_m \bar{x}_m \right)^2, \quad (4)$$

where \bar{x}_m is the estimated value of $g(x)$ predicted by model 'm', (s_m^2) is the within-model variance of the prediction (the square of the standard error) and w_m is the relative weight afforded to the model.

Confidence intervals were generated around the multimodel average, where 95% CI = $g(x) \pm 1.96(s_T)$. The upper CI limit was back-transformed to generate a 97.5th percentile or 'precautionary' scenario of the probability of occurrence. Regions of high uncertainty were identified by mapping the absolute difference between the upper and lower CI limits on a probability scale, and from the ratio of between-model to within-model variation on the $g(x)$ scale.

RESULTS

Species distribution and climate data

Over 300 records were obtained from 23 different countries and 260 different grid cells. Forty of these grid cells contained only 'urban' records, and were excluded from subsequent analyses as urban areas have unnatural levels of heating, cooling, food and water supply that are atypical of the surrounding areas from which our climate data were derived. In total, we were left with 220 occupied grid cells (Fig. 1).

All six climate variables were significantly correlated with one another. In particular, there was high colinearity amongst many of the temperature variables where Pearson correlation coefficients varied from 0.64 to 0.99 (Table S5).

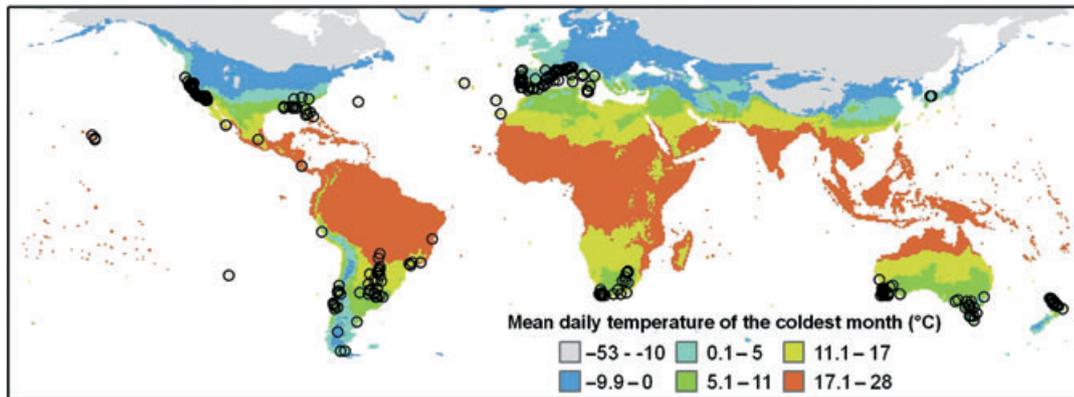


Figure 1 Records of Argentine ant occurrence (○) against a background of mean daily temperature during the coldest month (MINAVG).

Model goodness-of-fit and predictive ability

The various model specifications generated similar shaped ROC curves (Fig. 2a), with CP (AUC) values ranging from 0.957 (model *a*) down to 0.907 (model *f*). The cost-curves were also similar across models (e.g. Fig. 2b), although under the assumption of $R = 0.1$ the most cost-effective threshold for triggering a surveillance response varied from 0.040 (model *f*) to 0.092 (model *j*) (mean = 0.065).

Models *a*, *g* and *b* appeared to perform best as judged by their high CP and low cost of misclassification errors under self-validation (Fig. S1). As expected, performance under cross-validation was not as good, but also the relative performance of the models altered somewhat. Model *a* remained as the best model, but this time closely followed by models *b* and *c*. The latter two models are the same as *g* and *b*, but without the two-way interactions between variables. In general, the simpler models without interactions performed better in the cross-validation test, suggesting a degree of

over-fitting by the more complex models. Overall, support for model specifications *a*, *b*, *c*, *g* and *b* was relatively high as reflected in the model weights derived from the cross-validation tests (Table 1). By way of comparison, Akaike weights calculated from the standard AIC overwhelmingly supported model *a*. The weights derived from QAIC were sensitive to the value of the overdispersion parameter (\hat{c}). At low values (e.g. $\hat{c} < 3$, as suggested by empirical estimates) the QAIC weights were the same as those based on AIC, overwhelmingly supporting model *a*. At values between 13 and 55 models *c* and *b* were favoured, and at extremely high values ($\hat{c} > 55$) the simplest model, model *f* was favoured. At values of $\hat{c} \sim 12.5$ the QAIC weights come closest to those derived from cross-validation.

Climatic envelope

Model *c* was chosen as the best single model for deriving a parsimonious description of the species' climate envelope,

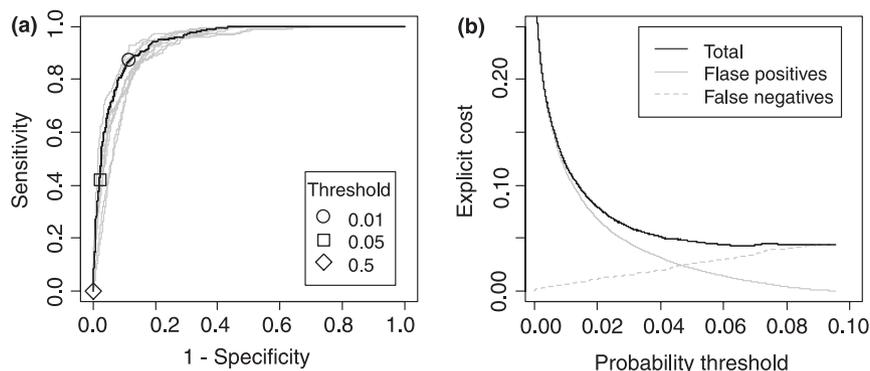


Figure 2 Threshold dependency of cross-classification. (a) Receiver's operator characteristic plot for all models (grey lines) and model *c* (black line), parameterized and self-validated with the complete data set of records. Symbols indicate the proportion of true positives (Sensitivity) and false positives (1 - Specificity) evaluated at three different thresholds. Area under the curve = concordance probability = 0.953. (b) Misclassification costs of model *c* when the cost of a false positive relative to the cost of a false negative is 0.1. At high classification thresholds > 0.15 , all true positive cells are misclassified as false negatives. The minimum cost for this model (0.042) occurs at an optimal threshold of 0.068; other models possessed optimal thresholds between 0.040 and 0.092.

Table 1 Relative model weights determined from two different performance criteria and two information criteria

Model	Performance measure		Information criterion	
	Concordance probability (AUC)	Total cost of misclassifications	AIC	QAIC
<i>a</i>	0.333	0.282	1.000	0.180
<i>b</i>	0.122	0.115	0.000	0.213
<i>c</i>	0.289	0.103	0.000	0.442
<i>d</i>	0.000	0.077	0.000	0.011
<i>e</i>	0.022	0.038	0.000	0.041
<i>f</i>	0.022	0.000	0.000	0.002
<i>g</i>	0.078	0.128	0.000	0.048
<i>h</i>	0.133	0.128	0.000	0.058
<i>i</i>	0.000	0.038	0.000	0.001
<i>j</i>	0.000	0.064	0.000	0.004
<i>k</i>	0.000	0.026	0.000	0.000

Performance-based weightings determined from cross-validation (see text for details). Information criteria determined from parameterization of the full data set. The Quasi-Akaike information criterion (QAIC) is illustrated with an overdispersion parameter of 15; values < 12 favour model *a*, values > 13 favour models *b* and *c*. When the overdispersion parameter equals one, QAIC reduces to AIC (Burnham & Anderson 2002). Model weights in bold account for over half of the total cumulative weighting for each measure.

due to the high CP values achieved in cross-validation. The complete data set was used to obtain the most accurate parameterization of this model (Table S6a). The fitted model *c* suggested that ideal environmental conditions for Argentine ants occur when the average temperature during the coldest month (MINAVG) is between 7 and 14 °C, the daily maximum during the hottest month (MAXMAX) averages between 19 and 30 °C and rainfall is between 50 and 150 cm year⁻¹ (Fig. 3). The models (*g*–*k*) that incorporated interactions between temperature and rainfall suggested that temperature optima increase as precipitation increases. Single-factor optima predicted by model *a* were difficult to interpret due to some opposing coefficients for highly correlated climate variables (Table S6b).

Potential geographical range

Our consensus or 'best' estimate of probability of occurrence was obtained from a back-transformation of the weighted model-average of $g(x)$ values. With this model, central Madagascar, Taiwan, high-elevation Ethiopia and Yemen, Saint Helena, the Chatham Islands of New Zealand, the Norfolk and Lord Howe Islands of Australia and a number of other oceanic islands currently believed to be uninvaded by Argentine ants, are identified as being at high risk of future establishment (Fig. 4a).

With a knowledge of the total variance associated with the estimates of $g(x)$ we were able to construct a pessimistic or precautionary estimate of potential range based upon the 97.5 percentile for probability of occurrence (Fig. 4b). Earlier results based on assumptions of cost–benefit ratios suggest that we should be concerned about sites where the predicted probability of occurrence is > 0.05. The central estimate predicts that 926 half-degree grid cells have a 0.05 or greater probability of supporting Argentine ants (Fig. 4a), whereas the 97.5 percentile scenario suggests that as many as 3258 of the 67 419 terrestrial grid cells are at risk (Fig. 4b), an approximate 3.5-fold increase in area. The 95% CI for probability of occurrence illustrates that greatest uncertainty exists over predictions for southern China, Angola, Bolivia and islands of the east coast of Australia (Fig. 4c). In general, the variation between models was two or more times that of the within-model variation, except in the areas where all models agreed that the probability of occurrence was virtually zero (Fig. 4d, Fig. 5). Although uncertainty in predictions may be regarded as a 'bad thing' its quantification paradoxically allows management decisions to be made with greater certainty than would be possible from mean predictions alone.

DISCUSSION

The predictions presented here show a close concordance with the currently known distribution of Argentine ants, suggesting that the environmental variables of temperature and rainfall play an important role in setting broad-scale limits to establishment. The predictions can be visualized as ellipses in 'climate-space' (Fig. 3) or as a distribution in geographic space (Fig. 4). The calculation of confidence intervals for each location should be particularly useful for those involved in risk management, where knowing the likelihood and magnitude of a 'worst-case scenario' is just as important as knowing the 'most-probable scenario'. The performance of the models developed here compares well with that of another model for predicting Argentine ant distributions (Roura-Pascual *et al.* 2004), although there are some interesting points of difference between the fine detail of the two models.

Roura-Pascual *et al.* (2004) used the same source climate data, but supplemented it with additional fine scale (1 × 1-km resolution) topographic data and moderate scale (8 × 8-km resolution) satellite data on vegetation cover, to create an initial set of 33 environmental predictors. Their models used a wider range of mathematical functions which were allowed to 'evolve' to fit the data according to the Genetic Algorithm for Rule-set Prediction (Stockwell & Peters 1999). The evolving rule sets were trained and validated on two checkerboard subsets of the native range, before being used to predict the potential exotic range.

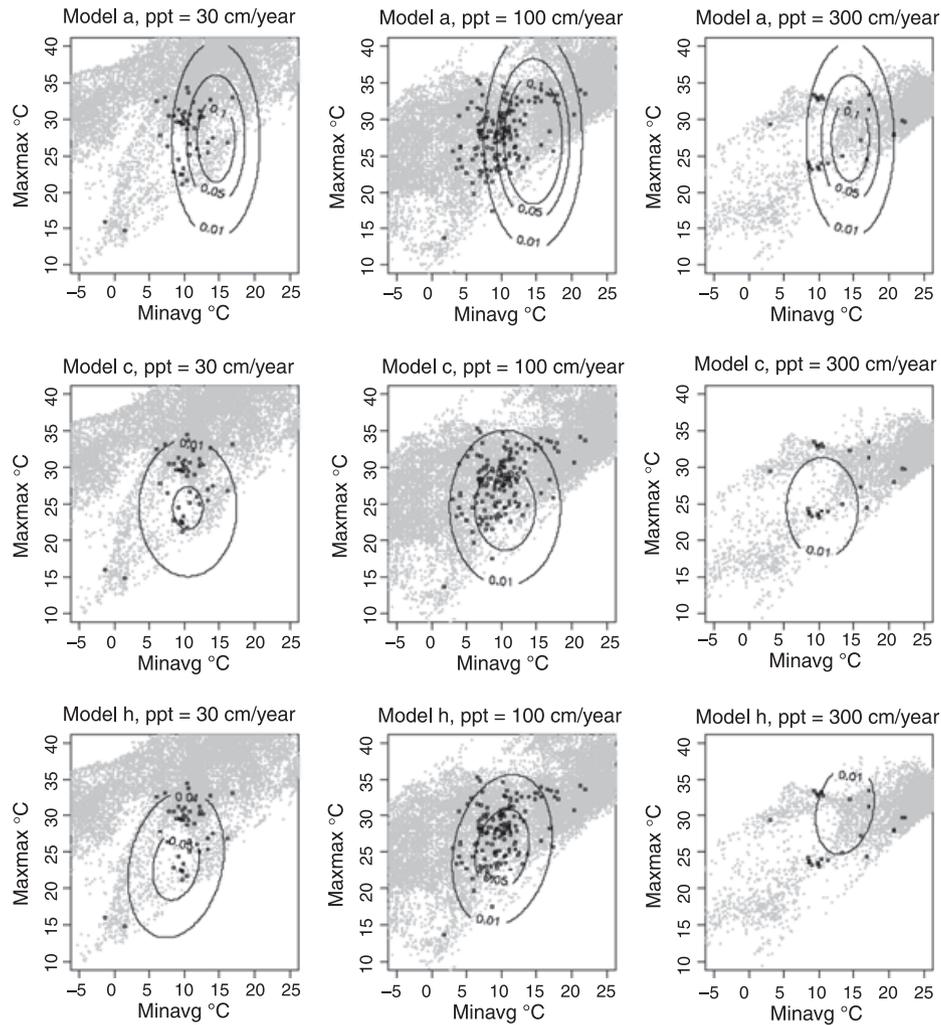


Figure 3 Climate envelopes of models *a*, *c* and *b* (top to bottom). Contours of probability of occurrence (0.01, 0.05 and 0.1) evaluated as a function of MINAVG and MAXMAX temperatures (see Table S3), at three different levels of precipitation: 30, 100 and 300 cm year⁻¹. Grey dots indicate sites lacking a record of Argentine ants, black squares indicate recorded presences, precipitation ranges are (from left to right) 0–50, 50–150 and 150+ cm year⁻¹. Other climate variables for model *a* set close to their optima: MINMIN = 5 °C, MAT = 15 °C and MAXAVG = 20 °C, although logically the following must be true: MINMIN ≤ MINAVG ≤ MAT ≤ MAXAVG ≤ MAXMAX. The three low temperature outliers in the first two columns are all records from southern Chile.

This approach, represents a more stringent test of a model's predictive ability, but if the limits to the native range are heavily modified by biotic interactions, or if certain combinations of acceptable environmental conditions do not exist within the native range, then the extrapolation from native to exotic areas may not be robust. In contrast, the approach adopted in this paper, using native and exotic ranges, is not robust if the model is trained on areas where the species has not had sufficient opportunity to occupy a large fraction of the potentially occupiable range.

There were also some locational records unique to each study. Due to these differences in input data, model

structure and model-fitting we view the two studies as complementary; and when the results are compared they provide further insight into the (un)certainty of model predictions.

Spatial scale

The model presented here was constructed at a resolution of half-degree latitude by longitude (*c.* 50 × 50 km at 40° latitude), and therefore its predictions can only be properly interpreted at this scale. If primary data had been available at a finer resolution, then it is possible that a tighter climate envelope would have been produced. For example, it is

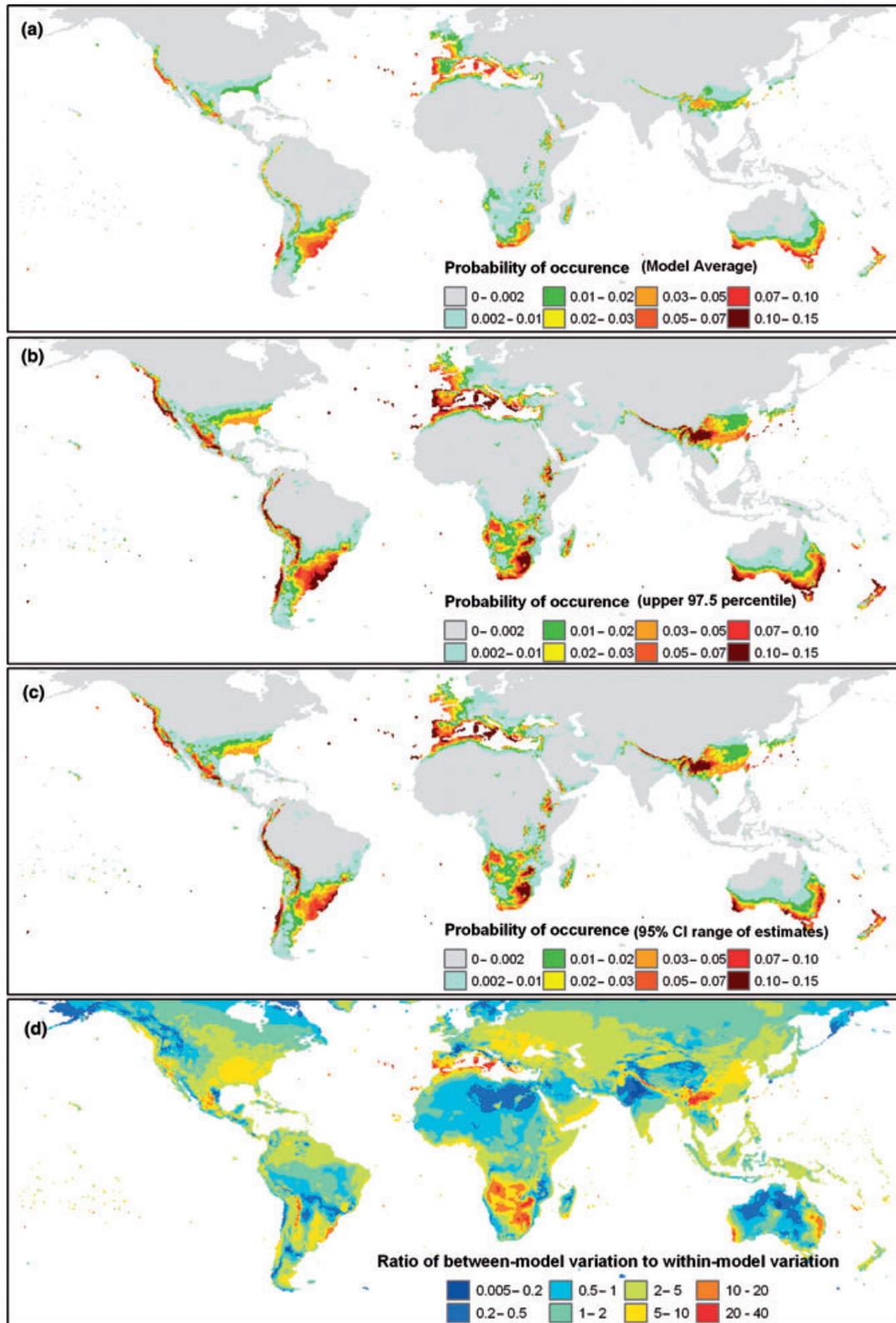


Figure 4 (a) Mean probability of occurrence according to the multimodel weighted average; (b) upper 97.5 percentile for the probability of occurrence; (c) 95% CI; (d) ratio of between-model variation to within-model variation.

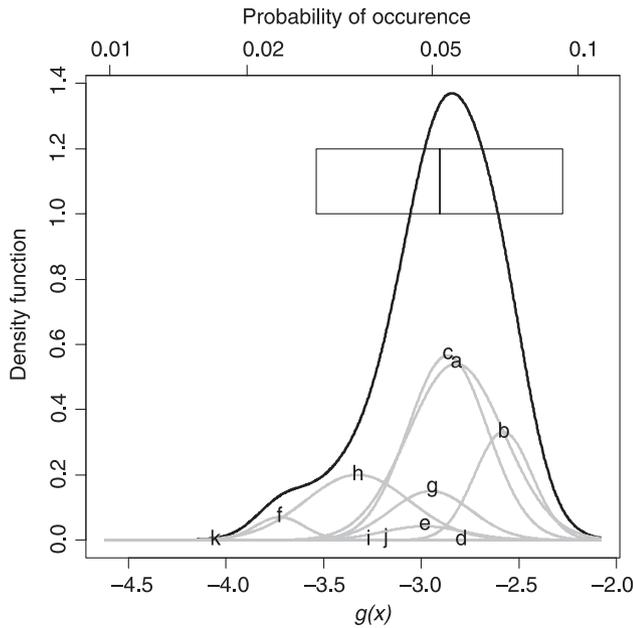


Figure 5 Probability density functions for the predicted probability of Argentine ant occurrence at Wellington, New Zealand ($41^{\circ}15' S$, $174^{\circ}45' W$). Thin lines (models *a–k*) follow a normal distribution on the scale of $g(x)$ (lower x -axis), with their density proportional to model weight. Thick line (a summation of the individual models) is the multimodel estimate. Box indicates the multimodel mean and 95% CI: 0.052, 0.028–0.093 probability of occurrence (upper x -axis).

likely that within the relatively 'warm dry' occupied grid cells, Argentine ants are present within cooler, moister microsites, while in the relatively 'cool' occupied grid cells (such as New Zealand) the ants are restricted to warmer microsites (such as open habitats). Proximity to water is known to be an important factor in hot-dry climates, particularly at scales of a kilometre or less (Holway *et al.* 2002a). At finer scales, habitat data would be an important characteristic for refining potential distribution maps, and would most likely interact with climate (Pearson *et al.* 2004).

Spatial autocorrelation

One of the basic assumptions of logistic regression is that the observations are independent from one another. When the sample units are cartographic grid cells, spatial autocorrelation in the predictor and response variables will violate this assumption (Lennon 2000). This has a variety of knock-on consequences, for example, standard errors within a model may be underestimated and model selection cannot rely on information-theoretic criteria. Cross-validation, based on a coarse-scale geographical partitioning of the data, is a powerful technique for making progress in the presence of spatial autocorrelation (Guisan & Thuiller

2005). Other promising approaches include explicitly modelling the correlation structure of the data in a mixed effects model. If within-model variances have been underestimated (for whatever reason), this provides yet further imperative to incorporate between-model variation.

Several studies have found improved model fit by including terms that explicitly generate spatial autocorrelation in the response variable (e.g. the spatial autologistic model of Augustin *et al.* 1996). We did not include such terms as this would bias models against predicting presences in outlying or isolated areas of suitable climate (Miller & Franklin 2002), and identification of these areas (e.g. uninhabited islands) was one of the main purposes of this study.

Measuring model performance

Many traditional measures of the performance of logistic regression models, require the predicted probabilities to be classified into presences and absences (Fielding & Bell 1997). A measure such as CP or AUC, is attractive because it does not require the definition of a probability threshold. However, using CP is equivalent to judging models on an unstated (and possibly inappropriate) assumption of the relative costs of false positives and false negatives. If these costs can be made explicit a cost minimization approach should be adopted, which has the added advantage of ensuring a connection to management practice (Zweig & Campbell 1993).

The cost of a false negative will always be extremely difficult to quantify. It depends on the economic, social, environmental and ecological costs that would be incurred in the event of a successful invasion. While there are many documented cases of the high cost of invasive species, it is also true that not every invasive species is damaging in every location, and often the extent of the cost is modified by many other unpredictable contingencies (Simberloff & von Holle 1999; Williamson 1999). The cost of a false positive is quantified by the unnecessary effort that is expended in border surveillance and response to an incursion, against a species that could never establish. Although this may seem easier to quantify, any single surveillance procedure will typically be effective to varying degrees against a whole suite of potential invasive species. A further complication arises if there are options for more than one level of surveillance (e.g. high alert and low alert), in which case 'probability of occurrence' should be classified into the corresponding number of categories and a multistate misclassification cost matrix developed. Further work in this area is advisable.

Biotic interactions

Although previous attempts to predict the invasibility of species based upon climate matching have often met with

little success (Williamson 1999), climate remains as one of the few practical predictors available (e.g. Lester 2005). Many of the 'mistakes' of climate matching might be explained by a failure to account for biotic interactions such as competition and facilitation (Davis *et al.* 1998; Case *et al.* 2005). However, we suspect that this is less of a problem for predicting Argentine ant distributions as they are often competitively dominant against other ant species (Holway 1999; Vega & Rust 2001), and they are capable of forming mutualistic associations with a wide variety of novel homopteran hosts, despite a lack of co-evolutionary history (Lester *et al.* 2003). Nonetheless, it is known that the outcome of competitive interactions can be mediated by climatic conditions (Holway *et al.* 2002a). In Australia, it has been suggested that the high diversity of native *Iridomyrmex* ants, combined with relatively hot, dry conditions, are restricting the spread of Argentine ants (Walters & Mackay 2003, 2004).

CONCLUSIONS

Our predictions indicate that there are still some areas of the globe that are currently uninvaded, but nonetheless possess a climate that would make the establishment of Argentine ants highly likely should they disperse to these regions (e.g. China, Taiwan, Madagascar and many oceanic islands). We cannot say whether Argentine ants are lacking from these areas because they have never been transported there (i.e. lack of opportunity), because they have arrived but cannot survive there (i.e. the model predictions are wrong), or indeed perhaps the Argentine ant is already present, but not recorded (i.e. the data are incomplete). Many islands harbour high levels of endemic biodiversity, thus the arrival of a significant pest ant could have serious ecological consequences (Holway *et al.* 2002b), similar to the 'invasional meltdown' precipitated by the arrival of the yellow crazy ant, *Anoplolepis gracilipes*, and associated homopterans on Christmas Island (O'Dowd *et al.* 2003).

The method we have presented here has a number of potential advantages for providing management advice. We highlight the distinction between measures of model performance that provide the most robust ranking of invasion risk (e.g. CP or AUC), compared with measures more directly related to the relative costs of false-positive vs. false-negative predictions.

The quantification and communication of uncertainty is fundamental to any risk-assessment exercise (Ormsby 2005). This is especially true where a precautionary approach is advisable. In the case of invasive species a precautionary approach would advocate using the upper limit of predicted establishment potential as the basis for surveillance and management. In contrast, translocation programmes for species of conservation concern would be better served

basing their decisions around the lower limit, so as to minimize the chance of costly mistakes. In the face of the continued spread of invasive species, loss of biodiversity and the threat of future climate change, the assessment of uncertainty in bioclimatic range predictions should become an essential ingredient of today's decision making.

ACKNOWLEDGEMENTS

We thank Steve Shattuck and Paul Wild for contributing unpublished records; Joanne Love for assistance with the data entry and processing; Darryl McKenzie for suggesting a model averaging approach, and Kirsti Abbott, Paul Krushelnicky, Shirley Pledger, Rudi Schnitzler and three referees for commenting on earlier versions of this article. Financial support was provided by the Victoria University of Wellington Research Fund (URF).

REFERENCES

- Augustin, N.H., Muggleston, M.A. & Buckland, S.T. (1996). An autologistic model for the spatial distribution of wildlife. *J. Appl. Ecol.*, **33**, 339–347.
- Augustin, N.H., Cummins, R.P. & French, D.D. (2001). Exploring spatial vegetation dynamics using logistic regression and a multinomial logit model. *J. Appl. Ecol.*, **38**, 991–1006.
- Burnham, K.P. & Anderson, D.R. (2002). *Model Selection and Multimodel Inference: A Practical Information-theoretic Approach*. Springer-Verlag, New York, NY, USA.
- Case, T.J., Holt, R.D., McPeck, M.A. & Keitt, T.H. (2005). The community context of species' borders: ecological and evolutionary perspectives. *Oikos*, **108**, 28–46.
- Chatfield, C. (1995). Model uncertainty, data mining and statistical inference. *J. R. Stat. Soc. Ser. A*, **158**, 419–466.
- Christian, C.E. (2001). Consequences of a biological invasion reveal the importance of mutualism for plant communities. *Nature*, **413**, 635–639.
- Clark, J.S. (2005). Why environmental scientists are becoming Bayesians. *Ecol. Lett.*, **8**, 2–14.
- Cole, F.R., Medeiros, A.C., Loope, L.L. & Zuehlke, W.W. (1992). Effects of Argentine ant on arthropod fauna of Hawaiian high-elevation shrubland. *Ecology*, **73**, 1313–1322.
- Davis, A.J., Jenkinson, L.S., Lawton, J.H., Shorrocks, B. & Wood, S. (1998). Making mistakes when predicting shifts in species range in response to global warming. *Nature*, **391**, 783–786.
- Elith, J., Burgman, M.A. & Regan, H.M. (2002). Mapping epistemic uncertainties and vague concepts in predictions of species distribution. *Ecol. Modell.*, **157**, 313–329.
- Fielding, A.H. & Bell, J.F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.*, **24**, 38–49.
- Franklin, J. (1998). Predicting the distribution of shrub species in southern California from climate and terrain-derived variables. *J. Veg. Sci.*, **9**, 733–748.
- Guisan, A. & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.*, **8**, 993–1009.

- Guisan, A., Edwards, T.C. Jr & Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecol. Modell.*, 157, 89–100.
- Hanley, J.A. & McNeil, B.J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143, 29–36.
- Harrell, F.E. Jr (2001). *Regression Modeling Strategies: with Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer-Verlag, New York, NY, USA.
- 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 Zeal. Entomol.*, 26, 91–100.
- Hastie, T.J., Tibshirani, R.J. & Friedman, J. (2001). *Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer, New York, NY, USA.
- Holway, D.A. (1999). Competitive mechanisms underlying the displacement of native ants by the invasive Argentine ant. *Ecology*, 80, 238–251.
- Holway, D.A., Suarez, A.V. & Case, T.J. (2002a). Role of abiotic factors in governing susceptibility to invasion: a test with Argentine ants. *Ecology*, 83, 1610–1619.
- Holway, D.A., Lach, L., Suarez, A.V., Tsutsui, N.D. & Case, T.J. (2002b). The causes and consequences of ant invasions. *Annu. Rev. Ecol. Syst.*, 33, 181–233.
- Hosmer, D.W. Jr & Lemeshow, S. (2000). *Applied Logistic Regression*. Wiley, New York, NY, USA.
- Johnson, J.B. & Omland, K.S. (2004). Model selection in ecology and evolution. *Trends Ecol. Evol.*, 19, 101–108.
- Kay, R. & Little, S. (1987). Transformations of the explanatory variables in the logistic regression model for binary data. *Biometrika*, 74, 495–501.
- Krushelnicky, 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. *Divers. Distrib.*, 11, 319–331.
- Lehmann, A., Overton, J.M. & Leathwick, J.R. (2002). GRASP: generalized regression analysis and spatial prediction. *Ecol. Modell.*, 157, 189–207.
- Lennon, J.J. (2000). Red-shifts and red herrings in geographical ecology. *Ecography*, 23, 101–113.
- Lester, P.J. (2005). Determinants for the successful establishment of exotic ants in New Zealand. *Divers. Distrib.*, 11, 279–288.
- Lester, P.J., Baring, C.W., Longson, C.G. & Hartley, S. (2003). Argentine and other ants (Hymenoptera: Formicidae) in New Zealand horticultural ecosystems: distribution, hemipteran hosts, and review. *New Zeal. Entomol.*, 26, 79–89.
- Mack, R.N., Simberloff, D., Lonsdale, W.M., Evans, H., Clout, M. & Bazzaz, F.A. (2000). Biotic invasions: causes, epidemiology, global consequences, and control. *Ecol. Appl.*, 10, 689–710.
- Markin, G.P. (1970). Foraging behaviour of the Argentine ant in a California citrus grove. *J. Econ. Entomol.*, 63, 740–744.
- Miller, J. & Franklin, J. (2002). Modeling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial dependence. *Ecol. Modell.*, 157, 227–247.
- New, M., Hulme, M. & Jones, P.D. (1999). Representing twentieth century space-time climate variability. Part 1: development of a 1961–90 mean monthly terrestrial climatology. *J. Clim.*, 12, 829–856.
- Newell, W. & Barber, T.C. (1913). The Argentine ant. *US Bur. Entomol.*, 122, 1–98.
- O'Dowd, D.J., Green, P.T. & Lake, P.S. (2003). Invasional 'melt-down' on an oceanic island. *Ecol. Lett.*, 6, 812–817.
- Ormsby, M. (2005). Risk analysis for biosecurity in New Zealand. *Biosecurity*, 64, 24–25.
- Pearce, J. & Ferrier, S. (2000). Evaluating the predictive performance of habitat models developed using logistic regression. *Ecol. Modell.*, 133, 225–245.
- Pearce, J.L., Cherry, K., Drielsma, M., Ferrier, S. & Whish, G. (2001). Incorporating expert opinion and fine-scale vegetation mapping into statistical models of faunal distribution. *J. Appl. Ecol.*, 38, 412–424.
- Pearson, R.G., Dawson, T.P. & Liu, C. (2004). Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. *Ecography*, 27, 285–298.
- R Development Core Team (2005). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, Available at <http://www.R-project.org> (last accessed 1 August 2006).
- Roura-Pascual, N., Suarez, A.V., Gomez, C., Pons, P., Touyama, Y., Wild, A.L. *et al.* (2004). Geographical potential of Argentine ants (*Linepithema humile* Mayr) in the face of global climate change. *Proc. R. Soc. Lond. B*, 271, 2527–2534.
- Simberloff, D. & von Holle, B. (1999). Positive interactions of nonindigenous species: invasional meltdown?. *Biol. Invasions*, 1, 21–32.
- Sing, T., Sander, O., Beerenwinkel, N. & Lengauer, T. (2004). ROCr: An R Package for Visualizing the Performance of Scoring Classifiers. Available at <http://rocr.bioinf.mpi-sb.mpg.de> (last accessed 1 August 2006).
- Stockwell, D. & Peters, D. (1999). The GARP modelling system: problems and solutions to automated spatial prediction. *Int. J. Geogr. Inf. Sci.*, 13, 143–158.
- 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. *Proc. Natl Acad. Sci. USA*, 98, 1095–1100.
- Vega, S.J. & Rust, M.K. (2001). The Argentine ant – a significant invasive species in agricultural, urban and natural environments. *Sociobiology*, 37, 3–25.
- Walters, A.C. & Mackay, D.A. (2003). An experimental study of the relative humidity preference and survival of the Argentine ant, *Linepithema humile* (Hymenoptera, Formicidae): comparisons with a native *Iridomyrmex* species in South Australia. *Insectes Soc.*, 50, 355–360.
- Walters, A.C. & Mackay, D.A. (2004). Comparisons of upper thermal tolerances between the invasive Argentine ant (Hymenoptera: Formicidae) and two native Australian ant species. *Ann. Entomol. Soc. Am.*, 97, 971–975.
- Wild, A.L. (2004). Taxonomy and distribution of the Argentine ant, *Linepithema humile* (Hymenoptera: Formicidae). *Ann. Entomol. Soc. Am.*, 97, 1204–1215.
- Williamson, M. (1999). Invasions. *Ecography*, 22, 5–12.
- Wintle, B.A., McCarthy, M.A., Volinsky, C.T. & Kavanagh, R.P. (2003). The use of Bayesian model averaging to better

represent uncertainty in ecological models. *Conserv. Biol.*, 17, 1579–1590.

Zweig, M.H. & Campbell, G. (1993). Receiver-operator characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin. Chem.*, 39, 561–577.

SUPPLEMENTARY MATERIAL

The following supplementary material is available online from <http://www.Blackwell-Synergy.com>:

Table S1 Description of the 12 climate variables.

Table S2 Argentine ant records and their data sources.

Table S3 Model specifications.

Table S4 Confusion matrix.

Table S5 Correlations between climate variables.

Table S6 Parameter values and climatic optima for each model.

Figure S1 Model performance measures: concordance probability and total cost of misclassification.

Figure S2 Correlograms for the climate variables, ant data and model residuals.

Editor, Nicholas Gotelli

Manuscript received 21 December 2005

First decision made 19 January 2006

Second decision made 9 June 2006

Manuscript accepted 3 July 2006