PREDICTING INTRODUCTIONS AND RANGE EXPANSIONS OF THE MONK PARAKEET WITH ECOLOGICAL NICHE MODELING AND LANDSCAPE GENETICS

by

CORENTIN LUC BOHL

A dissertation submitted to the Graduate Faculty in Biology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2013

This manuscript has been read and accepted for the Graduate Faculty in Biology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Date

Chair of Examining Committee
Dr. Jason Munshi-South

Date

Executive Officer
Dr. Laurel A. Eckhardt

Dr. Robert Anderson

Dr. Ana Carnaval

Dr. Stéphane Boissinot

Dr. George Amato

Supervising Committee

The City University of New York

ABSTRACT

PREDICTING INTRODUCTIONS AND RANGE EXPANSIONS OF THE MONK PARAKEET WITH ECOLOGICAL NICHE MODELING AND LANDSCAPE GENETICS

by

CORENTIN LUC BOHL

Adviser: Professor Jason Munshi-South

The ability to predict species future geographic distributions is an important challenge in biogeography and conservation biology, with critical implications for pressing environmental issues, including the potential spread of invasive species. This research examines a two-step framework to build accurate predictions of the invasive potential of the monk parakeet (*Myiopsitta monachus*). This species, native to temperate South-America, has established several stable populations worldwide, and shares many of the typical traits of high-risk invaders. The proposed framework aims to 1) identify areas where the species is likely to thrive, and 2) determine which of these suitable areas the species can likely disperse to. Objective 1 requires identifying the environmental conditions suitable to the species, for which I used ecological niche modeling (Chapter 1 & 2). Objective 2 addresses the ability of the species to conquer new adjacent favorable areas via dispersal, a problem that can be addressed with landscape genetics (Chapter 3). In Chapter 1, I developed a null model approach to evaluate the performance and significance of ecological niche models. The results highlight the importance of accounting for both discrimination and overfitting and correctly estimating significance. In Chapter 2, I tested the effect of different model calibration strategies on transferability (the ability to predict independent data in different geographic regions). I used this information to make predictions about the global invasive potential of the monk parakeet. The best prediction was obtained with native calibration records and complex model settings. This prediction indicates several areas with conditions suitable for monk parakeets, including areas adjacent to existing

introduced populations. In Chapter 3, I integrated ecological niche modeling and landscape genetics to make predictions about the landscape features that affect monk parakeet dispersal. I tested these predictions with genetic data from an introduced population in Florida, and assessed their significance with null models. Estimating resistance to dispersal with ecological niche modeling produced results equivalent to evaluating a range of alternative hypotheses with a stepwise regression model. The results indicate that monk parakeet may not be limited by distance and most landscape features and are likely to expand to adjacent suitable areas.

ACKNOWLEDGMENTS

I would like to express my most sincere appreciation to my advisor Dr. Jason Munshi-South and to Dr. Robert Anderson for their continued support and excellent guidance throughout every step of this project. Dr. Anderson collaborated with me on the first two chapters, Dr. Munshi-South on the third one, and they both provided extensive feedback on the entire document. They have conveyed the right balance of enthusiasm and professionalism that not only helped me bring this project to completion, but also gave an example to look up to for my future career. I would also like to extend this appreciation to the current and past members of their laboratories for their support, feedback, and friendship.

I would like to thank the members of my advisory committee for their helpful advice, encouragements, and for many stimulating conversations. I am particularly grateful to Dr. Munshi-South, Dr. Boissinot, and Dr. Amato for giving me a boost of confidence at a time when I was having serious doubts about my ability to finish my degree; and I cannot thank enough Dr. Amato for so generously providing resources and training at the Sackler Institute for Comparative Genomics.

I am grateful to Dr. Frank Grasso and the current and past members of the BioMimetic and Cognitive Robotics laboratory for countless hours of observations which helped me gather important information about the behavior of monk parakeets in New York City.

Finally, I would like to express my deepest gratitude to my partner Maja Musum, my parents Sophie and Xavier Bohl, my siblings Alice, Estelle, and Antonin, and to Pauline, Arnaud, Greg, Donnie, and David. Without their love, endless support, and friendship this dissertation would never have been possible.

TABLE OF CONTENTS

Page

CHAPTERS

CHAPTER 1: Null model approaches to quantifying the performance and significance of ecological niche models

CHAPTER 2: Transferability of ecological niche modeling predictions: a test using global monk parakeet invasion

CHAPTER 3: Estimating resistance to dispersal with ecological niche modeling and evaluating significance with null resistance models

APPENDICES

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1:

Null model approaches to quantifying the performance and significance of ecological niche models

ABSTRACT

The growing importance of ecological niche models in ecology and evolutionary biology calls for unbiased estimation of model performance and significance. Two of the most common evaluation measures, the AUC and the omission error rate, require assumptions that are likely to be violated in many situations. In such cases, null model approaches provide a powerful way to estimate a relevant null distribution and correctly assess significance. I modified existing null model tests and addressed this issue using occurrence data from the invasive monk parakeet (*Myiopsitta monachus*) and the presence-background modeling technique Maxent. Specifically, I tested the ability of 20 models built with different settings and sets of calibration records to predict the same spatially independent set of evaluation records. I quantified model performance with measures of discriminatory ability and overfitting based on AUC and omission error rate. I estimated the null distributions of these measures for each experiment with null models. Furthermore, for the omission error rate, I evaluated the agreement regarding statistical significance between null models and the commonly used binomial test. As quantifications of performance, measures of overfitting generally agreed with each other, but they provided different information than the measure of discrimination. Conclusions based on null models did not match those from binomial tests, which vastly overestimated significance. The results indicate that null model tests of significance provide a suitable alternative to the binomial test, which often may be very sensitive to Type I statistical error. More generally, they confirm the importance of assessing model performance with a variety of measures that together assess both overfitting and discriminatory ability, and of correctly assessing statistical significance.

INTRODUCTION

The proliferation of ecological niche modeling (ENM) papers and the wide range of modeling techniques recently developed testify for a growing importance of ENM predictions in ecology and evolutionary biology. Areas of application include potential introductions of invasive species, response to climate change, estimation of past distributions, critical areas for endangered species, and many others. This situation calls for unbiased, statistically sound, and easily interpretable methods to evaluate ENM predictions. Note that I use ENM to refer collectively to species distribution modeling and ENM, although these approaches are not strictly equivalent (R. P. Anderson, 2012; Peterson et al., 2011, pp 271, 279). As with any statistical approach, evaluation of ENMs requires choosing the right statistic for the right question, and calculating an unbiased estimate of statistical significance. Rather than providing an exhaustive list of the statistics available (see reviews in Peterson et al., 2011, pp 150-176 and Liu *et al.* 2011), I discuss two of the most common measures: the area under the curve of the receiver operating characteristic plot (AUC / ROC), and the false negative rate, or omission error rate (OR henceforth). For these measures, I first discuss what aspect of model performance they quantify. Then, I examine their random expectations and how their statistical significance can be estimated. I argue that some estimates require assumptions that are likely to be violated in many situations. I discuss null model alternatives to estimate relevant null distributions and correctly assess significance. Finally, I present an empirical test case with monk parakeets (*Myiopsitta monachus*) in order to: 1) compare model performance measured with discrimination vs. overfitting criteria, and 2) evaluate the agreement between a modification of existing null model tests of significance and the commonly used binomial test for the omission error rate.

Model performance: discriminatory ability vs. overfitting:

Different statistics measure different aspects of model performance, for example goodness-of-fit, discriminatory ability, and overfitting (Lobo, Jiménez-Valverde, & Real, 2008; Peterson et al., 2011, pp150-176; Warren & Seifert, 2011). Goodness-of-fit is a parametric evaluation of how well the actual values of a model output match reality. Discriminatory ability refers to the ability of a model to distinguish suitable from unsuitable areas and is assessed with rank-based non-parametric approaches. Overfitting is the tendency of a model to describe the random error (or any bias in the sample) rather than the true relationship between the calibration records and the predictor variables. Hence, overfit models predict very well the calibration data but perform poorly on other data sets. Therefore, it is desirable to maximize discriminatory ability and/or goodness-of-fit while minimizing overfitting, with appropriate measures for each (Radosavljevic & Anderson, *in press*).

AUC is typically considered as a measure of overall discriminatory ability (Peterson et al., 2011, pp 169-176, but see (Jiménez-Valverde 2011). High AUC values indicate that presences are predicted more strongly (i.e., ranked higher) than absences (or pseudo-absences, or background points), across a range of all possible thresholds. In contrast, OR can measure overfitting in the sense that high OR values indicate that many presences are incorrectly classified as falling into unsuitable areas at a chosen threshold because the prediction is too tightly fitted to the calibration conditions (Anderson *et al.* 2003). Specifically, ORs indicate overfitting if the observed OR is higher than that expected theoretically for the given threshold rule applied and for an unbiased sample (Radosavljevic & Anderson, *in review*). For example, some model outputs provide direct theoretical expectations of omission rates (e.g., cumulative output of Maxent; Phillips, Anderson, & Schapire, 2006) and many others allow empirical calibration data to be used to set a threshold that holds an expected level of omission on evaluation data (e.g., %-based calibration omission rules; Peterson et al., 2011, p 122).

Null Hypothesis and significance:

Once a measure of performance is calculated, it is important to assess the significance of the obtained statistics correctly. Tests of significance cannot substitute for tests of performance, since a significant model is not necessarily a “good” model. For instance, a given binary map can achieve statistical significance even if it suffers unacceptably high omission, especially if it predicts as suitable only a small proportion of the study region (Anderson *et al.* 2002; Peterson *et al.* 2011, p 136). Nevertheless, no statistical inference can be made from measures of model performance unless they are accompanied by an appropriate estimate of significance. The prediction of any ENM forms a hypothesis of where the species is likely to be found (e.g., probability of occurrence or index of environmental suitability, depending on the assumptions, system, data, and method). Hence, it must be measured against the null hypothesis that the obtained prediction is no better than random predictions. Therefore, the random expectations need to be defined correctly, and departure from them measured with an unbiased estimate.

Concerning AUC, the performance of a random prediction is represented by the y=x (true positive rate = false positive rate) diagonal across the ROC plot, and thus an AUC value of 0.5 (Fawcett 2006). This expectation is valid for ENM predictions only with true presence/absence data, although the underlying assumptions must be carefully considered (Peterson et al., 2011, pp170-172, but see pp175-175 and Peterson, Papeş, & Soberón, 2008). The common interpretation is that a random prediction across varying thresholds should yield true positive rates and false positive rates that are both equally proportional to the corresponding fractional predicted areas. In other words, larger predictions include more presences and absences than smaller predictions by chance alone.

With background or pseudo-absence records instead of true absences, it is still possible to calculate an AUC, because with a large sample of background or pseudo-absence records the false positive rate can be approximated with the fractional predicted area at any given threshold (Anderson *et al.* 2003; Phillips *et al.* 2006; Peterson *et al.* 2008, 2011, p 171). The null hypothesis for this type of data is not the same as with true absences. The expected AUC of 0.5 for a random prediction may be valid in some situations (Phillips *et al.* 2006). However, in most cases, the null expectation will be biased towards lower or higher values, depending on the system and choice of calibration, evaluation, and background or pseudo-absence data (Peterson *et al.* 2011, p 171-172). Nevertheless, using the fractional predicted area as an approximation of the false positive rate unequivocally assumes that for a random prediction, the true positive rate is proportional to the fractional predicted area.

Departure from the random expectation of 0.5 is often judged arbitrarily using rules of thumb (Peterson et al., 2011, p172; Swets, 1988), sometimes also accounting for the standard error of AUC values obtained via partitioning of the calibration data. Although these generalities apply only to true presence/absence data, they are often wrongly used with background, or pseudo-absence data (e.g., Guisan *et al.* 2007; Hoffman *et al.* 2008; Beaumont *et al.* 2009; Reside *et al.* 2011). Alternatively, with any kind of comparison data (absences, pseudo-absences, background), a p-value can also be estimated using bootstrap replicates of the calibration or evaluation data (e.g., Peterson, Papeş, & Soberón, 2008), or with null model approaches (Raes & ter Steege 2007; Beale *et al.* 2008).

For the OR, the theoretical random expectation (not to be confused with the theoretical expectation for an unbiased sample, as discussed earlier) is the inverse of the fractional predicted area corresponding to the threshold chosen (i.e.: if 60% of the area is predicted as suitable, then the OR for a random sample should be 40%). Hence, departure from this expectation can be assessed with the p-value obtained with a binomial test (or a chi-square approximation) where the probability of predicting a presence is equal to the fractional predicted area (Anderson, 2002; Peterson et al., 2011, p168).

Limitations of spatial null hypotheses:

For OR, AUC with background or pseudo-absence data, and in some cases AUC with true absence data, the null hypothesis is established using a purely spatial criterion to define randomness. The proportion of records predicted as suitable in a random sample is expected to be equal to the number of pixels predicted as suitable out of the total number of pixels available in the study region. This assumption is often not reasonable because it does not account for structure in the variation of environmental condition across the study region. This potential structure may cause the model prediction error (i.e., the range and probability of values obtainable by chance alone) to vary considerably across the pixels of the study region. Some pixels may be more susceptible than others to either type I or type II statistical errors (i.e., rejecting a true null hypothesis or retaining a false one). There is no reason to assume that the model error is always uniformly (or randomly) distributed spatially, in other words that all the pixels of the study region have the same probability of being either over- or under-predicted (Pontius & Schneider 2001; Lobo *et al.* 2008). An appropriate null hypothesis should take into account the environmental structure present across the study region (homogeneity, spatial autocorrelation, etc.). Therefore, the null distribution of any test of ENM performance is specific to the species, sample size, region of interest, and modeling technique chosen. One way to work around this problem would be to perform the evaluation in environmental space (see Peterson et al., 2011, pp 155-156, 176-178): the fraction of predicted area could be measured in environmental rather than geographic space. However, this approach would imply accounting for the uneven representation of environmental space in geography. It would also require measuring correctly the multivariate environmental space (e.g. with principal components analysis or Mahalanobis distance) using environmental variables that can be highly correlated, categorical, or not normally distributed.

Null models assessment of significance:

Another alternative is to estimate the unknown null distribution of the desired statistic using null models. This approach consists of using replicated models with random sets of input records instead of the real species presences (Raes & ter Steege 2007; Beale *et al.* 2008). The number of input records, the model settings, and the comparison data (absence, pseudo-absence, or background records) used for each replicate run are exactly the same as the real model. The only difference is that each replicate is based on a different set of random calibration records sampled from the same region as the comparison data. In Raes & ter Steege (2007), the random records were sampled purely randomly across space. Instead, Beale et al. (2008) imposed a sampling method intended to preserve the same degree of spatial autocorrelation as the real species occurrences. In either case, to assess significance the test statistic obtained with the real model can then be compared to the null distribution of values derived from models made with random calibration records. The advantage of these approaches is that the null models are subject to the same confounding and latent variables as the real model. Therefore, these approaches provide a powerful way to distinguish the true signal of the species' niche from the error associated with the spatial structure of environmental variables across the study region, the degree of spatial autocorrelation of the species' distribution (in the Beale *et al.* method), and modeling choices (comparison records, algorithm, model settings, etc.). Despite their advantages, these approaches are rarely used in the current literature, and there is little assessment of how they perform against other approaches or across different measures of performance such as discrimination and overfitting.

Objectives:

Here, I examine how different measures of discriminatory ability and overfitting perform across a range of experiments, and evaluate the agreement between a null model approach assessment of significance and the commonly used binomial test for the omission error rate. I performed these comparisons using presence data from the invasive monk parakeet (*Myiopsitta monachus*) and the presence-background modeling technique Maxent (Phillips *et al.* 2006; Phillips & Dudík 2008; Elith *et al.* 2011). I calculated significance using modifications of existing null model approaches (Raes & ter Steege 2007; Beale *et al.* 2008). Contrary to earlier methods, I measured the performance of the random models with the same independent evaluation data as the real model. This approach offers the advantage of creating a null distribution of obtainable performance for the particular set of real species records chosen for evaluation, rather than for just any set of random records.

METHODS

Species data:

I gathered 2997 occurrence records of the morphologically distinctive monk parakeet from various sources (GBIF, and Species Link databases, records from the primary literature, and personal observations). This parakeet is native to temperate South America and has established several stable and well documented populations worldwide (Muñoz & Real 2006; Pruett-Jones *et al.* 2011). I ignored records for the sub-species *M. monachus luchsi* because it is not found in introduced populations, and is geographically, genetically, and behaviorally distinct form the other sub-species (Russello *et al.* 2008). I filtered these records in both geographic and environmental space to exclude outliers unlikely to be part of well-established monk parakeet populations. First, environmental outliers were identified based on their distance from the multivariate arithmetic mean, using the Mahalanobis distance, and with a 1% alpha level (Rousseeuw & van Zomeren 1990; Farber & Kadmon 2003; Calenge *et al.* 2008). Second, geographic outliers were identified based on the criterion that any presence must be part of an existing population and thus potentially connected to other presences via dispersal (Waples & Gaggiotti, 2006). I considered 106 km as the best estimate of dispersal capabilities for this species because it is the maximum reported dispersal distance from population genetic analyses (Gonçalves da Silva *et al.* 2010), and because it corresponds almost exactly to the cutoff of the 1% upper tail of the distribution of the log of nearest-neighbor distances between the records in this study. Outliers identified with these methods were excluded from the subsequent analyses unless they were well documented in the primary literature. In addition, to reduce the degree of spatial-autocorrelation (that likely derives at least in part from biases in sampling effort) and match the resolution of the environmental data, I kept only one record per ~10 km^2 raster cell.

I assigned the resulting 797 records to different native and introduced “populations” worldwide. For the native range, I used the limit of the known distribution (Forshaw & Cooper

1989; Ridgely *et al.* 2007). For the invaded regions I used the aforementioned criterion that occurrences from the same population must be potentially connected via dispersal. I selected four of these populations as calibration sets to create the different models: the native population in Argentina and adjacent countries (AR, n = 121), and the three largest introduced populations (Spain, ES, n = 114; the wider New York City metropolitan area, NY, n = 96; and the southern part of Florida, FL, n = 119). The 347 remaining records, including the populations in Illinois, Texas, Louisiana, Puerto Rico, the northern part of Florida, and a number of smaller populations in the rest of the world, were reserved as a spatially independent evaluation set. I opted for spatially independent evaluation because it offers a more realistic test of performance than randomly selected evaluation records; the latter, being more sensitive to environmental spatial autocorrelation, typically result in inflated estimates (Veloz 2009; Hijmans 2012; Radosavljevic & Anderson, *in review*).

Environmental data:

There is evidence showing that monk parakeet presence is influenced by a combination of climatic and anthropogenic factors, in particular those related to winter survival (Weathers & Caccamise 1975; Caccamise & Weathers 1977; Tamara & Arnheim 1996; South & Pruett-Jones 2000; Butler 2005; Muñoz & Real 2006; Strubbe & Matthysen 2009). Furthermore, monk parakeets seem to prefer open areas rather than dense forest, and favor cultivated lands (Forshaw & Cooper 1989; Burger & Gochfeld 2005). Consequently, I considered a combination of 24 climatic, anthropogenic, and land cover predictors at 5 arc-minutes resolution. These include the 19 Worldclim bioclimatic variables (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005); three land cover variables from the Harmonized World Soil Database: percent forested land, percent grass/scrub land, and percent cultivated land (Fischer et al., 2008); and two anthropogenic predictors: the human population count (Gridded Population of the World, Version 3, 2005), and the Global Human Influence Index (Last of the Wild Project, Version 2, 2005; Sanderson et al., 2002).

ENM methods and experimental design:

As an example, I chose the algorithm Maxent (Phillips *et al.* 2006; Phillips & Dudík 2008; Elith *et al.* 2011), which has been shown highly effective yet susceptible to differences in modeling settings (Elith *et al.* 2006, 2010; Warren & Seifert 2011; Anderson & Gonzalez 2011; Syfert *et al.* 2013; Cao *et al.* 2013). I created sets of models for each of the calibration populations independently (AR, ES, NY, FL) and a set for all the calibration records pooled together (ARESNYFL). In each case, I constructed different models by varying the types of feature classes (Linear, Quadratic, Product, Threshold, and Hinge) and the value of the regularization multiplier (RM) used in Maxent. Feature classes determine the flexibility of the shape of the response that can be modeled. Regularization multipliers control the strength of penalties for increasingly complex models. At one end of the spectrum, I considered settings that should lead to simple models, made using linear and quadratic features and a RM of 3 (LQ-3). At the other extreme, I utilized all feature classes and a RM of 0.5 (LQPTH-0.5), which should produce complex models.

I also examined intermediate settings optimized independently for each of the calibration sets (an approach also termed "tuning" or "smoothing"; Anderson & Gonzalez 2011). For this, I considered two alternative methods: 1) internal optimization*,* where the evaluation was performed within the same region as the calibration records, using a 10-fold random split cross-validation procedure (Peterson et al., 2011, pp 157-159), and 2) external optimization*,* where the evaluation was performed with spatially independent data, using in turn each of the calibration populations to build the models and the other three to test them. In both cases, I examined 42 different model settings with the following combinations of Linear, Quadratic, Product, Threshold, and Hinge features: LQ, LQH, LQP, LQT, LQPH, LQPT, and LQPTH; and the following RM values: 0.5, 1, 1.5, 2, 2.5, and 3. The optimal settings were selected using the evaluation statistics described in the next section.

In summary, I performed a total of 20 experiments: 4 models with different settings (simple, complex, internally optimized, and externally optimized) for each of the 5 sets of calibration records. Except for varying the value of the regularization multiplier (RM) and the type of features used, I used the default settings of Maxent, version 3.3.3k. The background data for each model corresponded to all the pixels within a 106 km of the input calibration records (VanDerWal *et al.* 2009). This distance is the maximum reported dispersal distance for monk parakeet (Gonçalves da Silva *et al.* 2010b) and is thus of good indication of the potential accessible area for this species (Anderson & Raza 2010; Barve *et al.* 2011).

Model evaluation: discrimination, overfitting, and significance:

I assessed the ability of models to predict the 347 records of the evaluation set. To allow direct comparison of the evaluation statistics across the different experiments, I calculated the values over the same background area corresponding to a 106km distance buffer around all the filtered occurrence records (i.e. both calibration and evaluation records). I also tried considering the buffered region around only the 347 records of the evaluation set, but this method made it impossible to calculate meaningful AUC_{DIFF} (see below) and resulted in only marginal differences in pattern for the other statistics (not shown).

I quantified model performance with measures of discriminatory ability and overfitting. For discriminatory ability, I considered the AUC applied to the evaluation data (AUC_{TEST}; Phillips et al., 2006) because, despite its criticisms, it remains valid for comparisons of the same species in the same study region (Lobo *et al.* 2008; Peterson *et al.* 2011). I measured overfitting with the omission error rate of the evaluation data (OR) using a 10 percentile calibration presence fixed threshold. This thresholding rule should theoretically give an OR of 10% with an unbiased sample (Liu *et al.* 2005). Conversely, models that are too tightly fitted to the calibration data should select for a restrictive threshold and yield a high OR with independent evaluation data. I also considered a threshold-independent measure of overfitting: AUC_{DIFF} , the difference between the AUCs calculated with model input records (AUC_{TRAIN}) and evaluation records

(AUC_{TEST}) (Warren & Seifert, 2011). Overfit models should yield high AUC_{TRAIN} values and low AUC_{TEST} values, and conversely, models with good independent predictive ability should result in comparable values of both, and thus a low AUC_{DIFF}. For simplicity, I considered the absolute value of this difference since I obtained a few negative, albeit small, values of AUC_{DIFF}.

I used null models to determine if the obtained values of these statistics were significantly better than random predictions. My approach is similar to those described in the introduction (Raes & ter Steege 2007; Beale *et al.* 2008). As in Raes & ter Steege (2007), I sampled the null model input records randomly across space. I preferred this approach to that of Beale *et al.* (2008) because null models created with records with the same spatial characteristics as the real species may be more likely to predict the species' distribution than null models built with records taken randomly across space (Peterson *et al.* 2009; Thomas 2010). Consequently, the Beale et al. method may make it more difficult to reach significance (i.e., more susceptible to type II statistical error), as indicated in later studies (Araújo *et al.* 2009; Peterson *et al.* 2009; Thomas 2010; but see Beale et al., 2009). However, unlike both of these studies, I measured the performance of the random models with the same independent evaluation data as the real model. For each combination of Maxent settings and set of calibration records considered in my analyses, I ran one model based on the real species data, and 1,000 corresponding replicate models based on random records to create null distributions of OR, AUC_{DIFF}, and AUC_{TEST} values. I calculated the p-value of the obtained statistics using a parametric one-tailed Z-test and a non-parametric approach based on the rank of the value obtained for real species data compared with the corresponding null values. For comparison, I also calculated p-values for OR using a binomial test (Anderson, 2002; Peterson et al., 2011, p168). These analyses were all performed in R (R Core Team 2012) and can be replicated with the script included in the Appendix A.

Comparison analyses:

I determined how the conclusions regarding model performance and significance changed depending on the evaluation statistics used and the way significance was assessed using three sets of pairwise comparisons: 1) all possible comparisons between the three measures of performance (AUC_{TEST}, AUC_{DIFF}, and OR); 2) all possible comparisons between the p-values calculated with null models (p_AUC_{TEST}, p_AUC_{DIFF}, and p_OR); and 3) the comparison between the p-values for OR calculated with the null model approach vs. the binomial approach (p_OR vs. p_binomial). I performed these comparisons using the values obtained for each statistic across the 20 experiments.

For each set of comparisons, I first calculated the Spearman's rank correlation coefficient. I expected a high positive correlation between AUC_{DIFF} and OR since they should in theory both describe overfitting. Conversely, I did not expect AUC_{TEST} to be correlated either AUC_{DIFF} or OR, since it should measure a different aspect of model performance. I expected the same trend for the p-values: a high correlation between p_AUC_{DIFF} and p_OR, and no correlation between p_AUC_{TEST} and the overfitting measures. I also expected a high correlation between p_OR and p_binomial since they both measure significance for the same performance measure.

Then, for the comparison involving p-values, I determined whether the decision regarding the null hypothesis was concordant across the different measures of performance. Specifically, I calculated the percentage of experiments for which the pair of measures compared led to the same conclusion regarding the null hypothesis (i.e. either both rejecting, or both retaining the null). As with the correlations, I expected a high agreement between p_AUC_{DIFF}, and p_OR and a low agreement between p_AUC_{TEST} and either of the former two measures. I also expected a high agreement between p_OR and p_binomial. In addition, I performed χ^2 tests of independence (with Yate's correction for continuity) to assess if the decision regarding the null was independent of the measures of performance used. A significant χ^2 test would indicate that the measures compared do not lead to the same conclusion regarding the null. Thus, I expected significant χ^2 tests between p_AUC$_{TEST}$ and the p-values of the overfitting measures, but not between p_AUC$_{DIFF}$, and p_OR, or between p_OR and p_binomial.

In order to decide whether to retain or reject the null for each experiment, I used an alpha level of 0.05 and considered two scenarios: a) an analysis of the statistical conclusions that would be reached independently for each experiment; b) the statistical conclusion for the 20 experiments analyzed together. The second scenario implies accounting for the inflated risk of Type I error due to the evaluation of 20 different null hypotheses. Thus, I used an unmodified alpha level of 5% for the first scenario, and Holm's (1979) sequential Bonferroni correction for the second scenario.

RESULTS

Discrimination vs. overfitting, correlations:

As expected, I found that AUC$_{DIFF}$ and OR presented very similar trends (Figure 1.1 & Table 1.1) and were highly correlated (rho = 0.90). In contrast, AUC$_{TEST}$ displayed a markedly different trend than the measures of overfitting (Figure 1.1 & Table 1.1), and was only moderately correlated with AUC$_{DIFF}$ (rho = - 0.60) and OR (rho = -0.45). This pattern was the same for the corresponding p-values calculated with the null model approach (Figure 1.1 & Table 1.1), with a high correlation between p_AUC$_{DIFF}$ and p_OR and moderate correlations between p_AUC$_{TEST}$ and both p_AUC$_{DIFF}$ and p_OR (Table 1.2). Note that I only report the p-values calculated with the Z-test because in most cases the null values were approximately normally distributed and the non-parametric approach resulted in almost identical p-values (not shown).

Discrimination vs. overfitting, agreement regarding the null:

Similarly, the agreement regarding the null hypothesis was very high between p_AUC$_{DIFF}$ and p_OR, and poor between either of those measures and p_AUC$_{TEST}$ (Table 1.2). The χ^2 tests confirm that p_AUC$_{TEST}$ led to significantly different conclusions regarding the null than AUC$_{DIFF}$

and OR (Table 1.2). In contrast, the null hypothesis that p_AUC$_{DIFF}$ and p_OR led to the same conclusions could not be rejected (Table 1.2). These conclusions were the same regardless of whether correction for multiple tests was applied to the alpha level or not. In general, the p-values were very low for p_AUC$_{TEST}$ (high significance) and very high for p_AUC$_{DIFF}$ and p_OR (poor significance), with little correspondence to the actual values obtained for AUC$_{TEST}$, AUC$_{DIFF}$, and OR (Figure 1.1 & Table 1.1). Unlike p_AUC$_{TEST}$, p_AUC$_{DIFF}$ and p_OR rarely led to rejecting the null hypothesis (Figure 1.2).

Null models vs. binomial test:

Finally, for OR, conclusions based on randomization approaches did not match those from binomial tests. The OR p-values calculated using a binomial test were consistently low and almost always led to rejecting the null (Figure 1.2). Despite a moderate correlation to the values obtained with the null model approach (Table 1.2), these two methods led to very different conclusions (Figure 1.1, Table 1.1 & 1.2). The agreement between p_OR and p_binomial was very low and the χ^2 tests confirmed that the binomial test and null models led to significantly different conclusions regarding the null hypothesis (Table 1.2).

DISCUSSION

Discriminatory ability vs. overfitting:

The results confirm that despite their different nature (one is threshold-independent but not the other), AUC$_{DIFF}$ and OR (with a 10% of calibration presences threshold) measure overfitting. These statistics were indeed highly correalted (both the actual values, and their corresponding p-values), and almost systematically led to the same conclusion regarding the null hypothesis. Conversely, there was only a moderate relationship between performances in overfitting and discriminatory ability, with little agreement concerning the decision to retain or reject the null hypothesis. In fact, the p-values were generally very low for AUC$_{TEST}$, and very high for AUC$_{DIFF}$ and OR. AUC$_{TEST}$ may therefore be more susceptible to Type I error *(rejecting the null when the null is true)* while AUC$_{DIFF}$ and OR may be more subject to Type II error *(retaining the null when* *the null is false)*. Consequently, while it may be sufficient to consider only one statistic to assess overfitting, model performance should always be assessed with both discriminatory ability and overfitting criteria.

Performance vs. significance:

The results confirm that it is misleading to rely on the value of a statistic without comparison with a relevant null hypothesis. Models with poor discriminatory and overfitting performance were generally not significant. However, many models with apparently good performance were not significant either. In such cases, significance may have been reached had more statistical power been available. However, as in Merckx *et al.* (2011; observations made with AUC values), for several experiments the null distributions were shifted towards higher performance values (high values of AUC_{TEST} and low values of AUC_{DIFF} and OR). This observation indicates that depending on the system and modeling choices, high performance can be routinely obtained by chance alone. Therefore, inferences made from performance values without careful examination of an appropriate null hypothesis may lead to biased or incomplete conclusions. For instance, the values of the overfitting measures support the notion that simpler model settings provide more extrapolative ability (Jiménez-Valverde *et al.* 2008; Rodda *et al.* 2011; Heinänen *et al.* 2012). This conclusion would however be incomplete because the corresponding p-values of these statistics were very high (close to 1). In many cases simpler model settings led to generally larger prediction, regardless of whether real or random data are used to train the models. Such a situation makes it very difficult to achieve enough statistical power to reject the null. Thus, while simpler model settings may provide better transferability, a greater number of test records are necessary to reach that conclusion with enough statistical confidence.

Furthermore, the relationship between significance and performance is not straightforward: some values of overfitting and discriminatory ability allowed the rejection of the null hypothesis for some experiments but not others. This observation should encourage circumspection when comparing the performance of different predictions, even within the same species. Moreover, the results confirm that rules of thumb to interpret the AUC values (Swets 1988) are not appropriate for using AUC with presence/background or presence/pseudo-absence data, and that departure from the theoretical expectation of 0.5 does not indicate significance with such data (Peterson et al., 2011, pp171-172). Unfortunately, these points are still often misunderstood in the literature. The random expectation of 0.5 may often lead to misleading results even with true presence/absence data. The assumptions on which they rest are probably not met in many situations. In such cases, null models applied to presence/absence data will likely reveal significant departure from these expectations. Therefore, I advocate that researchers systematically include appropriate tests of statistical significance along with measures of model performance.

Null models vs. binomial test:

The fact that null models can produce relatively high performance values of implies that the predictive performance may be affected by factors independent of the species distribution. Therefore, tests of statistical significance that do not control for these potential confounding factors may lead to misleading results. The binomial test for OR may be particularly affected by this problem. This test rests on the assumption that the pixels of the study region are independent observations. This assumption is probably violated in many situations because of existing structure in the variation of environmental condition across the study region. Here, this method consistently produced extremely low p-values, even when the null model approach yielded non-significant values. This test thus probably results in high rates of Type I error *(rejecting the null when the null is true)*. Other methods based on bootstrap replicates of the calibration or evaluation data (e.g., Peterson, Papeş, & Soberón, 2008) may be equally susceptible to this problem since they account for error inherent to the evaluation and/or calibration data but not for sources of error associated with modeling choices (number and nature of predictor variables, spatial structure and autocorrelation of these variables, size and shape of background area, model algorithm and settings). Null model approaches provide a more robust alternative, since the models based on random occurrences are subjected to the same potential confounds as the real model. These methods are thus appropriate tests to assess if the species' environmental response is reliably different from the background noise or latent patterns independent of the species distribution, except admittedly for issues of sampling bias (Raes & ter Steege 2007; Merckx *et al.* 2011).

As indicated earlier, my approach for evaluating the null models is different than earlier methods (Raes & ter Steege 2007; Beale *et al.* 2008). Those studies compared the performance of models calibrated and evaluated with real species data to the performance of null models calibrated and evaluated with null data. This approach is thus somewhat flawed becaused it compares statistics calculated with different sets of records. Instead, I measured the performance of all models (real and null) with the same spatially independent evaluation records. The null distributions thus created are therefore an indication of the degree of discriminatory ability and overfitting that can be obtained by chance for the particular set of real species records chosen for evaluation, rather than for just any set of random points. Furthermore, Raes & ter Steege advocated not using independent evaluation data at all and instead comparing the AUC_{TRAIN} (ability to discriminate the input records) of the random models to that of the real model. Their method may favor unrealisticaly complex model settings since performance is measured on the same records used to create the models.

Conclusion:

This study offers a compelling case that overfitting and discriminatory ability account for different aspects of model performance, that high performance does not imply high significance, and that different methods to estimate significance can lead to very different results. These results might only apply to monk parakeets, the study regions and environmental predictors selected here, the specific thresholding rule used for OR, and a task which involves making predictions in a different region than where the models were trained. Despite these conservative conclusions, the results highlight the need to measure model performance with both overfitting and discriminatory ability criteria, and to correctly assess statistical significance. I recommend being particularly critical of p-values calculated with the binomial test, which is probably very sensitive to Type I error. Instead, I encourage researchers to consider null model tests of significance such as the one used in this study. These tests are more appropriate because they contrast the species' signal with the background noise resulting from factors associated with modeling decisions and the structure of environmental condition in the study region.

Table 1.1: Performance and significance of 20 experiments with different calibration records and model settings. The values of the evaluation statistics ($\mathrm{AUC}_{\mathrm{TEST}}$, $\mathrm{AUC}_{\mathrm{DIFF}}$, and OR with a fixed threshold of 10% of calibration presences) are presented along with their corresponding p-values estimated from 1000 replicate null-models (p_$\mathrm{AUC}_{\mathrm{TEST}}$, p_$\mathrm{AUC}_{\mathrm{DIFF}}$, and p_OR). The p_binomial column shows the p-values calculated using a binomial test for OR. Significant p-values with a 5% decision rule are bolded; those that remain significant after accounting for multiple comparisons with Holm's sequential Bonferroni correction are flagged with asterisks.

Population	Settings	Features-RM	$\mathrm{AUC}_{\mathrm{TEST}}$	p_$\mathrm{AUC}_{\mathrm{TEST}}$	$\mathrm{AUC}_{\mathrm{DIFF}}$	p_$\mathrm{AUC}_{\mathrm{DIFF}}$	OR	p_OR	P_binomial
ARESNYFL	Most specific	LQPTH-0.5	0.8101	**0.0001***	0.1343	0.1233	0.3170	0.1184	**0.0000***
ARESNYFL	Optimal Internal	LQPH-0.5	0.7617	**0.0050***	0.1657	0.4519	0.3516	0.3539	**0.0000***
ARESNYFL	Optimal External	LQPH-1.5	0.7836	**0.0064***	0.1265	0.3604	0.2853	0.3011	**0.0000***
ARESNYFL	Most general	LQ-3	0.8393	**0.0210***	0.0140	0.0896	0.1095	0.2302	**0.0000***
AR	Most specific	LQPTH-0.5	0.7479	**0.0020***	0.1203	**0.0394**	0.3026	**0.0142**	**0.0000***
AR	Optimal Internal	LQPT-1	0.8038	**0.0018***	0.0845	0.0572	0.2882	**0.0436**	**0.0000***
AR	Optimal External	LQP-1.5	0.7922	**0.0065***	0.0450	0.0987	0.1556	0.0500	**0.0000***
AR	Most general	LQ-3	0.7705	0.0504	0.0555	0.1293	0.1210	0.1439	**0.0000***
ES	Most specific	LQPTH-0.5	0.7314	**0.0361**	0.2595	0.3503	0.9798	0.8864	0.7915
ES	Optimal Internal	LQPH-1.5	0.7252	**0.0322***	0.1943	0.3412	0.7291	0.7347	**0.0000***
ES	Optimal External	LQP-2.5	0.7447	**0.0154***	0.0241	0.0788	0.1758	0.0888	**0.0000***
ES	Most general	LQ-3	0.7522	**0.0370**	0.0476	0.0857	0.2709	0.2966	**0.0000***
FL	Most specific	LQPTH-0.5	0.7440	**0.0381**	0.2420	0.4595	0.9885	0.9194	0.9892
FL	Optimal Internal	LQPH-1.5	0.8002	**0.0145***	0.1796	0.4367	0.9366	0.9766	**0.0464**
FL	Optimal External	LQT-0.5	0.7669	**0.0037***	0.2211	0.5365	0.9827	0.9858	0.8482
FL	Most general	LQ-3	0.7175	0.0719	0.2385	0.7325	0.8386	0.9892	**0.0004***
NY	Most specific	LQPTH-0.5	0.7024	0.0723	0.2867	0.4304	0.9107	0.8179	**0.0000***
NY	Optimal Internal	LQPH-0.5	0.8054	**0.0140***	0.1852	0.2660	0.8127	0.7192	**0.0000***
NY	Optimal External	LQP-0.5	0.7535	**0.0178***	0.0950	0.1955	0.3948	0.2827	**0.0000***
NY	Most general	LQ-3	0.7170	0.1072	0.2647	0.6166	0.8357	0.9106	**0.0000***

Table 1.2: Comparisons between the significance of various measures of performance. First, all possible comparisons are made between pairs of p-values obtained with the null-model approach for each measure of performance (p_AUC$_{TEST}$, p_AUC$_{DIFF}$, and p_OR). Then, the p-value calculated using a binomial test for OR (p_binomial) is compared to the one obtained with the null-model approach (p_OR). The columns show the Spearman's rank correlation coefficient, the percentage of agreement regarding the null hypothesis with and without correction for multiple tests, and the p-value of χ^2 tests of independence (indicating the probability that the measures compared lead to the same conclusion regarding the null), again with and without correction for the 20 tests. Correlations and χ^2 tests of independence that are significant are indicated in bold. Note that measures of overfitting generally agree with each other, but provide different information than the measure of discrimination. Additionally, note that for OR conclusions based on randomization approaches do not match those from binomial tests.

		% Agreement		**χ2 test p-value**	
Comparison	**Spearman rho**	**α = .05, uncorrected**	**Holm's correction**	**α = .05, uncorrected**	**Holm's correction**
p_AUC$_{TEST}$ vs. p_AUC$_{DIFF}$	0.4060	25%	35%	**0.0000**	**0.0033**
p_AUC$_{TEST}$ vs. p_OR	**0.4860**	30%	35%	**0.0000**	**0.0033**
p_AUC$_{DIFF}$ vs. p_OR	**0.9110**	95%	100%	0.3959	1.0000
p_binomial vs. p_OR	**0.7110**	25%	20%	**0.0000**	**0.0000**

Figure 1.1: Model performance and significance as a function of model settings and calibration population. Darker shades of gray indicate higher performance. The top three graphs show the actual values of the evaluation statistics, while the bottom three show their corresponding p-values obtained with the null model approach. For each graph, the rows show the different calibration populations, and the columns represent the type of model settings, in order of decreasing complexity. Optimal Internal and Optimal External refer to the settings optimized for each population (see Methods for details). Significant p-values at the uncorrected 5% level are indicated with one asterisk, those that remain significant after Holm's sequential Bonferroni correction are flagged with two asterisks. The p-values calculated with the binomial approach for OR (not shown here) were all significant except for the cases indicated with "ns" in the bottom right graph (although the one for the FL records with optimal internal settings was significant without correction for multiple tests).

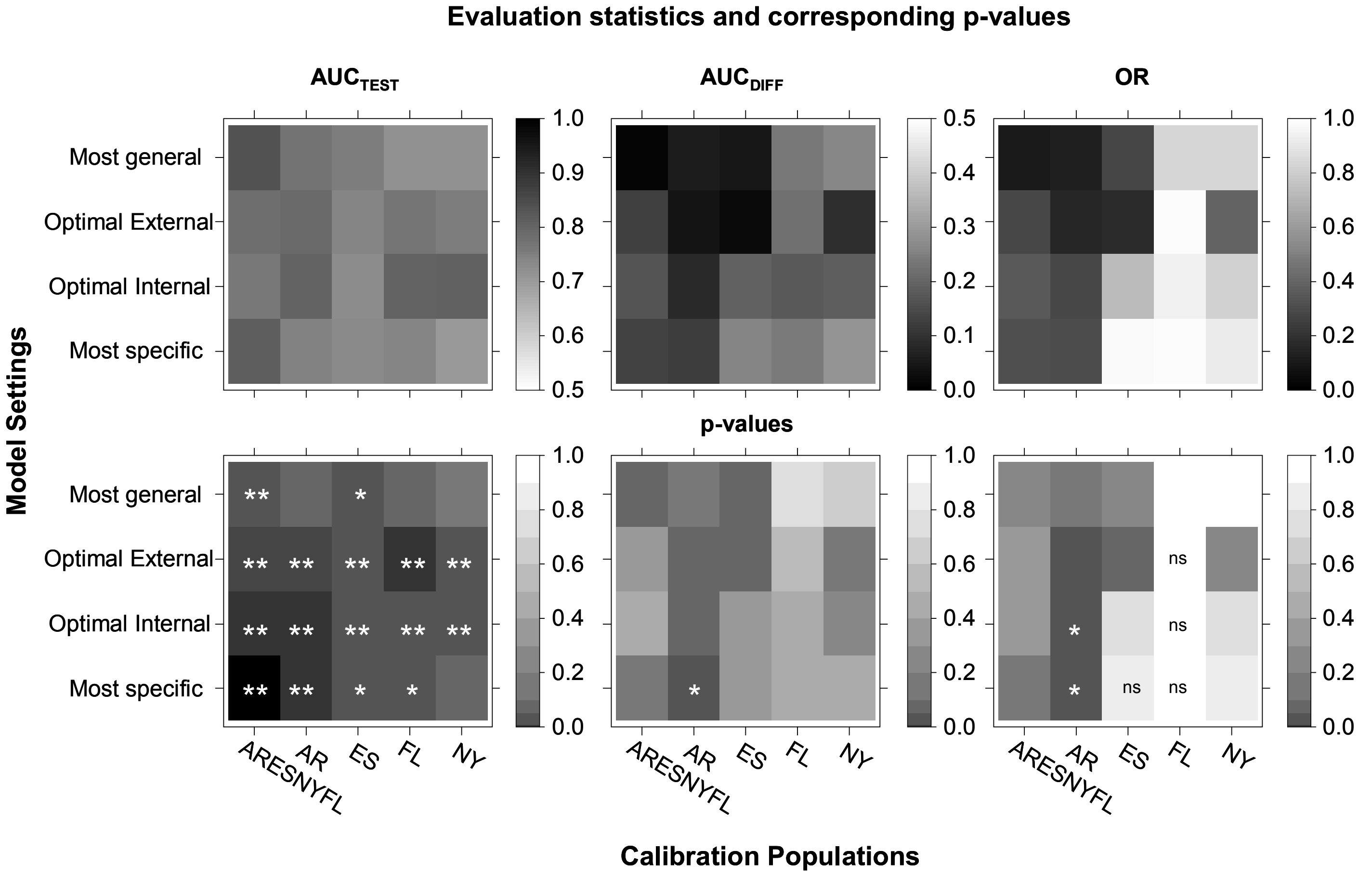
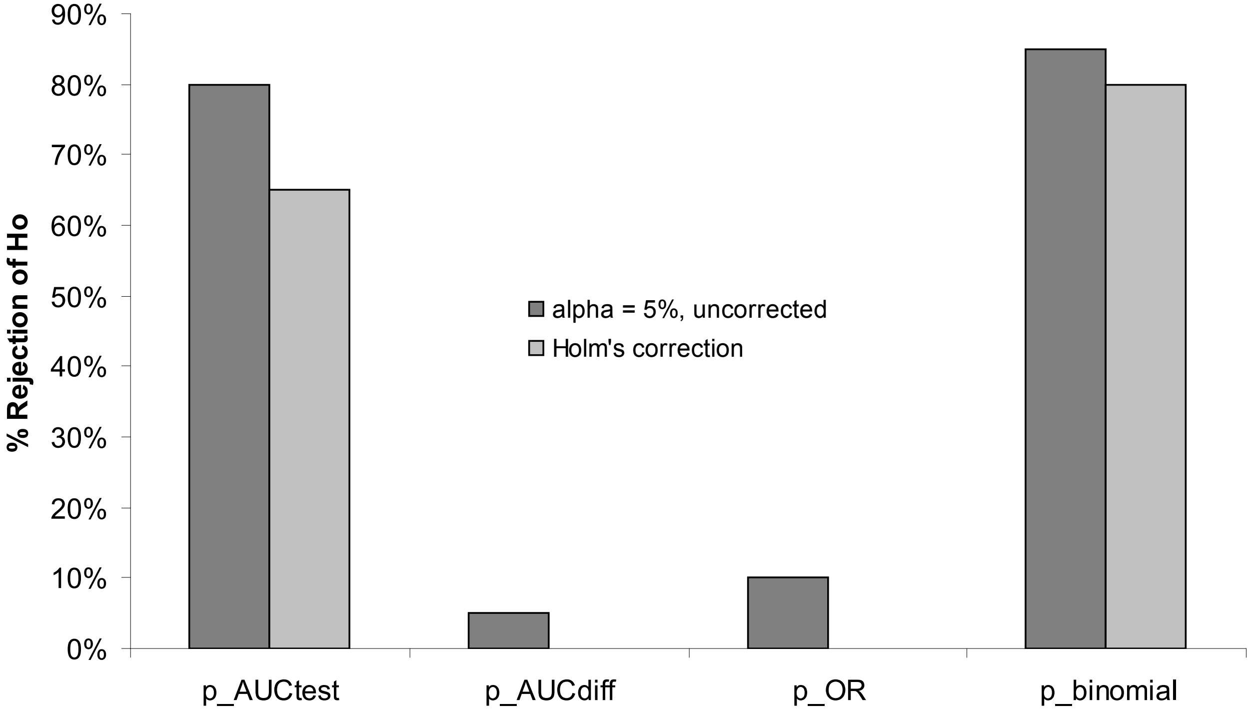


Figure 1.2: Percent of experiments that lead to the rejection of the null hypothesis for each measure of significance with an alpha level of 5%, using either no correction for multiple tests, or Holm’s sequential Bonferroni correction.



CHAPTER 2:

Transferability of ecological niche modeling predictions: a test using global monk parakeet invasions

ABSTRACT

Obtaining predictions of species potential distributions that can be accurately projected across space and time is a major challenge in the field of ecological niche modeling. Invasive species provide a unique opportunity to address this problem because they occupy areas subjected to unique combinations of biotic, abiotic, and historical conditions which can be used a natural experiments to test the transferability of ecological niche model predictions. I examined the issue of transferability with the monk parakeet (*Myiopsitta monachus*), a species originally native to temperate South America that has established several stable populations worldwide, and the presence-background modeling technique Maxent. I examined the effect of the choice of calibration records (native vs. introduced, vs. both combined) and settings of Maxent on transferability. Specifically, I tested the ability of 20 models built with different settings and sets of calibration records to predict the same spatially independent set of evaluation records. I quantified model performance with measures of discriminatory ability and overfitting, and estimated significance using null models. Finally, I examined the implications of these analyses for the worldwide invasive potential of the monk parakeet. The results show that for this species, the best strategy to create reliable transferable predictions is to model the niche of the species from the pattern of occurrences in the native range using complex model settings. In contrast to previous studies, models using both native and introduced records provided good discriminatory ability but did not perform better than null models for the measures of overfitting. Simple model settings also resulted in apparently good transferability, but the predictions were not significantly better than models based on random geographic records. The only model performing better than random models across all performance criteria revealed several areas at risk of new monk

parakeet introductions, and indicated that most of the currently well established populations are likely to expand to surrounding areas.

INTRODUCTION

Increasing human mobility and international trade are responsible for unprecedented rates of species introductions worldwide (Jenkins, 1996; Peterson, 2003). Although few species establish self-sustaining populations (Munoz and Real, 2006; Peterson, 2003), invasive species pose major threats to habitat stability, biodiversity, ecosystem functioning, resource availability, and public health (Ricciardi *et al.* 2000; Russello *et al.* 2008; Peterson *et al.* 2011). Consequently, preventing future introductions and managing existing invasive populations have become major concerns in disciplines as diverse as conservation biology, agriculture, transportation, economics, and epidemiology (Peterson, 2003; Ricciardi et al., 2000). A key element to achieve these goals is the ability to develop accurate predictions of potential invasions. The predominant approaches to creating such predictions are ecological niche and/or species distribution modeling techniques (ENM / SDM; Anderson 2012; Peterson et al. 2011). These approaches are based on the concept that the distribution of a species is constrained by a set of ecological conditions (i.e. its niche), and therefore a species will only be able to invade geographical areas that match these conditions. ENM estimates the set of environmental conditions that constitute the niche of a species from the pattern of occurrence within the species' current distribution. The model can then be projected to other geographic regions to identify areas where the species is currently absent but could potentially exist (Peterson, 2003).

There is however considerable debate over the "transferability" of such predictions (Peterson *et al.* 2007; Duncan *et al.* 2009; Varela *et al.* 2009; Mandle *et al.* 2010; Rodda *et al.* 2011; Pagel & Schurr 2012). How accurate are predictions created with information from one area and projected to another region with substantially different biotic and abiotic conditions? What is the best modeling strategy to arrive at reliable predictions? These considerations are equally relevant to predictions across space as to predictions across time (Duncan *et al.* 2009).

Therefore, transferability is perhaps one of the most important challenges in the ENM field because of its implications for biological invasions and climate change studies.

Invasive species provide a unique opportunity to address this problem because they occupy areas subjected to different ranges of environmental, biotic, and historical conditions. Hence, invaded regions can be used has a natural experiments with pseudo-manipulation of the native conditions to test hypotheses about model transferability. In fact, these natural experiments across space are necessary tests of how well models might transfer across time for climate change studies.

The monk parakeet (*Myiopsitta monachus*), native to temperate South America, is an ideal candidate for this approach because it has already established stable and growing populations around the world (Figure 2.1; Muñoz & Real 2006; Pruett-Jones *et al.* 2011) and shares many of the typical traits of high-risk invaders: high propagule pressure, long distance dispersal capabilities, tolerance to a wide range of environmental conditions, and tendency to prosper in human-dominated environments (Russello *et al.* 2008; Strubbe & Matthysen 2009; Gonçalves da Silva *et al.* 2010b). Monk parakeets also cause significant economic losses through their habit of nesting on electrical infrastructure, and they are a serious agricultural pest in their native range (Russello *et al.* 2008; Gonçalves da Silva *et al.* 2010b). Finally, most introduced populations are the result of multiple independent introductions via the pet trade over the period of several decades (i.e., high propagule pressure) and have not suffered from loss of genetic diversity (Domènech *et al.* 2003; Russello *et al.* 2008; Gonçalves da Silva *et al.* 2010b). Therefore, it is unlikely that introduced populations have undergone rapid evolutionary change and I can reasonably assume that the ecological niche of this species is conserved from native to introduced regions.

Here, I investigate the effect of two modeling decisions that can impact transferability: the choice of calibration records (native vs. introduced, vs. both combined); and the choice of model settings. I address these issues from both conceptual and methodological standpoints and test the predictions with global monk parakeet presence data and the presence-background modeling technique Maxent (Phillips *et al.* 2006; Phillips & Dudík 2008; Elith *et al.* 2011). Finally, I examined the implications of these analyses for the worldwide invasive potential of monk parakeets.

Choice of calibration records:

Although a few studies of introduced species have opted for creating models based on data from invaded areas (e.g., Fitzpatrick, Weltzin, Sanders, & Dunn, 2007), most models have been calibrated using information from the native range and then projected onto remote regions. Existing introductions were then used to validate these predictions. The rationale for this approach is that species distributions are more likely to be at equilibrium with their abiotic environments in their native ranges than in recently invaded regions (Peterson *et al.* 2011, p 217). This rationale is questionable, because non-equilibrium dynamics are common in ecology, especially considering the current rate and intensity of human disturbances (Pagel & Schurr 2012). For example, the current monk parakeet native distribution is the result of a relatively recent (~ 100 years) population expansion facilitated by large-scale plantation of *Eucalyptus* (Gibson 1920; Forshaw & Cooper 1989; Navarro *et al.* 1992). Nonetheless, species have had more opportunities to sample their environment in their native ranges than in introduced areas. Hence, the portion of the environmental space occupied by species in invaded areas is often largely constrained compared to what is observed in native ranges (Strubbe *et al.* 2013). Data from native ranges are thus more likely to be representative of species' niches than data from invaded regions.

Nevertheless, good model transferability is not guaranteed even when the equilibrium assumption is met (as in Varela *et al.* 2009, with prediction across time). Different regions, or time periods, present unique combinations of biotic and abiotic conditions, as well as unique historical and landscape limitations to movement and dispersal, which all impact the ability of a species to occupy the full extent of its fundamental ecological niche (Pulliam 2000; Soberon &

Peterson 2005; Soberón & Nakamura 2009; Peterson *et al.* 2011). However, models based on the pattern of presence (or presence/absence) of a species can only model the portion of the niche that is occupied rather than the complete fundamental niche. Given that species may not occupy the same portion of their ecological niche in different regions, models based on occurrences from one region will only have a limited ability to predict occurrences in other regions (or time periods). Several studies have indeed shown evidence of differences in occupied environmental space between native and introduced areas, and hence poor ability to predict invaded ranges from the pattern of native occurrences (Broennimann *et al.* 2007; Beaumont *et al.* 2009; Rödder & Lötters 2010).

Given that the portion of the niche occupied in recently invaded areas is often largely constrained compared to the native range (Strubbe *et al.* 2013), models based on native occurrences should depict a wider portion of the fundamental niche than models based on invaded areas. Nevertheless, introduced records are still representative of the fundamental niche of the species (provided they are not from transient or sink populations). Therefore a model using all the records available in the native and introduced ranges should model a wider portion of the fundamental niche. Consequently, several researchers have argued, and in some cases demonstrated, that the best strategy to predict potential invasions is to incorporate information from both native and introduced ranges (Broennimann & Guisan 2008; Beaumont *et al.* 2009; Heikkinen *et al.* 2012; Gallien *et al.* 2012).

Based on these considerations, I make the following predictions:

(i) – Models based on native occurrences should provide better overall transferable predictions than models based on records in invaded areas.

(ii) – Models incorporating all available occurrence records (both native and introduced) should provide better transferable predictions than models using records from individual restricted geographic regions (in either native or introduced ranges).

This study tests these predictions by comparing the performance of models built using data from different native and introduced populations of the monk parakeet.

Choice of model settings:

Another challenge linked to creating accurate predictions is to choose the appropriate modeling algorithm and model calibration settings. There is indeed evidence that different algorithms, and different model settings, will perform better than others at different predictive tasks such as transferability (Peterson *et al.* 2007, 2011; Warren & Seifert 2011; Rodda *et al.* 2011; Anderson & Gonzalez 2011; Heikkinen *et al.* 2012). I focus here on the ENM algorithm Maxent (Phillips *et al.* 2004, 2006; Elith *et al.* 2011) because compared with other techniques it has performed well in both overall predictive ability (Elith *et al.* 2006, 2010), and transferability (Heikkinen *et al.* 2012). Nevertheless, when used with default settings, Maxent (and other algorithms) tend to produce poorly transferable predictions (Warren & Seifert 2011; Rodda *et al.* 2011). Some have argued that simpler, more general settings should be preferred to achieve better transferability (Jiménez-Valverde *et al.* 2008; Rodda *et al.* 2011; Heinänen *et al.* 2012), while others have found that either too complex or overly simplified models perform equally bad, and recommend using intermediate settings (Warren & Seifert 2011). Therefore, the secondary purpose of this study is to compare the ability of different model calibration strategies to produce transferable predictions. I compared predictions obtained with three types of model settings: simple, complex, and fitted specifically to each set of calibration records in order to optimize model discriminatory ability while minimizing overfitting.

Objectives:

The final intent of this study is to create reliable worldwide predictions of environments suitable for monk parakeets. This problem requires identifying the modeling strategy that results in the best transferability. In summary, this study addresses the following goals: 1) test prediction *(i)* and *(ii)* by comparing the transferability of predictions created with models based on different sets of occurrences (native occurrences, records in invaded areas, and records from both native and invaded areas combined); 2) determine which Maxent settings are best suited to create transferable predictions; and 3) identify the model providing the best transferability for monk parakeets and use it to predict the worldwide invasive potential of this species.

METHODS

Although I analyzed the results with a different intent, the following methods are based on the same the same data and experiments as in Chapter 1. Therefore, I only provide a summary of the relevant methods and refer the reader to Chapter 1 and Appendix A for additional details. In this section, I emphasize some points that are more relevant to this chapter than the first one, particularly those that relate to transferability and the monk parakeet biology.

Species data:

Following the procedures described in Chapter 1, I selected the records of four monk parakeet “populations” as calibration sets to create the different models (Figure 2.1): the native population in Argentina and adjacent countries (AR, n = 121), and the three largest introduced populations (Spain, ES, n = 114; the wider New York City metropolitan area, NY, n = 96; and the southern part of Florida, FL, n = 119). The 347 remaining records were reserved as a spatially independent evaluation set (Figure 2.1). Note that I only included the southern part of the species’ range in Florida (about half of the records in that population) to artificially induce an extreme violation of the equilibrium assumption.

Environmental data:

Frostbite resulting from the combination of cold and wet winters is a common cause of mortality for this parakeet and other parrot species in temperate regions (Tamara & Arnheim 1996; Butler 2005). This finding is supported by Strubbe and Matthysen (2009), who modeled the distribution of monk parakeets in their native range and Europe and found that the best predictor was the number of frost days per year, closely followed by the log of human population density, and to a lesser extent high mean winter NDVI (an index of plant productivity, hence food availability). These birds tolerate cold temperature but compensate with a higher metabolic rate (Weathers & Caccamise 1975; Caccamise & Weathers 1977), and consequently require more food during cold periods. However, introduced populations in northern cities may predominantly depend on seeds from bird feeders during winter months (South & Pruett-Jones 2000). Thus, human population density may be a better indication of winter food availability than winter NDVI. In fact, this species is associated with human presence in both native and invaded areas. This trend is supported by the study described earlier (Strubbe & Matthysen 2009), as well as another modeling study where human influence was found to account for almost 64% of the Spanish distribution (Muñoz & Real 2006). Furthermore, this species prefers open areas over dense forest, and favors cultivated lands (Burger & Gochfeld 2005).

Therefore, I considered a combination of 24 climatic, anthropogenic, and land cover predictors at 5 arc-minutes resolution, which is approximately equivalent to the daily foraging range of this bird (Forshaw & Cooper 1989; Spreyer & Bucher 1998). These predictors are: the 19 Worldclim bioclimatic variables (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005); three land cover variables from the Harmonized World Soil Database: percent forested land, percent grass/scrub land, and percent cultivated land (Fischer et al., 2008); and two anthropogenic predictors: the human population count (Gridded Population of the World, Version 3, 2005) and the Global Human Influence Index (Last of the Wild Project, Version 2, 2005; Sanderson et al., 2002).

Analyses:

To test for prediction *(i) (native occurrence models provide better transferability than models with records from invaded regions)*, I created models for the native population (AR), and for each of the three "calibration" invaded populations (ES, NY, FL), and compared their ability to predict the occurrences of the evaluation set. To test for predictions *(ii) (models using all available occurrences across separate regions provide better transferability than models with records from individual regions)*, I created an additional set of models using all the records of the calibration set (ARESNYFL), and again assessed their ability to predict the evaluation set compared with the previous models based on individual populations.

These predictions may be affected by the degree of similarity between the environmental conditions present across the different native and introduced ranges (i.e. cross-predictions should be more accurate among set of records with similar environmental conditions). I therefore assessed the ranges of environmental conditions experienced by each population using a principal component analysis. I calculated the PCA with the values of all background and species presence records for the 24 standardized environmental variables using singular decomposition of the data matrix with the ‘prcomp’ function in the R ‘stats’ package (R Core Team 2012). I plotted the resulting scores for each population with their 95% confidence ellipses along the first two principal components.

To address the second objective of this study *(which model settings provide better transferability?)*, I examined four different types of model settings for each set of calibration occurrences by varying the types of features (Linear, Quadratic, Product, Threshold, and Hinge) and the value of the regularization multiplier (RM) used in Maxent (Phillips *et al.* 2004, 2006; Elith *et al.* 2011). I considered the simple, complex, internally optimized, and externally optimized settings described in Chapter1. In total, I performed 20 experiments: 4 models with different settings (simple, complex, internally optimized, and externally optimized) for each of the 5 sets of calibration records (AR, ES, NY, FL, and ARESNYFL).

Finally, to address the third objective of this study (assess the global invasive potential of the study species), I identified the model significantly providing the best transferability and projected it on the rest of the world. I evaluated the reliability of the resulting prediction by calculating the absolute difference between the predictions with and without clamping. Clamping refers to treating the variables outside the range of environmental conditions present in calibration region as if they were at the limit of the calibration range (Phillips *et al.* 2004, 2006; Elith *et al.* 2011). Without clamping, Maxent extrapolates the response of these variables in environmental space. Areas of the map that present a wider range of environmental conditions than the calibration region may show different predicted suitability values with and without clamping. Therefore, the model prediction is less reliable in areas that show a large absolute difference than in areas where the difference is minimal. Lastly, to identify the variables with the largest importance in the final model, I performed standard Maxent analyses of variable importance (percent contribution, permutation importance, and jackknife tests).

Evaluation statistics:

I evaluated the models based on their ability to predict the 347 records of the evaluation set using the measures of performance described in Chapter 1. To quantify discriminatory ability (the ability of a model to distinguish suitable from unsuitable areas), I used the area under the receiver operating characteristic curve (AUC) applied to the evaluation data (AUC_{TEST}; Phillips et al., 2006). To quantify overfitting (the tendency of a model to describe the random error, or any bias in the sample, rather than the true relationship between the calibration records and the predictor variables), I considered the difference between the AUCs calculated using calibration and evaluation records (AUC_{DIFF}; Warren & Seifert, 2011); and the omission error rate of the evaluation data (OR) using a 10 percentile calibration presence fixed threshold (Liu *et al.* 2005). Overfitting is particularly detrimental to transferability because it tends to underestimate the niche of a species (Phillips & Dudík 2008; Anderson & Gonzalez 2011). I estimated the significance of the obtained statistics using the null-model approach described in Chapter 1 and with an alpha level of 0.05.

To assess performance across the three evaluation criteria, I also created a composite measure of performance giving equal importance to discriminatory ability and overfitting: Z-Tradeoff, the weighted average of the Z-scores obtained for each statistic (the Z score of AUC_{TEST} was counted twice in order to compensate for the fact that two different measures of overfitting were used but only one for discrimination). The Z scores for each statistic indicate how extreme the obtained values are relative to the distribution of values obtained across all experiments. Z scores are expressed in number of standard deviations and can thus be used to compare values from different distributions (Zar 1998).

RESULTS

Comparison of choice of calibration records:

Concerning the effect of calibration records on transferability, the models based on all calibration occurrences provided the best discriminatory ability (highest and most significant AUC_{TEST} values; Table 2.1 & Figure 2.2). They also did perform generally better than the models based on the introduced populations across all criteria. Nevertheless, they did not significantly outperform random models with the AUC_{DIFF} and OR statistics (Table 2.1 & Figure 2.2). The AR models generally performed better on these overfitting criteria, with lower associated p-values (Table 2.1 & Figure 2.2).

According to the Z-Tradeoff measure, the best performing model was obtained using all the calibration records (ARESNYFL), and the most general settings (LQ-3; Table 2.1). Although this model was significant for AUC_{TEST} ($p = 0.02$), its predictions were not significant across the other two criteria. The only model to produce significant predictions across the three criteria was based on native occurrences (AR), using the most complex settings (Table 2.1 & Figure 2.2).

Concerning the portion of environmental space occupied by each set of calibration population, the results of the PCA show that most of the records from introduced populations occupy small portions of the environmental space (Figure 2.3), with little overlap between the different introduced populations. In contrast, even if they do not cover as wide a range as the set combining all the calibration occurrences (ARESNYFL), the records from the native range are widely distributed along the environmental spectrum.

Comparison of model settings:

Concerning the effect of Maxent settings on transferability, the external optimization approach consistently selected simpler model settings than the internal optimization approach (Figure 2.2 & Table 2.1). Regarding the measures of overfitting, simpler settings resulted in lower AUC_{DIFF} and OR values than did more complex settings (Figure 2.2 & Table 2.1). Nevertheless, the values obtained with simpler settings were rarely significant (only one experiment for AUC_{DIFF}, and two for OR; Figure 2.2 &Table 2.1). For the experiments with significantly low overfitting measures, more complex settings (most complex and optimal internal) led to higher significance than simpler settings (Figure 2.2 & Table 2.1). However, for the three introduced populations, the external optimization approach generally yielded the best results regarding overfitting in terms of both performance and significance (Figure 2.2).

In terms of discriminatory ability, the AUC_{TEST} values were relatively similar across the different model settings (internal optimization performed better except for ES and ARESNYFL). Nevertheless, as with overfitting, the corresponding p-values were generally lower with more specific settings, especially for the AR and ARESNYFL experiments (Figure 2.2 & Table 2.1). As with overfitting, external optimization led to higher significance with the introduced populations (although internal optimization performed marginally better for NY; Figure 2.2 & Table 2.1). In general, higher significance was reached across all three measures with the most complex settings for the models based on native records and on all the calibration records combined, and with external optimization for the models based on records in invaded regions (Figure 2.2 & Table 2.1).

Worldwide predictions of environmental suitability and variable importance:

The worldwide projection of the prediction of the only significant model reveals several areas with high and reliable predicted suitability (Figure 2.4). These areas include eastern and northern India, Pakistan, coastal areas along the Mediterranean Sea, the southern part of the Red Sea, the Persian Gulf, the Baja Peninsula of Mexico, and a large portion of sub-Saharan Africa. Except for Spain, Italy, and Israel, monk parakeets have not been reported in these areas. The areas adjacent to existing introduced populations in Texas, central Mexico, and the eastern part of Florida are also fairly well predicted. This is also true of areas near existing populations in Illinois, Indiana, Long Island, England, Belgium, and Holland, although these predictions are less reliable (Figure 2.4).

The Maxent analyses of variable importance (variable contribution, permutation importance, and the jackknife tests) show that the variable with the highest percent contribution (15%), and the highest calibration gain when used on its own was the precipitation of the driest month (Table 2.2). The variables with the highest permutation importance were the annual mean temperature (11.4%), the minimum temperature of the coldest month (10.8%), and the mean temperature of the wettest quarter (10.1%; Table 2.2). Human population count was also very important: the second highest model gain contribution (13.5%), and the largest gain loss when excluded from the models (Table 2.2).

DISCUSSION

Implications for transferability:

The results of this study confirmed prediction *(i)*: only one model provided significantly better transferable predictions than random models, and that model was constructed using occurrences from the native range. In fact, even if their predictions were not significant across all performance criteria (Table 2.1), the other three models created with native records performed better than all the other models, except for the one with all occurrence records and simple settings (Table 2.1 & Figure 2.2).

As mentioned in the introduction, the probable explanation for this outcome is that the monk parakeet distribution is more likely to be at equilibrium with its environment in the native range than in invaded areas. Indeed, the models using only southern Florida records, which were selected to artificially induce an extreme violation of the equilibrium assumption, provided some of the worse predictions (Table 2.1). In addition, as in Strubbe *et al.* (2013), the introduced populations occupy more constrained portions of the environmental space than the native population. However, the native occurrences used in this study cover a wide portion of the environmental space realized worldwide, which could also explain why models for native occurrences predicted the introduced records well.

Nevertheless, this study did not reliably support prediction *(ii)*. The set of records comprising all the calibration occurrences (ARESNYFL) represent an even wider portion of the environmental space. Thus, if cross-prediction performance was only related to the portion of environmental space covered, I would expect prediction *(ii)* to be also true and the models based on all calibration records to provide the best predictions. However, similar to previous studies (Broennimann & Guisan 2008; Beaumont *et al.* 2009), the models calibrated with both native and introduced records indeed provided the best discriminatory ability. In contrast, their performances assessed with overfitting criteria were not significantly better than random models. Models calibrated with native records generally performed better and led to higher significance across all three evaluation statistics.

It is unlikely that this outcome is an issue of statistical power since the AR model that reached significance across the three evaluation criteria included fewer input records than the ARESNYFL models. It is possible that the observed differences in performance and significance may stem from factors such as the extent of the background area used for each model or the distinctive sampling bias present in different regions. However, based on the present results, I can only conclude that the model based on native records provided better transferability than models including non native records, probably because of non-equilibrium situations in invaded areas (especially in the FL dataset, which was artificially created to be out of equilibrium).

Concerning the effect of model complexity on transferability, these results demonstrate different conclusions when performance is used without considering significance. The values of the overfitting measures support the notion that simpler model settings provide more transferable predictions (Jiménez-Valverde *et al.* 2008; Rodda *et al.* 2011; Heinänen *et al.* 2012). This conclusion would however be incomplete, because the corresponding p-values of these statistics were very high (close to 1). In many cases, simpler model settings led to a generally larger prediction, regardless of whether real or random data were used to calibrate the models. Such a situation makes it very difficult to achieve enough statistical power to reject the null. Thus, while simpler model settings may provide more transferability, a greater number of test records are necessary to reach that conclusion with statistical confidence.

The results of this study are to some extent in agreement with to those of Heinänen et al. (2012) with Finnish water birds, and of Warren & Seifert (2011) who tested temporal transferability with 51 different species in California. As in Heinänen et al. (2012), Here, complex model settings provided better performance than simpler settings for models made with all the calibration records , but also for models using native occurrences that are comparatively diverse in environmental space. For the other individual populations in invaded ranges, the best results were obtained with optimal external settings. These settings were of intermediate complexity (as in Warren & Seifert 20011), but consistently fairly simple across the three introduced populations (as in Heinänen et al. 2012), and thus agree to some extent with both studies discussed above. Nevertheless, these results should be interpreted with care, as eventually only one model (based on the most complex settings) produced predictions that were significantly better than random across the three evaluation criteria.

Implications for monk parakeet invasive potential:

The results suggest that the conditions present in the native range seem to be more informative to predict future introductions and range expansions than those in invaded areas. The worldwide prediction of the only significant model reveals several areas currently not occupied by monk parakeets that are particularly at risk of successful introductions. The prediction also shows that areas adjacent to existing populations in Spain, Italy, and Israel are particularly at risk of invasions. These parakeets have already been sighted in several of these regions (Strubbe & Matthysen 2009). I did not consider those locations in this analysis because there was not enough information supporting well established populations. To a lesser extent, the populations in Texas, central Mexico, and eastern Florida are also likely to expand their ranges. This risk is also possible for the populations in Illinois, Indiana, Long Island, England, Belgium, and Holland, although the predictions there are less reliable. In contrast, the native range seems to include most available suitable areas.

Note that in some areas, the global prediction is not consistent with what is know about the biology of the monk parakeet. For instance, large portions of sub-Saharan Africa were predicted as highly suitable although it is very unlikely that monk parakeets could survive there. Conversely, some parts of Florida and Illinois are poorly predicted although monk parakeets have been documented there (Avery *et al.* 2002; Tillman *et al.* 2004; Pruett-Jones *et al.* 2011). Such inconsistencies can be explained in part by the fact that the effect of clamping on the prediction is only one aspect of the reliability of the prediction. Different methods to assess reliability will likely reveal different areas where the prediction should be interpreted with care. In addition, the variables considered in the model may not correlate with local environmental tolerance and resource availability in the same way across the world. For instance, climatic variables may be more indicative of local resource availability in the native range than in Northern cities where the parakeet may depend primarily on bird feeders. Therefore, the prediction would likely be improved with the inclusion of more proximal predictors of local resource availability than the ones I selected. Unfortunately, such predictors are rarely available on global scales.

In terms of predictor variable importance, the critical variables are all indicative of the extent of winter cold and humidity in the native range. These variables are related to factors affecting survival in the cold months: winter food availability and the risk of frostbite. The analyses of variable importance should be interpreted with caution considering that several of the predictor variables were highly correlated, especially the bioclimatic ones. Nevertheless, the outcomes seem likely to be correct since they match those of previous study (Weathers & Caccamise 1975; Caccamise & Weathers 1977; Tamara & Arnheim 1996; Butler 2005; Strubbe & Matthysen 2009). It is difficult to assess the benefits of human influence in the native range, but they could include those mentioned for invaded regions: winter food availability via birdfeeders (South & Pruett-Jones 2000), and influx of new migrants from the pet trade (Muñoz & Real 2006).

Conclusion:

According to the results presented here, the best strategy to create reliable transferable predictions for monk parakeets is to model the niche of the species from the pattern of occurrences in the native range. This finding is in support of prediction *(i)* but not prediction *(ii)*. Even though the models including both native and invaded occurrences provided the best and most significant discriminatory ability, their performance was not significant for measures of overfitting. In contrast to previous studies, the most complex model settings provided the highest performance and significance when transferred. Simple model settings resulted in apparently good performance, but the predictions were not significantly better than null models. Settings selected with the external optimization approach were consistently simpler than those selected with the internal optimization method. Simple model settings seemed to perform better when modeling with occurrences from invaded regions, but the predictions were also not significantly better than those of null models. Although these considerations about transferability may only apply to this parakeet, they highlight the need to measure model performance with both overfitting and discriminatory ability criteria, and to include assessment of statistical significance. The worldwide projection of the only model with significant predictions reveal several areas at risk of new introductions and show that most of the currently well established populations are likely to expand to surrounding areas.

Table 2.1: Performance and significance of 20 experiments with different calibration records and model settings evaluated with the same spatially independent records from other invaded areas. The values of the evaluation statistics (AUC$_{TEST}$, AUC$_{DIFF}$, and OR) are presented along with their corresponding p-values estimated from 1000 replicate null-models (p_AUC$_{TEST}$, p_AUC$_{DIFF}$, and p_OR). Significant p-values with alpha = 0.05 are indicated in bold. Z-Tradeoff is a composite measure of performance across the three evaluation criteria, giving equal importance to discriminatory power and overfitting. The table entries are sorted according to this measure, from high to low performance.

Population	**Settings**	**Features-RM**	**AUC$_{TEST}$**	**p_AUC$_{TEST}$**	**AUC$_{DIFF}$**	**p_AUC$_{DIFF}$**	**OR**	**p_OR**	**Z-Tradeoff**
ARESNYFL	Most general	LQ-3	0.8393	**0.0210**	0.0140	0.0896	0.1095	0.2302	1.6536
AR	Optimal External	LQP-1.5	0.7922	**0.0065**	-0.0450	0.0987	0.1556	0.0500	1.1224
AR	Most general	LQ-3	0.7705	0.0504	-0.0555	0.1293	0.1210	0.1439	0.8775
AR	Optimal Internal	LQPT-1	0.8038	**0.0018**	0.0845	0.0572	0.2882	**0.0436**	0.8667
ARESNYFL	Most specific	LQPTH-0.5	0.8101	**0.0001**	0.1343	0.1233	0.3170	0.1184	0.8104
ARESNYFL	Optimal External	LQPH-1.5	0.7836	**0.0064**	0.1265	0.3604	0.2853	0.3011	0.4921
ES	Optimal External	LQP-2.5	0.7447	**0.0154**	0.0241	0.0788	0.1758	0.0888	0.2926
ES	Most general	LQ-3	0.7522	**0.0370**	0.0476	0.0857	0.2709	0.2966	0.2669
NY	Optimal Internal	LQPH-0.5	0.8054	**0.0140**	0.1852	0.2660	0.8127	0.7192	0.2563
FL	Optimal Internal	LQPH-1.5	0.8002	**0.0145**	0.1796	0.4367	0.9366	0.9766	0.1082
NY	Optimal External	LQP-0.5	0.7535	**0.0178**	0.0950	0.1955	0.3948	0.2827	0.0787
ARESNYFL	Optimal Internal	LQPH-0.5	0.7617	**0.0050**	0.1657	0.4519	0.3516	0.3539	0.0506
AR	Most specific	LQPTH-0.5	0.7479	**0.0020**	0.1203	**0.0394**	0.3026	**0.0142**	0.0080
FL	Optimal External	LQT-0.5	0.7669	**0.0037**	0.2211	0.5365	0.9827	0.9858	-0.4803
ES	Optimal Internal	LQPH-1.5	0.7252	**0.0322**	0.1943	0.3412	0.7291	0.7347	-0.7944
FL	Most specific	LQPTH-0.5	0.7440	**0.0381**	0.2420	0.4595	0.9885	0.9194	-0.8467
ES	Most specific	LQPTH-0.5	0.7314	**0.0361**	0.2595	0.3503	0.9798	0.8864	-1.0543
FL	Most general	LQ-3	0.7175	0.0719	0.2385	0.7325	0.8386	0.9892	-1.0876
NY	Most general	LQ-3	0.7170	0.1072	0.2647	0.6166	0.8357	0.9106	-1.1562
NY	Most specific	LQPTH-0.5	0.7024	0.0723	0.2867	0.4304	0.9107	0.8179	-1.4645

Table 2.2: Maxent analyses of variable importance for the only significant model (native calibration records with complex settings).

Variable	**Percent contribution** (%)	**Permutation importance** (%)	**Gain without variable**	**Gain with variable alone**
BIO1 = Annual Mean Temperature	4.6	11.4	1.41	0.22
BIO10 = Mean Temperature of Warmest Quarter	2.3	2.5	1.41	0.13
BIO11 = Mean Temperature of Coldest Quarter	1.2	2.1	1.42	0.19
BIO12 = Annual Precipitation	3.3	6.9	1.41	0.23
BIO13 = Precipitation of Wettest Month	1.6	1	1.41	0.23
BIO14 = Precipitation of Driest Month	15	5.4	1.39	0.43
BIO15 = Precipitation Seasonality (Coefficient of Variation)	1.3	0.3	1.40	0.12
BIO16 = Precipitation of Wettest Quarter	2.2	3.2	1.41	0.26
BIO17 = Precipitation of Driest Quarter	7.4	6.5	1.41	0.39
BIO18 = Precipitation of Warmest Quarter	2.1	3.3	1.41	0.24
BIO19 = Precipitation of Coldest Quarter	1.3	1.5	1.42	0.30
BIO2 = Mean Diurnal Range (Mean of monthly (max temp – min temp))	6.5	1	1.42	0.25
BIO3 = Isothermality (BIO2/BIO7) (* 100)	9.3	0.5	1.41	0.29
BIO4 = Temperature Seasonality (standard deviation *100)	2.3	2.7	1.41	0.18
BIO5 = Max Temperature of Warmest Month	1.7	1.2	1.42	0.18
BIO6 = Min Temperature of Coldest Month	1.8	10.8	1.41	0.17
BIO7 = Temperature Annual Range (BIO5-BIO6)	0.7	0.2	1.42	0.14
BIO8 = Mean Temperature of Wettest Quarter	4.3	10.1	1.39	0.10
BIO9 = Mean Temperature of Driest Quarter	1.8	0	1.42	0.15
Percent Cultivated Land	4.9	4.9	1.37	0.08
Percent forest	2.6	9.4	1.37	0.05
Human Population Count	13.5	7.6	1.35	0.35
Percent Grass Land	3.6	5.8	1.38	0.15
Human Influence Index	4.7	1.5	1.41	0.24

Figure 2.2: Maps of monk parakeet presences. The bottom right map shows the calibration and evaluation records in red and blue, respectively. Enlarged regions are shown in insets (a), (b), and (c). The symbols in the insets represent the different populations: AR = native range (Argentina and surrounding countries), ES = Spain, NY = the wider New York metropolitan area, FL = South Florida, FLn = North Florida, IL = Illinois, LA = Louisiana, TX = Texas, PR = Puerto Rico, nd = records not assigned to any population. In the insets, the evaluation records are indicated in blue, and the calibration records in orange, aqua, green, and pink for AR, ES, NY, and FL, respectively.

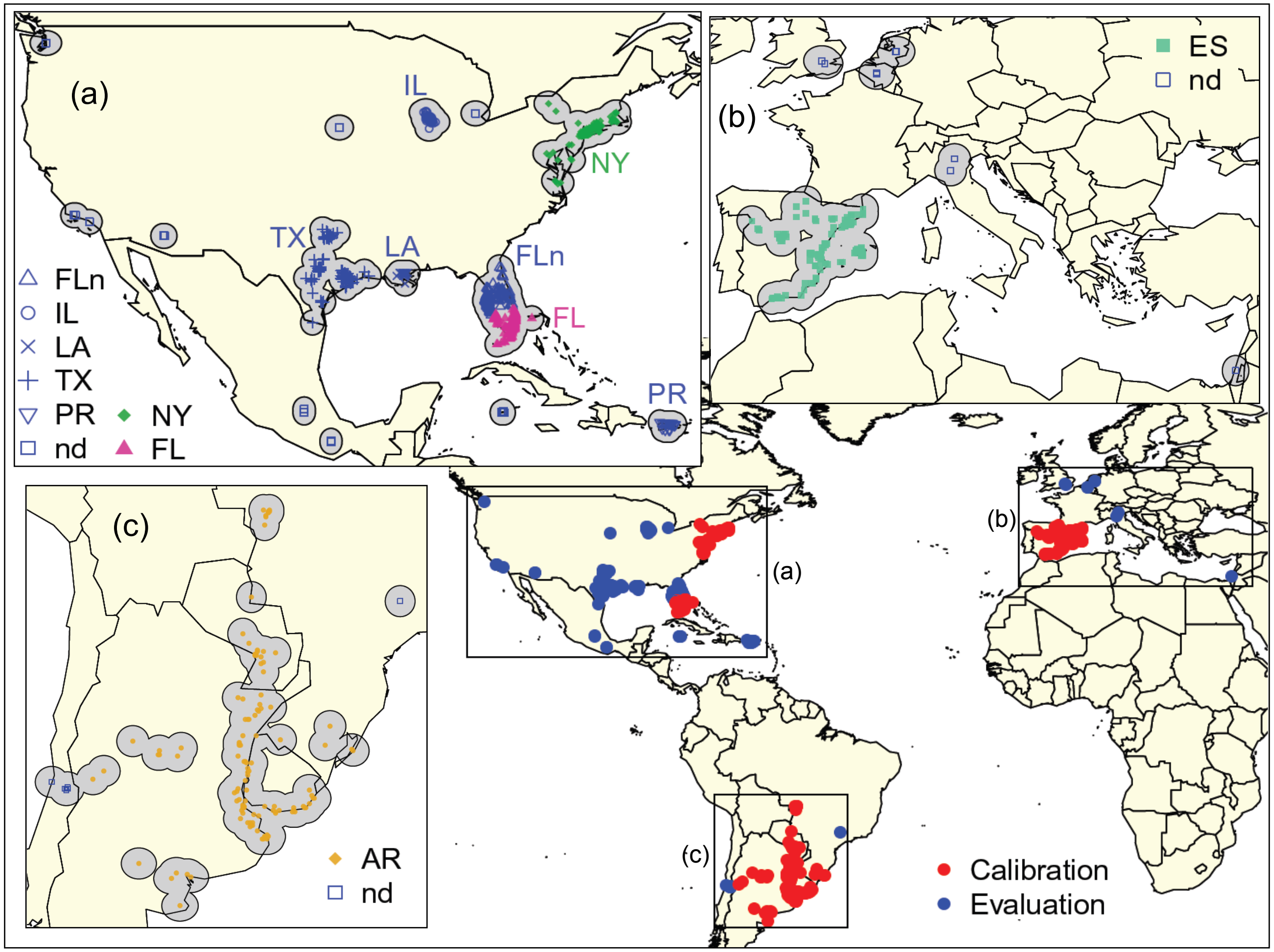


Figure 2.2: Model performance and significance as a function of model settings and calibration “population”. Darker shades of gray indicate higher performance. The top three graphs show the actual values of the evaluation statistics, while the bottom three show their corresponding p-values from 1000 replicate null-models. For each graph, the rows show the different calibration populations and the columns represent the type of model settings, in order of decreasing complexity. Optimal internal and Optimal external refer to the settings optimized for each population (see Methods for details). Asterisks indicate significant p-values with alpha = 0.05.

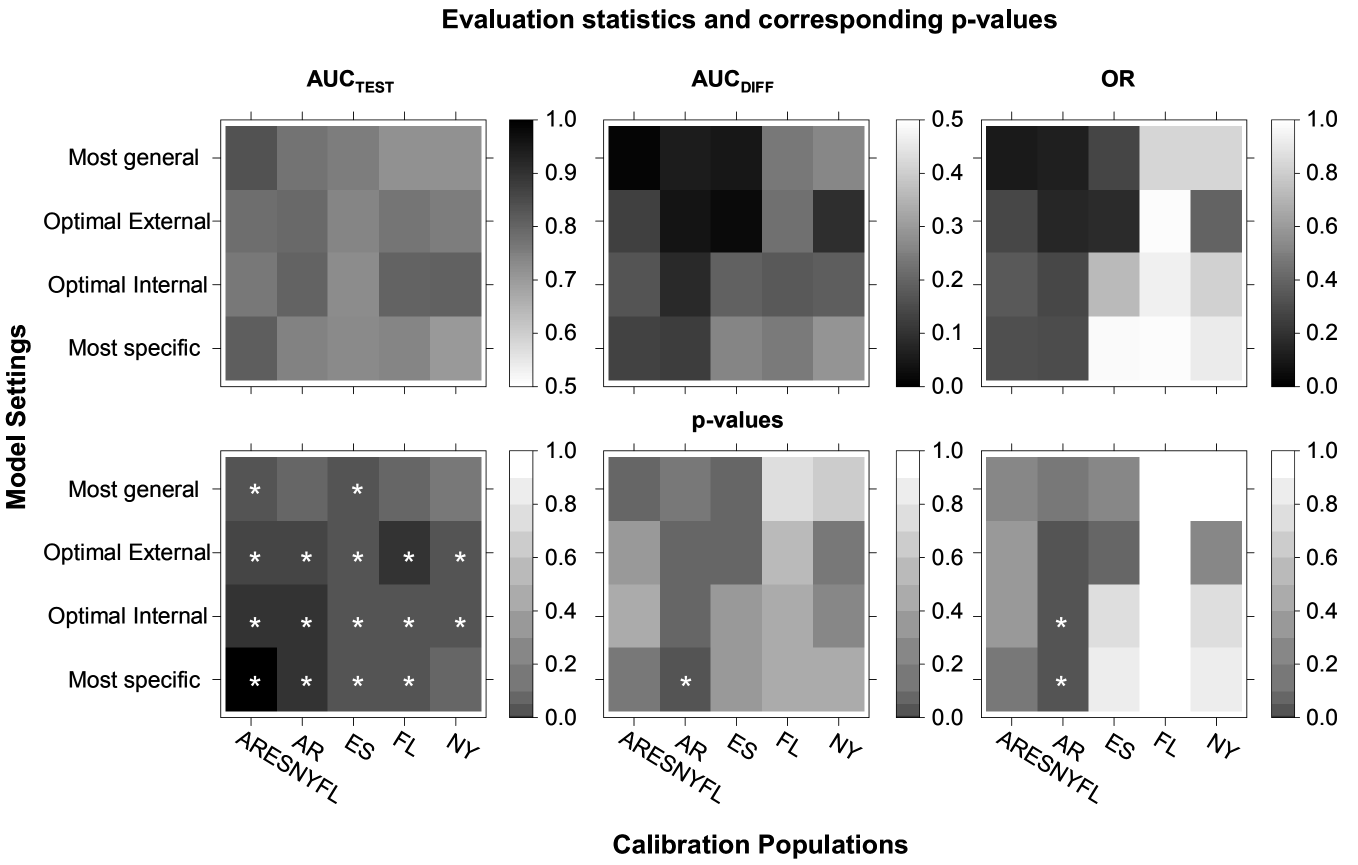


Figure 2.3: Environmental space showing the position of the monk parakeet occurrences (after filtering, see Methods of Chapter 1) along their first two principal components. The different populations are indicated with the same symbols and colors as in Figure 2.1. The records used for model calibration, along with their 95% confidence ellipses, are indicated as for Figure 2.1 in orange, aqua, green, and magenta for AR, ES, NY, and FL, respectively. The 95% confidence ellipse for all the calibration records combined is indicated in red (ARESNYFL). The spatially independent evaluation records are indicated in blue.

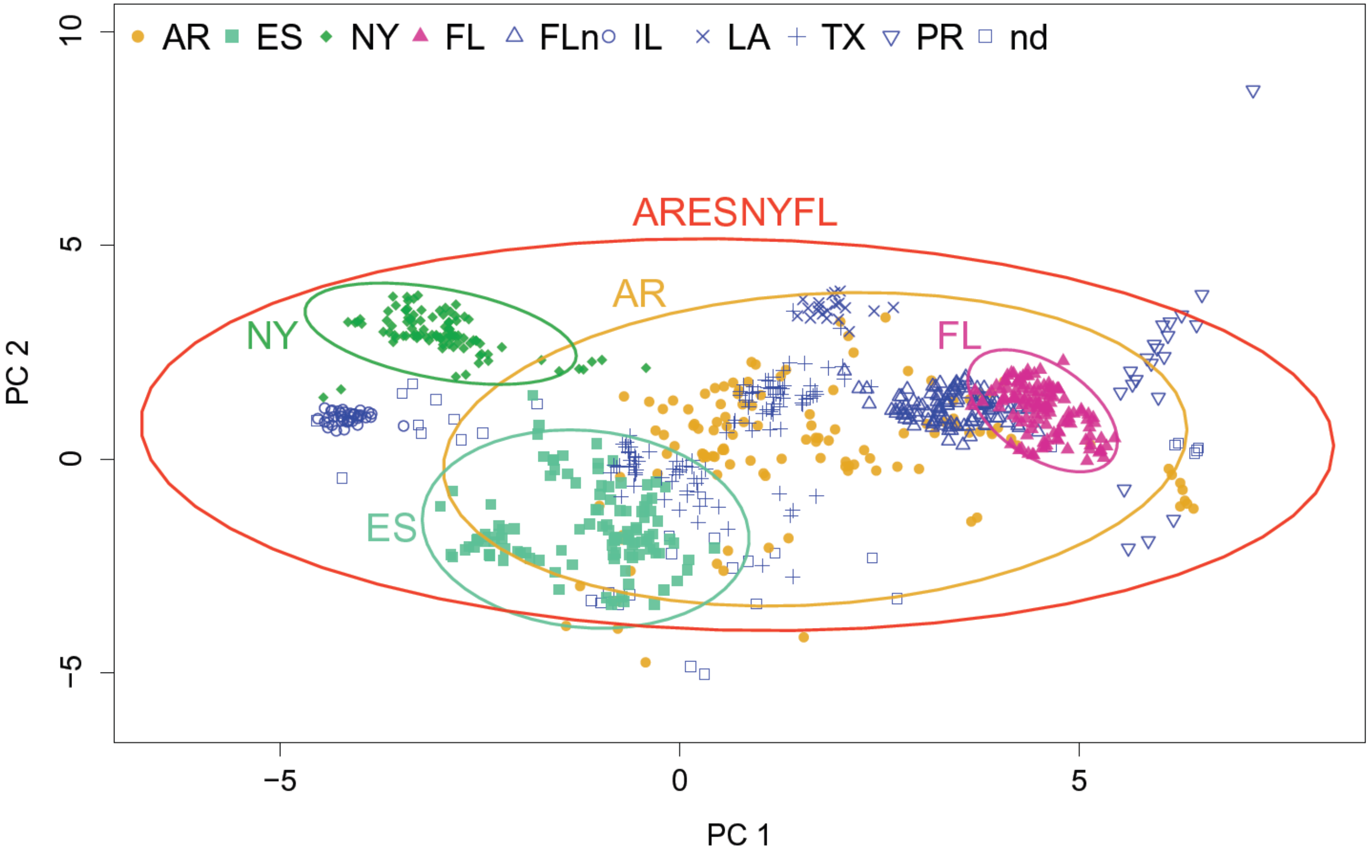
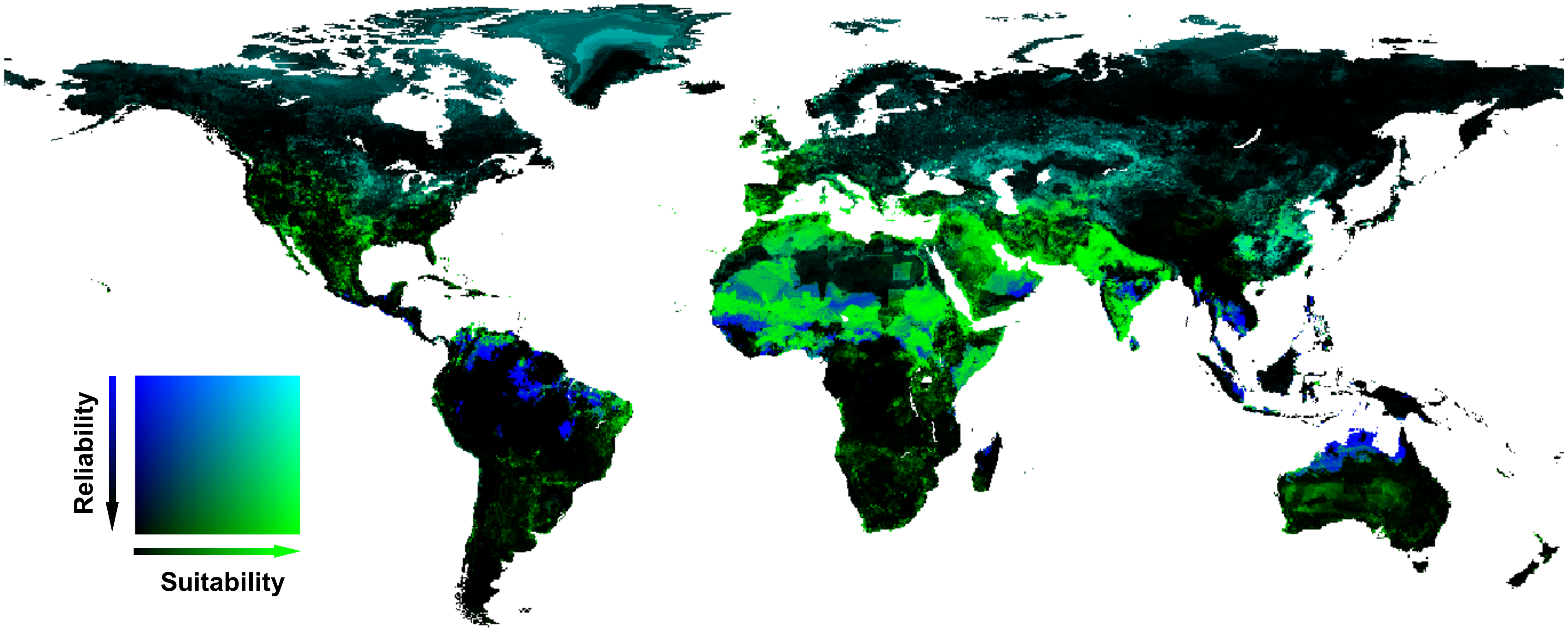


Figure 2.4: Worldwide prediction of the only significant model (native calibration records with complex settings). The green gradient represents the logistic probability of environmental suitability, from black = 0 to bright green = 1.The blue gradient shows the absolute difference between the predictions with and without clamping (i.e. treating the variables outside the range of environmental conditions present in calibration region as if they were at the limit of the calibration range), and hence the reliability of the predictions with regard to environmental extrapolation. Black indicates a difference of 0, and bright blue a difference of 1.Thus, areas of low prediction and low reliability are shown in blue, areas of high prediction and low reliability are shown in aqua, areas of low prediction and high reliability are shown in black, and areas of high prediction and high reliability are shown in green.



CHAPTER 3:

Estimating resistance to dispersal with ecological niche modeling and evaluating significance with null resistance models

ABSTRACT

I propose an ecological niche modeling approach to estimate the effect of landscape features on dispersal. This approach estimates landscape resistance by modeling the “dispersal niche” of a species using records of documented species presence and variables related to dispersal. Unlike most landscape genetics approaches, this method does not rely on iterative correlation to genetic data to estimate landscape resistance to dispersal. The proposed approach uses independent data, objective criteria, and repeatable methods to formulate a priori landscape hypotheses. Genetic data can be used only as a final step to evaluate the quality of the predicted estimates. I tested this approach with an introduced population of monk parakeets (*Myiopsitta monachus*) in Florida, and the ecological niche modeling algorithm Maxent. As a comparison, I evaluated multiple hypotheses based on expert opinion with stepwise regression models. I also implemented a null model approach to evaluate the significance of associations between effective and genetic distances. Although the monk parakeet population in Florida presented little genetic spatial structure beyond high within-nest relatedness, the niche modeling approach produced results equivalent to the stepwise regression model. This method offers a promising alternative to estimate resistance to dispersal that does not involve comparing a wide range of alternative hypotheses. In addition, I confirmed that partial Mantel tests and the derived multiple regressions on distance matrices may suffer from highly inflated rates of Type I error, even when the effect of spatial autocorrelation is removed. Null models may be a better alternative to evaluate the significance of association between genetic and effective distance matrices while controlling for isolation by distance. Finally, the results imply that although monk parakeets may to some extent avoid dispersing through areas with low canopy cover in the

Miami metropolitan area, they are probably not limited by distance and most landscape features, and may therefore continue to expand their range to adjacent ecologically suitable areas.

INTRODUCTION

Understanding how local heterogeneity in landscape features influences species dispersal is an important challenge in evolutionary and conservation biology, with critical implications for pressing environmental issues. This information can be used to manage the spread of pathogens and species invasions (Aylor 2003; Margosian *et al.* 2009; Fitzpatrick *et al.* 2012), and design wildlife corridors that promote genetic diversity (Beier *et al.* 2008). Furthermore, dispersal information is also critical to studies aiming to predict current and future species distributions (Anderson *in press.*; Hartig *et al.* 2012). These studies typically use climatic models to predict areas with conditions that are suitable for the species (Elith & Leathwick 2009). Dispersal information can be used to determine if a species is likely to reach those suitable areas, or it can be incorporated directly in the model to provide more realistic predictions (Hartig *et al.* 2012; Cabral & Kreft 2012; Algar *et al.* 2013). The effect of landscape on dispersal is typically assessed with landscape genetics (Manel 2003; Storfer *et al.* 2007; Balkenhol *et al.* 2009a). Here, I first review current landscape genetics approaches and then propose an alternative method that utilizes an ecological niche modeling algorithm (ENM; Elith & Leathwick 2009; Peterson *et al.* 2011) to directly model a species' dispersal preferences using presence data and variables related to dispersal. Finally, I present an empirical test case with monk parakeets (*Myiopsitta monachus*) in order to test this "dispersal niche modeling" approach.

Landscape Genetics:

The discipline of landscape genetics aims to understand the effect of landscape features on micro-evolutionary processes - such as local adaptation, genetic drift, and gene flow - by comparing patterns of genetic variation to landscape heterogeneities (Manel 2003; Storfer *et al.* 2007; Balkenhol *et al.* 2009a). This approach can therefore be useful to estimate how landscape features affect dispersal, especially when direct data on dispersal capabilities are difficult to acquire. However, gene flow, adaptation, and drift often act jointly in shaping genetic variation. Therefore, it can be difficult to disentangle the effect of landscape features on dispersal and resulting gene flow, as opposed to their effect on local genetic differentiation due to drift or adaptation. This distinction can be addressed by selecting appropriate landscape data and analytical methods. To examine the effect of landscape features on drift or adaptation, data for conditions occurring at the sampling sites may be sufficient. Differences between sites can then be related to genetic variation, typically through ordination analyses (Jombart *et al.* 2009; Guillot *et al.* 2009; Legendre & Fortin 2010). Conversely, data from landscape features “between” the sites is more relevant to estimates of dispersal than data “at” the sites. For instance, elevational differences between the sampling sites may be useful to detect altitudinal clines of genetic variation (as in Fang et al., 2012). However, it tells us little about the effect of altitude on dispersal. Depending on the topography, and the dispersal path, the effective changes in altitude experienced during dispersal may be much larger than the absolute difference between the sites.

Therefore, estimating dispersal patterns with landscape genetics requires data from the entire surrounding landscape, rather than only from the occurrence records. In addition, it requires the formulation of two types of hypotheses: (a) a landscape hypothesis stating how the landscape features promote or hinder dispersal; and (b) a dispersal route hypothesis stating how the organism moves through the landscape (e.g., straight paths, least cost paths, random walks, correlated random walks, Lévy flights; Mårell, Ball, & Hofgaard, 2002; Mcrae & Beier, 2007). These two hypotheses result in a dispersal cost distance matrix which can then be compared to a genetic distance matrix. Then, the null hypothesis that the genetic and dispersal cost distance matrices are independent is typically evaluated with a Mantel test, or a derived form (partial Mantel test, multiple regressions on distance matrices; Mantel 1967; Hubert & Golledge 1981; Smouse *et al.* 1986; Lichstein 2006; Legendre & Fortin 2010). Despite its limitations (Oden & Sokal 1992; Raufaste & Rousset 2001; Balkenhol *et al.* 2009b; Jaquiéry *et al.* 2010; Guillot & Rousset 2013; Graves *et al.* 2013), the Mantel test is by far the most employed test in landscape genetics studies (Storfer *et al.* 2010). Typically, landscape hypotheses are also contrasted to a simple isolation by distance (IBD) hypothesis that states that the population genetic differences can be explained by geographic distance alone. IBD assumes that all landscape features have the same effect on dispersal. Although unrealistic in many situations, IBD should not be overlooked as a null hypothesis (Balkenhol *et al.* 2009a).

One of the main challenges of this approach is arriving at realistic hypotheses concerning both the landscape and dispersal routes. I focus here on landscape hypotheses. Discussions about modeling organism movement through the landscape can be found elsewhere (Mårell *et al.* 2002; McRae 2006; Mcrae & Beier 2007; McRae *et al.* 2008; Barton *et al.* 2009; Smouse *et al.* 2010). Landscape hypotheses are typically formulated in the form of resistance surfaces: each pixel of the map is assigned a value for its resistance to movement that is proportional to some landscape feature(s). Some hypotheses can be relatively simple to model, such as IBD with a uniform resistance surface, complete barriers to dispersal with infinite resistance, and panmixia with infinite conductance, the reciprocal of resistance (Cushman *et al.* 2006; Landguth *et al.* 2010). However, assigning intermediate resistance values corresponding to the effect of multiple landscape variables presents a challenge.

Optimally, resistance values should be calibrated using independent observational or experimental data on dispersal capabilities, such as from radio-telemetry studies (Epps *et al.* 2007; Stevenson *et al.* 2013). However, very few studies use such data (reviewed in Spear *et al.* 2010). Instead, the most common approach remains to make a few educated guesses (i.e. “expert modeling”) and select the one that best explains the genetic data (Spear *et al.* 2010; Graves *et al.* 2013), preferably within a causal modeling framework (Legendre 1993; Cushman *et al.* 2006; Cushman & Landguth 2010). An alternative approach is to start with guesses and iteratively change the resistance values to obtain the highest correlation with the genetic data

(Shirk *et al.* 2010). However, even a very large number of alternative hypotheses (as in Wang *et al.* 2009) may only represent a small portion of the multivariate parameter space and thus detect local rather than global optima (Shirk *et al.* 2010; Graves *et al.* 2013). A large number of initial guesses may alleviate that problem but it also increases the chance of obtaining spurious correlations. Necessary corrections for the resulting inflation of Type I error may dramatically reduce statistical power. As an alternative, Graves *et al.* (2013) examined with simulated data an optimization algorithm to efficiently search the parameter space and find a global optimum if one exists. However, their simulations showed that their obtained optimal resistance values were often far from the true values, and that causal modeling failed to support the real values. Finally, unless spatially and genetically independent data are used to evaluate the optimized resistance surface, optimization approaches can lead to overfitting resistance values to random noise, or sampling bias, rather than describing the true relationship between the landscape and genetic data.

Dispersal Niche Modeling:

I propose here an alternative approach that makes use of an ecological niche modeling algorithm (ENM; Elith & Leathwick 2009; Peterson *et al.* 2011), species presence data, and variables related to dispersal to produce realistic resistance values without relying on the genetic data. The rationale for this approach is that a species is more likely to be sighted, captured, or collected in places where it can freely move than in landscapes that are detrimental to its movement. Therefore, the species' preferences can be modeled by comparing landscape data from sites where the species' presence has been documented, to sites where it has not been documented but that are potentially available to the species (Barve *et al.* 2011). The proposed "dispersal niche modeling" approach is a simple extension of ENM, but rather than selecting environmental predictors that are relevant for the normal biological functioning of the species, it requires landscape variables that are relevant for dispersal, such as wind direction, terrain slope, road density, river networks, or canopy cover (Manel 2003; Cushman *et al.* 2006;

Spear *et al.* 2010). These predictors must also be considered at a resolution that can appropriately capture the landscape features that may impede or facilitate the species' movement. Hence, this method models the "dispersal niche" rather than the ecological niche of the species. As in ENM, the quality of the presence data (correct identification, accurate geographic coordinates, no sampling bias) is crucial to obtain accurate predictions (Syfert *et al.* 2013). However, unlike ENM the presence records need not represent true source populations. Documenting that a species was present at a particular location is enough to show that it was able to move in that landscape.

Depending on the type of data and method used, ENM model outputs typically correspond to a continuous prediction: a probability of occurrence, or an index of suitability. The model prediction can thus be directly translated to continuous resistance values: high suitability indicating low resistance to movement, and vice versa. Alternatively, isolation by barrier (IBB) hypotheses can be tested by converting the model output to binary predictions based on a specific threshold. Areas with predicted values below the threshold can represent complete barriers to dispersal and areas with values above the threshold indicate infinite conductance.

Objectives:

I tested the proposed "dispersal niche modeling" approach with the ENM algorithm Maxent (Phillips *et al.* 2006; Phillips & Dudík 2008; Elith *et al.* 2011) and different sets of presence records from introduced populations of monk parakeets (*Myiopsitta monachus*) in the continental US, Florida, and the Miami metro area. I evaluated the quality of the resulting resistance to dispersal hypotheses with genetic data for the Miami population. As a comparison, I examined multiple guesses of resistance values with a forward stepwise regression model (Legendre *et al.* 1994; Zapala & Schork 2006).

In both cases, I used multiple regression on distance matrices (MRM; Hubert & Golledge 1981; Legendre *et al.* 1994; Lichstein 2006; Zapala & Schork 2006) to assess the ability of the landscape hypotheses to explain the genetic data, and to evaluate their significance. MRM is essentially an extension of partial Mantel tests and may therefore suffer from the same limitations: notably, highly inflated rates of Type I error in the presence of spatial autocorrelation (Oden & Sokal 1992; Raufaste & Rousset 2001; Guillot & Rousset 2013). Therefore, I also implemented a null model approach to assess statistical significance. This approach involves creating a set of resistance surfaces with random values. These null surfaces are then used to calculate effective distance and test for association with the genetic distance matrix. The measure of association corresponding to the real hypothesis is then compared to the resulting null distribution to assess significance.

In summary, I 1) created a set of “dispersal niche models”, 2) examined their association with pairwise genetic data, 3) calculated significance of associations between distance matrices using permutations and a null model approach, and 4) repeated steps 1-3 using multiple guesses of resistance values and a forward stepwise regression model. I follow my interpretation of the results by a discussion of the merits and limitations of the “dispersal niche modeling” approach. Finally, I discuss the implications of the results regarding the invasive potential of the monk parakeet.

METHODS

Study region and landscape data:

I considered presence data from the entire continental US to calibrate the dispersal niche models (Figure 3.1a). However, I restricted the landscape genetics analyses to a smaller rectangular region comprising the monk parakeet records with genetic information in the wider Miami metropolitan area (Florida, USA); with a minimum distance of 10 km from any record (Figure 3.1b). For both regions, I considered high resolution (30 m^2) land cover variables from the National Land Cover Database: percent tree canopy density (Homer *et al.* 2007), percent developed impervious surface, and land cover classification (Fry *et al.* 2011). These variables have been used successfully to explain patterns of dispersal and gene flow in other organisms (Greenwald *et al.* 2009; Murphy *et al.* 2010; Sackett *et al.* 2011; Munshi-south 2012).

Species presence and genetic data:

The monk parakeet (*Myiopsitta monachus*) is native to temperate South America and has established several stable and well documented populations in the US (Figure 3.1a; Pruett-Jones *et al.* 2011). This species is easy to detect and identify because of its distinctive morphology, loud calls, and unique nesting behavior (Forshaw & Cooper 1989). I gathered 2,325 presence records for the entire continental US (Figure 3.1a), including 1,045 from Florida, from various sources: the GBIF and Species Link databases, records from the primary literature, and personal observations. The genetic data were those reported in Gonçalves da Silva *et al.* (2010b), and deposited in the Dryad repository: DOI:10.5061/dryad.1682 (Gonçalves da Silva *et al.* 2010a).

These data include genotypes for 10 species-specific microsatellite loci and geographic coordinates for 149 monk parakeets in the Miami metro area (Figure 3.1b). I excluded 2 individuals from the landscape genetics analyses because the recorded coordinates are most likely inaccurate since they do not refer to a valid land location. However, I included all 149 genotypes to calculate the population allele frequencies. For these data, Gonçalves et al reported a mean allelic richness of 0.6 (± 2 SD), and no significant deviation from Hardy-Weinberg equilibrium (observerd/expected heterozigosity : H_O = 0.62 ± 0.15 SD / H_E =0.64 ± 0.13). In addition, they found evidence for high propagule pressure and that the population in Florida has not suffered from loss of genetic diversity compared to native populations, in agreement with an earlier study using mitochondrial markers (Russello *et al.* 2008). Therefore, these data are appropriate to obtain reasonable estimates of pairwise relatedness between the parakeet samples in the Miami metropolitan area.

Dispersal niche models:

I used the presence-background ENM algorithm Maxent (Phillips *et al.* 2006; Phillips & Dudík 2008; Elith *et al.* 2011) to create three dispersal niche models: one with occurrence data from the entire continental US (n = 2325; Figure 3.1a), one based on the records in Florida (n =

1045), and the last one based only on the records with genetic data in the Miami metro area (n = 147; Figure 3.1b). I refer to these models as US, FL, and MIA, respectively. For each model, I selected the comparison background records based on the criterion that they must be potentially accessible to the species (Anderson & Raza 2010; Barve *et al.* 2011). Hence, I randomly sampled these records within 106 km of the calibration records (VanDerWal *et al.* 2009), which is the maximum reported dispersal distance for monk parakeet (Gonçalves da Silva *et al.* 2010b). The resulting samples included 86,172 background records for the US, 29,896 for FL, and 29,777 for MIA.

I optimized the models following the procedure described in Chapter 1 with the following modification: I used a 5-fold cross-validation procedure with spatially independent partitions (Radosavljevic & Anderson, *in press*). I opted for spatially independent evaluation because it offers a more realistic test of performance than randomly selected evaluation records; the latter, being more sensitive to environmental spatial autocorrelation, typically results in inflated estimates (Veloz 2009; Hijmans 2012; Radosavljevic & Anderson, *in review*). I selected the partitions of presence and corresponding background records with Voronoi tessellations (Aurenhammer 1991; Appendix B). I selected the optimal model settings according to both discriminatory and overfitting criteria, controlling for statistical significance with null models (as in Chapter 1, the optimal settings were the one providing the best tradeoff between high AUC_{TEST}, low AUC_{DIFF}, and low omission rate, provided these statistics were significant as determined with 1000 null model replicates). Additional details and R scripts for these procedures are included in Appendices A and B, statistics of the Maxent models are provided in Appendix C.

Landscape hypotheses:

I expressed the landscape hypotheses in the form of conductance surfaces, with conductance values ranging from 0 to 100. I used conductance rather than resistance for computational reasons (see van Etten 2012). First, I considered an IBD hypothesis by creating a uniform surface with conductance values of 1. Then, I created a set of conductance surfaces corresponding to the dispersal niche model predictions by projecting the optimized models on the smaller study region in Miami. I converted the Maxent logistic outputs (with suitability values ranging from 0 to 1) to conductance values by multiplying the predicted values for each map pixel by 100.

I also created isolation by barrier (IBB) hypotheses by converting the continuous model outputs to binary predictions using the following thresholds: 0.5, the midpoint of the potentially obtainable prediction values; the minimum predicted value of the MIA records, which ensures that all the records used for the landscape genetics analysis are in suitable areas; and the value of the tenth percentile predicted value for the MIA records, which accounts for possible outliers in the MIA sample. Areas with predicted values below the threshold were assigned conductance values of 0. Areas with predicted values above the threshold were assigned conductance values of 100. Thus, I created 12 different conductance surfaces: one continuous and three IBB scenarios for each of the three dispersal niche models (US, FL, and MIA).

As a comparison, I created conductance surfaces individually for each land cover variable. For the percent imperviousness and the percent canopy cover, I considered a set of continuous transformations to convert the percent values to conductance values. I used 14 different transforms covering a range of potential relationships (see Appendix C). For land cover classifications, I assigned conductance values equal to the frequencies of monk parakeet presences observed for each land cover class. I created three such surfaces corresponding to the frequencies of the US, FL, and MIA records. For each of the three land cover variables, I also created three IBB scenarios. I determined the barriers either with an educated guess (e.g. low canopy cover, or high imperviousness), or based on the frequencies of documented monk parakeet presences (Appendix C). In summary, I considered 53 alternative landscape hypotheses: a uniform surface, 12 niche model predictions, and 40 continuous and barrier scenarios based on individual land cover variables.

Dispersal cost distance:

I calculated effective dispersal cost using least cost paths with Dijkstra's algorithm (Dijkstra 1959) as implemented in the R package ‘gdistance’ version 1.1-4 (van Etten 2012). Each conductance surface raster cell was connected to its 8 adjacent neighbors. The effective cost to move between pairs of adjacent cells was obtained by adding 0.01 to the mean of their conductance values and taking the reciprocal of that number. I added 0.01 to avoid complete barriers to dispersal. Instead, movement between adjacent cells with conductance values of 0 (barriers) was ~100 times more costly than movement between cells with values of 1 (the geographic distance reference), and ~10,000 times more costly than between cells with the maximum value of 100. I calculated least cost paths between all pairs of genotyped records and expressed the total cumulative costs along the paths in meters.

Genetic distance:

I estimated genetic distance with pairwise relatedness estimators calculated with the program COANCESTRY version 1.0.1.2 (Wang 2011). This program implements seven relatedness estimators, including two likelihood estimators that allow for inbreeding. I initially considered all 7 estimators but they were highly correlated and led to the same conclusions regarding relationships between effective distance and relatedness. Therefore, I only considered the Lynch & Ritland (1999) estimator in the final analyses because it provided the best fit. In addition, Gonçalves da Silva *et al.* (2010b) reported no significant deviation from Hardy-Weinberg equilibrium at any of the 10 loci. However, they reported significant signatures of null alleles for two loci, and significant linkage disequilibrium for two others. I performed preliminary analyses with and without these loci. Despite small differences in relatedness estimates, the conclusions regarding the dispersal cost distances were globally the same. Therefore, I included all loci in the final analyses.

Statistical Analysis:

I evaluated the quality of the dispersal niche model and alternative landscape hypotheses with the following statistical analyses. First, I used a simple Mantel test to determine if least cost path distance on a uniform landscape was equivalent to Euclidian geographic distance. Then, I used parametric multiple regression on distance matrices (MRM; Hubert & Golledge 1981; Legendre *et al.* 1994; Lichstein 2006; Zapala & Schork 2006) to evaluate the effect of geographic (Euclidian) and dispersal cost (effective) distances on relatedness. Mantel and MRM tests were calculated using the R package 'ecodist' version 1.2.7 (Goslee & Urban 2007). All distances (Euclidian and effective) were ln(x+1) transformed to remove skew and provide a better linear fit (as in Lichstein 2006; Martiny *et al.* 2011; Li *et al.* 2011). Since the distance matrices were not completely independent, for any MRM that included more than one independent distance matrix, I estimated the "pure" contribution of each distance matrix as in Krasnov *et al.* (2010; see also Legendre *et al.* 2005; Peres-Neto *et al.* 2006; Li *et al.* 2011). The pure contribution of each independent matrix represents the difference between the R^2 obtained with all the independent matrices and without the matrix of interest. I expressed the pure contribution as a percent of the explainable variation (the R^2 of the entire MRM model).

For the dispersal niche models, I ran MRM tests for each of the 12 effective distance hypotheses, first with the effective distance matrix alone, then with both effective and Euclidian distance matrices to control for the effect of geographic distance. For the other 40 alternative hypotheses, I used a forward stepwise MRM model (Legendre *et al.* 1994; Zapala & Schork 2006) to select the landscape hypotheses that best explained relatedness. I started with all 40 effective distance matrices and the Euclidian distance matrix individually, and added matrices sequentially based on the largest increases in R^2, provided that the F statistic, R^2, and the regression coefficient(s) were significant (see below). However, whenever a distance matrix corresponding to a landscape hypothesis was included in the model, I excluded all other distance matrices corresponding to the different transforms of the same land cover variable.

This procedure ensured that the final full model included only one distance matrix for each of the three land cover variables. Note that I used R^2 rather than the adjusted R^2 or the Akaike information criterion because these statistics cannot be directly assessed given that the observations in pairwise distance matrices are not independent (Legendre *et al.* 1994).

For all Mantel and MRM tests, I evaluated significance using 10^5 permutations of the relatedness matrix and a Bonferroni corrected alpha level of 0.05. Therefore, the MRM F statistic, R^2, and the regression coefficients were considered significant if their p values were lower than 0.05/53 = 0.00094. As an alternative, I implemented a null model approach to evaluate the significance of association between relatedness and distance matrices. This method can be applied to MRM as well as simple or partial Mantel tests. It involves: 1) creating *k* null conductance surfaces with random values ranging from 0 to 100, 2) applying to these surfaces the same transform (or barrier reclassification) as the real landscape hypotheses, 3) calculating pairwise least cost path distances for each null surface to create a set of *k* null distance matrices, 4) running *k* MRM (or Mantel) tests with the null matrices in place of the real distance matrix, and 5) comparing the R^2 (or any other statistic) obtained with the real distance matrix against the null distribution of *k* values obtained with the null matrices. Here, I estimated significance of the R^2 values with a parametric Z test.

Because this method was computationally intensive, I only used 99 null replicates and applied it to test the significance of a restricted set of 4 alternative hypotheses: the geographic distance matrix, the matrices for best dispersal niche models (with and without geographic distance), and the effective distance matrix that contributed the largest increase in R^2 in the full stepwise model. For clarity, I refer to the p-values obtained with permutation as p_{PERM}, and to those obtained with the null model approach as p_{NULL}. I also investigated the validity of permutation tests for MRM, Mantel, and partial Mantel tests (correcting for geographic distance) by testing the association between the relatedness matrix and each of the null matrices with permutation tests. The association between the real relatedness matrix and the null matrices should not be significant given that the null matrices correspond to random landscape resistance values. I calculated the significance of the MRM, Mantel, and partial Mantel tests with 10^5 permutations.

Finally, to determine how geographic and effective distance may affect the genetic structure of the monk parakeet at different distance lags, I examined the fine-scale spatial structure of relatedness by modifying the “mgram” function in the “ecodist” R package (Goslee & Urban 2007) and calculated the average relatedness at different distance lags. I estimated the 95% confidence interval of the mean with 1,000 permutations of the relatedness matrix. I performed this analysis for the four distance matrices described in the preceding paragraph. I considered 11 distance lags: the within nest distance of 0 m, and 10 distance classes selected to include approximately the same number of observations.

RESULTS

Isolation by distance:

The Euclidean geographic distance matrix and the least cost path distance matrix based on a uniform conductance surface were highly correlated (Mantel R = 0.99, $p_{PERM} < 0.00001$). I detected a small but highly significant pattern of IBD ($R^2 = 0.0734$, p_{PERM} & $p_{NULL} < 0.00001$; Table 3.1). In fact, the distance matrices corresponding to the 53 alternative hypotheses were all significantly associated to the relatedness matrix according to the MRM tests, with consistently high p values ($p_{PERM} < 0.00001$).

Dispersal niche models:

Concerning the dispersal niche models, the predictions of all models resulted in highly significant associations to relatedness ($p_{PERM} < 0.00001$), whether effective distance was analyzed alone or jointly with geographic distance. Except for two IBB scenarios with the MIA models, all models yielded similar R^2, which were consistently higher with the inclusion of geographic distance (Table 3.2). When effective distance was analyzed alone, the MRM with the largest R^2 corresponded to the continuous prediction of the MIA model (Tables 3.1 & 3.2;

Figure 3.2). This association was highly significant according to both permutation and null model tests (p_{PERM} & $p_{NULL} < 0.00001$; Table 3.1). When effective distance was analyzed jointly with geographic distance, the largest R^2 was obtained with the IBB prediction of the FL model with a threshold of 0.5 (Table 3.1 & 3.2; Figure 3.2). This matrix also had the highest individual contribution to the total R^2 (Table 3.2). This association was highly significant according the permutations test ($p_{PERM} < 0.00001$), but not with the null models test ($p_{NULL} = 0.38$). The MRM with the smallest contribution of geographic distance included the prediction of MIA model.

Stepwise regression model:

For the stepwise MRM model, the full model included the following distance matrices in this order: 1) the IBB hypothesis with a quasi barrier below 20% canopy cover and high conductance above (Figure 3.2), 2) the IBB hypothesis with a quasi barrier above 50% impervious surface and high conductance below, 3) the geographic distance, and 4) the land cover classification hypothesis with conductance matching the frequency of the FL records (Table 3.1; Appendix C). This model yielded a larger R^2 than any other MRM test in this study and was highly significant ($R^2 = 0.0940$, $p_{PERM} < 0.00001$; Table 3.1). In fact, the second highest R^2 (excluding dispersal niche models) was obtained when the first independent distance matrix (the IBB for canopy cover) was tested alone ($R^2 = 0.0808$, $p_{PERM} < 0.00001$; Table 3.1). This effective distance matrix also had the largest 'pure' contribution in the full model (Table 3.1). However, the MRM for the full stepwise model was not significant according to the null model approach ($p_{NULL} > 0.99999$; in fact, R^2 was significantly *lower* than the null replicates), unlike the MRM with the IBB canopy matrix alone ($p_{NULL} < 0.00001$; Table 3.1).

Null resistance vs. permutation tests:

According to the permutation tests of significance, every single null distance matrix had a significant association to the relatedness matrix. This observation was true regardless of which test was used: MRM, Mantel, and partial Mantel tests (correcting for geographic distance) were all consistently highly significant ($p_{PERM} < 0.00001$). The R^2 null values were approximately normally distributed, with a small range of variation (see examples in Figure 3.3). The null distance matrices were also highly correlated to geographic distance (mean Mantel R = 0.94; $p_{PERM} < 0.00001$).

Fine-scale spatial genetic structure:

The analysis of the fine-scale spatial structure of relatedness revealed a high average within-nest relatedness ($r = .36$, $p_{PERM} < 0.001$) which rapidly decreased to values close to 0 at higher distance lags (Figure 3.4). This analysis produced the same results with geographic and effective distances. The scales of effective least cost path distances were different but the general pattern was the same (Figure 3.4). Note that I only plotted these trends for the three distance matrices that were significantly associated with relatedness according to both permutation and null model tests.

DISCUSSION

The results indicate a weak but highly significant pattern of IBD and that several landscape hypotheses could explain more of the genetic variation than IBD alone. Among those hypotheses, the full stepwise regression model yielded the highest proportion of explained variance. This model included two IBB hypotheses (quasi barriers below 20% canopy cover, and above 50% impervious surface), the geographic distance, and a continuous hypothesis based the frequency of land cover classes with monk parakeet presences in FL. When considered in conjunction with geographic distance, dispersal niche models produced very similar results (Table 3.1). These results indicate that the dispersal niche model approach may offer a promising alternative to estimate resistance to dispersal.

However, although highly significant with permutations tests, neither the stepwise regression model, nor the best dispersal niche model with geographic distance, was significant according to the null model approach. The continuous prediction of the MIA dispersal niche model was significant according to both permutations and null models. It also corresponded to the MRM that was the least improved by the inclusion of geographic distance (Table 3.2). Nonetheless, it explained less of the variation in pairwise relatedness than IBD alone (Table 3.1). Conversely, the effective distance matrix with the largest contribution to the full stepwise model produced the largest R^2 that was significant with both methods (Table 3.1). This matrix corresponded to the IBB hypothesis with a quasi barrier below 20% canopy cover and high conductance above. This result implies that monk parakeets may avoid moving in areas with very low tree density, which is consistent with my personal observations.

The fine-scale analyses of the spatial structure of relatedness showed that individuals from the same nests are on average closely related (Figure 3.4), but that there is no colony or population structure beyond the nests. This pattern was the same with geographic and effective distances, and can explain the weak association between those distances and relatedness. Therefore, the pattern of genetic variation in the Miami population can probably be explained by the fact that monk parakeets avoid nesting close to the natal nest (Gonçalves da Silva *et al.* 2010b), unlike what was previously documented (Martin & Bucher 1993). While monk parakeets may to some extent avoid areas with low canopy cover, they are probably strong enough flyers that they are not limited by distance and many landscape features.

Null models vs. Permutations:

Although I implemented the null model method for only a restricted number of hypotheses, in several of those cases null models and permutations produced very different results. Unlike null models, permutations consistently resulted in very high significance, regardless of whether MRM, Mantel, or partial Mantel tests were used. In fact, these tests were also highly significant when null distance matrices were used in place of the real independent matrices. This observation supports previous findings that Mantel and partial Mantel tests have highly inflated rates of Type I error (Oden & Sokal 1992; Raufaste & Rousset 2001; Guillot & Rousset 2013), and that this limitation also applies to MRM.

However, the null distance matrices were also highly associated to the geographic distance matrix, which indicates that they preserve the geographic distance signal with some random noise. This finding is not surprising considering that the null matrices were created by changing the landscape conductance but not the distance between the presence records. An alternative null model approach could be to keep the same conductance values as the landscape hypotheses considered, but use random geographic records instead of the real species presence (as in Chapter 1). Null geographic records could also be selected to preserve the spatial autocorrelation present in the real sample (as in Beale *et al.* 2008). This approach could be useful to keep the landscape features and their spatial autocorrelation constant, but change the distance between the records. As implemented here however, null models preserve the geographic locations and replace real landscape features (including the existing spatial autocorrelation) with random ones. Given that the effective least cost path distances on a uniform landscape were highly correlated with geographic distances, adding random variation between the presence records should only weaken this correlation, but not remove it completely.

Given that I detected a significant pattern of IBD (with both permutations and null models), it is reasonable that Mantel tests showed significant associations between the null distance matrices and the pairwise relatedness. However, it does not explain why the partial Mantel tests were also significant. Partial mantel tests are commonly used to test for the association between two matrices, while controlling for the effect of a third one (Cushman & Landguth 2010). According to this interpretation, partialling out the geographic distance signal from the null distance matrices should leave only random noise, and thus no association to the relatedness matrix. Inflated Type I error with partial Mantel tests is typically attributed to spatial autocorrelation (Oden & Sokal 1992; Raufaste & Rousset 2001; Guillot & Rousset 2013). This explanation is not appropriate here since the null conductance surfaces did not preserve the landscape structure.

Jaquiéry *et al.* (2010) demonstrated with simulations that Inflated Type I error with permutation tests may in fact be related to the strength of the landscape signal. In addition, this issue is not limited to partial correlations but applies also to regressions and approximate Bayesian computations. The power and accuracy of these methods is good only when there is a clear contrast between the conductivity of different landscape features and for features that strongly impede dispersal (Jaquiéry *et al.* 2010). This explanation certainly applies to the results and could explain why the hypothesis that yielded the largest significant R^2 (barrier below 20% canopy cover) also produced some of the sharpest contrast in conductivity across the landscape (Figure 3.2). On the contrary, the dispersal niche models produced relatively homogenous predictions with large portions of the study region with high conductivity (Figure 3.2). These patterns were therefore probably much harder to differentiate from IBD than the sharper canopy IBB scenario.

Nevertheless, I showed that for at least one of the dispersal niche models a significant pattern was detected with null models. With this approach, a significant association indicates that the landscape signal was significantly stronger than a random effect with the same geographic signal. Therefore, this null model approach may be a better alternative than the partial mantel test (and related methods) to evaluate a landscape hypothesis while controlling for IBD, especially when the landscape signal is weak.

Dispersal niche models:

Despite the weak relationship between geographic and effective distance on relatedness in the Miami monk parakeet population, the results indicate that the dispersal niche model approach produced results equivalent to the stepwise regression model. However, the stepwise regression model evaluated 41 different hypotheses based on multiple guesses of conductance values, and provided no protection for increased complexity. As discussed in the introduction, such an approach has a number of limitations (Graves *et al.* 2013). Thus, the dispersal niche models may offer a promising alternative to estimate resistance to dispersal. The most important advantage of this approach is that it does not require iterative correlation to the genetic data to estimate the resistance (or conductance) values and select the important landscape variables. Instead, it uses independent data to arrive at these values and the genetic data is only used for final hypothesis testing, in agreement with earlier recommendations (Spear *et al.* 2010; Graves *et al.* 2013). A dispersal niche model prediction forms an a priori hypothesis formulated with objective criteria and a repeatable method. Such an approach is arguably preferable to iteratively testing many “educated guesses”. Furthermore, unlike other methods such as telemetry studies, this approach does not require intensive behavioral studies. Species presence and environmental data are often readily accessible online for many taxa and geographic regions (Peterson *et al.* 2011, pp 62-95). As an extension of ENM, this approach can also benefit from all the theoretical and practical advances of this rapidly developing field. These advances can help researchers follow the best practices to select appropriate species data, environmental predictors, modeling algorithm, modeling settings, and evaluation methods (Elith & Leathwick 2009; Peterson *et al.* 2011). Finally, similarly to ENM, this approach can be used to make predictions in the future or in the past under different scenarios of urbanization, or climate change (Peterson *et al.* 2011, pp 200-208).

However, this approach also suffers from several limitations. First, aside from the current study there is to my knowledge no evidence that ecological niche modeling methods may be used to estimate dispersal preferences. Depending on the species, data, and spatial scales, this approach may not be appropriate. Studies with other species, different landscape variables, and simulated data will be necessary to determine the practical limits of this approach. In addition, as mentioned earlier the quality of the presence data (correct identification, accurate geographic coordinates, no sampling bias) is crucial to obtain accurate predictions (Syfert *et al.* 2013). Locations where the species is observed (sighted or captured) may not necessarily represent what is suitable for dispersal but rather what is suitable for other biological functions (foraging, courtship, mating, or nesting).

In addition, presence data is sensitive to sampling bias related to detection probability. For instance, a bird species may favor dense canopy to move/disperse but it may be much more likely to be sighted in open areas. This risk is probably small with monk parakeets given that they can be easily identified with their loud calls and active social behavior (Forshaw & Cooper 1989). Nevertheless, for the 2325 presence records used in this study, monk parakeets were predominantly documented in areas with less than 20% canopy cover (Appendix C). This may seem in opposition to the result of the landscape genetics analysis. However, percent canopy cover had a very low contribution in the Maxent models (Appendix C). This observation suggests that monk parakeets were documented more often in areas with low canopy because those areas are more abundant in the study region, rather than because of behavioral preferences. As a result, the dispersal niche models predicted a large influence of impervious surfaces (Appendix C) but may have underestimated the positive effect of increasing canopy cover.

Finally, the niche models predictions produced here were relatively homogenous with large portions of the study region with high conductivity. The power and accuracy of most tests of matrix association is dramatically reduced with conductance surfaces with such characteristics (Jaquiéry *et al.* 2010). Therefore, it may be necessary to choose modeling algorithms, or types of model outputs that offer more contrast in the predictions. For instance, here I could have chosen to multiply the predicted suitability by 10,000 rather than 100, or I could have opted for more restrictive thresholds for the IBB predictions. However, these decisions should be made a priori and based on biologically meaningful criteria.

Implications for monk parakeets' invasive potential:

Monk parakeets were for a long time suspected of displaying strong natal philopatry (i.e., dispersing close to the natal nest; Emlen 1990; Martin & Bucher 1993) which may have limited their rate of range expansion in invaded areas (Butler 2005). On the contrary, Gonçalves da Silva *et al.* (2010b) demonstrated based on parental assignment analyses that long distance dispersal was frequent in both native and introduced populations. The fine-scale spatial genetic structure analysis presented here suggests that there is no colony or population structure beyond high within-nest relatedness. Thus, monk parakeet may in fact avoid nesting close to the natal nest. Furthermore, the current results and those of Gonçalves da Silva *et al.* (2010b) indicate that monk parakeets have strong dispersal capabilities and are probably not limited by distance and most landscape features. They may to some extent avoid areas with very low canopy cover but this pattern is probably not strong enough to have any real consequences on their invasive potential. Therefore, it is likely that monk parakeets are capable of expanding their range to the adjacent ecologically suitable areas identified in Chapter 2. Unlike previously suggested (Butler 2005), whether these expansions will materialize may depend more on restrictions linked to local resource availability and eradication effort rather than on the monk parakeet dispersal potential. In Chapter 2, the ENM predictions revealed several areas adjacent to most of the currently well established populations that are particularly at risk of invasion. Considering the rapid rates of ranges expansions and populations growths documented in Spain (Domènech *et al.* 2003; Muñoz & Real 2006; Conroy & Senar 2009) and the US (Van Bael & Pruett-Jones 1996; Street *et al.* 2007; Pruett-Jones *et al.* 2011), and that most populations have benefitted from strong propagule pressure (Gonçalves da Silva *et al.* 2010b), it is very likely that monk parakeets will continue to expand their range in the future.

Conclusion:

Despite its limitations, the dispersal niche modeling approach offers many advantages over the traditional approach of examining multiple guesses of resistance values. Although the monk parakeet population in Florida presented little genetic spatial structure beyond high within-nest relatedness, the niche modeling approach produced equivalent results as a stepwise regression model considering a wide range of hypothesis. Furthermore, the results confirmed that partial Mantel tests and the derived multiple regressions on distance matrices may suffer from highly inflated rates of Type I error, even when the effect of spatial autocorrelation is removed. The null model approach implemented here may be a better alternative to evaluate the significance of association between genetic and effective distance matrices while controlling for isolation by distance. Finally, the results imply that although monk parakeets may to some extent avoid areas with low canopy cover, they are probably not limited by distance and most landscape features, and may therefore continue to expand their range to adjacent ecologically suitable areas.

Table 3.1: Results of the multiple regressions on distance matrices (MRM) tests evaluated with both the permutation and null model methods, with the Lynch-Ritland pairwise relatedness used as the dependent matrix, and the ln(x+1) transformed pairwise geographic (Geo) and least cost paths distances used as the independent matrices (IM). Note that the DNM model with geographic distance explains a proportion if variance equivalent to that of the full stepwise regression model. However, neither of these MRM tests was significant according to null resistance tests. The only MRM tests to be significant with both permutation and null resistance tests were the ones for the geographic distance, the MIA DNM model, and the IBB hypothesis with a quasi-barrier below 20 percent canopy cover.

MRM model	IM	Coefficient	Contribution	R^2	F
Geo	-	-	-	**0.0734***	849.71*
Intercept = 0.23*	Geo	-0.02*	-	-	-
Full model	-	-	-	0.0940*	278.31*
Intercept = 0.32*	Canopy-IBB-20%	-0.02*	6.83%	-	-
	Imperv-IBB-50%	0.01*	1.13%	-	-
	Geo	-0.03*	5.30%	-	-
	Landcov-FL-freq	0.02*	1.32%	-	-
Canopy	-	-	-	**0.0808***	943.50*
Intercept = 0.27*	Canopy-IBB-20%	-0.02*	NA	-	-
DNM	-	-	-	**0.0481***	541.87*
Intercept = 0.14*	DNM-MIA	-0.02*	-	-	-
DNM + Geo	-	-	-	0.0906*	534.19*
Intercept = 0.31*	DNM-FL-IBB-0.5	0.04*	18.97%	-	-
	Geo	-0.06*	51.35%	-	-

IBB = Isolation By Barrier; Canopy-IBB-20% = quasi-barrier below 20 percent canopy cover; Imperv-IBB-50% = quasi-barrier above 50 percent impervious surface; Landcov-FL-freq = conductance based on the frequency of land cover classes with monk parakeet presences in Florida. DNM = Dispersal Niche Model; DNM-MIA = DNM based on monk parakeet presences in Miami. DNM-FL-IBB-0.5 = DNM based on presences in Florida, with a 0.5 threshold applied. Contribution = difference between the R^2 obtained with all independent matrices and without the matrix of interest, as a percentage of total R^2. Asterisks indicate significant values according to permutations tests. R^2 values that are significant with null resistance tests are bolded. Significance assessed with a Bonferroni corrected alpha level of 0.05/53 = 0.00094.

Table 3.2: R^2 of the multiple regressions on distance matrices (MRM) for the dispersal niche models predictions (DNM). Lynch-Ritland pairwise relatedness was used as the dependent matrix. The ln(x+1) transformed least cost paths distances for the niche model predictions were used as the independent matrices (IM), either alone (DNM), or jointly with the ln(x+1) transformed geographic distance (DNM + Geo). Note that the MIA model provided the best fit alone, while the FL model with a barrier applied below a threshold of 0.5 provided the best fit when accounting for geographic distance (FL-IBB-0.5).

	IBB	**MRM Model**		**Contribution**	
IM	**Threshold**	**DNM**	**DNM + Geo**	**DNM**	**Geo**
MIA	None	0.0481	0.0890	17.57%	46.00%
MIA-IBB-0.5	0.5	0.0239	0.0852	13.87%	71.95%
MIA-IBB-10%	0.44	0.0259	0.0848	13.49%	69.45%
MIA-IBB-min	0.26	0.0445	0.0904	18.79%	50.76%
FL	None	0.0472	0.0893	17.85%	47.21%
FL-IBB-0.5	0.5	0.0441	0.0906	18.97%	51.35%
FL-IBB-10%	0.58	0.0450	0.0899	18.40%	50.01%
FL-IBB-min	0.55	0.0445	0.0902	18.67%	50.63%
US	None	0.0472	0.0893	17.81%	47.08%
US-IBB-0.5	0.5	0.0441	0.0905	18.94%	51.27%
US-IBB-10%	0.62	0.0449	0.0901	18.58%	50.18%
US-IBB-min	0.57	0.0445	0.0903	18.77%	50.76%

MIA, FL, US = DNM predictions based on monk parakeet presences in Miami, Florida, and the US, respectively; IBB = continuous prediction converted to binary Isolation By Barrier scenario based on following Maxent logistic threshold: “0.5” (fixed threshold of 0.5), “10%” (the tenth percentile predicted value of MIA presences), and “min” (the minimum predicted value of MIA presences). Contribution = difference between the R^2 obtained with the MRM for DNM + Geo and without the matrix of interest, as a percentage of total R^2. All R^2 are significant with permutations tests ($p < 0.00094$).

Figure 3.1: Monk parakeet presence records: (a) in the continental US, (b) for the genotyped individuals in the Miami metro Area. The black rectangle shows the extent of landscape considered for the least cost path analyses.

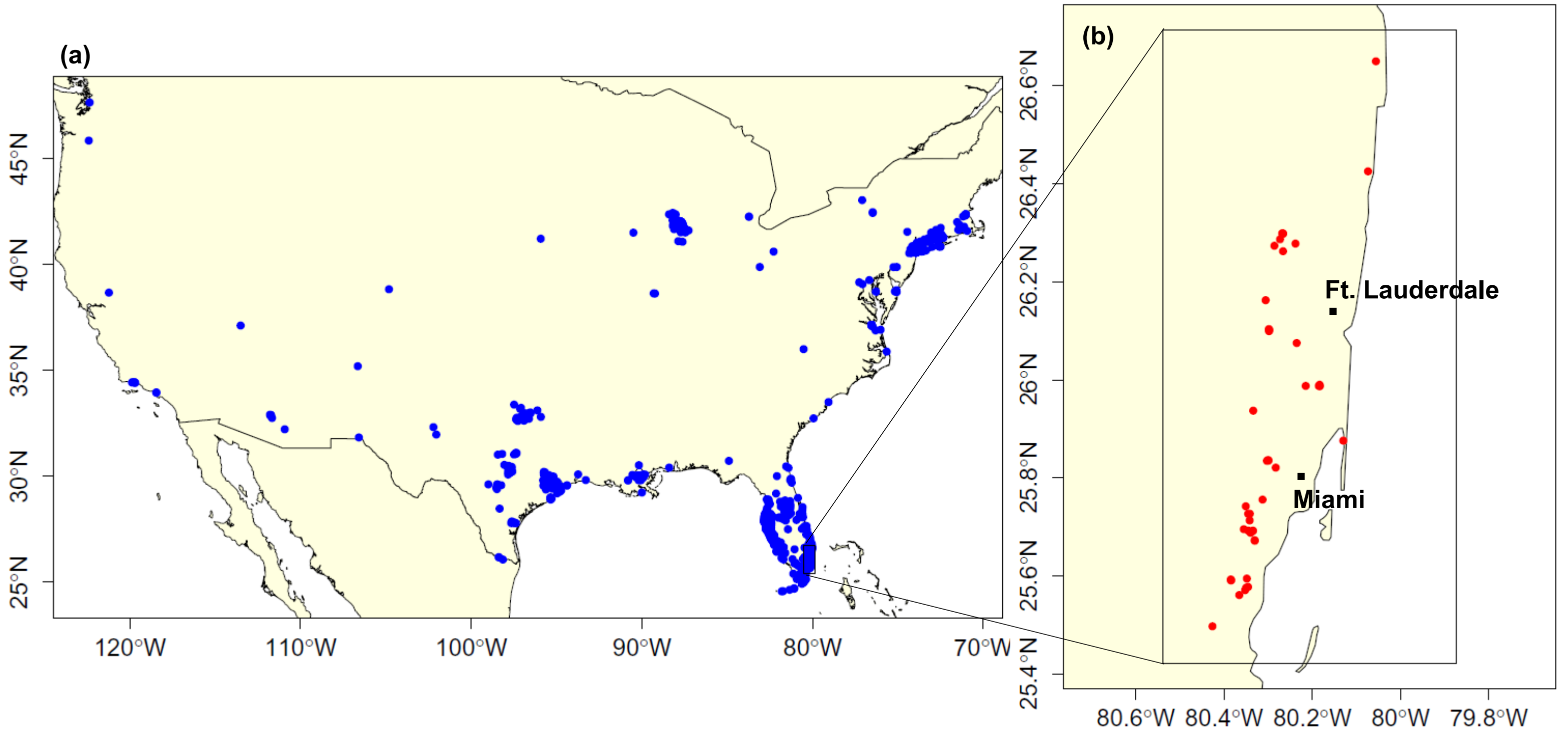


Figure 3.2: Examples of conductance surfaces: (a) the continuous prediction of the MIA dispersal niche model, (b) the binary (IBB) prediction of the FL dispersal niche model with a threshold of 0.5, and (c) the IBB hypothesis with a quasi barrier below 20% canopy cover and high conductance above (0 = quasi-barrier to dispersal, 100 = high conductance).

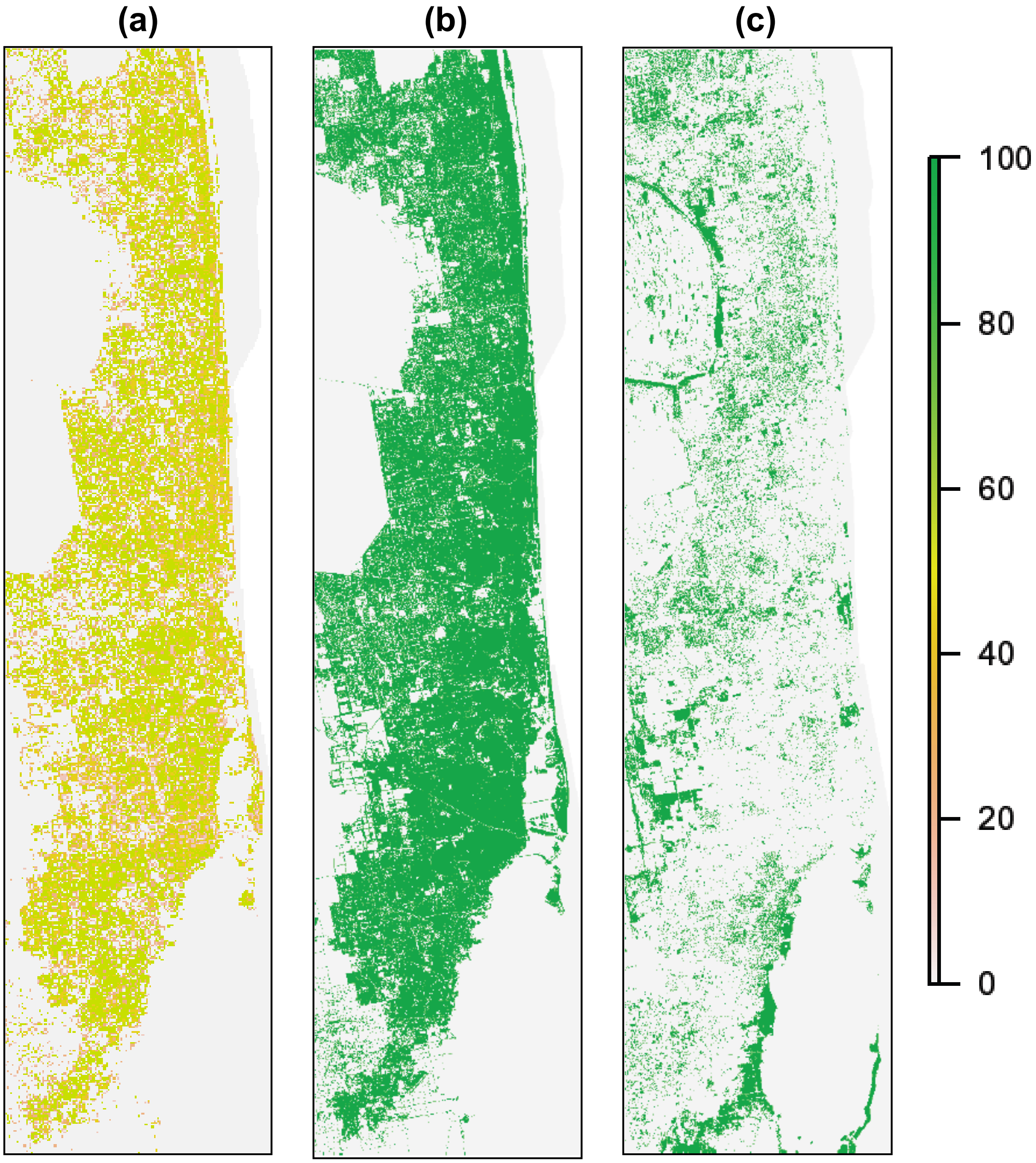


Figure 3.3: Examples of R^2 null distributions obtained from 99 MRM tests with null distance matrices. For each null replicate, the pairwise Lynch-Ritland relatedness was used as the dependent matrix, and a ln(x+1) transformed null distance matrix as the independent matrix, either alone (a), or jointly with the ln(x+1) transformed geographic distance (b). Null distance matrices were obtained by calculating pairwise least cost paths distances through landscapes with random conductance values. Note that the null R^2 are approximately normally distributed within a very narrow range of values.

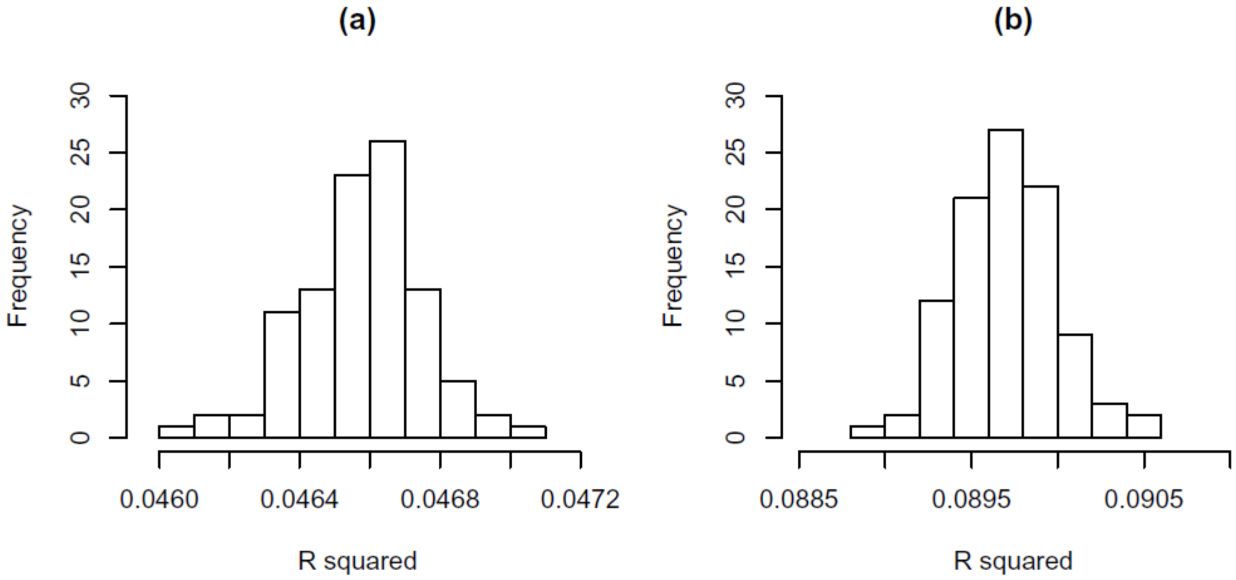
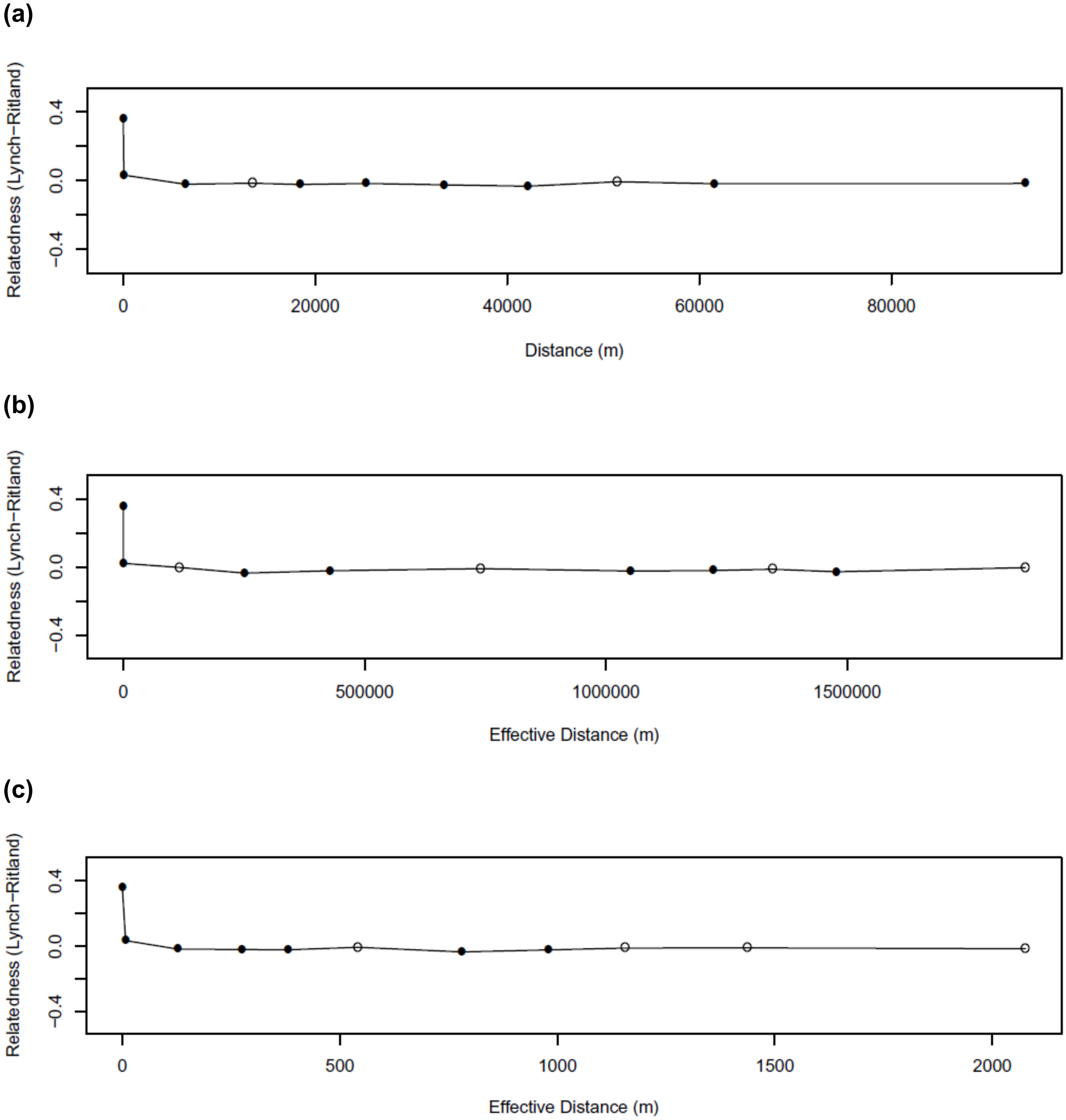


Figure 3.4: Average pairwise relatedness at different distance lags for: (a) the pairwise geographic distance, and the pairwise least cost path distance through (b) a landscape with low conductance below 20% canopy cover and high conductance above, and (c) the landscape corresponding to the prediction of the MIA dispersal niche model. Filled symbols are significant (p < .05) based on 1,000 permutations of the relatedness matrix. Distance lags were chosen to include the within nest distance of 0 m and 10 distance classes with approximately the same number of observations.



APPENDIX A

R scripts for the null model approach and the Maxent settings optimization procedure:

```
## This R script includes: 1) the main function to evaluate the performance of a model with
AUCtest, AUCdiff, and OR for a specified threshold, and to calculate significance with null
models. The progress is displayed on screen and the results can optionally be saved to a file
after each Maxent iteration. 2) A script to automate in parallel the optimization procedure
described in the text.
## Both scripts may require the following libraries, and options to be loaded in R:
Sys.setenv(NOAWT=TRUE) #only for certain macs before loading rJava
library(rJava) # required to run Maxent within dismo
#optional, to set the memory allocated to java, hence maxent (before loading dismo)
options(java.parameters = "-Xmx2g" )
library(dismo) # should also load automatically the required packages sp and raster
## 1) The main null model function:
mxt.nulltest(train, bg, n = 9, proj=bg, c.proj=TRUE, args = "addsamplestobackground", threshold
= .1, abs.diff=TRUE, test=NULL, group=NULL,  bg.group=NULL,save=NULL)
## The inputs for this function are:
## - train : a dataframe of the environmental values for the species calibration presences: one
row for each presence, one column for each predictor. Note that categorical variable must be
indicated with as.factor()
## - n : the number of null replicates
## - bg : the same as “train” but for the background points (the comparisons records for the
model)
## - proj : a dataframe in the same format as “train” and “bg” which corresponds to the pixels
over which the statistics will be calculated. By default this is the same as “bg” but it may be
useful when the models are projected to a different region
```

```
##  - c.proj : TRUE, if “proj” includes both the calibration and evaluation records, FALSE, if it
includes only the evaluation records (note that when FALSE is selected, AUCdiff will be
meaningless)
## - args: a list of Maxent arguments
## - abs.diff : TRUE or FALSE, whether the absolute value of AUCdiff should be returned
## - threshold : the percentile of calibration records used to set the threshold to calculate the
omission error rate.
## - test: the same as “train” but for the evaluation records. If it is not provided, a cross
validation test procedure will be used.
## - group : (only if test is missing), a vector of length nrow(train) which indicates the partitions
for the cross validation procedure. If omitted the presence records are assigned to partitions
randomly (with a minimum of 10 records per partitions, and a maximum of 10 partitions)
## - bg.group : (only if test is missing) the same as “group” but for the background points. If
omitted these are assigned to partitions randomly.
## - save: the path for the .csv file where the results should be saved (optional) after each
Maxent run (for both the real data and the null replicates)
## The function returns an object with the following slots:
## @summary : a data.frame with the performance of and significance summary statistics
## @random.reps : an array with the performance statistics for the null replicates
## Note: when a save file path is not provided, the results of this function should be saved to a
R object for further analyses.

mxt.nulltest = function(train, bg, n = 9, proj=bg, c.proj=TRUE, args =
"addsamplestobackground", threshold = .1, abs.diff=TRUE, test=NULL, group=NULL,
bg.group=NULL,save=NULL) {
  require(ROCR)
```

```
# the following sub-function runs the Maxent models and calculates the evaluation statistics
 f.real= function(x, p, args, proj, c.proj, bg, train, test, abs.diff, threshold) {
  tmpdir=paste(tempdir(),runif(1,0,1),sep="/")
  dir.create(tmpdir, showWarnings = TRUE, recursive = FALSE)
  mod=maxent(x,p, args=args,path=tmpdir)
  test.bb = predict(mod,proj)
  if(c.proj==TRUE) train.bb=test.bb else train.bb=predict(mod,bg)
  testpp = predict(mod,test)
  trainpp = predict(mod,train)
  combinedtest = c(testpp, test.bb)
  labeltest = c(rep(1,length(testpp)), rep(0,length(test.bb)))
  combinedtrain = c(trainpp, train.bb)
  labeltrain = c(rep(1,length(trainpp)), rep(0,length(train.bb)))
  predtest =  prediction(combinedtest, labeltest)
  predtrain =  prediction(combinedtrain, labeltrain)
  AUC= function(x) {performance(x, "auc")@y.values[[1]]}
  AUCtest = AUC(predtest)
  AUCtrain = AUC(predtrain)
  AUCdiff= AUCtrain - AUCtest
  if(abs.diff==FALSE) AUCdiff=AUCdiff else AUCdiff=abs(AUCdiff)
  r=length(testpp[testpp<quantile(trainpp,threshold)])
  t=length(testpp)
  a=length(test.bb[test.bb>=quantile(trainpp,threshold)])/length(test.bb)
  bintest=binom.test(t-r,t,p=a,alternative = "g")
  OR_binom_p=bintest$p.value
  OR= r/t
```

```
    stats= c(AUCtrain, AUCtest, AUCdiff, OR, OR_binom_p)
    unlink(tmpdir,recursive=T,force=T)
    return(stats)}
 # the rest of the function iterates the sub-function above for each of the real and null models and (optionally) saves the results
  if(missing(test)) {
    if(missing(group)) {
      if(nrow(train)<100) {k=trunc(nrow(train)/10)} else {k=10}
      group=kfold(train,k)} else {group=group
      k=length(unique(group))}
    if(missing(bg.group)) bg.group=kfold(bg, length(unique(group))) else bg.group=bg.group
    lbl=vector()
    for(i in 1:k) {lbl[[i]]=paste("Real",i,sep="_")}
    a=k-1
    for(i in 1:k) {for(j in 1:n) {lbl[[a+i+j]]=paste("Null",i,j,sep="_")}
                   a=a+n-1}
res=array(,dim=c(k*(n+1)+5,5),dimnames=list(c(lbl,"Mean_Real","Mean_Null","Std.Dev_Null","Z_score","p_value"),c("AUCtrain","AUCtest","AUCdiff","OR","OR_binom_p")))
    cat("Replicate",colnames(res),"  % Completion","\n", sep ="\t")
null=array(,dim=c(n,5,k),dimnames=list(NULL,c("AUCtrain","AUCtest","AUCdiff","OR","OR_binom_p"),NULL))
    a=k-1
    for (i in 1:k) {
      k.train = train[group != i,]
      k.test = train[group == i,]
      k.bg = bg[bg.group != i,]
```

```
      x=rbind(k.train, k.bg)
      p=c(rep(1,nrow(k.train)),rep(0,nrow(k.bg)))
      reps.x= list()
      reps =list()
      for (j in 1:n) {
        s = sample(nrow(k.bg),nrow(k.train))
        reps[[j]]=k.bg[s,]
        if(any(args=="noaddsamplestobackground")) {
          reps.x[[j]] = rbind(k.bg[s,], k.bg)
          reps.p = p} else {
            reps.x[[j]] = rbind(k.bg[s,], k.bg[-s,])
            reps.p = c(rep(1,length(s)),rep(0,(nrow(k.bg)-length(s))))}
      }
      res[i,]=f.real(x, p, args, proj, c.proj, bg, train=k.train, test=k.test, abs.diff, threshold)
cat(rownames(res)[i],res[i,],"  ",round((i*(n+1)-n)/((n+1)*k+1)*100,digits=2),"%","\n", sep ="\t")
for (j in 1:n) {
  null[j,,i]=f.real(reps.x[[j]],p= reps.p, args, proj, c.proj, bg, train=reps[[j]], test=k.test, abs.diff,
threshold)
  res[a+i+j,]=null[j,,i]
  cat(rownames(res)[a+i+j],res[a+i+j,],"  ",round((i*(n+1)-n+j)/((n+1)*k+1)*100,digits=2),"%","\n",
sep ="\t")
  if(!missing(save)) write.csv(res,save)
}
a=a+n-1}
res[k*(n+1)+1,]=apply(res[1:k,],2,mean)
means.k.null = rowMeans(null, dims=2)
```

```
res[k*(n+1)+2,]=colMeans(means.k.null)
res[k*(n+1)+3,]= apply(means.k.null,2,sd)
res[k*(n+1)+4,]=(res[k*(n+1)+1,]-res[k*(n+1)+2,])/res[k*(n+1)+3,]
res[k*(n+1)+5,1:2]=(1-pnorm(res[k*(n+1)+4,1:2]))
res[k*(n+1)+5,3:5]=(pnorm(res[k*(n+1)+4,3:5]))
cat(rownames(res)[k*(n+1)+5],res[k*(n+1)+5,],"  100 %","\n", sep ="\t")
real=res[-((k+1):(k*(n+1))),]
  } else {
    x=rbind(train, bg)
    p=c(rep(1,nrow(train)),rep(0,nrow(bg)))
    reps.x= list()
    reps =list()
    for (i in 1:n) {
      s = sample(nrow(bg),nrow(train))
      reps[[i]]=bg[s,]
      if(any(args=="noaddsamplestobackground")) {
        reps.x[[i]] = rbind(bg[s,], bg)
        reps.p = p} else {
          reps.x[[i]] = rbind(bg[s,], bg[-s,])
          reps.p = c(rep(1,length(s)),rep(0,(nrow(bg)-length(s))))
        }
    }
    lbl=vector()
    for (i in 1:n) {lbl[[i]]=paste("Null",i,sep="_")}
res=array(,dim=c(n+5,5),dimnames=list(c("Real",lbl,"Mean_Null","Std.Dev_Null","Z_score","p_v
alue"),c("AUCtrain","AUCtest","AUCdiff","OR","OR_binom_p")))
```

```
   cat("Replicate",colnames(res),"  % Completion","\n", sep ="\t")
   res[1,]=f.real(x, p, args, proj, c.proj, bg, train=train, test=test, abs.diff, threshold)
   cat(rownames(res)[1],res[1,],"  1 %","\n", sep ="\t")
   for(i in 1:n) {res[i+1,]=f.real(reps.x[[i]],p= reps.p, args, proj, c.proj, bg, train=reps[[i]], test=test,
abs.diff, threshold)
               cat(rownames(res)[i+1],res[i+1,],"  ",round((i+1)/(n+2)*100,digits=2),"%","\n", sep
="\t")
               if(!missing(save)) write.csv(res,save)
   }
   null=res[2:(n+1),]
   res[n+2,]=apply(null,2,mean)
   res[n+3,]=apply(null,2,sd)
   res[n+4,]=(res[1,]-res[n+2,])/res[n+3,]
   res[n+5,1:2]=1-pnorm(res[n+4,1:2])
   res[n+5,3:5]=pnorm(res[n+4,3:5])
   cat(rownames(res)[n+5],res[n+5,],"  100 %","\n", sep ="\t")
   real=res[-(2:(n+1)),]
  }
if(!missing(save)) write.csv(res,save)
Results<- setClass("Results", representation(summary="data.frame", random.reps="array"))
res=new("Results")
res@summary = data.frame(real)
res@random.reps = null
return(res)}
```

```
## 2) The script to automate in parallel the optimization procedure
# Create a list of maxent arguments corresponding to those settings 42 different settings
a="addsamplestobackground"
b="noautofeature"
h="nohinge"
t="nothreshold"
p="noproduct"
LQ=lapply(seq(.5,3,.5), function(x) {c(a,b,p,t,h,paste("betamultiplier",x,sep="="))})
LQP=lapply(seq(.5,3,.5), function(x) {c(a,b,t,h,paste("betamultiplier",x,sep="="))})
LQT=lapply(seq(.5,3,.5), function(x) {c(a,b,p,h,paste("betamultiplier",x,sep="="))})
LQH=lapply(seq(.5,3,.5), function(x) {c(a,b,p,t,paste("betamultiplier",x,sep="="))})
LQPT=lapply(seq(.5,3,.5), function(x) {c(a,b,h,paste("betamultiplier",x,sep="="))})
LQPH=lapply(seq(.5,3,.5), function(x) {c(a,b,t,paste("betamultiplier",x,sep="="))})
LQTH=lapply(seq(.5,3,.5), function(x) {c(a,b,t,h,paste("betamultiplier",x,sep="="))})
LQPTH=lapply(seq(.5,3,.5), function(x) {c(a,b,paste("betamultiplier",x,sep="="))})
args=c(LQ,LQP,LQT,LQH,LQPT,LQPH,LQTH,LQPTH)
# create a list of the files where the results for each combination of settings will be saved
(replace “your_path/your_file” by the desired path and file name)
LQs=lapply(seq(.5,3,.5),function(x) {paste("your_path/your_file_LQ",x,".csv",sep="")})
LQPs=lapply(seq(.5,3,.5),function(x) {paste("your_path/your_file_LQP",x,".csv",sep="")})
LQTs=lapply(seq(.5,3,.5),function(x) {paste("your_path/your_file_LQT",x,".csv",sep="")})
LQHs=lapply(seq(.5,3,.5),function(x) {paste("your_path/your_file_LQH",x,".csv",sep="")})
LQPTs=lapply(seq(.5,3,.5),function(x) {paste("your_path/your_file_LQPT",x,".csv",sep="")})
LQPHs=lapply(seq(.5,3,.5),function(x) {paste("your_path/your_file_LQPH",x,".csv",sep="")})
LQTHs=lapply(seq(.5,3,.5),function(x) {paste("your_path/your_file_LQTH",x,".csv",sep="")})
LQPTHs=lapply(seq(.5,3,.5),function(x) {paste("your_path/your_file_LQPTH",x,".csv",sep="")})
```

```
save=c(LQs,LQPs,LQTs,LQHs,LQPTs,LQPHs,LQTHs,LQPTHs)
# load the following libraries required for parallel execution
library(foreach)
library(doSNOW)
# make a cluster corresponding to how many parallel versions of R should be run (avoid using
every single core in your computer)
cl <- makeCluster(6, "SOCK")
registerDoSNOW(cl)
# run the foreach loop in parallel
maxent.runs <- foreach(i = 1:length(args), .packages = c("dismo", "rJava","ROCR")) %dopar% {
  mxt.nulltest(train, bg, n = 1000, proj=bg, c.proj=TRUE, args = args[[i]], group=group,
bg.group=bg.group,save=save[[i]])}
stopCluster(cl)
# the results should have been saved to the specified files but they can also be viewed in R, or
as a summary with the following lines of code:
summary=lapply(maxent.runs, function(x) x@summary)
names(summary)=save
##END
```

APPENDIX B

R script to partition a study region with a Voronoi tessellation:

```
## This function partitions a study region using a Vornoi tessellation and identifies the presence
and background records located in each partition. The Voronoi tessellation is created by
sampling a k points in the study region, where k is the number of partitions desired. The
sampling can be random, regular, stratified, clumped, or any other sampling method as defined
in the ‘spsample’ function from the ‘sp’ package.
voronoikfold(pres.pts,bg.pts,k,bg.poly='hull',type="regular",lonlat=F)
## The function requires the following inputs:
## pres.pts : an object of class SpatialPoints-class, the coordinates of the presence points
## bg.pts : an object of class SpatialPoints-class, the coordinates of the background points
## k : the number of desired partitions
## bg.poly : the extent of the study regions. By default, the function uses the convex hull around
all the presence points. If a distance is specified, it will calculate a buffered region around all the
presence points with a radius equal to the specified distance. Any other extent can be provided,
as long as it is an object of class SpatialPolygons or a SpatialPolygonsDataFrame.
## type: the sampling method, see ‘spsample’ from the ‘sp’ package. The default is “regular” to
create partitions of approximately equal sizes.
## lonlat : TRUE or FALSE (the default), whether the presence, background, and study region
extent, have a longitude/latitude coordinate reference system
## The function returns an object with the following slots:
## @polygons : an object of class SpatialPolygons which can be plotted to show the Voronoi
tessellation produced
## @group = a vector indicating in which partition each of the presence record is located
## @bg.group = a vector indicating in which partition each of the backround point is located
voronoikfold= function(pres.pts,bg.pts,k,bg.poly='hull',type="regular",lonlat=F) {
```

```
  require(deldir)
  require(sp)
  require(raster)
  require(rgeos)
  require(dismo)
  require(maptools)
  if(is.numeric(bg.poly)){
    c <- circles(pres.pts, d=bg.poly, lonlat=lonlat)
    if (!rgeosStatus()) gpclibPermit()
    cp=unionSpatialPolygons(c@polygons,rep(1,length(pres.pts)))
    bgextpoly=as(extent(bg.pts), "SpatialPolygons")
    y = gIntersection(cp,bgextpoly,byid=TRUE)} else {
      if(any(class(bg.poly)=="SpatialPolygons"|class(bg.poly)=="SpatialPolygonsDataFrame")) {y=bg.poly} else {
        if(bg.poly=='hull') {
          hull = convHull(pres.pts)
          y = hull@polygons} else {
            stop('bg.poly must be either a distance in meter, a SpatialPolygons or a SpatialPolygonsDataFrame')}}}
  kc=NULL
  while(length(kc)!=k) kc=spsample(y, k, type=type)
  bb = bbox(bg.pts)
  rw = as.numeric(t(bb))
  z <- deldir(kc@coords[,1], kc@coords[,2],rw=rw)
  w <- tile.list(z)
  polys <- vector(mode='list', length=length(w))
```

```
  for(i in seq(along=polys)) {
    pcrds = cbind(w[[i]]$x, w[[i]]$y)
    pcrds = rbind(pcrds, pcrds[1,])
    polys[[i]] <- Polygons(list(Polygon(pcrds)), ID=as.character(i))
  }
  SP <- SpatialPolygons(polys)
  proj4string(SP)=proj4string(pres.pts)
  Results<- setClass("Results", representation(polygons="SpatialPolygons", group="vector",
bg.group="vector"))
  res=new("Results")
  res@polygons = SP
  res@group = over(pres.pts,SP)
  res@bg.group = over(bg.pts,SP)
  return(res)
}
##END
```

APPENDIX C

Supplemental documentation for Chapter 3:

This Appendix includes statistics of the optimized Maxent models (Table C.1); the transforms applied to convert the percent canopy cover and percent impervious surface land cover variables to conductance surfaces (Table C.2 & Figure C.1); and the frequency of monk parakeet presence observed for the three land cover variables, with matching conductance for the corresponding IBB hypotheses (Table C.3).

Table C.1: Statistics of the optimized Maxent models based on monk parakeet presence records in the Miami metro area (MIA), Florida (FL), and the entire continental United States (US).

	Maxent Model		
Statistic	**MIA**	**FL**	**US**
#Training samples	51	1045	2325
Regularized training gain	2.27	0.86	0.88
Unregularized training gain	2.35	0.89	0.90
Iterations	300	500	500
Calibration AUC	0.96	0.83	0.84
#Background points (includes training samples)	29783	30022	86344
Canopy contribution (%)	0.12	0.19	1.86
Impervious contribution (%)	93.01	62.87	91.67
Land cover class contribution (%)	6.86	36.94	6.47
Canopy permutation importance (%)	2.10	2.23	8.36
Impervious permutation importance (%)	92.17	11.45	70.27
Land cover class permutation importance (%)	5.73	86.33	21.37
Entropy	8.03	9.46	10.49
Prevalence (average of logistic output over background sites)	0.05	0.20	0.20
Minimum MIA records logistic threshold	0.26	0.55	0.57
10 percentile MIA records logistic threshold	0.44	0.58	0.62

Table C.2: Transforms applied to convert continuous land cover variables (percent canopy cover and percent impervious surface) to conductance surfaces, with the corresponding hypothesis names. The legend refers to the illustration of these transforms in Figure C.1.

		Hypothesis Name	
Transform	**Legend**	**% Canopy Cover**	**% Impervious surface**
y = x	a	Canopy	Imperv
y = x^2 / 100	b	Canopy_sqr	Imperv_sqr
y = x^0.5 * 10	c	Canopy_sqroot	Imperv_sqroot
y = ln(x+1) / ln(101) * 100	d	Canopy_ln	Imperv_ln
y = exp(x*ln(101)/100) – 1	e	Canopy_exp	Imperv_exp
y = abs(100/(1+x) - 100)	f	Canopy_recInv_r	Imperv_recInv_r
y = 100/(101-x)	g	Canopy_rec_r	Imperv_rec_r
y = 100 – x	h	Canopy_r	Imperv_r
y = (100-x)^0.5 * 10	i	Canopy_sqroot_r	Imperv_sqroot_r
y = (100-x)^2 / 100	j	Canopy_sqr_r	Imperv_sqr_r
y = ln(101-x) / ln(101) * 100	k	Canopy_ln_r	Imperv_ln_r
y = exp((100-x)*ln(101)/100) - 1	l	Canopy_exp_r	Imperv_exp_r
y = abs(100/(101-x) - 100)	m	Canopy_recInv	Imperv_recInv
y= 100/(x+1)	n	Canopy_rec	Imperv_rec

Figure C.1: Transforms applied to convert continuous land cover variables to conductance surfaces. Legend in Table C.2.

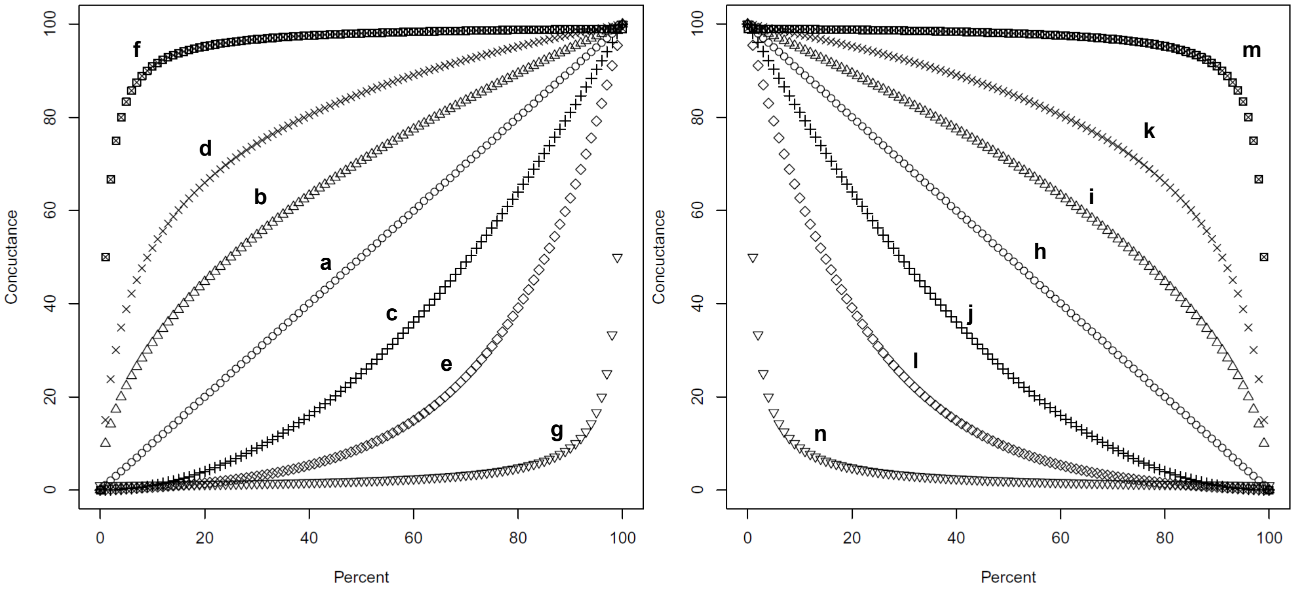


Table C.3: Frequency of monk parakeet presences observed in the Miami metropolitan area (MIA), Florida (FL), and the entire continental United States (US) for: (a) the percent canopy cover, (b) the percent impervious surface, and (c) the land cover classification. Conductance values are included for the corresponding Isolation By Barrier (IBB) hypotheses.

(a)

% Canopy cover	**Frequency**			**IBB Conductance**		
Interval	**MIA**	**FL**	**US**	**IBB-20%**	**IBB-20-80%**	**IBB-80%**
[0,10]	82.35%	74.35%	74.84%	100	0	0
(10,20]	13.73%	6.03%	3.96%	100	0	0
(20,30]	1.96%	4.88%	4.00%	0	100	0
(30,40]	1.96%	2.97%	3.23%	0	100	0
(40,50]	0.00%	1.63%	2.54%	0	100	0
(50,60]	0.00%	1.44%	1.72%	0	100	0
(60,70]	0.00%	1.53%	1.89%	0	100	0
(70,80]	0.00%	1.34%	2.45%	0	100	0
(80,90]	0.00%	1.63%	2.28%	0	0	100
(90,100]	0.00%	4.21%	3.10%	0	0	100

(b)

% Impervious surface	**Frequency**			**IBB Conductance**		
Interval	**MIA**	**FL**	**US**	**IBB-20%**	**IBB-50%**	**IBB-80%**
[0,10]	0.00%	37.32%	34.54%	100	100	100
(10,20]	11.76%	7.85%	7.31%	100	100	100
(20,30]	15.69%	9.47%	7.96%	0	100	100
(30,40]	31.37%	10.72%	9.85%	0	100	100
(40,50]	17.65%	8.42%	8.04%	0	100	100
(50,60]	13.73%	8.04%	8.09%	0	0	100
(60,70]	5.88%	5.36%	6.71%	0	0	100
(70,80]	3.92%	4.88%	5.72%	0	0	100
(80,90]	0.00%	3.92%	6.11%	0	0	0
(90,100]	0.00%	4.02%	5.68%	0	0	0

(c): continued on following page.

Table C.3: continued from previous page.

(c)

	Frequency			IBB Conductance		
Land Cover Class	**MIA**	**FL**	**US**	**IBB-1**	**IBB-2**	**IBB-3**
Open Water	0.00%	4.88%	4.60%	0	0	100
Developed, Open Space	3.92%	18.47%	16.39%	100	100	100
Developed, Low Intensity	74.51%	29.76%	26.32%	100	100	100
Developed, Medium Intensity	21.57%	18.95%	21.25%	100	100	100
Developed, High Intensity	0.00%	7.85%	11.96%	0	100	100
Barren Land (Rock/Sand/Clay)	0.00%	1.44%	1.63%	0	0	0
Deciduous Forest	0.00%	0.00%	2.11%	0	0	0
Evergreen Forest	0.00%	0.19%	0.99%	0	0	0
Mixed Forest	0.00%	0.10%	0.26%	0	0	0
Shrub/Scrub	0.00%	0.57%	0.90%	0	0	0
Grassland/Herbaceous	0.00%	0.38%	1.46%	0	0	0
Pasture/Hay	0.00%	0.77%	1.25%	0	0	0
Cultivated Crops	0.00%	0.96%	0.90%	0	0	0
Woody Wetlands	0.00%	10.24%	6.24%	0	100	100
Emergent Herbaceous Wetlands	0.00%	5.45%	3.74%	0	0	100

REFERENCES

Algar, A.C., Mahler, D.L., Glor, R.E. & Losos, J.B. (2013). Niche incumbency, dispersal limitation and climate shape geographical distributions in a species-rich island adaptive radiation. *Global Ecology and Biogeography*, **22**, 391–402.

Anderson, R.P. (2012). Harnessing the world's biodiversity data: promise and peril in ecological niche modeling of species distributions. *Annals of the New York Academy of Sciences*, **1260**, 66–80.

Anderson, R.P., Gomez-Laverde, M. & Peterson, A.T. (2002). Geographical distributions of spiny pocket mice in South America: insights from predictive models. *Global Ecology and Biogeography*, **11**, 131–141.

Anderson, R.P. & Gonzalez, I. (2011). Species-specific tuning increases robustness to sampling bias in models of species distributions: An implementation with Maxent. *Ecological Modelling*, **222**, 2796–2811.

Anderson, R.P., Lew, D. & Peterson, A.T. (2003). Evaluating predictive models of species ' distributions : criteria for selecting optimal models. *Ecological Modelling*, **162**, 211–232.

Anderson, R.P. & Raza, A. (2010). The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with montane rodents (genus Nephelomys) in Venezuela. *Journal of Biogeography*, **37**, 1378–1393.

Araújo, M.B., Thuiller, W. & Yoccoz, N.G. (2009). Reopening the climate envelope reveals macroscale associations with climate in European birds. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, E45–6; author reply E41–3.

Aurenhammer, F. (1991). Voronoi diagrams---a survey of a fundamental geometric data structure. *ACM Computing Surveys*, **23**, 345–405.

Avery, M.L., Greiner, E., Lindsay, J.R., Newman, J.R. & Pruett-Jones, S. (2002). Monk parakeet management at electric utility facilities in south Florida. *Proceedings 20th Vertebrate Pest Conference* (eds R.M. Timm & R.H. Schmidt), pp. 140–145. University of California, Davis.

Aylor, D.E. (2003). SPREAD OF PLANT DISEASE ON A CONTINENTAL SCALE: ROLE OF AERIAL DISPERSAL OF PATHOGENS. *Ecology*, **84**, 1989–1997.

Van Bael, S. & Pruett-Jones, S. (1996). Exponential population growth of monk parakeets in the United States. *The Wilson Bulletin*, **108**, 584–588.

Balkenhol, N., Gugerli, F., Cushman, S. a., Waits, L.P., Coulon, A., Arntzen, J.W., Holderegger, R. & Wagner, H.H. (2009a). Identifying future research needs in landscape genetics: where to from here? *Landscape Ecology*, **24**, 455–463.

Balkenhol, N., Waits, L.P. & Dezzani, R.J. (2009b). Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. *Ecography*, **32**, 818–830.

Barton, K.A., Phillips, B.L., Morales, J.M. & Travis, J.M.J. (2009). The evolution of an “intelligent” dispersal strategy: biased, correlated random walks in patchy landscapes. *Oikos*, **118**, 309–319.

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

Beale, C.M., Lennon, J.J. & Gimona, A. (2009). European bird distributions still show few climate associations. *Proceedings of the National Academy of Sciences*, **106**, E41–E43.

Beale, C.M., Lennon, J.J. & Gimona, A. (2008). Opening the climate envelope reveals no macroscale associations with climate in European birds. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 14908–12.

Beaumont, L.J., Gallagher, R. V., Thuiller, W., Downey, P.O., Leishman, M.R. & Hughes, L. (2009). Different climatic envelopes among invasive populations may lead to underestimations of current and future biological invasions. *Diversity and Distributions*, **15**, 409–420.

Beier, P., Majka, D.R. & Spencer, W.D. (2008). Forks in the road: choices in procedures for designing wildland linkages. *Conservation biology : the journal of the Society for Conservation Biology*, **22**, 836–51.

Broennimann, O. & Guisan, A. (2008). Predicting current and future biological invasions: both native and invaded ranges matter. *Biology letters*, **4**, 585–9.

Broennimann, O., Treier, U. a, Müller-Schärer, H., Thuiller, W., Peterson, A.T. & Guisan, A. (2007). Evidence of climatic niche shift during biological invasion. *Ecology letters*, **10**, 701–9.

Burger, J. & Gochfeld, M. (2005). Nesting behavior and nest site selection in monk parakeets (Myiopsitta monachus) in the Pantanal of Brazil. *Acta Ethologica*, **8**, 23–34.

Butler, C.J. (2005). Feral Parrots in the Continental United States and United Kingdom: Past, Present, and Future. *Journal of Avian Medicine and Surgery*, **19**, 142–149.

Cabral, J.S. & Kreft, H. (2012). Linking ecological niche, community ecology and biogeography: insights from a mechanistic niche model (S. Higgins, Ed.). *Journal of Biogeography*, **39**, 2212–2224.

Caccamise, D.F. & Weathers, W.W. (1977). Winter nest microclimate of Monk Parakeets. *The Wilson Bulletin*, **89**, 346–349.

Calenge, C., Darmon, G., Basille, M., Loison, a & Jullien, J.-M. (2008). The factorial decomposition of the Mahalanobis distances in habitat selection studies. *Ecology*, **89**, 555–66.

Cao, Y., DeWalt, R.E., Robinson, J.L., Tweddale, T., Hinz, L. & Pessino, M. (2013). Using Maxent to model the historic distributions of stonefly species in Illinois streams: The effects of regularization and threshold selections. *Ecological Modelling*, **259**, 30–39.

Conroy, M.J. & Senar, J.C. (2009). Modeling Demographic Processes In Marked Populations (D.L. Thomson, E.G. Cooch & M.J. Conroy, Eds.). *Environmental and Ecological Statistics*, **3**, 491–510.

Cushman, S. a & Landguth, E.L. (2010). Spurious correlations and inference in landscape genetics. *Molecular ecology*, 3592–3602.

Cushman, S. a, McKelvey, K.S., Hayden, J. & Schwartz, M.K. (2006). Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. *The American naturalist*, **168**, 486–99.

Dijkstra, E.W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, **1**, 269–271.

Domènech, J., Carrillo, J. & Senar, J.C. (2003). Population size of the Monk Parakeet (Myiopsitta monachus) in Catalonia. *Revista Catalana d'Ornitologia*, **20**, 1–9.

Duncan, R.P., Cassey, P. & Blackburn, T.M. (2009). Do climate envelope models transfer? A manipulative test using dung beetle introductions. *Proceedings. Biological sciences / The Royal Society*, **276**, 1449–57.

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

Elith, J., Kearney, M. & Phillips, S.J. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution*, **1**, 330–342.

Elith, J. & Leathwick, J.R. (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 677–697.

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

Emlen, S.T. (1990). Observations on a captive colony of Quaker Parakeets. *AFA Watchbird*, **27**, 26–29.

Epps, C.W., Wehausen, J.D., Bleich, V.C., Torres, S.G. & Brashares, J.S. (2007). Optimizing dispersal and corridor models using landscape genetics. *Journal of Applied Ecology*, **44**, 714–724.

Van Etten, J. (2012). gdistance: distances and routes on geographical grids. R package version 1.1-4. http://CRAN.R-project.org/package=gdistance.

Fang, Z., Pyhäjärvi, T., Weber, A.L., Dawe, R.K., Glaubitz, J.C., González, J.D.J.S., Ross-Ibarra, C., Doebley, J., Morrell, P.L. & Ross-Ibarra, J. (2012). Megabase-scale inversion polymorphism in the wild ancestor of maize. *Genetics*, **191**, 883–94.

Farber, O. & Kadmon, R. (2003). Assessment of alternative approaches for bioclimatic modeling with special emphasis on the Mahalanobis distance. *Ecological Modelling*, **160**, 115–130.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, **27**, 861–874.

Fischer, G., Nachtergaele, F., Prieler, S., van Velthuizen, H.T., Verelst, L. & Wiberg, D. (2008). Global Agro-ecological Zones Assessment for Agriculture (GAEZ 2008).

Fitzpatrick, M.C., Preisser, E.L., Porter, A., Elkinton, J. & Ellison, A.M. (2012). Modeling range dynamics in heterogeneous landscapes: invasion of the hemlock woolly adelgid in eastern North America. *Ecological Applications*, **22**, 472–486.

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.

Forshaw, J.M. & Cooper, W.T. (1989). *Parrots of the world*, Third [Revn. Lansdowne Editions, Willoughby, Australia.

Fry, J., Xian, G.J., In, S., Dewitz, J., Homer, C., Yang, L., Barnes, C., Herold, N. & Wickham, J. (2011). Completion of the 2006 National Land Cover Database for the Conterminous United States. *PE&RS*, **77**, 858–864.

Gallien, L., Douzet, R., Pratte, S., Zimmermann, N.E. & Thuiller, W. (2012). Invasive species distribution models - how violating the equilibrium assumption can create new insights. *Global Ecology and Biogeography*, **21**, 1126–1136.

Gibson, E. (1920). Further Ornithological Notes from the Neighbourhood of Cape San Antonio, Province of Buenos Ayres. Part III. Phœnicopterldae- Rheidae. *Ibis*, **62**, 1–97.

Gonçalves da Silva, A., Eberhard, J.R., Wright, T.F., Avery, M.L. & Russello, M.A. (2010a). Data from: Genetic evidence for high propagule pressure and long-distance dispersal in monk parakeet (Myiopsitta monachus) invasive populations. Dryad Digital Repository. doi:10.5061/dryad.1682.

Gonçalves da Silva, A., Eberhard, J.R., Wright, T.F., Avery, M.L. & Russello, M.A. (2010b). Genetic evidence for high propagule pressure and long-distance dispersal in monk parakeet (Myiopsitta monachus) invasive populations. *Molecular ecology*, **19**, 3336–50.

Goslee, S.C. & Urban, D.L. (2007). The ecodist Package for Dissimilarity-based Analysis of Ecological Data. **22**.

Graves, T. a, Beier, P. & Royle, J.A. (2013). Current approaches using genetic distances produce poor estimates of landscape resistance to interindividual dispersal. *Molecular ecology*, 1–16.

Greenwald, K.R., Gibbs, H.L. & Waite, T. a. (2009). Efficacy of land-cover models in predicting isolation of marbled salamander populations in a fragmented landscape. *Conservation biology : the journal of the Society for Conservation Biology*, **23**, 1232–41.

Gridded Population of the World, Version 3 (GPWv3): Population Count Grid. (2005).

Guillot, G., Leblois, R., Coulon, A. & Frantz, A.C. (2009). Statistical methods in spatial genetics. *Molecular ecology*, **18**, 4734–56.

Guillot, G. & Rousset, F. (2013). Dismantling the Mantel tests (L. Harmon, Ed.). *Methods in Ecology and Evolution*, **4**, 336–344.

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.

Hartig, F., Dyke, J., Hickler, T., Higgins, S.I., O'Hara, R.B., Scheiter, S. & Huth, A. (2012). Connecting dynamic vegetation models to data - an inverse perspective. *Journal of Biogeography*, **39**, 2240–2252.

Heikkinen, R.K., Marmion, M. & Luoto, M. (2012). Does the interpolation accuracy of species distribution models come at the expense of transferability? *Ecography*, **35**, 276–288.

Heinänen, S., Erola, J. & Numers, M. (2012). High resolution species distribution models of two nesting water bird species: a study of transferability and predictive performance. *Landscape Ecology*, **27**, 545–555.

Hijmans, R.J. (2012). Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology*, **93**, 679–688.

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

Hoffman, J.D., Narumalani, S., Mishra, D.R., Merani, P. & Wilson, R.G. (2008). Predicting Potential Occurrence and Spread of Invasive Plant Species along the North Platte River, Nebraska. *Invasive Plant Science and Management*, **1**, 359–367.

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, **6**, 65–70.

Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., Herold, N., McKerrow, A., VanDriel, J.N. & Wickham, J. (2007). Completion of the 2001 National Land Cover Database for the Counterminous United States. *PE&RS*, **73**, 337–341.

Hubert, L.J. & Golledge, R.G. (1981). A heuristic method for the comparison of related structures. *Journal of Mathematical Psychology*, **23**, 214–226.

Jaquiéry, J., Broquet, T., Hirzel, a H., Yearsley, J. & Perrin, N. (2010). Inferring landscape effects on dispersal from genetic distances: how far can we go? *Molecular ecology*.

Jiménez-Valverde, A. (2011). Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Global Ecology and Biogeography*, **5**, no–no.

Jiménez-Valverde, A., Lobo, J.M. & Hortal, J. (2008). Not as good as they seem: the importance of concepts in species distribution modelling. *Diversity and Distributions*, **14**, 885–890.

Jombart, T., Pontier, D. & Dufour, a-B. (2009). Genetic markers in the playground of multivariate analysis. *Heredity*, **102**, 330–41.

Krasnov, B.R., Mouillot, D., Shenbrot, G.I., Khokhlova, I.S. & Poulin, R. (2010). Deconstructing spatial patterns in species composition of ectoparasite communities: the relative contribution of host composition, environmental variables and geography. *Global Ecology and Biogeography*, 515–526.

Landguth, E.L., Cushman, S. a, Schwartz, M.K., McKelvey, K.S., Murphy, M. & Luikart, G. (2010). Quantifying the lag time to detect barriers in landscape genetics. *Molecular ecology*, 4179–4191.

Last of the Wild Project, Version 2, 2005 (LWP-2): Global Human Influence Index (HII) Dataset. (2005).

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

Legendre, P., Borcard, D. & Peres-Neto, P.R. (2005). ANALYZING BETA DIVERSITY: PARTITIONING THE SPATIAL VARIATION OF COMMUNITY COMPOSITION DATA. *Ecological Monographs*, **75**, 435–450.

Legendre, P. & Fortin, M.-J. (2010). Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular ecology resources*, **10**, 831–44.

Legendre, P., Lapointe, F.J. & Casgrain, P. (1994). Modeling brain evolution from behavior: a permutational regression approach. *Evolution*, 1487–1499.

Li, Q., Yang, X., Soininen, J., Chu, C.-J., Zhang, J.-Q., Yu, K.-L. & Wang, G. (2011). Relative importance of spatial processes and environmental factors in shaping alpine meadow communities. *Journal of Plant Ecology*, **4**, 249–258.

Lichstein, J.W. (2007). Multiple regression on distance matrices: a multivariate spatial analysis tool. *Plant Ecology*, **188**, 117–131.

Liu, C., Berry, P.M., Dawson, T.P. & Pearson, R.G. (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, **3**, 385–393.

Liu, C., White, M. & Newell, G. (2011). Measuring and comparing the accuracy of species distribution models with presence-absence data. *Ecography*, **34**, 232–243.

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

Lynch, M. & Ritland, K. (1999). Estimation of Pairwise Relatedness With Molecular Markers. *Genetics*.

Mandle, L., Warren, D.L., Hoffmann, M.H., Peterson, A.T., Schmitt, J. & von Wettberg, E.J. (2010). Conclusions about Niche Expansion in Introduced Impatiens walleriana Populations Depend on Method of Analysis (J.C. Stout, Ed.). *PLoS ONE*, **5**, e15297.

Manel, S. (2003). Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, **18**, 189–197.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer research*, **27**, 209–20.

Mårell, A., Ball, J.P. & Hofgaard, A. (2002). Foraging and movement paths of female reindeer: insights from fractal analysis, correlated random walks, and Lévy flights. *Canadian Journal of Zoology*, **80**, 854–865.

Margosian, M.L., Garrett, K.A., Hutchinson, J.M.S. & With, K.A. (2009). Connectivity of the American Agricultural Landscape: Assessing the National Risk of Crop Pest and Disease Spread. *BioScience*, **59**, 141–151.

Martin, L.F. & Bucher, E.H. (1993). Natal dispersal and first breeding age in Monk Parakeets. *The Auk*, **110**, 930–933.

Martiny, J.B.H., Eisen, J.A., Penn, K., Allison, S.D. & Horner-devine, M.C. (2011). Drivers of bacterial β -diversity depend on spatial scale.

McRae, B.H. (2006). Isolation By Resistance. *Evolution*, **60**, 1551–1561.

Mcrae, B.H. & Beier, P. (2007). Circuit theory predicts gene flow in plant and animal populations. *Sciences-New York*.

McRae, B., Dickson, B., Keitt, T. & Shah, V. (2008). Using circuit theory to model connectivity in ecology, evolution, and conservation. *Ecology*, **89**, 2712–2724.

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

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

Munshi-south, J. (2012). Urban landscape genetics : canopy cover predicts gene flow between white-footed mouse (Peromyscus leucopus) populations in New York City. *New York*.

Murphy, M. a, Dezzani, R., Pilliod, D.S. & Storfer, a. (2010). Landscape genetics of high mountain frog metapopulations. *Molecular ecology*, **19**, 3634–49.

Navarro, J.L., Martella, M.B. & Bucher, E.H. (1992). Breeding season and productivity of Monk Parakeets in Cordoba, Argentina. *The Wilson Bulletin*, **104**, 413–424.

Oden, N.L. & Sokal, R.R. (1992). An investigation of three-matrix permutation tests. *Journal of Classification*, **9**, 275–290.

Pagel, J. & Schurr, F.M. (2012). Forecasting species ranges by statistical estimation of ecological niches and spatial population dynamics. *Global Ecology and Biogeography*, **21**, 293–304.

Peres-Neto, P.R., Legendre, P., Dray, S. & Borcard, D. (2006). Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology*, **87**, 2614–25.

Peterson, A.T., Barve, N., Bini, L.M., Diniz-Filho, J.A., Jiménez-Valverde, A., Lira-Noriega, A., Lobo, J., Maher, S., de Marco, P., Martínez-Meyer, E., Nakazawa, Y. & Soberón, J. (2009). The climate envelope may not be empty. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, E47; author reply E41–3.

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.

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

Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M. & Araújo, M.B. (2011). *Ecological Niches and Geographic Distributions* (S.A. Levin & H.S. Horn, Eds.). Princeton University Press.

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

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

Phillips, S.J., Dudík, M. & Schapire, R.E. (2004). A maximum entropy approach to species distribution modeling. *Twenty-first international conference on Machine learning - ICML '04*, 83.

Pontius, R. & Schneider, L. (2001). Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. *Agriculture, Ecosystems & Environment*, **85**, 239–248.

Pruett-Jones, S., Appelt, C.W., Sarfaty, A., Vossen, B., Leibold, M. a. & Minor, E.S. (2011). Urban parakeets in Northern Illinois: A 40-year perspective. *Urban Ecosystems*.

Pulliam, H.R. (2000). On the relationship between niche and distribution. *Ecology Letters*, **58**, 1103–361.

R Core Team. (2012). R: A Language and Environment for Statistical Computing.

Raes, N. & ter Steege, H. (2007). A null-model for significance testing of presence-only species distribution models. *Ecography*, **30**, 727–736.

Raufaste, N. & Rousset, F. (2001). ARE PARTIAL MANTEL TESTS ADEQUATE? *Evolution*, **55**, 1703–1705.

Reside, A.E., Watson, I., VanDerWal, J. & Kutt, A.S. (2011). Incorporating low-resolution historic species location data decreases performance of distribution models. *Ecological Modelling*, **222**, 3444–3448.

Ricciardi, A., Steiner, W.W.M., Mack, R.N. & Simberloff, D. (2000). Toward a global information system for invasive species. *BioScience*, **50**, 239–244.

Ridgely, R.S., Allnutt, T.F., Brooks, T., McNicol, D.K., Mehlman, D.W., Young, B.E. & Zook, J.R. (2007). *Digital Distribution Maps of the Birds of the Western Hemisphere, version 3.0.* NatureServe, Arlington, Virginia, USA.

Rodda, G.H., Jarnevich, C.S. & Reed, R.N. (2011). Challenges in Identifying Sites Climatically Matched to the Native Ranges of Animal Invaders (A. Hector, Ed.). *PLoS ONE*, **6**, e14670.

Rödder, D. & Lötters, S. (2010). Explanative power of variables used in species distribution modelling: an issue of general model transferability or niche shift in the invasive Greenhouse frog (Eleutherodactylus planirostris). *Die Naturwissenschaften*, 781–796.

Rousseeuw, P.J. & van Zomeren, B.C. (1990). Unmasking Multivariate Outliers and Leverage Points. *Journal of the American Statistical Association*, **85**, 633–639.

Russello, M. a, Avery, M.L. & Wright, T.F. (2008). Genetic evidence links invasive monk parakeet populations in the United States to the international pet trade. *BMC evolutionary biology*, **8**, 217.

Sackett, L.C., Cross, T.B., Jones, R.T., Johnson, W.C., Ballare, K., Ray, C., Collinge, S.K. & Martin, A.P. (2011). Connectivity of prairie dog colonies in an altered landscape: inferences from analysis of microsatellite DNA variation. *Conservation Genetics*, **13**, 407–418.

Sanderson, E.W., Jaiteh, M., Levy, M. a., Redford, K.H., Wannebo, A. V. & Woolmer, G. (2002). The Human Footprint and the Last of the Wild. *BioScience*, **52**, 891.

Shirk, A.J., Wallin, D.O., Cushman, S. a, Rice, C.G. & Warheit, K.I. (2010). Inferring landscape effects on gene flow: a new model selection framework. *Molecular ecology*, **19**, 3603–19.

Smouse, P.E., Focardi, S., Moorcroft, P.R., Kie, J.G., Forester, J.D. & Morales, J.M. (2010). Stochastic modelling of animal movement. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **365**, 2201–11.

Smouse, P.E., Long, J.C. & Sokal, R.R. (1986). Multiple Regression and Correlation Extensions of the Mantel Test of Matrix Correspondence. *Systematic Zoology*, **35**, 627.

Soberón, J. & Nakamura, M. (2009). Niches and distributional areas: concepts, methods, and assumptions. *Proceedings of the National Academy of Sciences of the United States of America*, **106 Suppl** , 19644–50.

Soberon, J. & Peterson, A.T. (2005). Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity informatics*, **2**, 1–10.

South, J.M. & Pruett-Jones, S. (2000). Patterns of flock size, diet, and vigilance of naturalized Monk parakeets in Hyde Park, Chicago. *The Condor*, **102**, 848–854.

Spear, S.F., Balkenhol, N., Scribner, K.I.M., Fortin, M.-J. & McRae, B.H. (2010). Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analysis. *Molecular ecology*, 3576–3591.

Spreyer, M.F. & Bucher, E.H. (1998). Monk Parakeet (Myiopsitta monachus). *The birds of North America, No. 322* (eds A. Poole & F. Gill), Philadelphia, PA.

Stevenson, C.D., Ferryman, M., Nevin, O.T., Ramsey, A.D., Bailey, S. & Watts, K. (2013). Using GPS telemetry to validate least-cost modeling of gray squirrel (Sciurus carolinensis) movement within a fragmented landscape. *Ecology and Evolution*, n/a–n/a.

Storfer, A., Murphy, M. a, Evans, J.S., Goldberg, C.S., Robinson, S., Spear, S.F., Dezzani, R., Delmelle, E., Vierling, L. & Waits, L.P. (2007). Putting the "landscape" in landscape genetics. *Heredity*, **98**, 128–42.

Storfer, A., Murphy, M. a, Spear, S.F., Holderegger, R. & Waits, L.P. (2010). Landscape genetics: where are we now? *Molecular ecology*, **19**, 3496–514.

Street, E., Indsay, J.A.R.L., Power, F., Company, L., Boulevard, U. & Beach, J. (2007). Population viability analysis of monk parakeets in the United States and examination of alternative management strategies. *Wildlife Research*, **1**, 35–44.

Strubbe, D., Broennimann, O., Chiron, F. & Matthysen, E. (2013). Niche conservatism in non-native birds in Europe: niche unfilling rather than niche expansion. *Global Ecology and Biogeography*, n/a–n/a.

Strubbe, D. & Matthysen, E. (2009). Establishment success of invasive ring-necked and monk parakeets in Europe. *Journal of Biogeography*, **36**, 2264–2278.

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

Syfert, M.M., Smith, M.J. & Coomes, D. a. (2013). The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models. *PloS one*, **8**, e55158.

Tamara, K. & Arnheim, R. (1996). Perruches à collier (Psittacula krameri) victimes des conditions climatiques en région bruxelloise. *Aves*, **33**, 128–129.

Thomas, C.D. (2010). Climate, climate change and range boundaries. *Diversity and Distributions*, **16**, 488–495.

Tillman, E.A., Genchi, A.C., Lindsay, J.R., Newman, J.R. & Avery, M.L. (2004). Evaluation of Trapping to Reduce Monk Parakeet Populations at Electric Utility Facilities Electric Utility Facilities. *Proceedings 21st Vertebrate Pest Conference* (eds R.M. Timm & W.P. Gorenzel), pp. 126–129. University of California, Davis.

VanDerWal, J., Shoo, L.P., Graham, C. & Williams, S.E. (2009). Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know? *Ecological Modelling*, **220**, 589–594.

Varela, S., Rodríguez, J. & Lobo, J.M. (2009). Is current climatic equilibrium a guarantee for the transferability of distribution model predictions? A case study of the spotted hyena. *Journal of Biogeography*, **36**, 1645–1655.

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

Wang, J. (2011). Coancestry: a Program for Simulating, Estimating and Analysing Relatedness and Inbreeding Coefficients. *Molecular ecology resources*, **11**, 141–5.

Wang, I.J., Savage, W.K. & Bradley Shaffer, H. (2009). Landscape genetics and least-cost path analysis reveal unexpected dispersal routes in the California tiger salamander (Ambystoma californiense). *Molecular ecology*, **18**, 1365–74.

Waples, R.S. & Gaggiotti, O. (2006). What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular ecology*, **15**, 1419–39.

Warren, D.L. & Seifert, S.N. (2011). Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications*, **21**, 335–342.

Weathers, W.W. & Caccamise, D.F. (1975). Temperature regulation and water requirements of the monk parakeet, Myiopsitta monachus. *Oecologia*, **18**, 329–342.

Zapala, M. a & Schork, N.J. (2006). Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 19430–5.

Zar, J.H. (1998). *Biostatistical Analysis*. Prentice Hall, Englewood Cliffs, NJ.