FRONTIERS IN INVASIVE SPECIES DISTRIBUTION MODELING (iSDM): ASSESSING EFFECTS OF ABSENCE DATA, DISPERSAL CONSTRAINTS, STAGE OF INVASION AND SPATIAL DEPENDENCE

by

Tomáš Václavík

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Geography and Urban Regional Analysis

Charlotte

2011

Approved by:

____
Dr. Ross K. Meentemeyer

____
Dr. John Chadwick

____
Dr. Helene Hilger

____
Dr. John A. Kupfer

____
Dr. Wei-Ning Xiang

ABSTRACT

TOMÁŠ VÁCLAVÍK. Frontiers in invasive species distribution modeling (iSDM): Assessing effects of absence data, dispersal constraints, stage of invasion and spatial dependence. (Under the direction of DR. ROSS K. MEENTEMEYER)

Successful management of biological invasions depends heavily on our ability to predict their geographic ranges and potential habitats. Species distribution modeling (SDM) provides a methodological framework to predict spatial distributions of organisms but the unique aspects of modeling invasive species have been largely ignored in previous applications. Here, three unresolved challenges facing invasive species distribution modeling (iSDM) were examined in an effort to increase prediction accuracy and improve ecological understanding of actual and potential distributions of biological invasions. The effects of absence data and dispersal constraints, stage of invasion, and spatial dependence were assessed, using an extensive collection of field-based data on the invasive forest pathogen *Phytophthora ramorum*. Spatial analyses were based on a range of statistical techniques (generalized linear models, classification trees, maximum entropy, ecological niche factor analysis, multicriteria evaluation) and four groups of environmental parameters that varied in space and time: atmospheric moisture and temperature, topographic variability, abundance and susceptibility of host vegetation, and dispersal pressure. Results show that incorporating data on species absence and dispersal limitations is crucial not only to avoid overpredictions of the actual invaded range in a specific period of time but also for ecologically meaningful evaluation of iSDMs. When dispersal and colonization cannot be estimated explicitly, e.g. via dispersal kernels of propagule pressure, spatial dependence measured as spatial autocorrelation at multiple scales can serve as an important surrogate for dynamic processes that explain ecological mechanisms of invasion. If the goal is to identify habitats at potential risk of future spread, the stage of invasion should be considered because it represents the degree to which an organism is at equilibrium with its environment and limits the extent to which occurrence observations provide a sample of the species ecological niche. This research provides insight into several key principles of the SDM discipline, with implications for practical management of biological invasions.

DEDICATION

To my grandparents.

ACKNOWLEDGEMENTS

I would like to thank all our collaborators, researchers and field assistance from UNC Charlotte, UC Davis, UC Berkeley, Sonoma State University, Cambridge University, Oregon State University, Oregon Department of Forestry, USDA Forest Service, Phytosphere Research and the California Oak Mortality Task Force for their remarkable research on sudden oak death and incredible effort put into field work and data collection. I am deeply grateful to all my friends and colleagues from the Center for Applied GIScience and the Department of Geography and Earth Sciences at UNC Charlotte for their helpful comments, critiques and discussions about species distribution modeling and many other topics.

I am particularly grateful to my advisor Ross Meentemeyer for his enthusiasm, encouragement and guidance through the past several years. I also thank my doctoral committee, John Chadwick, Helene Hilger, John Kupfer, and Wei-Ning Xiang, for their insightful suggestions to improve this manuscript. Last but not least, I am immensely thankful to my wife Markéta for providing me with constant love and motivation; and my family for tolerating my rather long absence.

This research was supported by grants from the National Science Foundation (EF-0622677) and USDA Forest Service–Pacific Southwest Research Station (08-JV-11272138-055).

INTRODUCTION

Biological invasions of non-native species and emerging infectious diseases constitute one of the major threats to our biodiversity and natural ecosystems (Vitousek *et al.*, 1996; Daszak *et al.*, 2000). Globalization of human activities, drastic land transformation, and climate change facilitate the establishment and spread of invasive organisms both on regional and global scales (Mack *et al.*, 2000; Foley *et al.*, 2005; Svenning & Condit, 2008). The detrimental impacts of invasions generate unprecedented environmental and economic costs by impeding our efforts to sustain agricultural production and maintain healthy forest ecosystems (Pimentel *et al.*, 2000; Hoffmeister *et al.*, 2005). As the efficacy of invasion control and eradication treatments depend heavily on early detection and preventive strategies, there is great incentive to be able to accurately predict the potential and actual distribution of invasive species across landscapes (Ibanez *et al.*, 2009).

Species distribution modeling (SDM) based on characterizing the ecological niche of organisms has become a vital methodological framework for predicting species ranges and assessing impacts of human activities and changing environmental conditions on natural ecosystems. In principle, SDMs draw statistical inference about the environmental conditions associated with species occurrences and extrapolate this information to identify other geographic locations that posses similar characteristics (Franklin, 1995; Guisan & Zimmermann, 2000; Guisan & Thuiller, 2005). In the last years, substantial growth in the use of SDMs to predict biological invasions has occurred. Even though recent studies identified the most pressing challenges that need to be addressed for this science to move forward (Araujo & Guisan, 2006; Guisan *et al.*, 2006; Austin, 2007;

Franklin, 2010b), the specificities of modeling biological invasions have been largely ignored or inappropriately taken into account in many modeling efforts.

Invasive species distribution models (iSDMs) face special intertwined challenges because the ecological theory and assumptions underlying SDMs typically do not apply to invasive species or because issues such as spatial autocorrelation are inherent to all geographical data and modeling. The purpose of this dissertation research is to assess the implications of selected modeling challenges in the context of spatial analyses of biological invasions. To achieve this goal I utilized a unique collection of several extensive datasets on the occurrence of the invasive forest pathogen *Phytophthora ramorum* that causes the emerging infectious disease sudden oak death (SOD). I used multi-temporal and multi-scale data on this real example of a biological invasion but also a virtually simulated species to examine the effects of absence data, dispersal constraints, stage of invasion, and spatial dependence in iSDMs.

This dissertation is organized in such a way that each individual chapter addresses a particular modeling challenge and represents a stand-alone publication-style (or already published) article. Chapter 1 deals with the issue of species absence data that are often ignored in modeling efforts because they are unavailable or believed to be difficult to interpret in presumably suitable habitats. To overcome this obstacle, modelers use presence-only methods or generate pseudo-absences for use in traditional presence/absence approaches (Elith *et al.*, 2006; Chefaoui & Lobo, 2008). I examine the hypothesis that true absence data, when accompanied by dispersal constraints, improve the performance and ecological understanding of iSDMs that aim to predict the actual distribution of biological invasions. Using data on *P. ramorum* in California, I evaluate the impact of presence-only, true-absence, and pseudo-absence data on prediction accuracy in models that do or do not account for dispersal constraints estimated by measures of propagule pressure. Chapter 1 has been published in the journal *Ecological Modelling*.

Chapter 2 provides a critical insight into the standard working postulate of SDM that species are at equilibrium with their environment (Guisan & Thuiller, 2005). Previous studies have shown that such assumption may be incorrect even for native taxa (Pearson & Dawson, 2003; Svenning & Skov, 2004) but is especially violated by invasive species because their geographical ranges are restricted by dispersal and colonization processes. The stage of invasion, representing the degree of equilibrium, may limit the extent to which occurrence observations provide a sample of the species ecological niche. Therefore, I assume the stage of invasion will have a profound effect on the performance of iSDMs that aim to predict the potential distribution of biological invasions. Using data on both a real invasive organism *P. ramorum* and a simulated species *Phytophthora virtualis* in Oregon, I test the hypotheses that the accuracy of potential distribution predictions will be lower and the extent of predicted range smaller when models are calibrated with occurrence data from early stages of invasion.

Chapter 3 addresses the issue of different origins and scales of spatial autocorrelation (SAC) in biogeographical data that complicate the analyses of species distributions (Diniz-Filho *et al.*, 2003; Dormann, 2007b). SAC may be particularly important for iSDMs because biological invasions are strongly influenced by dispersal and colonization processes that typically create highly structured distribution patterns. Using a compilation dataset on the incidence of *P. ramorum* in the western United States, I examine the efficacy of a multi-scale framework that accounts for different origins of SAC and compares non-spatial models with models that account for SAC at multiple levels. In this chapter I show that, apart from being vital to avoid problems in multivariate statistical analyses, spatial pattern may be an important surrogate for dynamic processes that explain ecological mechanisms of invasion.

Chapter 4 represents an example of an iSDM application for prioritizing landscape contexts for early detection and eradication of invasion outbreaks. Two spatial predictive models of *P. ramorum* establishment and spread risk are developed for Oregon forest ecosystems. Models are based on three primary parameters that vary in space and time: atmospheric moisture and temperature, abundance and susceptibility of host vegetation, and dispersal pressure. First, a heuristic model is built using multi-criteria evaluation to identify large-scale areas at *potential* risk of pathogen invasion. Second, using field data for calibration, a machine-learning method, maximum entropy, is applied to predict the *actual* distribution of the epidemic. These spatially-explicit models of potential and actual distribution of *P. ramorum* invasion in Oregon provide a better picture of threatened forest resources across the state and are actively used by forest managers to guide detection surveys and eradication strategies. Chapter 4 has been published in the journal *Forest Ecology and Management*.

TABLE OF CONTENTS

CHAPTER 1: INVASIVE SPECIES DISTRIBUTION MODELING (ISDM). ARE ABSENCE DATA AND DISPERSAL CONSTRAINTS NEEDED TO PREDICT ACTUAL DISTRIBUTIONS? 1

1.1 Abstract 1

1.2 Introduction 2

1.3 Methods 7

1.3.1 Target species and presence/absence data 7

1.3.2 Environmental predictor variables 9

1.3.3 Dispersal constraints 10

1.3.4 Models 12

1.3.5 Presence-only models 12

1.3.6 Presence/absence models 14

1.3.7 Assessment of model performance 15

1.4 Results 17

1.5 Discussion 20

1.6 Conclusions 25

1.7 Acknowledgements 25

Tables 27

Figures 28

CHAPTER 2: EQUILIBRIUM OR NOT? MODELING POTENTIAL DISTRIBUTION OF INVASIVE SPECIES IN DIFFERENT STAGES OF INVASION 33

2.1 Abstract 33

2.2 Introduction 34

2.3 Methods 37

2.3.1 Field-based species data 37

2.3.2 Virtual species data 37

2.3.3 Environmental predictor variables 39

2.3.4 Building models of potential distribution 41

2.3.5 Evaluating effects of non-equilibrium 42

2.4 Results 43

2.5 Discussion 45

2.6 Acknowledgements 49

Figures 51

CHAPTER 3: ACCOUNTING FOR MULTI-SCALE SPATIAL AUTOCORRELATION IMPROVES PERFORMANCE OF INVASIVE SPECIES DISTRIBUTION MODELING (iSDM) 57

3.1 Abstract 57

3.2 Introduction 58

3.3 Methods 62

3.3.1 Study system – *Phytophthora ramorum* 62

3.3.2 Species data 63

3.3.3 Environmental predictors 64

3.3.4 Statistical methods 65

3.3.4.1 Spatially invariant models 65

3.3.4.2 Incorporating broad-scale spatial structure 66

3.3.4.3 Incorporating fine-scale spatial structure 67

3.3.4.4 Incorporating multi-scale spatial structure 68

3.3.5 Assessment of model performance 69

3.3.6 Investigation of SAC effects 69

3.4 Results 70

3.4.1 Model performance 70

3.4.2 Explanatory power and relative importance of variables 72

3.5 Discussion 73

3.5.1 Model performance and implications for iSDMs 73

3.5.2 Explanatory power and relative importance of variables 75

3.5.3 Method consideration 76

3.6 Conclusions 77

3.7 Acknowledgements 78

Tables 80

Figures 81

CHAPTER 4: PREDICTING POTENTIAL AND ACTUAL DISTRIBUTION OF SUDDEN OAK DEATH IN OREGON. PRIORITIZING LANDSCAPES FOR EARLY DETECTION AND ERADICATION 87

4.1 Abstract 87

4.2 Introduction 88

4.3 Methods 93

4.3.1 Field data collection 93

4.3.2 Host species mapping 94

4.3.3 Climate and topography surfaces 95

4.3.4 Model of potential distribution 96

4.3.4.1 Ranking host vegetation 96

4.3.4.2 Ranking climate factors 97

4.3.4.3 Developing heuristic model 98

4.3.5 Model of actual distribution 99

4.3.5.1 Developing statistical model 100

4.3.6 Evaluating the models 101

4.4 Results 102

4.4.1 Predicted geographic patterns of potential invasion 102

4.4.2 Predicted geographic patterns of actual invasion 105

4.4.3 Model evaluation 106

4.5 Discussion 106

4.6 Acknowledgements 112

Tables 113

Figures 116

CONCLUSIONS 120

REFERENCES 124

CHAPTER 1: INVASIVE SPECIES DISTRIBUTION MODELING (iSDM). ARE ABSENCE DATA AND DISPERSAL CONSTRAINTS NEEDED TO PREDICT ACTUAL DISTRIBUTIONS?

1.1 Abstract

Species distribution models (SDMs) based on statistical relationships between occurrence data and underlying environmental conditions are increasingly used to predict spatial patterns of biological invasions and prioritize locations for early detection and control of invasion outbreaks. However, invasive species distribution models (iSDMs) face special challenges because (i) they typically violate SDM's assumption that the organism is in equilibrium with its environment, and (ii) species absence data are often unavailable or believed to be too difficult to interpret. This often leads researchers to generate pseudo-absences for model training or utilize presence-only methods, and to confuse the distinction between predictions of potential vs. actual distribution. We examined the hypothesis that true absence data, when accompanied by dispersal constraints, improve prediction accuracy and ecological understanding of iSDMs that aim to predict the actual distribution of biological invasions. We evaluated the impact of presence-only, true-absence and pseudo-absence data on model accuracy using an extensive dataset on the distribution of the invasive forest pathogen *Phytophthora ramorum* in California. Two traditional presence/absence models (Generalized Linear Model and Classification Trees) and two alternative presence-only models (Ecological Niche Factor Analysis and Maximum Entropy) were developed based on 890 field plots of pathogen occurrence and several climatic, topographic, host vegetation and dispersal variables. The effects of all three possible types of occurrence data on model performance were evaluated with receiver operating characteristic (ROC) and omission/commission error rates. Results show that prediction of actual distribution was less accurate when we ignored true-absences and dispersal constraints. Presence-only models and models without dispersal information tended to over-predict the actual range of invasions. Models based on pseudo-absence data exhibited similar accuracies as presence-only models but produced spatially less feasible predictions. We suggest that true-absence data are a critical ingredient not only for accurate calibration but also for ecologically meaningful assessment of iSDMs that focus on predictions of actual distributions.

1.2 Introduction

Scientists have long sought a predictive understanding of the geographical distribution of ecological entities (species, populations, ecosystems). Species distribution models (SDMs) have provided a popular analytical framework for predicting species distributions by relating geo-located observations of occurrence to environmental variables that contribute to a species' survival and propagation (Franklin, 1995; Guisan & Zimmermann, 2000). This relation is based on statistically or theoretically derived response functions that characterize the environmental conditions associated with the ecological niche of a given organism (Austin, 2007). When applied in a geographic information system (GIS), SDMs can produce spatial predictions of occurrence likelihood at locations where information on species distribution was previously unavailable. Recent advancements in geospatial and statistical modeling methodologies along with growing availability of species data have enabled SDMs to increasingly tackle a range of pressing ecological problems, such as managing rare and endangered species and predicting species' responses to climate change and human modifications of habitat structure (Guisan & Thuiller, 2005). Due to globalization and extensive land transformations that facilitate the transfer and establishment of non-native organisms, SDM methods are also being increasingly used to predict spatial patterns of biological invasions and prioritize locations for early detection and control of invasion outbreaks (Peterson & Vieglais, 2001; Fonseca *et al.*, 2006; Lippitt *et al.*, 2008; Meentemeyer *et al.*, 2008a; Strubbe & Matthysen, 2009).

Invasive species distribution models (iSDMs) face two special challenges because the ecological theory and assumptions underlying SDMs typically do not apply to invasive species. The first challenge is that, by definition, the assumption of equilibrium between organisms and their environment is violated, and potential dispersal limitations of the invader are often ignored. As most SDMs implicitly rely on ecological niche concepts (Grinnell, 1917; Hutchinson, 1957), they assume that species occur at all locations where the environmental conditions are favorable and that dispersal is not a limiting factor (Jeschke & Strayer, 2008). However, invasive species are often absent at particular locations not because of low habitat quality but because the species has not dispersed to that site due to stochastic events, geographical barriers and dispersal constraints (Higgins *et al.*, 1999; Araujo & Pearson, 2005; Araujo & Guisan, 2006). Although dispersal limitations, more than biotic interactions, stochastic events or abiotic factors, are known to play a major role in the spread of invasions (Hastings *et al.*, 2005; Soberon & Peterson, 2005; Araujo & Guisan, 2006), few studies to date have tested empirically the benefits of including dispersal constraints in iSDMs (Meentemeyer *et al.*, 2008a).

The second challenge is that absence data are typically not used to develop or evaluate iSDMs. In practice, absence data are often cited as unavailable or they are ignored due to a perceived difficulty interpreting the meaning of absences at presumably suitable habitats. To overcome the obstacle of lacking data on species absence, a variety of presence-only profile techniques have been introduced and tested comprehensively for a number of native taxa (Segurado & Araujo, 2004; Elith *et al.*, 2006; Tsoar *et al.*, 2007). Nevertheless, application of presence-only techniques to iSDM is complex because the environmental space profiling tends to predict potential distribution of invasion rather than actual distribution (Guo et al., 2005; Jimenez-Valverde et al., 2008); and rigorous evaluation of distribution predictions is limited when the absence component is missing (Hirzel *et al.*, 2006). Alternatively, modelers often generate pseudo-absence data by sampling environmental conditions at locations where the organism is not recorded (Lutolf et al., 2006), but there is always the possibility of introducing false-negative errors into a model. To avoid collecting pseudo-absence data in potentially suitable locations where the species of interest may actually occur, methods have been proposed which utilize pseudo-absences that are heuristically determined to be outside the organism's ecological domain (Engler *et al.*, 2004; Chefaoui & Lobo, 2008). However, information on the absence of an organism at favorable sites can be useful in iSDMs when dispersal parameters are incorporated and the goal is to predict the actual distribution of an invader (Meentemeyer et al. 2008). A further limitation of the pseudo-absence approach is that pseudo-absence data are typically used in both model calibration and evaluation, thus verifying the goodness of fit of the training data, rather than the true predictive capability of the model (Zaniewski *et al.*, 2002; Engler *et al.*, 2004; Lutolf *et al.*, 2006; Chefaoui & Lobo, 2008). To our knowledge, the assumptions of using presence-only and pseudo-absence data in iSDMs have never been tested with extensive true-absence data; such information is needed to advance ecological conceptualization of SDMs for biological invasions.

As a consequence of ignoring equilibrium assumptions and true-absence data in SDMs, we believe that the conceptualization of the potential versus actual distribution is often confused in the practice of species distribution modeling in general, but especially for biological invasions (Soberon, 2007; Hirzel & Le Lay, 2008; Jimenez-Valverde *et al.*, 2008; Peterson *et al.*, 2008; Phillips, 2008). Here, we emphasize that a clear distinction should be drawn between the potential and actual distribution in the iSDM framework. While the potential distribution is a hypothetical concept that refers to locations where an invader could exist based on suitable environmental factors, the actual distribution refers to locations where the invader actually exists at a specific time, as constrained by environmental and dispersal limitations. This distinction is relevant because SDMs of invasive organisms often assume the potential distribution is being modeled (Peterson *et al.*, 2003; Davis, 2004; Guo *et al.*, 2005; Chen *et al.*, 2007; Giovanelli *et al.*, 2008; Lopez-Darias *et al.*, 2008; Rodder *et al.*, 2008; Strubbe & Matthysen, 2009), although it has been argued that all SDMs *de facto* quantify the actual distribution, as calibration data represent samples of the current range constrained by biotic, geographic and dispersal limitations (Guisan & Thuiller, 2005; Phillips *et al.*, 2006). The applicability of models that aim to predict potential distribution of invasions is wide, including projections of geographical distribution of species under climate change (Berry et al., 2002; Thomas et al., 2004; Pearson, 2006a; Engler et al., 2009) or understanding the behavior of invaders in novel landscapes (Peterson, 2003; Peterson *et al.*, 2003; Sutherst & Bourne, 2009). However, a growing number of publications used SDMs to predict the actual distribution of biological invasions (e.g., Havel *et al.*, 2002; Meentemeyer *et al.*, 2008a). The issue of iSDM became an interesting frontier in ecological modeling due to its ability to predict extant consequences of an invasion at unsampled locations. Here, we use the framework defined by Meentemeyer et al. (2008a) and apply iSDMs to model the actual invasive distribution which can be used to target locations for early detection surveillance and invasion control, and to quantify the current extent of invasion spread.

In this study, we examine the hypothesis that true-absence data, when accompanied by dispersal information, improves the accuracy and ecological meaning of models designed to predict the actual distribution of a biological invasion. We use an extensive dataset on the occurrence of the invasive forest pathogen *Phytophthora ramorum* in California to evaluate two questions that address the impact of ignoring absence data and dispersal in iSDMs: (1) Do models calibrated with presence-only, true-absence or pseudo-absence data significantly differ in their performance? (2) Does incorporation of dispersal constraints improve model accuracy? We focus on the capability of iSDMs to predict the actual distribution of invasion because we believe it provides the best analytical framework for early detection and control of invasion outbreaks; and because predictions of actual distribution can be assessed using presence/absence observation data, whereas predictions of potential distribution cannot. To assess how the choice of different types of occurrence data affects prediction accuracy, we compared the performance of two common presence/absence modeling methods (using both true-absence and randomly generated pseudo-absence data) with two common presence-only methods. We further assessed the degree to which incorporating 'force of invasion' dispersal kernels influences performance of each model type (Hastings *et al.*, 2005; Allouche *et al.*, 2008; Meentemeyer *et al.*, 2008a). All models were evaluated based on presence and true-absence data using *k*-fold cross-validation, area under the curve (AUC), and commission/omission error rates. Research addressing the effects of including absence data and dispersal constraints on model performance is needed to improve spatial predictions of biological invasions and advance ecological conceptualization of species distribution modeling.

1.3 Methods

1.3.1 Target species and presence/absence data

We focused on modeling the actual distribution of the invasive pathogen *Phytophthora ramorum*, a generalist pathogen (Oomycota) causing the emerging infectious forest disease known as sudden oak death. Since its introduction in 1990s, the pathogen has reached epidemic levels in coastal forests of California and south-western Oregon, killing large numbers of oak (*Quercus* sp.) and tanoak (*Lithocarpus densiflorus*) trees (Rizzo & Garbelotto, 2003). The disease is thought to be primarily transmitted via infective spores formed on the leaves of foliar hosts, such as the evergreen tree bay laurel (*Umbellularia californica*), which are passively dispersed to nearby individuals via rain splash and from stand to stand via wind-blown rain (Rizzo & Garbelotto, 2003; Davidson *et al.*, 2005). To date, spread of the pathogen has been patchily distributed across approximately 10% of its geographical host range in California (Meentemeyer et al.

2008) with considerable forest area facing risk of infection due to widespread host availability and presumably suitable habitat conditions (Rizzo et al., 2005). A predictive understanding of *P. ramorum* distribution is needed to prioritize locations for early detection and control of invasion (Rizzo *et al.*, 2005; Meentemeyer *et al.*, 2008a). *P. ramorum* is an ideal target organism for our modeling purpose in this study because it is actively invading native habitats, it is moderately dispersal limited, and there are numerous susceptible habitats in California that are both close and far in distance to known sources of inoculum.

To obtain reliable occurrence data for calibration and assessment of our predictive models, we surveyed 890 early-detection field plots for the presence and absence of *P. ramorum* over the summers of 2003, 2004, and 2005 (described in Meentemeyer *et al.*, 2008a). Field plot locations were distributed in a stratified-random manner across five levels of habitat suitability defined by Meentemeyer et al. (2004), with variable proximities to infected sites previously confirmed by the California Department of Food and Agriculture (CDFA). A minimum distance of 400 m between individual plots was enforced to avoid sampling within the scale at which the disease is known to be clustered (Kelly & Meentemeyer, 2002).

At each plot location, we established two 50 × 10 m "L-shaped" transects to determine the occurrence of *P. ramorum*. Along each transect up to 25 necrotic leaves were collected from five of the most visually symptomatic individuals from over a dozen foliar host species (Meentemeyer et al. 2008). Symptomatic samples were processed and cultured in the laboratory on a selective media for *Phytophthora* species (Hayden et al., 2004) and as an additional test any negative cases were resampled with a polymerase chain reaction (PCR)-based molecular assay, using primers designed to amplify *P. ramorum* DNA (Ivors et al., 2004). The pathogen was only considered absent at a location if there was no positive culture isolation and no PCR detection of pathogen DNA in the leaf samples. This sampling design enabled the collection and discrimination of reliable presence (n=78) and true-absence (n=812) data on *P. ramorum* invasion across the entire state of California.

To examine the effect of pseudo-absence data on model performance, we randomly selected 812 pseudo-absence locations from the same range of susceptible host vegetation as used for the real plot data described above, not allowing the locations to occur within 400 m of one another and the plots (Fig. 1). We generated the same number of pseudo-absences as true-absences to avoid potential bias caused by different levels of prevalence in the presence/absence datasets (Manel et al., 2001). Although some studies suggest that pseudo-absence data should be limited to areas with clearly unsuitable environmental conditions (Zaniewski et al., 2002; Engler et al., 2004), invasive species are inherently absent at many environmentally favorable locations (Pulliam, 2000; Austin, 2002). Therefore, we purposely distributed pseudo-absence data across all levels of environmental suitability in an effort to produce models reflecting the actual distribution of the invasion.

1.3.2 Environmental predictor variables

We calculated a set of eight environmental variables that we hypothesized would predict the actual distribution of *P. ramorum* in California. To characterize moisture and temperature conditions known to affect foliar plant pathogens (Woods et al., 2005), we derived four climate variables from the parameter elevation regression on independent slopes model (PRISM; Daly et al., 2001) at 800 m spatial resolution. Maximum and minimum temperature, precipitation and relative humidity were aggregated to provide 30-year monthly average values between December to May, the reproductive season for *P. ramorum* in California (Davidson et al., 2005). We also mapped elevation and derived two topographic variables, solar insolation index (SII) and topographic moisture index (TMI), using a U.S. Geological Survey 90-m digital elevation model. The SII was calculated for each cell as the potential mean solar radiation in the rainy season using the cosine of illumination angle on slope equation (Dubayah, 1994). The TMI was calculated as the natural log of the ratio between the upslope contributing drainage area and the slope gradient of a grid cell (Moore et al., 1991). Finally, we mapped the spatial distribution of the key infectious host bay laurel (*Umbellularia californica*) using data summarized in Meentemeyer et al. (2004). This species is considered to be the most epidemiologically important host for *P. ramorum* because it produces large amounts of inoculum (Davidson et al., 2005; Anacker et al., 2008) and it is associated with oak and tanoak mortality (Kelly & Meentemeyer, 2002; Maloney *et al.*, 2005).

1.3.3 Dispersal constraints

To incorporate the effect of dispersal constraints on the actual distribution of *P. ramorum*, we quantified the potential force of invasion on each field plot (Hastings *et al.*, 2005; Meentemeyer *et al.*, 2008a) and included it as an additional predictor variable into the models. The force of invasion (F_i) was calculated as a negative exponential dispersal kernel:

$$F_i = \sum_{k=1}^{N} \exp\left(\frac{-d_{ik}}{a}\right) \tag{1}$$

where d_{ik} is the Euclidean distance between each potential source of invasion k and target plot i. The parameter a modifies the form of the dispersal kernel where low values of a indicate high dispersal limitation and high values of a indicate low dispersal limitation (Havel *et al.*, 2002; Meentemeyer *et al.*, 2008a). The optimal value of a was selected based on the goodness of fit of the best generalized linear model based on true-presence/true-absence data, to which F_i with varied values of a was iteratively added (Meentemeyer *et al.*, 2008a). We used the negative exponential dispersal kernel because previous research has shown that this kernel adequately describes dispersal characteristics of rain-splash dispersed plant pathogens (McCartney & Fitt, 1985; Fitt *et al.*, 1989).

Empirically calculating negative exponential dispersal kernel from distribution data is a common method to represent force of invasion in models of spatial spread of invasions (Havel *et al.*, 2002; Hastings *et al.*, 2005; Meentemeyer *et al.*, 2008a). However, it can be used only when data allow it. Since true-absence species data are required to fit the optimal form of the dispersal kernel, the negative exponential dispersal kernel was only applied in true-absence data models. For the presence-only and pseudo-absence data models, we implemented prevailing best practice conditions and necessarily used a simplified version of force of invasion according to a method suggested by Allouche et al. (2008). Here, we calculated a cumulative distance metric that incorporates dispersal limitations in iSDMs without explicitly estimating the dispersal characteristics of the organism (Allouche et al., 2008). The cumulative distance (D_i) sums the inverse of the squared Euclidean distances d_{ik} between each potential source of invasion k and target plot i:

$$D_i = \sum_{k=1}^{N} \left(\frac{1}{(d_{ik})^2} \right) \qquad (2)$$

We calculated both force of invasion terms based on negative exponential dispersal kernel and inverse cumulative distance using the distance from our early-detection sample plots to all sources of inoculum confirmed by the California Department of Food and Agriculture in 2005. These reference data maintained by the California Oak Mortality Task Force (COMTF; Kelly & Tuxen, 2003) are independent from our sample plots used to calibrate the models.

1.3.4 Models

We used four commonly applied modeling methods to evaluate the impact of presence-only, true-absence and pseudo-absence data on prediction of the actual distribution of *P. ramorum* in California. For each of the three data assumption types, we used both parametric and non-parametric techniques to model the relative likelihood of pathogen occurrence, in order to account for variations between different algorithm families (Elith & Burgman, 2003; Elith *et al.*, 2006). To evaluate each model under normal practice conditions, model calibration and variable selection were conducted on an individual basis. To test the importance of dispersal limitation, we developed models based on: (i) the environmental variables only, and (ii) the combination of environmental variables and dispersal constraints (hybrid models).

1.3.5 Presence-only models

1.3.5.1 Ecological niche factor analysis (ENFA)

In the multidimensional space of ecological variables, ENFA compares the distribution of locations where the focal species was identified to a reference set describing the whole study area (Hirzel et al., 2002). Similar to principal component analysis (PCA), it computes uncorrelated factors that explain a major part of the ecological distribution of the species. Two types of factors with biological significance are extracted: (i) marginality describes how the species optimum differs from the global mean of environmental conditions in the study area; (ii) specialization (tolerance) factors sorted by decreasing amount of explained variance describe how species variance compares to the global variance. Using the BIOMAPPER software (Hirzel et al., 2007) version 4.0, we calculated correlations between variables prior ENFA analyses and removed predictors with correlation coefficients greater than 0.5. The number of retained factors was determined based on their eigenvalues compared to the "broken-stick" distribution (McArthur, 1957), and ranged between 2 and 4 factors with 91-95% of explained variability. We computed the final prediction maps using the Medians algorithm. Recommended Box-Cox transformation of predictor variables produced poorer results than raw data and was thus not used in the final models.

1.3.5.2 Maximum entropy (MAXENT)

MAXENT is a machine-learning method that estimates distributions of organisms by finding the probability distribution of maximum entropy (i.e., the most uniform) given the constraint that the expected value of each environmental predictor under this estimated distribution matches the empirical average of sample locations (Phillips et al., 2006). We iteratively weighted each environmental variable to maximize the likelihood to reach the optimum probability distribution, and then divided it by a scaling constant to ensure a predicted range between 0 and 1 (Elith & Burgman, 2003). We utilized the MAXENT software version 3.2.1 using a maximum of 500 iterations and the logistic output, and employing the regularization procedure in order to compensate for the tendency of the algorithm to overfit calibration data (Phillips et al., 2006)

1.3.6 Presence/absence models

1.3.6.1 Generalized linear model (GLM)

GLM is an extension of common multiple regression that allows for modeling non-normal response variables (McCullagh & Nelder, 1989). Most frequently used for SDM is the logistic model that employs a maximum likelihood parameter optimization technique to model the log odds of a binary response variable (Franklin, 1995; Miller, 2005). Using both true-absence and pseudo-absence species data, we fitted all models in JMP 7.0 (SAS Institute Inc., Cary, NC) specifying a binomial error distribution and logit-link function. The logit transformation of the probability (p_i) that a susceptible plot becomes invaded was calculated as:

$$\operatorname{logit}(p_i) = \log \frac{p_i}{1 - p_i} = \beta_o + \sum_{j=1}^{8} \beta_j x_j + \beta F_i \qquad (3)$$

where β is the regression coefficient, x_1, x_2, ...x_8 are the set of environmental variables, and F_i is the force of invasion. We tested all possible subsets of variables using the combination of manual selection and stepwise regression with p-to-enter and/or p-to-remove equal to 0.05 and 0.10. The best model selection was conducted based on logit R^2 (also known as the uncertainty coefficient U) and negative log-likelihood ratio test (LRT) (Johnson & Omland, 2004). We focused on LRT over the Akaike's information criterion (AIC) because previous SDM studies showed that LRT outperformed AIC, producing more parsimonious models (Maggini et al., 2006; Austin, 2007). Pairwise interaction terms were also tested for significance; higher order combinations of variables were not explored.

1.3.6.2 Classification trees (CT)

CT is a non-parametric, data-driven method that recursively partitions data into homogeneous groups based on identification of a specific threshold for each environmental predictor variable (Franklin, 1995; De'ath & Fabricius, 2000; Miller & Franklin, 2002). We produced a tree of hierarchical decision rules using IDRISI 15 (The Andes Edition, Clark Labs/Clark University, 2006, Worcester, MA) to split data into “mostly present” and “mostly absent” classes using both true-absence and pseudo-absence species data. We used the Gini splitting rule that measures the impurity of pixels at a given node and thus attempts to find the largest homogeneous class and isolate it from the rest of the dataset (Eastman, 2006). To avoid the likely overfit of calibration data, we auto-pruned the final tree, eliminating leaves with pixel counts less or equal to 3%. The proportion of observations correctly classified at each terminal node represents the approximate degree of membership of unsampled data associated with the same ecological factors defined by the node (Miller, 2005). This degree of membership is then analogous to the probability of occurrence defined by, e.g., a GLM model.

1.3.7 Assessment of model performance

For each of the four methods, we assessed spatial predictions of *P. ramorum* actual distribution with true-presence/true-absence data, using *k*-fold cross-validation technique, area under the curve (AUC) of the receiver operating characteristic (ROC), and simple threshold assessment based on the commission/omission errors minimizer. Although some SDM studies in the past applied resubstitution techniques (for review see, e.g., Araujo et al., 2005), in which the same data used for calibration are used to verify the models, an independent evaluation or data splitting is recommended to ensure a degree of independence from the events used to make the predictions (Guisan & Zimmermann, 2000; Araujo & Guisan, 2006; Jeschke & Strayer, 2008). We employed *k*-fold cross-validation, dividing the occurrence dataset into *k* independent partitions, using *k*-1 for model calibration and the left-out partition to evaluate the models with AUC, while repeating this procedure *k* times (Hirzel *et al.*, 2006). Having a large dataset (n=890) and 9 predictor variables, we used the heuristic recommended by Fielding and Bell (1997) that approximates the training (calibration) dataset to consist of 75% of samples, i.e., *k*=4.

For each model, we calculated AUC of the ROC function to provide a threshold and prevalence independent measure of models' performance (Fielding & Bell, 1997). ROC compares a rank map of predicted species occurrence against a boolean map of true occurrence and plots the true positive rate (sensitivity) as a function of false positive rate (1-specificity or commission error) at each possible threshold (Pontius & Schneider, 2001). The area under the plotted line is the AUC statistic that provides a single discrimination measure, equivalent to the non-parametric Wilcoxon test, across all possible ranges of thresholds (Lobo et al., 2008). In order to avoid rank ordering that can lead to locations of the same likelihood value being calculated at different thresholds and thus introducing potential bias in the ROC curve (Lippitt et al., 2008; Lobo et al., 2008), we also used simple threshold assessment based on model efficiency (Jimenez-Valverde & Lobo, 2007; Freeman & Moisen, 2008). Assuming equal weights being placed on presences and absences in iSDM, the only correct threshold needed to efficiently transform predicted probabilities to binary presence/absence predictions is the one that minimizes the difference between commission and omission error rates. We calculated the error minimizer for each possible threshold *i* as:

$$\text{Error minimizer} = \text{Min}[x_i - y_i] \qquad (4)$$

where x_i is the commission error rate at threshold *i* and y_i is the omission error rate at threshold *i*. Neither commission nor omission errors were preferred because the aim was to model the actual distribution for the purpose of prioritizing areas for early detection and eradication, and to evaluate practicable current impacts rather than hypothetical potential surfaces. If the omission error rate was high, model prediction would result in overly conservative scenario, where positive sites go undetected. If the commission error rate was high, even marginally suitable areas far from current sources of infection would be predicted, resulting in increased costs of needless sampling and eradication efforts in the field (Meentemeyer *et al.*, 2008a). In addition to commission/omission error rates, we report the total area predicted by each model to illuminate potential over- or under-prediction of actual distribution. Finally, we assessed all models developed with pseudo-absence locations using both true- and pseudo-absence data to investigate the degree of uncertainty introduced in the evaluation process when true-absence data are ignored.

1.4 Results

Application of each of the twelve models in the GIS produced probability maps of actual *P. ramorum* distribution in 2005 (Fig.2). The mean and variability of AUC values obtained via cross-validation with true-presence/true-absence data showed marked differences in models' performances (Fig. 3). The most accurate models were GLM (AUC=0.90) and CT (AUC=0.89) based on presence/true-absence data with a combination of both environmental factors and dispersal constraints. The least accurate were CT models based on environment-only factors with true-absence (AUC=0.73) and pseudo-absence data (AUC=0.65); all other models exhibited accuracies over 0.78 of the AUC statistic. Conversion of the continuous probability maps to a binomial distribution of predicted presence/absence also shows that models using true-absences with dispersal constraints were the most efficient: CT (commission/omission error rate=0.135 at 0.051 threshold) and GLM (commission/omission error rate=0.192 at 0.206 threshold) (Table 1). The highest error rates resulted from models based on pseudo-absences with environment-only variables: CT (commission/omission error rate=0.346 at 0.034 threshold) and GLM (commission/omission error rate=0.308 at 0.161 threshold). In addition, models that used true-absences for calibration had lower variability of AUC from cross-validation results (e.g. SD=0.018 for GLM with dispersal constraints) than models based on presence-only data or pseudo-absences (e.g. SD=0.083 for ENFA; SD=0.089 for CT).

Incorporating dispersal constraints significantly increased the explanatory capacity of most models. Hybrid models were always more accurate than their corresponding environment-only equivalents, with the exception of GLM based on pseudo-absence data where the cumulative distance was not significant in any of the cross-validation runs and therefore not used for final prediction. However, the effect of dispersal constraints varied considerably for different types of modeling groups. When dispersal constraints were omitted, the overall accuracy of modeling groups decreased in the following order: presence-only models, presence/true-absence models, presence/pseudo-absence models. However, the presence-only models, on average, outperformed the models based on presence/absence data because of the good performance by MAXENT (AUC=0.85; commission/omission error rate=0.231 at 0.357 threshold), while ENFA had AUC=0.78 and poorer efficiency (commission/omission error rate=0.290 at 0.390 threshold) than both models using true-absences. In contrast, when dispersal constraints were taken into account, the predictive capacity of both models with true-absences improved from AUC of 0.73 to 0.89 (CT) and from 0.82 to 0.90 (GLM), and thus outperformed all models with presence-only and presence/pseudo-absence data.

Despite the differences in assessment results among different modeling methods, the general pattern of *P. ramorum* prediction was relatively consistent, exhibiting large areas of location agreement (Fig. 2). In general, presence-only models predicted larger areas of invasion than both presence-absence groups of models (Table 1), especially because of the high over-prediction of ENFA (13 678 km^2). Incorporating dispersal constraints resulted in a marked reduction of the predicted area for most models, with the exception of GLM based on pseudo-absences, in which dispersal constraints were insignificant, and CT based on true-absences, in which a slight increase in area was observed.

Finally, we found striking differences in assessment results when models developed with pseudo-absence data were cross-validated with pseudo-absence data, a commonly used modeling practice when true-absences are unavailable (Fig. 4; Table 1). In this assessment, the mean AUC values for GLM models increased from 0.80 to 0.90 and the error rate for thresholded predictions decreased from 0.308 to 0.180. Moreover, the variability of individual cross-validation runs decreased in contrast to those where true-absence data were used (decrease in SD=0.034). Similar results emerged for CT models; especially the environment-only CT model exhibited accrual in AUC from 0.65 to 0.81, reduction in error rate from 0.346 to 0.204, and decrease in variability of cross-validation runs (decrease in SD=0.012).

1.5 Discussion

In this study, we analyzed a unique set of survey data on the invasive forest pathogen *Phytophthora ramorum* to address the question whether true-absence data and dispersal constraints are needed to accurately predict the actual distribution of biological invasions. Our results demonstrated that the most accurate and efficient models were those that incorporated true-absence data in environmental models augmented by dispersal constraints. These findings support our hypothesis that the actual distribution of invasive species should be modeled using reliable presence/absence data and incorporating distribution restriction factors, such as dispersal limitations.

The primacy of models based on presence and true-absence data were consistent for all modeling algorithms if dispersal constraints were included. Contrary to our expectations, the results were not as clear for models when dispersal was omitted. Although we would expect both presence-only models to largely over-predict the actual range, MAXENT produced more accurate predictions than both true-absence models when force of invasion was not included. We suggest three possible explanations. First, the reason may be inherent to modeling algorithms of the presence-only models. Comparative studies confirmed excellent performance of MAXENT with small sample sizes and its tendency towards restricted predictions, while ENFA is prone to over-estimate species distributions (Zaniewski et al., 2002; Engler et al., 2004; Elith et al., 2006). Second, dispersal constraints appear to play a larger role in confining predictions than absence data alone. For instance, Allouche et al. (2008) demonstrated that, in some cases, models based on mere distance constraints may produce more accurate results than environment-based models. Third, presence-only models might have produced larger over-predictions if the target organism was in a later stage of invasion. The stage of invasion affects the extent to which species observations provide a sample of the ecological domain of the species (Araujo & Pearson, 2005; Pearson, 2006b). Since *P. ramorum* was introduced to California in the early 1990s and is still spreading, the field data from 2003-2005 likely provide a poor representation of all the conditions suitable for the pathogen, and thus fitted models project only a small portion of its ecological domain in geographical space.

Integration of dispersal constraints in the modeling process enhanced the performance of all models with the exception of GLM based on pseudo-absences, in which the force of invasion was statistically insignificant (p>0.05). The improvement for all types of modeling approaches indicates that the importance of dispersal limitations is not unique to a specific algorithm examined in this study. Dispersal constraints thus represent an important component in iSDMs accounting for limitations that prevent invasive species from colonizing places environmentally suitable but isolated or remote from already invaded locations (Allouche et al., 2008). The force of invasion term has been shown to not only improve the accuracy of spatially-explicit iSDMs but also illuminate the dispersal characteristics of the organism (Meentemeyer *et al.*, 2008a). For *P. ramorum*, the estimated dispersal kernel (a=58) indicated a moderate dispersal limitation. Such finding is consistent with studies that described the transfer of *P. ramorum* spores via rain splash and wind as highly localized (up to 10 m from the forest edge) (Davidson et al., 2005), although long-distance dispersal events during storms or facilitated by humans or vertebrates are possible (Rizzo *et al.*, 2005; Cushman & Meentemeyer, 2008). However, the optimization of a dispersal kernel for a specific organism requires true-presence and true-absence locations. Here, we demonstrate that when true-absence data are unavailable or ignored, parameterization of this force of invasion is prevented. The use of non-parameterized, distance-based functions, such as inverse squared cumulative distance, represents a possible alternative when true-absence data are lacking. This term does not account explicitly for species-specific dispersal characteristics but provides a mean of accounting for spatially autocorrelated factors that are not included as predictors in the models (Allouche et al., 2008). If the purpose of this research was to assess the performance of different modeling algorithms, the use of the same (non-parameterized) dispersal constraint for all models would provide more meaningful comparison. Since the purpose of our study was to compare different modeling strategies (with and without true-absence data), rather than modeling algorithms, we implemented prevailing best practice conditions and thus included the optimized dispersal kernel when data allowed it; otherwise the predictive capability of the presence-absence strategy would be artificially decreased. However, if potential bias in final predictions is to be avoided, it is highly desirable to use data completely independent from calibration and evaluation datasets to calculate both types of dispersal constraints. In addition, it is important to note that both types of dispersal constraints used in the study describe force of invasion based on distance metrics but do not explicitly integrate the effect of barriers or connectivity of landscape features on species dispersal.

Based on the accuracy statistics for pseudo-absence models comparable to those documented for ENFA, random selection of pseudo-absence data may be a valid approach for iSDMs when true-absence data are unavailable. Although previous studies suggested that more reliable pseudo-absence data can be derived from areas with unsuitable environmental conditions identified with the use of profile (presence-only) techniques (Zaniewski et al., 2002; Engler et al., 2004; Lutolf et al., 2006), this approach may only be appropriate under equilibrium conditions or when the goal is to model the potential distribution of the focal organism (Svenning & Skov, 2004; Hirzel & Le Lay, 2008). Random selection of pseudo-absence data from geographical spaces that are both near and distant to the ecological domain of the organism produce the most constrained prediction that is closer to the actual distribution (Thuiller *et al.*, 2004; Chefaoui & Lobo, 2008). If the goal is to achieve predictions closer to the potential distribution, not only should pseudo-absence data be selected from locations with unsuitable conditions, but also dispersal constraints should be omitted, or profile techniques used, in order to avoid inevitable reduction of the predicted range (Svenning & Skov, 2004; Hirzel & Le Lay, 2008; Lobo *et al.*, 2008). However, the potential distribution is a hypothetical concept and cannot be rigorously assessed with the use of observational presence/absence data.

Although critical issues about AUC have been recently brought to attention in the species modeling context, the ROC function remains a highly reliable technique for SDMs' assessment, if it is used to compare models for the same species at the same extent, and the measures of commission and omission errors and total predicted area are considered (Lobo et al., 2008; Peterson et al., 2008). However, the weakness of single-number accuracy measures is that they do not provide information on the spatial arrangement of correctly and incorrectly predicted occurrences (Pontius & Schneider, 2001; Lobo *et al.*, 2008). Verification of predicted pattern in final maps can render additional information about models performances. In this study, all maps showed pathogen's invasion consistently concentrated along the western coast of California. In general, predictions of models with dispersal constraints were more confined to the San Francisco Bay Area, Santa Cruz County and in Humboldt County. Models developed without dispersal constraints exhibited more dispersed ranges. ENFA and GLM predicted large areas of invasions along the northern coast of California in Mendocino and Humboldt Counties, where the invasion of sudden oak death has been documented (COMTF; Kelly & Tuxen, 2003). The GLM and CT models based on pseudo-absences predicted invasions along the southern coast in Santa Barbara, Ventura, Los Angeles, Orange and San Diego Counties, more than 500 km from the nearest documented invasion (COMTF; Kelly & Tuxen, 2003). This finding suggests that notwithstanding the similar accuracies of presence-only and pseudo-absence methods, the latter produced spatially less feasible predictions due to incorrect parameterization based on the spatial distribution of pseudo-absence data.

Although our analysis indicated that true-absences in combination with dispersal constraints enhance the performance of iSDMs, the acquisition of true-absence data may be desirable not only for model development. When models based on pseudo-absences were assessed with pseudo-absences, according to a common practice in SDM research (Zaniewski *et al.*, 2002; Engler *et al.*, 2004; Lutolf *et al.*, 2006; Chefaoui & Lobo, 2008), they appeared to be significantly more accurate and stable than when true-absences were used for evaluation (difference in AUC= 0.16 for environment-only CT and 0.11 for

GLM). If true-absences are missing, the accuracy measures can only indicate how well models discriminate data considered in the training process but reveals little about the real prediction capability. Therefore, we suggest that true-absence data are a critical ingredient not only for accurate calibration but also ecologically meaningful assessment of iSDMs that focus on predictions of actual distributions.

1.6 Conclusions

Despite the growing use of SDMs to predict current spatial patterns of biological invasions, the implications of ignoring absence data and dispersal limitations in iSDMs have been rarely taken into account. In this study, we assessed the effects of different types of occurrence data and incorporation of dispersal constraints on the accuracy of models predicting the actual distribution of the invasive pathogen *P. ramorum* in California. We provide empirical evidence that predictive models calibrated with true-absence data and augmented with dispersal information significantly improve their performance, and that true-absence data are also critically needed to meaningfully assess invasion predictions. Our results contribute to the broad ecological understanding and conceptualization of iSDMs and illustrate the procedures needed to increase the efficacy of spatial predictions of invasive organisms. If iSDMs should serve as effective tools for early-detection and management of invasive species in conservation practice, their accuracy and correct interpretation is crucial to minimize the ecological impact and economic cost of biological invasions.

1.7 Acknowledgements

We thank J. Amaris, C. Boylen, K. Buja, M. Cooper, G. Garner, A. Gauthier, E. Gordon, R. Hunter, A. Jirka, M. Kennedy, S. King, M. Kozanitas, E. McCowan, L.

Miller, S. Moyle, G. Neargarder, E. Sabo, H. Torpin, K. Ward, and K. Wells for assistance with field work; L. Bell, E. Lotz, and L. Fischer for logistical support in the field. We also thank C. Blomquist (CDFA) and J. Bienapfl (UC Davis) for pathogen identification, B. Anacker, W. Mark, and D. Rizzo for their previous work that made this study possible, and two anonymous reviewers as well as all members of the Center for Applied GIScience at UNC Charlotte for their helpful comments on the manuscript. This research was supported by grants from the National Science Foundation (EF-0622677) and USDA Forest Service (State and Private Forestry–Pacific Southwest Region).

Tables

Table 1: Simple threshold assessment for the most efficient models showing: the best threshold, minimized commission/omission error rate for assessment with true-absences, error rate for pseudo-absence models assessed with pseudo-absences, and the total area predicted as presence.

		------------ with dispersal constraints ------------				------------- environment-only models -------------			
Model group	**Model**	**Threshold**	**Error rate**	**Error rate (with PsAbs)**	**Area (km^2)**	**Threshold**	**Error rate**	**Error rate (with PsAbs)**	**Area (km^2)**
Presence-only	ENFA	0.250	0.270	-	8060	0.390	0.290	-	13678
	MAXENT	0.343	0.207	-	4388	0.357	0.231	-	5285
True-absence	GLM	0.206	0.192	-	4471	0.160	0.267	-	8861
	CT	0.051	0.135	-	4421	0.086	0.272	-	3724
Pseudo-absence	GLM-PsAbs	0.161	0.308	0.180	4925	0.161	0.308	0.180	4925
	CT-PsAbs	0.034	0.230	0.204	3263	0.034	0.346	0.204	8322

Figures

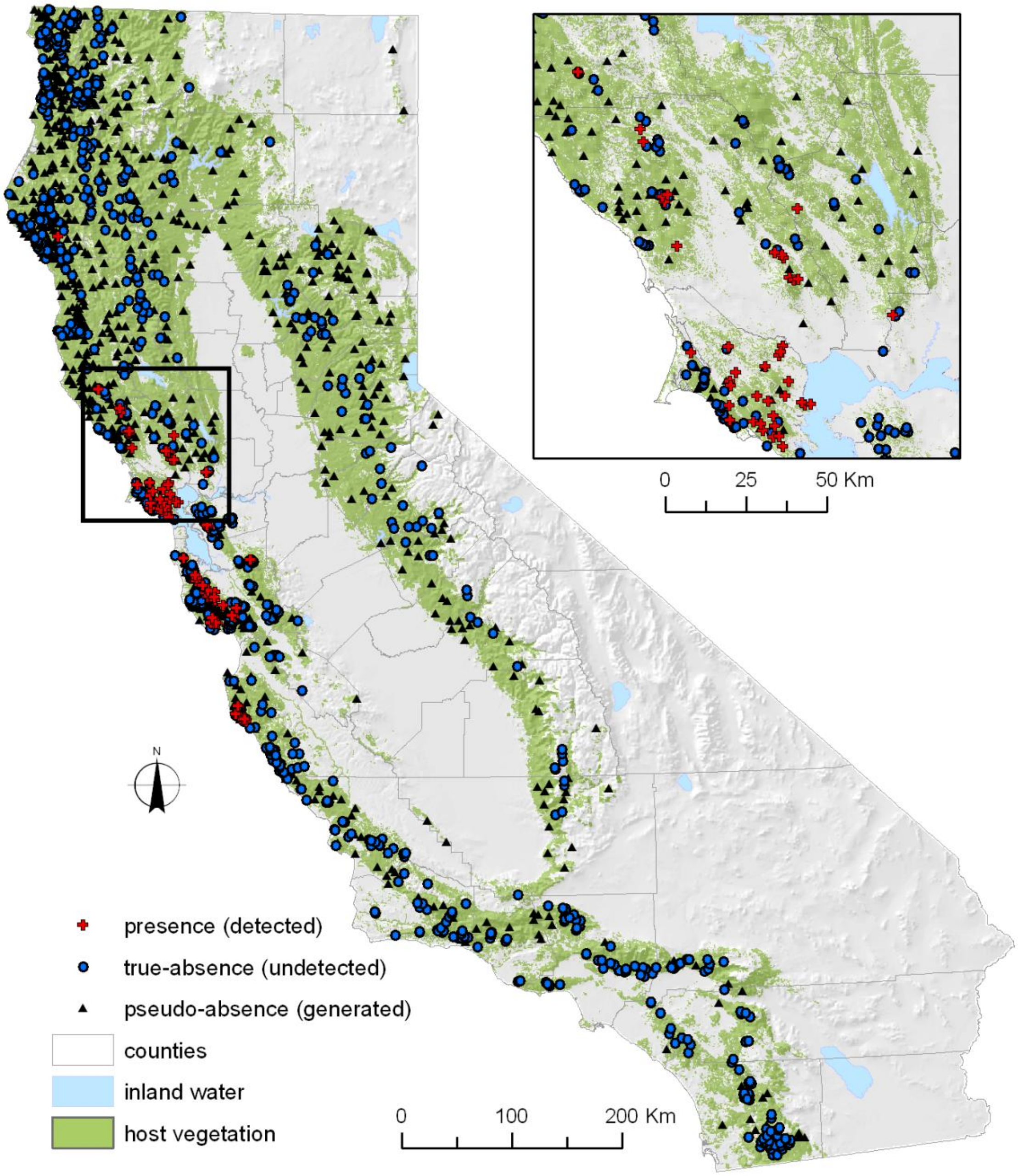
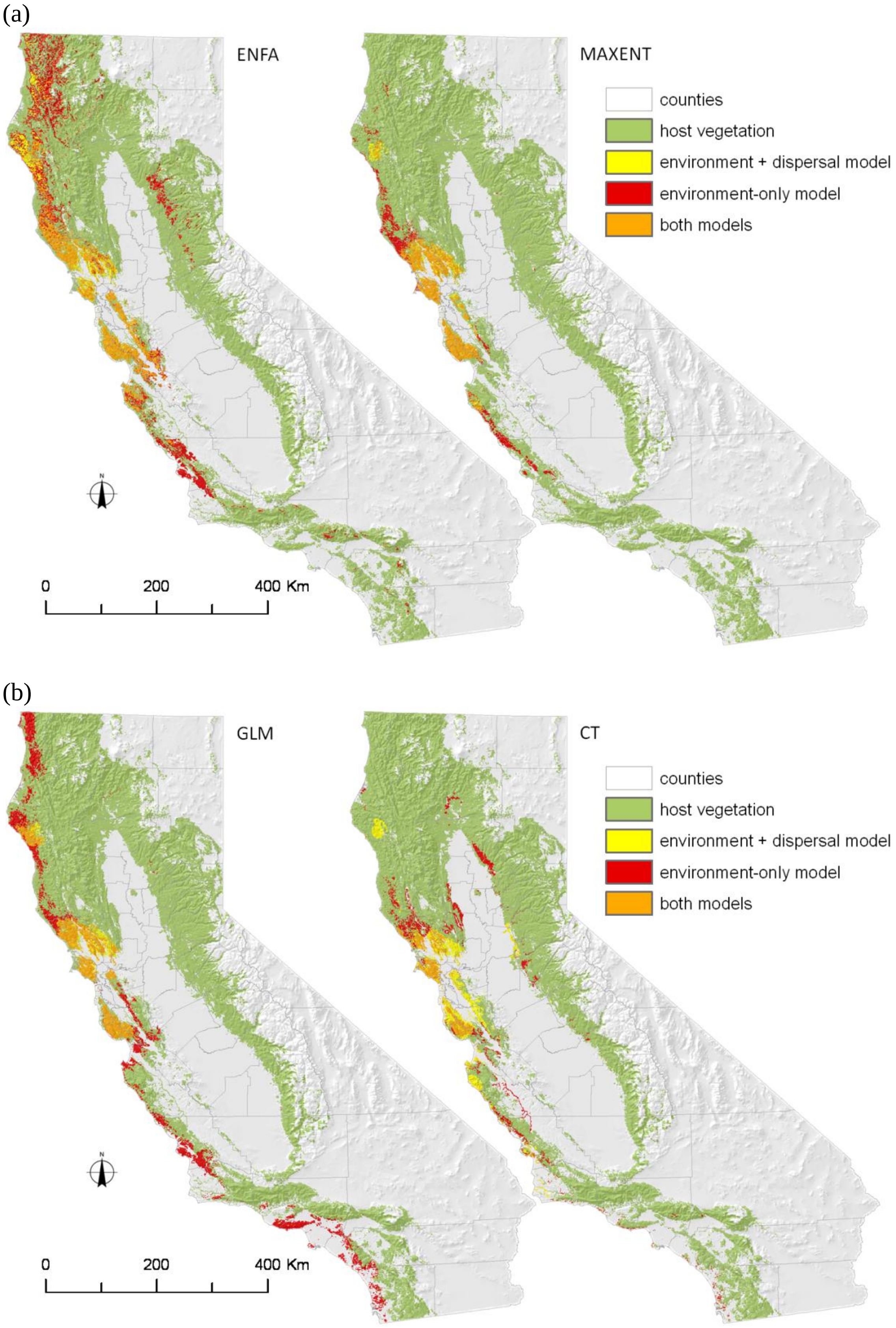


Figure 1: Map of 890 field plots surveyed for the presence of *Phythopthora ramorum* in California and distribution of 812 pseudo-absence points randomly generated in susceptible forest across a range of environmental conditions.


(a)
ENFA
MAXENT
□ counties
🟩 host vegetation
🟨 environment + dispersal model
🟥 environment-only model
🟧 both models
N
0
200
400 Km
(b)
GLM
CT
□ counties
🟩 host vegetation
🟨 environment + dispersal model
🟥 environment-only model
🟧 both models
N
0
200
400 Km

(c)

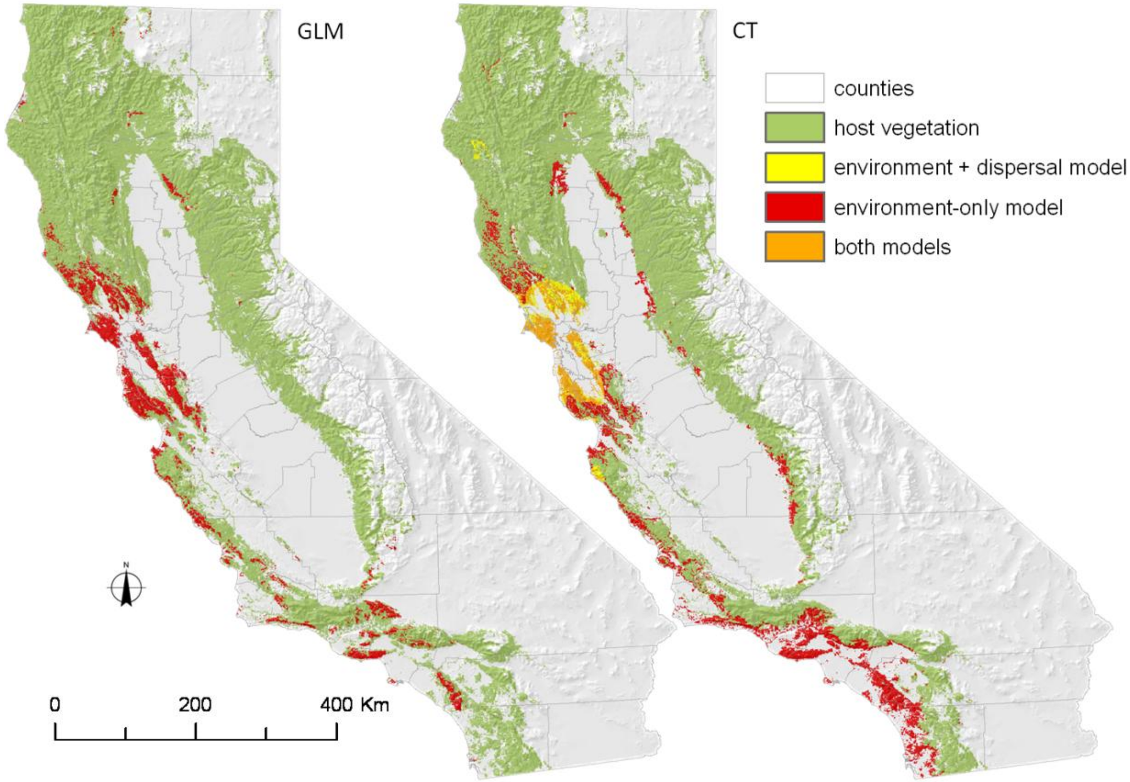


Figure 2: Thresholded maps of *Phythopthora ramorum* occurrence predicted by (a) presence-only models, (b) presence/true-absence models, and (c) presence/pseudo-absence models. Areas predicted with environment-only variables are depicted in red; areas predicted with combination of environmental variables and dispersal constraints are in yellow; areas predicted by both environment-only and hybrid models are in orange. Green color indicates susceptible host vegetation predicted as absence.

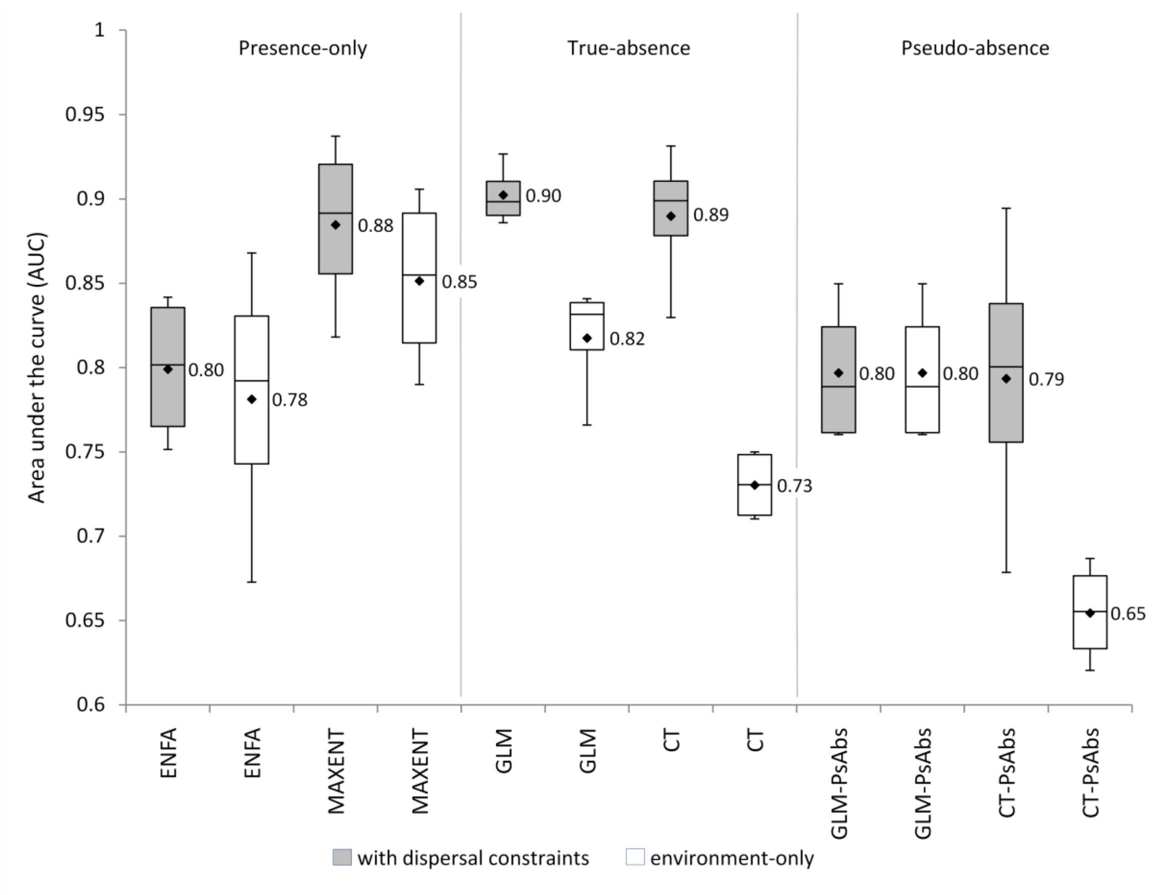


Figure 3: Model performances expressed by AUC for presence-only, presence/true-absence, and presence/pseudo-absence models. Each box-plot represents the results of all cross-validation runs using true occurrence data. The dot and number in box-plots is the mean AUC.

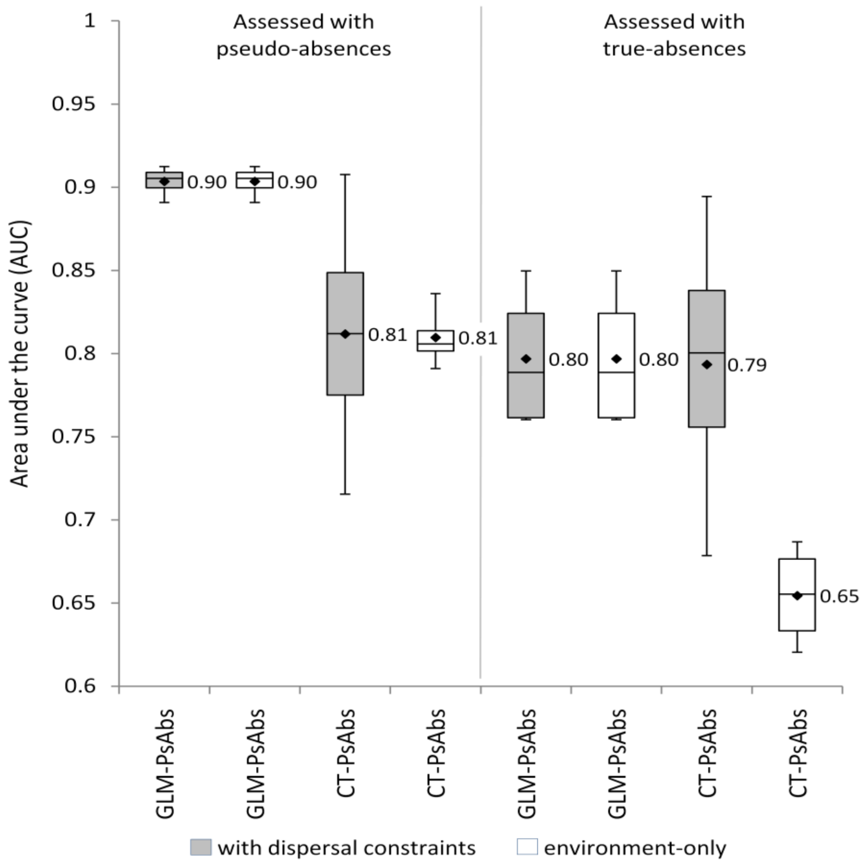


Figure 4: Differences in AUC for presence/pseudo-absence models when assessed with presence/pseudo-absence data or with presence/true-absence data. Each box-plot represents the results of all cross-validation runs; the dot and number in box-plots is the mean AUC.

CHAPTER 2: EQUILIBRIUM OR NOT? MODELING POTENTIAL DISTRIBUTION OF INVASIVE SPECIES IN DIFFERENT STAGES OF INVASION

2.1 Abstract

The assumption of equilibrium between organisms and their environment is a standard working postulate in species distribution models (SDMs). However, this assumption is typically violated in models of biological invasions where range expansions are highly constrained by dispersal and colonization processes. Here we examined how stage of invasion affects the extent to which occurrence data represent the ecological niche of organisms and in turn influence spatial prediction of species' potential distributions. We compiled occurrence data from 697 field plots collected over a nine-year period (2001–2009) of monitoring the spread of the invasive forest pathogen *Phytophthora ramorum*. Using these data we applied ecological niche factor analysis to calibrate models of potential distribution across different years of colonization. We accounted for natural variation and uncertainties in model evaluation by comparing findings for *P. ramorum* with three scenarios of varying equilibrium in a simulated virtual species, for which the "true" potential distribution was known. Results confirm our hypothesis that SDMs calibrated in early stages of invasion are less accurate than models calibrated under scenarios closer to equilibrium. In addition, SDMs that were developed in early stages of invasion tend to underpredict the potential range compared to models that are built in later stages of invasion. This research highlights the

consequences of ignoring equilibrium assumptions in the application of SDMs in conservation practice.

2.2 Introduction

Human activities continue to facilitate the spread of invasive species and infectious diseases through increased mobility and global trade, alteration of natural environments, and changes in global climate (Vitousek *et al.*, 1996; Mack *et al.*, 2000; Svenning & Condit, 2008). Successful management of biological invasions depends heavily on our ability to predict potential geographic ranges of invasive organisms and to identify factors that promote their spread (Sharma *et al.*, 2005). Species distribution models (SDMs), which draw statistical inference on drivers of species ranges from a snapshot of occurrence data, are being increasingly used to predict the spatial extent of invasions and identify at-risk habitats. However, modeling potential ranges of exotic organisms poses special challenges to modelers because invasive species distribution models (iSDMs) often require extrapolation to novel environments (Kearney, 2006; Jeschke & Strayer, 2008) and, without precautions, may violate the assumption of species equilibrium with environmental conditions (Vaclavik & Meentemeyer, 2009; Elith *et al.*, 2010; Robinson *et al.*, 2010).

The assumption of equilibrium is a standard working postulate for static SDMs (Guisan & Thuiller, 2005; Miller, 2010). An organism is considered to be at equilibrium with its environment if it occurs in all suitable locations, while being absent from all unsuitable ones (Araujo & Pearson, 2005). However, this condition is possible only under unlimited dispersal scenarios and very high extinction rates outside the extremes of favorable environmental factors (De Marco *et al.*, 2008). Non-equilibrium is more likely under ecological scenarios that involve biotic interactions, metapopulation dynamics, or slow colonization after recent environmental changes. Past studies have shown that non-equilibrium occurs for native species with slow range shifts (Pearson & Dawson, 2003; Svenning & Skov, 2004), but it may play an especially influential role in models of invasive organisms, where the departure from equilibrium is particularly high in earlier stages of invasion due to colonization time lag and dispersal limitations (Vaclavik *et al.*, 2010). The biogeographic processes underlying non-equilibrium have been widely discussed but questions of how far an organism is from equilibrium and what its consequences are on predictive models have received little attention in the SDM literature (Svenning & Skov, 2004; Araujo & Pearson, 2005; Pearson, 2006a; Roura-Pascual *et al.*, 2009).

Stage of invasion (i.e. the degree to which an invader is at equilibrium) may have a profound influence on iSDMs by affecting the extent to which occurrence observations match the species environmental niche (Jimenez-Valverde *et al.*, 2008; Zimmermann *et al.*, 2010). For example, if a species is in an early stage of invasion, sampling its emerging realized niche may only capture a small portion of the climatic and local habitat combinations that could be potentially inhabited by the species. Models fitted with such data would project only that small portion of the fundamental niche to the geographical space. On the other hand, organisms in early stages of range expansion exhibit strong range cohesion (sensu Rahbek *et al.*, 2007) that may prevent modelers from generating high commission error rates. Thus, there is no direct relationship between sampling in environmental space and geographical space because even poor or small samples can provide good approximations of ecological dimensions if the landscape is heterogeneous

(Hirzel & Le Lay, 2008). The complex relationship between sampling in geographical space and approximation of ecological dimensions requires that we test SDMs under various degrees of equilibrium. However, well-distributed multi-temporal data are rarely available for independent validation of expanding ranges of biological invasions (Sutherst & Bourne, 2009).

In this study, we use field data on the spread of the invasive forest pathogen *Phytophthora ramorum*, collected over the span of nine years (2001–2009), to examine how stage of invasion affects the extent to which occurrence data represent the ecological niche of an organism and in turn influences spatial prediction of potential distribution. To control for complications from natural variation and uncertainties in model assessment, we compare findings based on *P. ramorum* with those of a simulated virtual species, for which the "true" environmental niche is known. Assuming the pathogen will be closer to equilibrium the longer it has been established in the new environment (Araujo & Pearson, 2005), we test two hypotheses: (i) the accuracy of models will be lower and (ii) the predicted range smaller when models are calibrated with occurrence data from emerging stages of invasion. We employ ecological niche factor analysis (ENFA; Hirzel *et al.*, 2002) to develop iSDMs of potential distribution calibrated with *P. ramorum* data across different years of colonization and with virtual species data sampled under three hypothetical scenarios of equilibrium. Model performances are evaluated with crossvalidation procedures for the real species and contrasted with the "true" probability of invasion known for the simulated species. This research emphasizes one of the fundamental modeling principles that have profound implications for practical applications of iSDMs in conservation management.

2.3 Methods

2.3.1 Field-based species data

To assess the role of the stage of invasion with a real invasive species, we compiled data from extensive forest surveys of *P. ramorum* infection coordinated by the Oregon Department of Forestry and the USDA Forest Service in southwestern Oregon during 2001–2009 (Vaclavik *et al.*, 2010) (Figure 1). Each year, heterogeneous forest ecosystems were systematically surveyed using a combination of helicopter scanning and ground-based checks, identifying and recording apparent mortality and symptomatic trees in individual forest stands. Leaf samples were collected from symptomatic sites and tested for *P. ramorum* at Oregon State University and Oregon Department of Agriculture laboratories via traditional culturing and a polymerase chain reaction (PCR)-based molecular assays (Kanaskie *et al.*, 2009a). This procedure produced a reliable dataset of pathogen's occurrence in the study area (n_{total}=697, n_{2001}=27, n_{2002}=51, n_{2003}=50, n_{2004}=25, n_{2005}=42, n_{2006}=128, n_{2007}=152, n_{2008}=120, n_{2009}=102). We assumed the collection of data in the span of nine years would effectively capture different degrees of equilibrium and range expansion because *P. ramorum* is a highly virulent pathogen that is rapidly filling its niche potential (Hansen *et al.*, 2008; Meentemeyer *et al.*, 2008a). Moreover, the isolated nature of this disease outbreak makes this dataset ideal for examining the role of invasion stages in iSDMs.

2.3.2 Virtual species data

We simulated a hypothetical invasive species *Phytophthora virtualis* to test the stage of invasion effect in a controlled environment. The "true" potential distribution of *P. virtualis* was simulated as a combination of additive and multiplicative ecological niche requirements (sensu Meynard & Quinn, 2007; Elith & Graham, 2009) based on four real environmental variables for western Oregon: potential solar irradiation (PSI), precipitation, maximum temperature and host index (for details see below). The probability of invasion P_i at site i was calculated as a weighted average of species responses to environmental factors:

$$P_i = \frac{1}{\sum_{j=1}^{3} w_j}\left(w_1 f_1(PSI) + w_2 f_2(precipitation) + w_3 f_3(temperature)\, f_4(hostindex)\right) \quad (1)$$

where f is a function of environmental variables representing a partial probability of occurrence and w_j is the weight assigned to the partial probability of occurrence j. The additive part of the equation describes a situation in which environmental factors are substitutable, i.e., unsuitability of one condition can be compensated by suitability of remaining conditions (Hirzel *et al.*, 2001). The multiplicative part represents an interaction between environmental factors that are irreplaceable, i.e., unsuitability of one condition causes low probability of occurrence even though the other condition is in species optimum (Meynard & Quinn, 2007). To create a realistic species, we used two types of functions to model the responses of *P. virtualis* to environmental factors (Hirzel *et al.*, 2001; Santika & Hutchinson, 2009): (i) a Gaussian response with a symmetrically decreasing probability of occurrence around median optimum (for PSI and temperature), and (ii) a linear response with an increasing probability of occurrence towards an extreme optimum (for precipitation and host index). The final probability was rescaled to values between 0 and 1 using linear transformation to preserve the original shape of response functions (Meynard & Quinn, 2007).

We generated three “true” presence-absence scenarios of *P. virtualis* distribution to represent distinct stages of invasion. First, the “equilibrium” scenario was created by using probability values of each site as a success rate for a sample drawn from the binomial distribution. This procedure induced some stochasticity in the data, producing a situation under which a partially suitable site, e.g. P_i=0.7, has a 70% chance to be invaded and a 30% chance to be uninvaded (Elith & Graham, 2009). Second, the “intermediate” scenario was produced by subtracting d^2 from the probability value of each site, with *d* being the distance of that site to a randomly selected location in southwestern Oregon, representing a hypothetical site of introduction (Hirzel *et al.*, 2001). After rescaling the values to a 0–1range, the product was submitted to the same operation as the “equilibrium” scenario. Third, the “initial” scenario was generated by the same procedures as the “intermediate” scenario, with the exception of using $\sqrt{d}$ to subtract from the probability of invasion values. *P. virtualis* in our initial, intermediate, and equilibrium scenarios occupied 2%, 4%, and 8% of the study area respectively. We randomly sampled 1000 positive locations from each scenario to provide data for model calibration, as this number of sites is sufficient to fit models well and is consistent with sample sizes commonly used in SDM studies (Elith & Graham, 2009).

2.3.3 Environmental predictor variables

We used the same set of three climate, three topographical, and 14 host species predictor variables as in Vaclavik *et al.* (2010). For climate, we derived 30-year monthly averages (1971–2001) of maximum temperature, minimum temperature, and precipitation from the parameter elevation regression on independent slopes model (PRISM; Daly *et al.*, 2001), using 800 m resolution grids for each month in the pathogen’s reproductive season (December to May) in the western United States (Davidson *et al.*, 2005). For topography, we derived elevation, topographic moisture index (TMI), and potential solar irradiation (PSI) from the U.S. Geological Survey 30-m digital terrain model. We calculated TMI as the natural log of the ratio between the upslope contributing area and the slope gradient (Moore *et al.*, 1991), and PSI as the potential mean solar radiation in the reproductive season using the cosine of illumination angle on slope equation (Dubayah, 1994).

To characterize vegetation heterogeneity of *P. ramorum* habitat, we used geospatial vegetation data developed using gradient nearest neighbor (GNN) imputation (Ohmann & Gregory, 2002; Ohmann *et al.*, 2007; Pierce *et al.*, 2009). GNN was applied to canopy cover data from several thousand field plots installed in regional inventory, ecology, and fuel mapping programs to map the abundance of 14 host species susceptible to *P. ramorum* infection: *Arbutus menzeisii*, *Arctostaphylos* spp., *Frangula californica*, *Frangula purshiana*, *Notholithocarpus densiflorus*, *Lonicera hispidula*, *Pseudotsuga menziesii*, *Quercus chrysolepis*, *Quercus kelloggii*, *Rhododendron* spp., *Rubus spectabilis*, *Sequoia sempervirens*, *Umbellularia californica*, *Vaccinium ovatum*. For simplicity in simulating the environmental niche of our virtual species, we used the host index variable (instead of individual hosts species), combining all hosts into a single predictor based on their abundance, susceptibility, and spread potential (Meentemeyer *et al.*, 2004; Vaclavik *et al.*, 2010). All variables were prepared in IDRISI Taiga (Clark Labs/Clark University, Worcester, MA) and normalized by the Box-Cox transformation (Sokal & Rohlf, 1981).

2.3.4 Building models of potential distribution

Presence-only methods based on environmental profiling are recommended for predictions of potential distributions because they do not consider absence locations that could potentially overlap with environmentally suitable areas (Jimenez-Valverde *et al.*, 2008; Vaclavik & Meentemeyer, 2009; Lobo *et al.*, 2010). We chose the ecological niche factor analysis (ENFA; Hirzel *et al.*, 2002) that provides good generalization of species niche responses and employs algorithms specifically designed for species at the edge of their fundamental niche (Braunisch *et al.*, 2008). ENFA summarizes environmental predictors into few uncorrelated factors based on the analysis of species (i) marginality (the extent to which the species optimum departs from the most frequent conditions in the study area) and (ii) specialization (tolerance to conditions that are increasingly different from the species optimum).

For *P. ramorum,* we used infected sites from the period between 2001 and 2009 to calibrate ENFA models incrementally for each year t_x with data from t_1 to t_x. This procedure mimics a common conservation practice when predictive models are cumulatively updated as new invasion outbreaks are discovered. For *P. virtualis*, we calibrated three models with presence data sampled from each of the three stage of invasion scenarios. Here, we chose the same sample size (n=1000) for each scenario to control for uncertainties that can stem from different sizes of training datasets. Predictions of potential distributions were calculated in the Biomapper 4.0 software (Hirzel et al., 2007) using the median-extremum algorithm that assumes the species optimum is indicated by either the median or extreme values of environmental conditions (Braunisch *et al.*, 2008). The number of significant factors was determined based on their eigenvalues compared to the broken-stick distribution (McArthur, 1957), and ranged between 2 and 4 factors with 87–93% of explained variability for *P. ramorum*, and 2 and 3 factors with 92–97% of explained variability for *P. virtualis*.

2.3.5 Evaluating effects of non-equilibrium

The accuracy of modeling non-equilibrium species can be only partially evaluated with current occurrence data because potential distribution is a hypothetical concept and the true future distribution of species is unknown (Jimenez-Valverde *et al.*, 2008; Elith *et al.*, 2010). For example, the use of binary statistics (e.g. Cohen's kappa) is inappropriate because absence data are unavailable and/or include locations where environmental conditions are suitable but the invader is absent due to dispersal limitations and colonization time lag. We assessed and compared the accuracy of *P. ramorum* models calibrated for each stage of invasion by means of jack-knifed *k*-fold crossvalidation, with *k*=4 determined by the Huberty's heuristic (Fielding & Bell, 1997). Three evaluation indices were calculated for each replicate and characterized by their mean and standard deviation across replicates. The absolute validation index (AVI) was computed as the proportion of presence evaluation points falling above a fixed threshold. The contrast validation index (CVI) was calculated as the AVI minus the AVI of a null model predicting presence at random (Ayala *et al.*, 2009). The fixed threshold was determined according to Hirzel *et al.* (2006) through an inspection of the predicted-to-expected frequency curves (P/E, Boyce area-adjusted frequencies), finding the values for which presences were more frequent than expected by chance (P/E>1). To alleviate the threshold constraint, we used an additional statistic: the continuous Boyce index (Boyce *et al.*, 2002; Hirzel *et al.*, 2006). The index is based on "moving window" analysis across the range of predicted values and uses the Spearman rank correlation coefficient to measure the monotonic increase in the P/E frequency ratio with increasing habitat suitability. It varies from -1 to 1, with 0 indicating predictions indifferent from a random model.

The “true” potential distribution for *P. virtualis* was perfectly known, thus a direct evaluation was conducted using all locations in the study area. In order to avoid pseudo-replication due to spatial autocorrelation among sites and to assess model robustness, we generated ten validation datasets of the same size as the calibration dataset (n=1000). Since the threshold selection and evaluation of predictive models is more influenced by prevalence than sample size (McPherson *et al.*, 2004; Jimenez-Valverde & Lobo, 2006), the prevalence of 0.5 was forced to allow for 500 presence and 500 absence validation points sampled from the true “equilibrium” distribution scenario. For these independent datasets we computed means and standard deviations of (i) the Pearson correlation coefficients between predicted P_i for each scenario and the original “true” probability of invasion (Hirzel *et al.*, 2001), (ii) the area under the curve (AUC) of the receiver operating characteristic (ROC) to measure model ability to discriminate potentially occupied and unoccupied sites (Phillips & Elith, 2010), and (iii) the omission and commission error rates based on the probability threshold determined by P/E frequency curves (Hirzel *et al.*, 2006).

2.4 Results

Predictions of *P. ramorum* potential distributions (Figure 2) required substantial extrapolation because data from all stages of invasion were geographically restricted to the southwest portion of the study area. Results show that models calibrated with presences from later stages of invasion (closer to equilibrium) were more accurate and robust than models based on data from earlier stages (Figure 3). Boyce index calculated for each model as an average for all crossvalidation runs exhibited an increasing trend (B_{2001}=0.01, B_{2004}=0.49, B_{2007}=0.75, B_{2009}=0.86) and decreasing variability (SD_{2001}= 0.81, SD_{2004}= 0.55, SD_{2007}= 0.26, SD_{2009}=0.18) with progressing invasion. The same tendency was observed for AVI and CVI based on P/E>1 thresholds (AVI_{2001}=0.38, AVI_{2004}=0.53, AVI_{2007}=0.75, AVI_{2009}=0.83 and CVI_{2001}=0.37, CVI_{2004}=0.49, CVI_{2007}=0.69, CVI_{2009}=0.76), with some fluctuation in index levels for the first three years. The amount of land area predicted for the pathogen's potential distribution also increased with increasing degree of equilibrium (A_{2001}=1671 km^2, i.e. 1.5% of the total study area; A_{2004}=3920 km^2, i.e. 3.5%; A_{2007}=6980 km^2, i.e. 6.2%; A_{2009}=7938km^2, i.e. 7.1%; Figures 2 and 3), suggesting that larger datasets from later stages of invasion captured more combinations of environmental variables that constitute the fundamental niche of *P. ramorum*.

For *P. virtualis*, model predictions attempted to reconstruct the "true" potential distribution of this artificial species. Whilst all models correctly predicted high probabilities of invasion in the southwest portion of the study area (close to the hypothetical location of introduction), the "initial" and "intermediate" models failed to identify the second hot spot of habitat suitability in the northwest (Figure 4). This observation was statistically supported by Pearson correlations between predicted and "true" probabilities of invasion, with the highest coefficients estimated for the "equilibrium" scenario (ρ=0.7; Figure 5). Based on AUC values computed with independent validation datasets, all three models had good ability to discriminate between potentially occupied and unoccupied sites (AUC≥0.9 for all models), but the discrimination power and robustness increased with invasion stages closer to equilibrium (Figure 6). The amount of land area predicted for the pathogen's potential distribution also increased with increasing degree of equilibrium (Figure 5), similarly as in the case of *P. ramorum*. However, knowing the "true" fundamental niche of *P. virtualis*, we found that models developed in earlier stages of invasion had higher omission error rates than models in later stages (decrease in omission error rate by 0.14 from the "initial" to "equilibrium" scenario; Figure 7). All three models exhibited similarly low rates of commission errors (≤0.12).

2.5 Discussion

Theoretical issues of species non-equilibrium have been discussed in multiple studies (e.g. Araujo & Pearson, 2005; Guisan & Thuiller, 2005; Elith *et al.*, 2010; Robinson *et al.*, 2010). In the present study, we empirically examined its role in iSDMs using a unique set of field-based data on a real invasive organism in combination with artificially created species of known fundamental niche. Our results confirmed the hypothesis that iSDMs calibrated under non-equilibrium are less accurate and robust in predicting the habitat potentially prone to invasion than models calibrated under scenarios closer to equilibrium. In addition, iSDMs of species in early stages of invasion had a higher tendency to underpredict the potential range than models of species in later stages of invasion. Although *P. ramorum* data from all stages of invasion in Oregon reflected situations distant from potential equilibrium, the general trends in model accuracies and predicted extents were in close agreement between the real and virtual species.

Our findings are congruent with theoretical expectations that the full environmental niche of invasive species cannot be effectively captured with data from a realized distribution that is restricted by processes preventing full occupancy of suitable habitats (Thuiller *et al.*, 2004; Jimenez-Valverde *et al.*, 2008). Alternative strategies exist that involve building (i) correlative models based on the native range and projecting them to potential introduction sites (e.g. Peterson, 2003; Fitzpatrick *et al.*, 2007) or (ii) mechanistic models based on physiological responses of organisms to their environment (e.g. Meentemeyer *et al.*, 2004; Kearney *et al.*, 2008). However, the former approach requires extrapolation to novel environments with unique biotic interactions and may be misleading because recent studies provided evidence of niche shift during invasion (Broennimann *et al.*, 2007; Beaumont *et al.*, 2009). The latter approach requires substantial knowledge of organisms' physiological traits and the availability of such data for model parameterization is limited (Jeschke & Strayer, 2008; Kearney & Porter, 2009). Thus, correlative approaches based on snapshots of range expansion through time will likely remain prevailing (Elith *et al.*, 2010) and the effects of non-equilibrium will need to be considered.

Here we used a correlative approach based on *P. ramorum* data from an invaded range to examine the role of non-equilibrium. However, a mechanistic model for this pathogen has been previously developed. Vaclavik *et al.* (2010) predicted potential distribution of *P. ramorum* in Oregon using a heuristic model based on current knowledge of host susceptibility and pathogen reproduction and transmission with particular regard to spatial and temporal variability of host vegetation and climate. The area and levels of pathogen's establishment and spread risk closely match the predictions of our ENFA model calibrated with the full field-based dataset (2001–2009). This suggests that, although the realized distribution of *P. ramorum* covers only a fraction of its habitat potential, even a small sample from geographically restricted area can allow identification of a large portion of the species’ fundamental niche if environmental conditions in the occupied geographical space are highly heterogeneous. In addition, the agreement of both models indicates that there are fewer habitats potentially at risk from *P. ramorum* invasion in Oregon than in California (Meentemeyer *et al.*, 2004), due to less suitable climate and lower abundances of competent hosts.

The ecological niche factor analysis was applied in this study as a typical example of environmental envelope approaches that provide smooth responses to environmental factors. Comparative studies showed that many novel methods are able to fit complex functions to calibration data and provide more accurate predictions of species distributions (Elith *et al.*, 2006; Tsoar *et al.*, 2007). However, these approaches were evaluated with contemporary presence/absence data and thus measured predictions of the realized distribution rather than the potential distribution. Close model fit does not guarantee that the fundamental niche is successfully captured (Pearson *et al.*, 2006) and the use of profile techniques that enforce smooth responses have been advocated to estimate species potential ranges in new environments and future climates (Sutherst & Bourne, 2009; Elith *et al.*, 2010).

Complex methods, e.g. maximum entropy (MaxEnt), and methods based on both presence and absence data, e.g. generalized additive models (GAM), are more appropriate for modeling realized distributions of invasive species but these should explicitly incorporate measures of dispersal and colonization processes that constrain invaders from occupying all potentially suitable ranges (Vaclavik & Meentemeyer, 2009). However, the effects of the stage of invasion on predictions of realized distributions have not been sufficiently explored. We suggest that the degree of equilibrium will influence not only predictive performance but also the reliability of accuracy statistics due to varying relative occurrence area (ROA) in different stages of invasion (i.e. differences in the ratio between the extent of occurrence and the extent of the study region; Lobo *et al.*, 2008). Empirical examination of non-equilibrium in models of realized distribution represents a felicitous topic for future research of biological invasions.

Multi-modeling approaches have also been suggested to deal with non-equilibrium in invasive species, including methods that couple SDMs with dispersal, population, and landscape dynamics (Keith *et al.*, 2008; Franklin, 2010b; Smolik *et al.*, 2010). The advantage of dynamic models is that they explicitly account for spatio-temporal ecological processes and allow predictions of spread rates and levels of invasion risk at specific time intervals. For example, Prasad et al. (2010) modeled the spread of emerald ash borer by combining flight and human-assisted types of dispersal with landscape heterogeneity and anthropogenic factors in a spatially-explicit cellular automata framework to identify areas at high risk of infestation over the next 2–4 years. Meentemeyer et al. (2011) coupled stochastic epidemiological modeling with pattern-based biogeographical modeling to predict the spread of the sudden oak death pathogen through heterogeneous wildland forests over a 40-year period. These approaches have great potential to increase our understanding of invasion dynamics and account or compensate for the effects of non-equilibrium in SDMs. However, the technical challenges, hunger for data, and computational intensity associated with their implementation over large heterogeneous landscapes may limit their feasibility in many cases (Franklin, 2010b).

Invasive species and the ecological and economic costs associated with their spread are driving demand for tools to estimate species potential distributions. iSDMs are being increasingly used to address this task, and the evaluation of diverse SDM approaches has received enormous attention in recent ecological literature (e.g. Segurado & Araujo, 2004; Elith *et al.*, 2006; Meynard & Quinn, 2007; Tsoar *et al.*, 2007). However, it has been argued that we need to start asking “why” models differ in their performance and go back to the basic ecological principles of this discipline, rather than continue applying descriptive models to multiple species using black-box approaches (Austin *et al.*, 2006; Jimenez-Valverde *et al.*, 2008; Elith & Graham, 2009; Sutherst & Bourne, 2009). Our study highlighted one of the key ecological principles that should be carefully considered in applications of iSDMs; namely the effect of different stage of invasion scenarios on the ability of predictive models to identify species potential distributions. Our findings contribute to the ecological understanding of the SDM discipline and demonstrate that modeling efforts require caution when conducted under non-equilibrium scenarios. If iSDMs are to be used effectively in conservation practice, the effect of invasion stages needs to be considered in order to avoid underestimation of habitats at risk of future invasion spread.

2.6 Acknowledgements

We are grateful to A. Kanaskie and E. Hansen for their research and data collection on *P. ramorum* in Oregon, and J. Ohmann for providing us with the GNN vegetation data. We also thank M. Dorning and W. Dillon for their comments on the earlier version of the manuscript. This research was supported by grants from the National Science Foundation (EF-0622677) and USDA Forest Service–Pacific Southwest Research Station (08-JV-11272138-055).

Figures

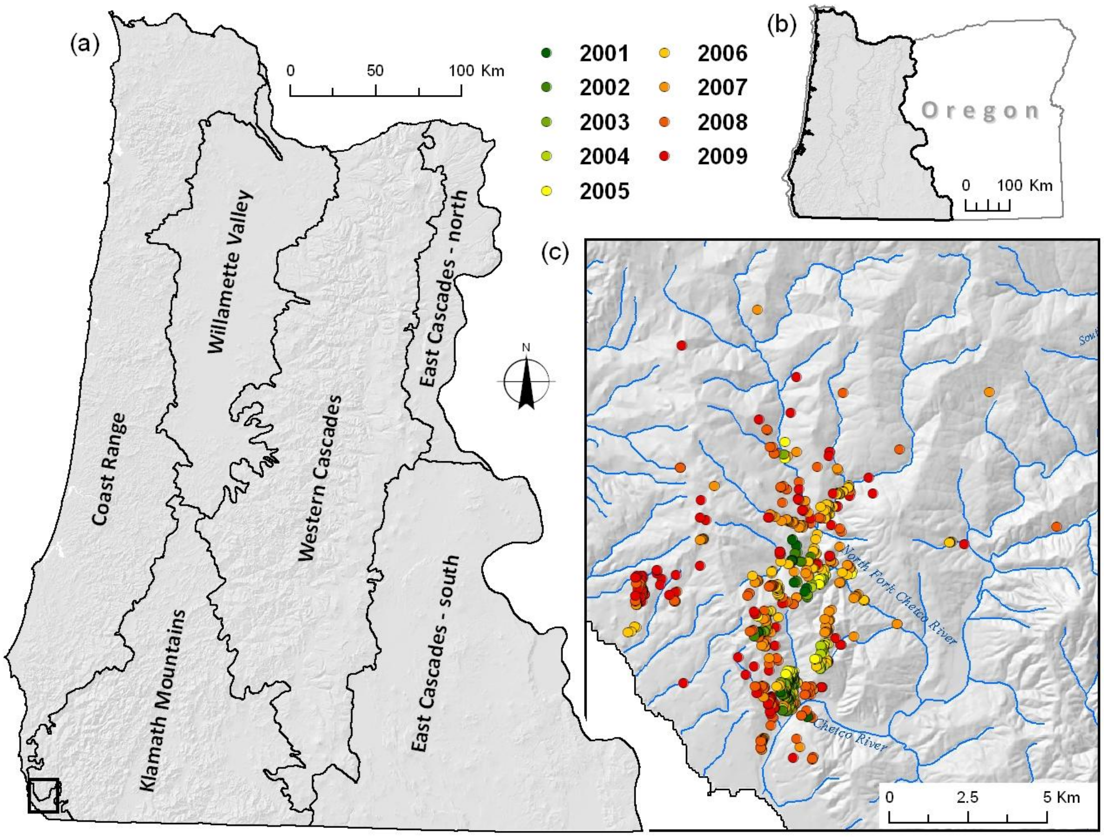


Figure 1: Study area (a) and its location (b) in western Oregon: six ecoregions with host vegetation potentially susceptible to *Phytophthora ramorum*. Inset (c) shows the distribution of 697 plots invaded by *P. ramorum* between 2001 and 2009.

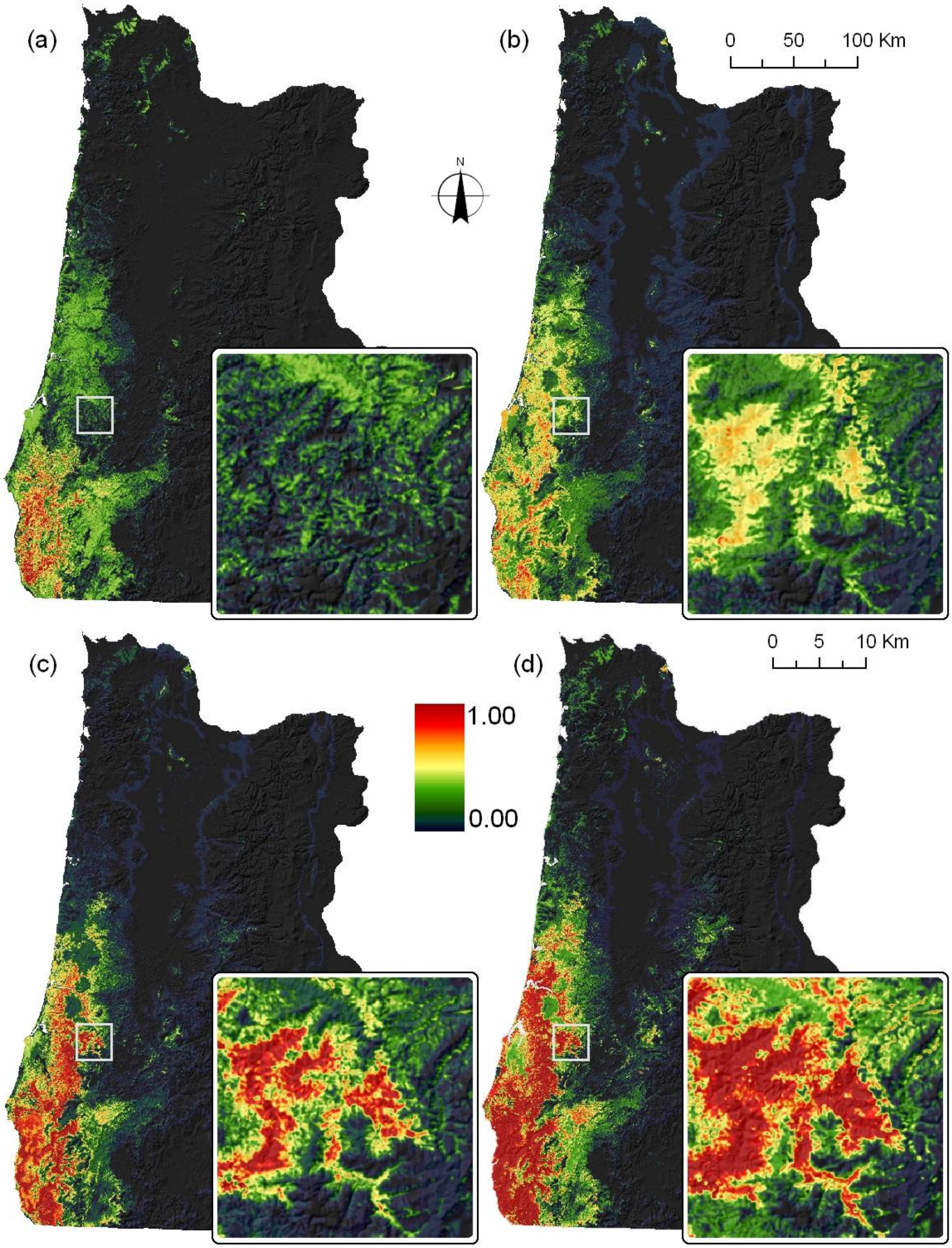


Figure 2: Predictions of *Phytophthora ramorum* potential distribution based on data from different stages of invasion. Four examples: (a) 2001, (b) 2001–2004, (c) 2001–2007, (d) 2001–2009. The range of values between 0 and 1 represents probabilities of invasion (P_i).

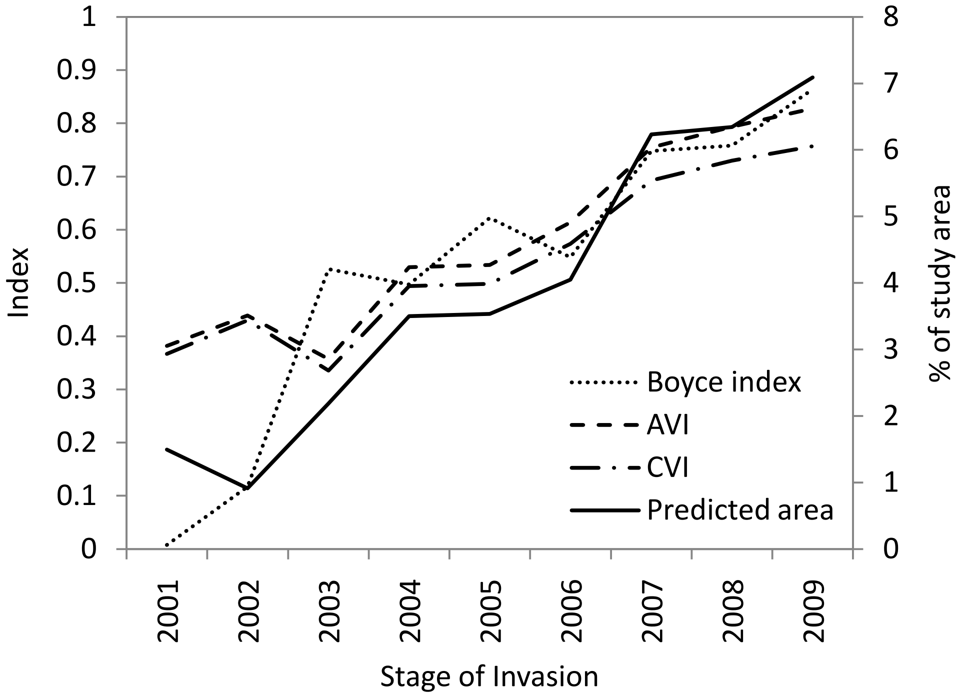


Figure 3: Model performance for *Phytophthora ramorum* expressed by Boyce index, absolute validation index (AVI), contrast validation index (CVI), and the total predicted area. Lines represent means across 4-fold crossvalidation replicates.

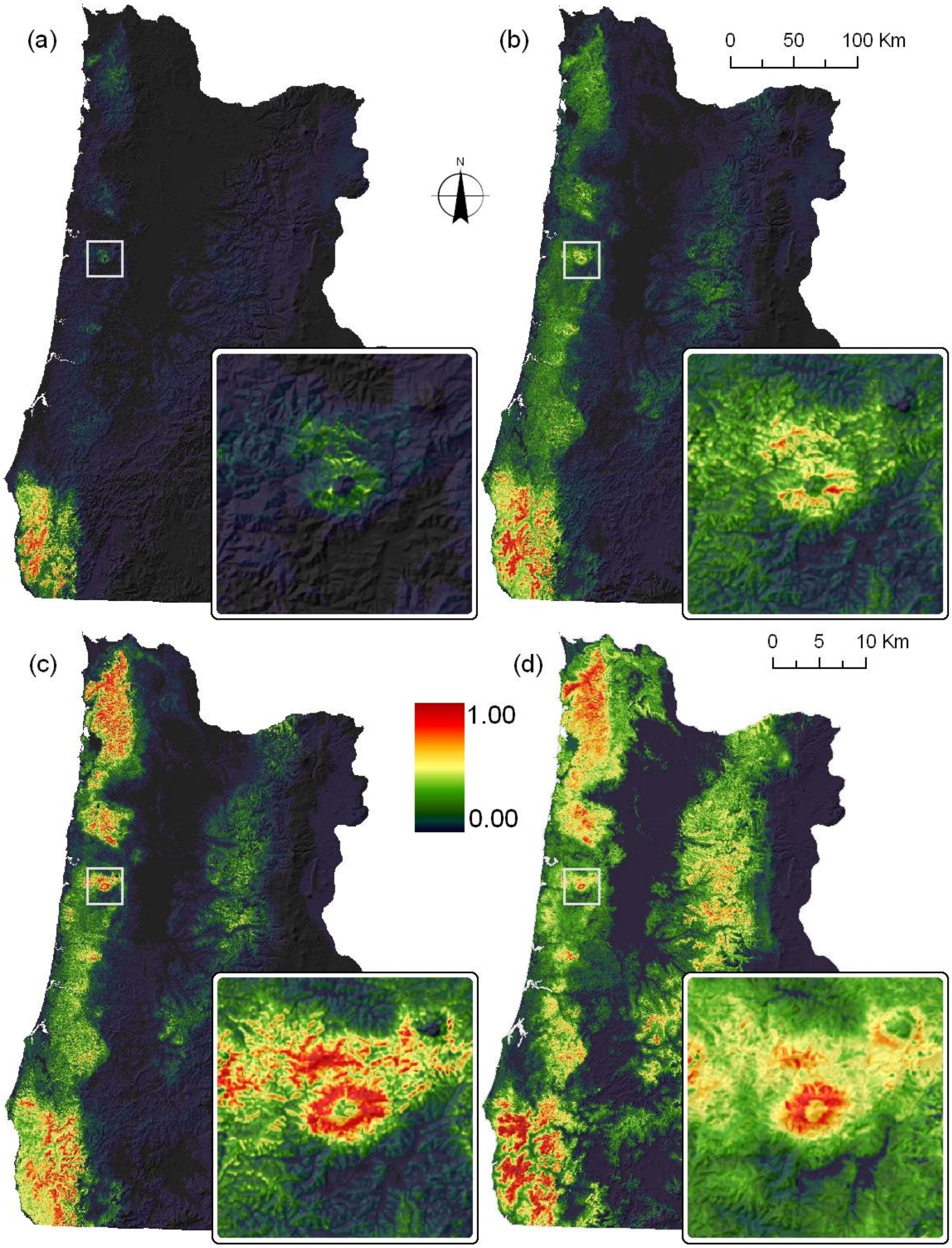


Figure 4: Predictions of *Phytophthora virtualis* potential distributions based on hypothetical data that simulate different degrees of equilibrium: (a) “initial” scenario, (b) “intermediate” scenario, (c) “equilibrium” scenario. Map (d) is the “true” potential distribution. The range of values between 0 and 1 represents probabilities of invasion (P_i).

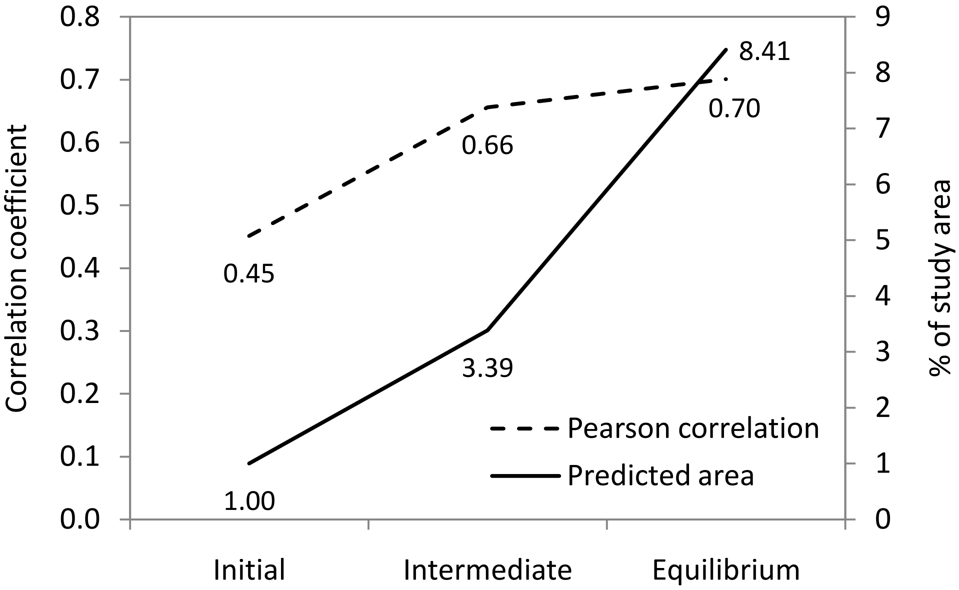


Figure 5: Model performance for *Phytophthora virtualis* expressed by the total predicted area and Pearson correlation coefficients between each predicted scenario and the “true” potential distribution. Results represent means calculated with ten independent evaluation datasets.

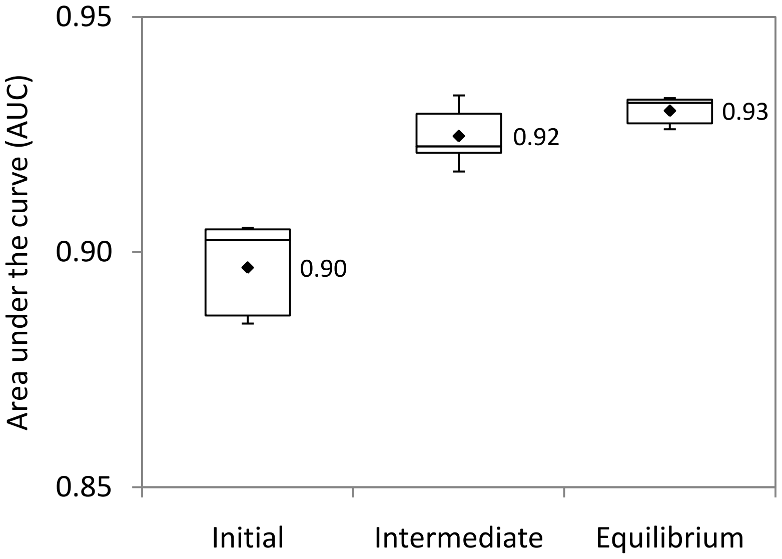


Figure 6: Discrimination power for models of *Phytophthora virtualis* expressed by area under the curve (AUC) of the receiver operating characteristic (ROC). Each box-plot represents results calculated with ten independent evaluation datasets; the dots are mean values.

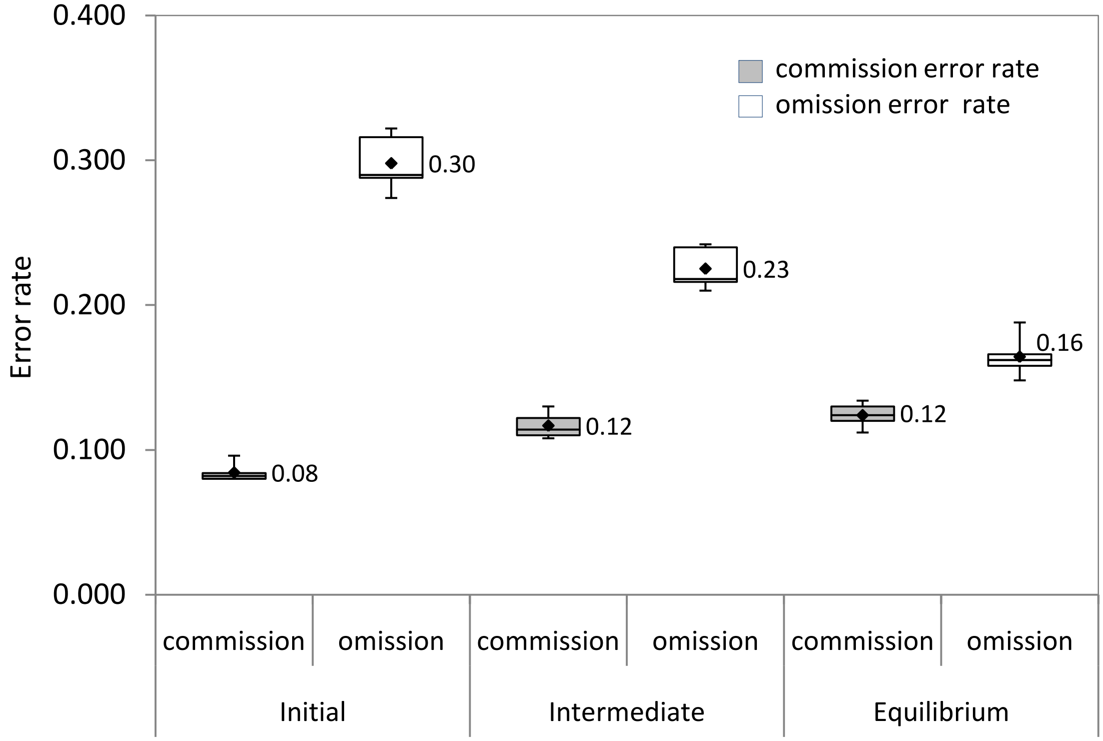


Figure 7: Threshold-dependent assessment for models of *Phytophthora virtualis* expressed by commission and omission error rates. Each box-plot represents results calculated with ten independent evaluation datasets; the dots are mean values.

CHAPTER 3: ACCOUNTING FOR MULTI-SCALE SPATIAL AUTOCORRELATION IMPROVES PERFORMANCE OF INVASIVE SPECIES DISTRIBUTION MODELING (iSDM)

3.1 Abstract

Analyses of species distributions are complicated by various origins of spatial autocorrelation (SAC) in biogeographical data. SAC may be particularly important for invasive species distribution models (iSDMs) because biological invasions are strongly influenced by dispersal and colonization processes that typically create highly structured distribution patterns. Here, we examined the efficacy of using a multi-scale framework to account for different origins of SAC and compared non-spatial models with models that accounted for SAC at multiple levels. We applied one conventional statistical method (GLM) and one non-parametric technique (MAXENT) to a large dataset on invasive forest pathogen *Phytophthora ramorum* in western North America (n=3787) to develop four types of models that either ignored spatial context or incorporated it at a broad scale using trend surface analysis, a local scale using autocovariates, or multiple scales using spatial eigenvector mapping. We evaluated model accuracies and amounts of explained spatial structure and examined the changes in variables' predictive power. Results show that accounting for different scales of SAC significantly enhanced predictive capability of iSDMs. Dramatic improvements were observed when fine-scale SAC was included, suggesting that local range-confining processes are driving *P. ramorum* spread. The importance of environmental variables was relatively consistent across all models, but the

explanatory power decreased in spatial models for factors with strong spatial structure. While accounting for SAC reduced the amount of residual autocorrelation for GLM but not for MAXENT, it still improved performance of both approaches, supporting our hypothesis that dispersal and colonization processes are important factors to consider in distribution models of biological invasions. We conclude that, apart from being vital to avoid problems in multivariate statistical analyses, spatial pattern may be an important surrogate for dynamic processes that explain ecological mechanisms of invasion and improve predictive performance of iSDMs.

3.2 Introduction

Species distribution models (SDMs) have long played a role in advancing biogeographical theory and are being increasingly applied to assess current impacts of environmental change on natural ecosystems and biodiversity (Elith & Leathwick, 2009; Franklin, 2010a). One of the most pressing challenges to moving forward the science of SDMs is accounting for the complexity of spatial autocorrelation (SAC) in ecological data and analyses of species distributions (Araujo & Guisan, 2006; Guisan *et al.*, 2006; Austin, 2007). While the origins and scales of spatial dependence in biogeographical observations are manifold, two types of factors are commonly recognized: a) exogenous factors associated with broad-scale spatial trends in underlying environmental conditions and b) endogenous factors caused by local-scale processes, such as dispersal limitation, metapopulation dynamics, or disturbance history (Storch *et al.*, 2003; Dormann, 2007b; Miller *et al.*, 2007). If ignored or misspecified, these factors can lead to autocorrelated residuals and violate the assumption of statistical independence in hypothesis testing and modeling of species distributions (Legendre, 1993; Lichstein *et al.*, 2002). The presence of SAC has been shown to inflate the significance of measured relationships in SDMs and bias model parameters when non-spatial models are applied to spatially structured data (Dormann, 2007b; Kuhn, 2007; Kissling & Carl, 2008). While ignoring SAC can potentially lead to flawed studies (Lennon, 2000; Beale *et al.*, 2007), SAC should also be thought of as a useful tool to examine species distribution processes at multiple spatial scales and improve spatial prediction (Dormann *et al.*, 2007).

Investigations of the role of SAC in models of species distributions have focused on native species (e.g. Miller, 2005; Segurado *et al.*, 2006) but SAC has not been examined in invasive species distribution models (iSDMs), a special type of SDMs that is being increasingly used to predict biological invasions and guide early detection (e.g. Lippitt *et al.*, 2008; Meentemeyer *et al.*, 2008a; Chytry *et al.*, 2009). Accounting for SAC in iSDMs may be particularly important for capturing fine-scale contagious processes of invasion that lead to geographically structured distributions and violate the assumption that species are in equilibrium with environmental controls (Vaclavik & Meentemeyer, 2009). The clustering of range expansion around introduction foci, especially in the early stages of invasion, leads to a mismatch between species' potential and realized distribution (Lobo *et al.*, 2010; Vaclavik *et al.*, 2010) and environmental predictors may exhibit explanatory power in statistical models simply because they are more similar at neighboring sites (Segurado *et al.*, 2006; De Marco *et al.*, 2008).

Invasive species are also constrained by dispersal limitations that prevent them from invading places that are environmentally suitable but isolated from already colonized locations. Although successful propagation is a major trait associated with species invasiveness (Pysek & Richardson, 2007), no organism produces globally dispersing offspring. The density of generated propagules and progeny usually declines with distance, resulting in spatially dependent occurrences (Dormann, 2007b). It has been suggested that dispersal processes responsible for SAC should be incorporated explicitly in distribution models (Franklin, 2010b). Previous studies attempted to account for dispersal limitations using dispersal kernels and distance decay formula to model propagule pressure as a function of distance (Havel *et al.*, 2002; Allouche *et al.*, 2008; Meentemeyer *et al.*, 2008a), or by using cellular automata to dynamically simulate species dispersal to new suitable habitats (Iverson *et al.*, 2004; Engler & Guisan, 2009). However, when data for estimating dispersal processes are unavailable or the knowledge of which factors lead to spatial patterns is absent, incorporating SAC at various scales may be crucial to account for dynamic processes in static iSDMs.

Still, most iSDMs, as well as SDMs in general, do not directly account for the effects of spatial dependence (Dormann, 2007b; Elith & Leathwick, 2009). A wider integration of SAC concepts in iSDMs would be facilitated by comprehensive studies that address three specific needs. First, the techniques that allow for explicit incorporation of SAC typically address only a single scale of spatial context. Although these methods, ranging from trend surface analysis (Lichstein *et al.*, 2002) and autoregressive models (Santika & Hutchinson, 2009) to geostatistical methods (Miller & Franklin, 2002; Miller, 2005) and geographically weighted regression (Kupfer & Farris, 2007), have recently experienced a surge in ecological applications, there is a gap in our knowledge on how multiple scales and origins of SAC affect the performance of iSDMs. Such knowledge is needed, especially when there is an indication that factors operating at various levels of geographical space (such as dispersal and colonization) are influential to species distributions (Elith & Leathwick, 2009).

Second, there is a lack of studies that use real and finely-resolved biogeographical data on invaders in relatively early stages of invasion. Rather, the majority of studies that have tested spatial models have used simulated datasets with simplistic inclusion of SAC through error terms (e.g. Dormann *et al.*, 2007; de Knegt *et al.*, 2010), although a few studies have used spatially-coarse biogeographical data for native taxa (e.g. Diniz-Filho *et al.*, 2003; Segurado *et al.*, 2006). We acknowledge the value of virtual data and controlled modeling environments to test statistical methods, but ultimately we need to know how SDM methods behave when confronted with real data on biological invasions.

Third, SAC complicates the key assumption of standard statistical analyses that residuals are independent and identically distributed. Previous studies that considered SAC were thus limited to correlative or regression based models (e.g. Dark, 2004; Bini *et al.*, 2009). As new machine-learning techniques are becoming prevalent in SDM applications (Elith & Leathwick, 2009), there is a need to study the implications of SAC not only in traditional statistical models but also in non-parametric methods.

In this study, we examine the effects of SAC on the performance of iSDMs in order to improve biogeographical prediction of an invasive organism. Expanding on suggestions by Guisan *et al.* (2006), we consider both global and local scales of spatial pattern in biogeographical analyses to assess the efficacy of using a multi-scale framework that accounts for different origins of SAC at various scales. We use an extensive dataset on the invasive plant pathogen *Phytophthora ramorum* to address the following research questions: (1) What are the effects of spatial autocorrelation on the accuracy and parameterization of predictive distribution models of an invasive organism? (2) How does the performance of spatially invariant models differ from models that deal with spatial dependence at a broad scale, a local scale, or multiple spatial scales? (3) Are parametric and non-parametric methods equally sensitive to spatial autocorrelation in biogeographical data? We focus on four approaches that either ignore spatial dependence or incorporate it at a broad scale using trend surface analysis, at a local scale using autocovariates, or across multiple scales using spatial eigenvector mapping. We evaluate model accuracies and amounts of explained SAC and examine the changes in variables' predictive power using both a conventional statistical method (generalized linear model; GLM) and a machine-learning technique (maximum entropy; MAXENT).

3.3 Methods

3.3.1 Study system – *Phytophthora ramorum*

The invasive plant pathogen *P. ramorum* is a new species of unknown origin that has spread across coastal forests in the western United States since the 1990's (Meentemeyer *et al.*, 2008a) and afflicted nurseries and plantations in several European countries (Brasier & Webber, 2010). In California and Oregon, *P. ramorum* causes the disease known as sudden oak death which has killed potentially millions of oak (*Quercus* spp.) and tanoak (*Notholithocarpus densiflorus*) trees and can attack over 40 other plant genera. Disease symptoms are expressed in two distinct forms: a canker infection that may cause tree mortality or a non-lethal foliar and twig infection (Rizzo & Garbelotto, 2003). Dispersal of the pathogen is driven by production of spores that form on the leaves of foliar hosts, such as bay laurel (*Umbellularia californica*), and are passively transmitted among trees and forest patches via rain splash and wind-blown rain

(Davidson *et al.*, 2005). To date, *P. ramorum* has established across approximately 10% of its host range (Fig. 1) with patchy distribution clustered around initial introductory foci (Condeso & Meentemeyer, 2007; Meentemeyer *et al.*, 2008a). The invasive character, relatively early stage of invasion, and spatially structured distribution at both local and state-wide scales makes *P. ramorum* an ideal organism for examining the effects of multi-scale SAC in iSDMs.

3.3.2 Species data

We compiled plot-level data on *P. ramorum* incidence collected in California and Oregon since 2001 for 3787 field plots established by researchers from the University of California Davis and University of North Carolina Charlotte (n=1370), Phytosphere Research (n=107), and Oregon Department of Forestry (n=1675). These monitoring projects were designed to gain baseline data on disease risk factors and to identify previously undetected invasion outbreaks. In addition, we acquired another large dataset managed by the California Oak Mortality Task Force (COMTF; Kelly *et al.*, 2004), which reports locations of *P. ramorum* infections (n=635) confirmed by the California Department of Food and Agriculture (see Meentemeyer *et al.*, 2008a). In each plot or reported location with symptomatic trees, canker or necrotic leaf tissues were isolated and cultured in the laboratory on pimaricin-ampicillin-rifampicin-pentachloro-nitrobenzene (PARP) agar, a selective media for *Phytophthora* species (Ivors *et al.*, 2004). As an additional test, samples that did not yield positive cultures were examined with a polymerase chain reaction (PCR)-based molecular assay, using primers designed to amplify *P. ramorum* DNA (Hayden *et al.*, 2004).The pathogen was considered present at a location if at least one sample yielded a positive culture or PCR detection of pathogen DNA. These procedures allowed a reliable collection and discrimination of presence (n=1673) and absence (n=2114) data on *P. ramorum* invasion between 2001 and 2009 (Fig. 1).

3.3.3 Environmental predictors

We quantified three climate, three topographical, and 11 host vegetation variables to characterize habitat heterogeneity and environmental conditions important for the establishment and spread of *P. ramorum*. For climate, we mapped precipitation, maximum temperature, and minimum temperature using 30-year monthly averages (1971–2001) derived from the parameter elevation regression on independent slopes model (PRISM; Daly *et al.*, 2001) at 800 m resolution. We aggregated the monthly grids to the rainy season (December to May), the pathogen's major reproductive period in California and Oregon (Davidson *et al.*, 2005), but also used maximum July and minimum January temperatures to account for the effects of climate extremes (Zimmermann *et al.*, 2009). For topography, we mapped elevation and derived potential solar irradiation (PSI) and topographic moisture index (TMI) from a U.S. Geological Survey 90-m digital elevation model. PSI was calculated as the mean potential direct radiation during the reproductive season using the cosine of illumination angle on slope equation (Dubayah, 1994). TMI characterizes the effect of local topography on soil moisture and was calculated as the natural log of the ratio between upslope drainage area and the slope gradient of a grid cell (Moore *et al.*, 1991). Lastly, we used CALVEG data summarized by Meentemeyer *et al.* (2004) for California and geospatial vegetation data developed using gradient nearest neighbor (GNN) imputation for Oregon (Ohmann & Gregory, 2002; Pierce *et al.*, 2009) to map the spatial distribution and abundance of two key infectious hosts, bay laurel and tanoak, as well as nine additional host species: *Arbutus menzeisii*, *Lonicera hispidula*, *Pseudotsuga menziesii*, *Quercus agrifolia*, *Quercus chrysolepis*, *Quercus kelloggii*, *Rhododendron* spp., *Sequoia sempervirens*, and *Vaccinium ovatum.*

3.3.4 Statistical methods

We applied one parametric and one non-parametric analytical method commonly used for prediction of species distributions. GLM is a parametric statistical approach that expands on common multiple regression, allowing for modeling non-normal response variables (McCullagh & Nelder, 1989). We used the logistic variant that employs maximum likelihood estimation to model the log odds of a binary response variable as a linear function of explanatory variables (Franklin, 1995; Guisan *et al.*, 2002). In contrast, MAXENT is a machine-learning algorithm that predicts species distributions by finding the probability distribution of maximum entropy that respects a set of constraints derived from sample locations. These constraints are calculated as functions of environmental variables with their means required to match the empirical average of occurrence sites (Phillips & Dudik, 2008).

3.3.4.1 Spatially invariant models

We fitted two spatially invariant models that took into account environmental variables but ignored the spatial context of input data. First, we modeled the probability of *P. ramorum* invasion as a function of the 18 environmental variables using GLM with a logit-link function and binomial error distribution. We calculated correlations between variables prior to statistical analysis and removed predictors with correlations higher than 0.5. We examined all possible subsets of explanatory variables and identified the best model based on logit R^2 (the uncertainty coefficient U), negative log-likelihood ratio test (LRT), and corrected Akaike Information Criterion (AICc) (sensu Burnham & Anderson, 2004). We tested pairwise interaction terms for significance but, for simplicity, higher order combinations of variables were not explored.

Second, we fitted a MAXENT model to estimate *P. ramorum* invasion by iteratively weighting each predictor variable to maximize the likelihood of reaching the optimum probability distribution. We used our presence locations for model calibration and absence locations as the background data representing the range of environmental conditions in the modeled region (Phillips *et al.*, 2009). The probability distribution was calculated as the sum of each weighted variable divided by a scaling constant to ensure the output range between 0 and 1. We selected 500 iterations for model convergence and employed the regularization procedure to prevent overfitting (Phillips & Dudik, 2008).

3.3.4.2 Incorporating broad-scale spatial structure

We modeled the geographical gradient in *P. ramorum* distribution via trend surface analysis (TSA) to account for a broad-scale spatially structured variation in the occurrence data (Maggini *et al.*, 2006). This approach assumes that a general spatial trend can be reasonably represented by a polynomial surface of closest fit to the observations, minimizing the difference between the interpolated value at a data location and its original value (Lichstein *et al.*, 2002). The trend surface was calculated as:

$$Z(U,V) = \alpha_{00} + \alpha_{10}U + \alpha_{01}V + \alpha_{20}U^2 + \alpha_{11}UV + \ldots + \alpha_{pq}U^pV^q \qquad (1)$$

where Z is the areally-distributed variable, in this case species presence and absence, αs are the polynomial coefficients, and U and V are the geographic coordinates (sensu Vaclavik & Rogan, 2009). We explored different degrees of polynomials and offered each resulting trend surface to the GLM and MAXENT models as an additional covariate, maintaining the same variable selection procedures as for spatially invariant versions.

3.3.4.3 Incorporating fine-scale spatial structure

We used an autocovariate (AC) method to account for fine-scale spatial variation in the data by estimating how much the response variable at every location reflects response values at surrounding locations (Dormann, 2007a). Although the computation of an autocovariate term is commonly done for lattice data without missing values, we applied this concept to irregularly distributed occurrence points, thus avoiding the need for applying the Gibbs sampler to optimize model performance (Augustin *et al.*, 1996). For every cell in our study area, we calculated the autocovariate as:

$$AC_i = \sum_{j \neq i}^{j \in N_i} w_j y_j , \qquad \text{with } w_j = d_{ij}{}^{-1} \tag{2}$$

where an autocovariate at location *i* is defined as a weighted sum of observation records *y* at already invaded locations *j* in a neighborhood determined by N_i (Miller *et al.*, 2007). We set the weight of the site *j* in relation to site *i* to be the inverse Euclidean distance and defined two types of neighborhoods N_i. First, we set the neighborhood size to the entire study area, including all calibration presence sites in the calculation. Second, we applied Ripley's K-function to the presence data to determine the amount of clustering at various scales (Ripley, 1976). The scale (i.e. the search radius) that exhibited the highest amount of spatial aggregation was used to define the second neighborhood size (Delmelle *et al.*, 2010). We incorporated each autocovariate as an additional explanatory variable both in MAXENT and the GLM while maintaining the same variable selection procedures.

3.3.4.4 Incorporating multi-scale spatial structure

We used spatial eigenvector mapping (SEVM) to account for spatial structure at multiple scales simultaneously. This method assumes that the spatial arrangement of data can be translated into a set of predictor variables, which capture spatial relationships among sites at different resolutions (Diniz-Filho & Bini, 2005). First, we constructed a pairwise matrix of Euclidean distances among geographical locations of our calibration presence sites: $D = [d_{ij}]$. We chose a truncation threshold t to calculate the connectivity matrix W based on the rule:

$$W = (w_{ij}) = \begin{cases} 0 & \text{if } i = j \\ 0 & \text{if } d_{ij} > t \\ \left[1 - (d_{ij} / 4t)^2\right] & \text{if } d_{ij} \le t \end{cases} \tag{3}$$

where t is the distance that maintains the connections among sample sites (Griffith & Peres-Neto, 2006). To give more weights to short-distance effects, we defined the truncation distance using the intercept of Moran's I correlograms (see below) for residuals from spatially invariant models. The connectivity matrix was submitted to a principal coordinate analysis of neighbour matrices to compute eigenvectors from the double-centred W matrix (Dray *et al.*, 2006). We extracted first 50 eigenvectors with positive eigenvalues that represented both global (high eigenvalues) and local (low eigenvalues) spatial variation in our data. As these spatial filters were mutually orthogonal, we entered them successively as additional predictors in GLM and MAXENT models, checking for model stability and retaining those filters that (i) exhibited a significant relationship with *P. ramorum* invasion and (ii) maximized the model fit (Diniz-Filho & Bini, 2005).

3.3.5 Assessment of model performance

To ensure the data used for evaluation of the eight models were independent from the data used for calibration, we followed the Huberty's heuristic (Fielding & Bell, 1997) and used *k*=4 to randomly divide our occurrence dataset into *k* independent partitions, using *k*-1 (75% of data) for model calibration and computation of spatial terms, and the remaining partition (25% of data) for model assessment, while repeating this procedure *k* times (Hirzel *et al.*, 2006). For each cross-validation run, we calculated the area under the curve (AUC) of the relative operation characteristic (ROC) to examine the true positive rate as a function of the false positive rate at each probability threshold predicted by the models (Pontius & Schneider, 2001). To compensate for the possibility that locations of the same likelihood values are calculated at different thresholds (Lobo *et al.*, 2008), we additionally calculated (i) omission and commission error rates at the threshold that maximized model specificity and sensitivity (Freeman & Moisen, 2008) and (ii) the true skill statistic (TSS), a prevalence-independent modification of the Kappa statistic, that measures the overall accuracy of presence-absence predictions while correcting for the accuracy expected to occur by chance (Allouche *et al.*, 2006).

3.3.6 Investigation of SAC effects

The amount of SAC in model residuals (i.e. Pearson residuals for GLM and observed occurrence–probability of occurrence for MAXENT) was quantified using spatial correlograms to identify Moran's *I* values for multiple lags. We used 200 permutations to test Moran's *I* significance and Bonferroni correction to adjust for repeated testing (Diniz-Filho & Bini, 2005). We investigated the changes in explanatory power of environmental variables via significance testing of parameter estimates (for

GLM) and jackknife testing of the relative contribution to model gain (for MAXENT; Phillips *et al.*, 2006). To examine potential shifts in the relative importance of explanatory variables, we ranked all variables in each model and compared them using Spearman correlation coefficients. A perfect positive correlation would indicate that incorporating SAC did not change the relative importance of environmental variables even when it changed the parameter estimates (Bini *et al.*, 2009). Finally, we calculated Pearson correlation coefficients between variables' global Moran's *I* and their change in explanatory power, in order to examine whether spatially dependent variables are more prone to experience a shift in their effect when a spatial component is included in the analysis.

3.4 Results

Application of the best performing models in the GIS produced geographic predictions of *P. ramorum* current invasion probability across heterogeneous forests in California and Oregon (Fig. 2). The following spatial components provided the best model fit in corresponding spatial models: (i) a trend surface based on a third order polynomial function, (ii) an autocovariate calculated from the neighborhood size N_i=8000 m defined by Ripley's K function, and (iii) a total of 23 spatial filters with positive eigenvalues computed from a matrix with truncation distance of 8000 m defined by Moran's *I* correlograms.

3.4.1 Model performance

Comparison of the AUC values shows that spatial models always produced more accurate predictions than spatially invariant versions (Fig. 3a). Accounting for a broad spatial trend (using TSA) significantly improved the performance of both GLM and

MAXENT models (increase in AUC by 0.085 and 0.062 respectively) but incorporating a fine-scale spatial context (using AC) generated higher predictive accuracies (increase in AUC by 0.137 and 0.115 respectively). The models with most accurate predictions were those that accounted for spatial structure at multiple spatial scales using SEVM (AUC values of 0.958 for GLM and 0.918 for MAXENT), although the improvement was not as dramatic when compared to autocovariate models (difference in AUC 0.012 for GLM and 0.014 for MAXENT). Converting predicted probabilities to binary presence/absence predictions showed almost identical trends in accuracies (Figs 3b and 3c), with SEVM models yielding the highest TSS values (0.801 for GLM and 0.711 for MAXENT) and lowest commission/omission error rates (0.10/0.09 for GLM and 0.15/0.12 for MAXENT). Spatially invariant models and TSA models produced generally higher commission error rates while AC and SEVM models had lower rates of both commission and omission errors (Fig. 3c).

Moran’s *I* correlograms indicated significant amounts of residual SAC in most models up to the lag of 8000 m (Fig. 4). The highest Moran’s *I* was measured at the lag of 300–400 m ($I = 0.95$, $P < 0.01$) for a spatially invariant GLM (Fig. 4a) and 200–300 m ($I = 0.95$, $P < 0.01$) for a spatially invariant MAXENT model (Fig. 4b). Incorporating spatial terms in GLM substantially decreased the amount of residual autocorrelation (TSA: $I = 0.55$, $P < 0.01$; AC: $I = 0.47$, $P < 0.01$) or removed it completely in the case of SEVM ($I = 0.02$, $P = 0.69$). There was no significant difference in residual SAC among spatially invariant and spatially structured MAXENT models, with the exception of SEVM, which experienced a slight decrease ($I = 0.79$, $P < 0.01$).

3.4.2 Explanatory power and relative importance of variables

Based on significance levels (P < 0.05) and jackknife tests of variable importance (regularized model gain > 0.1), elevation, PSI, precipitation, minimum temperature, maximum July temperature, and abundances of tanoak, bay laurel, coast live oak, and redwood were all significant predictors of *P. ramorum* presence in both types (GLM, MAXENT) of spatially invariant models and in TSA models. The same set of variables, but excluding live oak and redwood abundances, was selected when we accounted for fine-scale SAC in GLMs with AC or SEVM terms. In addition, the interaction term for tanoak abundance and precipitation was significant in all GLMs except for the SEVM model.

The differences in absolute values of standardized beta coefficients and the regularized model gain show that the explanatory power of most factors decreased when SAC was incorporated in the modeling process (Fig. 5). The highest coefficient differences were detected when fine-scale SAC was included via autocovariate terms. In GLM, the only variable that experienced an increase in explanatory power after accounting for all three levels of SAC was maximum temperature. The coefficient increase in tanoak in TSA and AC models was caused by the counteracting effects of its interaction with precipitation. In MAXENT models, all variables underwent a decrease in explanatory power when any level of spatial structure was added.

Relatively high Spearman correlation coefficients (0.77 – 0.93, P < 0.05) show that there were only minor shifts in relative importance of predictor variables when SAC was included (e.g. a swap of two factors in their order of importance) (Table 1). While the similarity in variable importance among spatially invariant GLM and MAXENT models was low and statistically insignificant (0.35, P = 0.36), adding a spatial component made the relative importance more consistent among corresponding spatial versions of GLM and MAXENT (0.57 – 0.79, P < 0.05). In addition, the correlation between global Moran's *I* of individual variables and their change in explanatory power showed that variables with strong spatial dependence were more likely to experience a shift in their effect when a spatial component was added in GLMs (Table 2). In MAXENT models, the amount of SAC in variables was not correlated with the magnitude of changes in their explanatory power.

3.5 Discussion

3.5.1 Model performance and implications for iSDMs

Comprehensive research addressing the effects of SAC on model performance is needed not only to improve spatial predictions of biological invasions but also to advance ecological conceptualization of SDM as a discipline. Our results confirm the hypothesis that spatially invariant models of *P. ramorum* invasion probability exhibit lower predictive accuracy than models that incorporate spatial structure. Improved model performance after a trend surface was included indicates that a broad-scale factor affecting the pathogen's distribution was likely omitted in our analysis. It has been suggested that where a species distribution is largely determined by environmental factors and the spatial structure in the response variable is dependent on spatial structure in predictor variables, a properly specified model fitted with correct covariates would reduce the need to account for SAC (Austin, 2002). We argue this situation is implausible for invasive organisms because their distributions are driven by factors other than environmental, and in most cases the perfect set of direct predictors is not available at a required biological accuracy and spatio-temporal resolution (Dormann, 2007b). As a result, the use of indirect gradients (e.g. temperature averages at a coarse spatial resolution) and surrogate factors (e.g. trend surfaces) remain important for explaining the distributions of invasive species.

Dramatic improvements in model performance and residual SAC were observed when we accounted for fine-scale spatial dependence. This finding illustrates the unique character of modeling non-equilibrium species and suggests that local range-confining processes have a strong influence on *P. ramorum* spread. Our conclusion is supported by high levels of SAC at relatively short lags (200–400 m), a scale at which the disease is known to be clustered (Condeso & Meentemeyer, 2007), and is consistent with studies that recognized dispersal limitations and propagule pressure to be essential for predicting biological invasions (e.g. Rouget & Richardson, 2003; Vaclavik & Meentemeyer, 2009). Indeed, previous studies of *P. ramorum* distribution using dispersal kernels (Meentemeyer *et al.*, 2008a; Ellis *et al.*, 2010) or simple metrics of distance to formerly invaded sites (Vaclavik *et al.*, 2010) showed that dispersal pressure was a better indicator of pathogen’s distribution than environmental factors. As the explicit estimation of dispersal is not always feasible, incorporating fine-scale SAC may represent an alternative way to constrain predictions, especially when invaders are in initial stages of colonization. The high commission error rates of spatially invariant and TSA models show that, when local spatial context is ignored, models largely over-predict the invaded range. This is a highly undesirable output for conservation management if the aim is to quantify the actual extent of invasion spread and target locations for early detection surveillance and control.

3.5.2 Explanatory power and relative importance of variables

Incorporating SAC altered the fitted species-environment relationships and changed the explanatory power of predictor variables. While the relative importance of environmental factors was generally consistent among spatial and invariant models, the explanatory power of most variables decreased when we accounted for spatial structure. The correlations for GLMs indicated that variables with strong spatial structure tend to show weaker effects when SAC is explicitly modeled. Although Bini *et al.* (2009) were unable to identify reliable predictors of coefficient shifts in spatial regression, our results corroborate findings of other studies (e.g. Lichstein *et al.*, 2002; Tognelli & Kelt, 2004), suggesting that ignoring SAC leads to overestimation of environmental effects.

The most marked changes in variable explanatory power were identified in AC models, probably due to their tendency to overcompensate for SAC. While an autocovariate approach is considered a viable option to increase prediction success, developers of this method cautioned against its use for inference (Augustin *et al.*, 1996). Although we do not know the 'true' parameter estimates for *P. ramorum*, our results coincide with findings by Dormann (2007a) that demonstrated consistent underestimation of the effects of environmental variables in autologistic models. Less dramatic shifts were exhibited by TSA and SEVM models. As both methods work as spatial filters, their results may sometimes be similar (Diniz-Filho & Bini, 2005), but TSA assumes a simple broad-scale gradient and cannot account for localized processes of invasion. Moreover, TSA is dependent on a sampling design that should be close to regular and requires the use of only one polynomial surface to avoid colinearity possibly hindering model selection.

SEVM produced the most accurate predictions in our analysis and removed SAC from residuals in GLMs, but the interpretation of parameter estimates can be problematic. Although parameters adjusted for SAC are universally considered more reliable, Bini *et al.* (2009) conclude that if coefficients vary among different algorithms used to model spatial structure and depend on their complex interactions with multivariate datasets, it becomes difficult to argue for the primacy of one variable over another. Some of these issues may be effectively resolved with simulation procedures (Dormann *et al.*, 2007; Kissling & Carl, 2008), but artificial datasets often include simplistic forms of SAC. Simulations that incorporate more realistic spatial patterns and compare them with real data on invasive species in different stages of invasion could be an interesting avenue for future research.

3.5.3 Method consideration

Results indicate that SAC may affect parametric statistical methods differently than non-parametric techniques. While incorporating spatial terms decreased problematic SAC in residuals of GLMs, the amount of residual SAC was nearly the same for MAXENT models. However, the assumption of independent and identically distributed residuals applies primarily to statistical methods based on probability theory, and the effects of SAC on variable selection and explanatory power in machine-learning approaches are unclear. In our analyses, the changes in predictive power in MAXENT models were not associated with the amount of spatial structure in variables as in GLMs. However, Spearman correlations showed that the relative variable importance became more consistent among corresponding GLM and MAXENT models when a spatial component was included.

Different sensitivities of parametric and non-parametric methods to SAC are consistent with theoretical expectations because non-parametric methods place fewer constraints on the shape of fitted species-environment relationships and apply stronger adjustments to predictions (Segurado *et al.*, 2006). While the complex nature of machine-learning methods has perhaps limited their use for ecological inference, they provide valuable tools for SDM applications focused more on prediction. In our case, the incorporation of SAC in MAXENT models substantially improved their prediction accuracy. This finding supports our assumption about the origins of SAC in the data and suggests that SAC may be a crucial surrogate for dispersal and colonization processes that determine the spatial distribution of biological invasions in a non-equilibrium state.

3.6 Conclusions

Identifying geographic distributions of invasive species and diseases is essential for eliminating their impacts on natural ecosystems and developing effective control strategies (Holdenrieder *et al.*, 2004). iSDM can be useful for management of biological invasions, but understanding the effects of multi-level spatial structure on predictive models is needed to improve their efficacy and interpretation. We analysed a unique set of empirical data on the incidence of the invasive forest pathogen *P. ramorum* to examine whether predictive performance of static distribution models can be enhanced by accounting for dynamic contagious processes of invasion via incorporating various levels of spatial structure. We conclude that:

1. Accounting for a multi-scale structure of SAC significantly enhanced predictive capability of iSDMs.

2. Model performance improved when a trend surface was incorporated, suggesting that a broad-scale environmental gradient was likely unexplained by considered predictors.
3. Models were even more accurate when they accounted for fine-scale spatial structure, indicating that local range-confining processes are driving *P. ramorum* invasion. When local spatial context was ignored, models tended to over-predict the invaded range.
4. The relative importance of environmental variables was generally consistent across all models, but the explanatory power decreased in spatial models for factors with strong spatial structure.
5. While accounting for SAC reduced the amount of residual autocorrelation for GLM but not for MAXENT, it still improved performance of both methods. This supports the assumption that dispersal and colonization processes are important for spatial distribution of invasive organisms.

Spatial autocorrelation has become a paradigm in biogeography and ecological modeling. Apart from being vital to avoid common problems in multivariate statistical analyses, we showed that spatial pattern may be an important surrogate for processes that explain the ecological mechanisms of invasion and improve the predictive performance of iSDMs.

3.7 Acknowledgements

We are deeply grateful to our collaborators, researchers and field assistance from UNC Charlotte, UC Davis, Sonoma State University, Phytosphere Research, Oregon Department of Forestry, Oregon State University, USDA Forest Service, and the

California Oak Mortality Task Force, for their incredible effort put into data collection and compilation of our datasets. We thank D. Rizzo, J.H. Cushman, N. Rank, A. Kanaskie, T. Swiecki, and J. Ohmann for their previous work that made this study possible and members of the Center for Applied GIScience at UNC Charlotte for their helpful comments on the manuscript. This research was supported by grants from the National Science Foundation (EF-0622677) and USDA Forest Service–Pacific Southwest Research Station (08-JV-11272138-055).

Tables

Table 1: Spearman correlations of relative variable importance among different versions of spatial and non-spatial models. High values signify high consistency in the order (rank) of variable importance. (GLM=generalized linear model, MAXENT=maximum entropy, TSA=trend surface analysis, AC=autocovariate, SEVM=spatial eigenvector mapping)

Model	**Model**	**Spearman ρ**	**Prob>|ρ|**
GLM	GLM-TSA	**0.8303**	0.0029
GLM	GLM-AC	**0.8333**	0.0102
GLM	GLM-SEVM	**0.8571**	0.0137
MAXENT	MAXENT-TSA	**0.9273**	0.0001
MAXENT	MAXENT-AC	**0.6646**	0.036
MAXENT	MAXENT-SEVM	**0.8303**	0.0029
GLM	MAXENT	0.35	0.3558
GLM-TSA	MAXENT-TSA	**0.7167**	0.0298
GLM-AC	MAXENT-AC	**0.7857**	0.0362
GLM-SEVM	MAXENT-SEVM	**0.6714**	0.0402

Table 2: Pearson correlations between the amount of SAC in environmental variables (measured by global Moran's *I*) and their shift in explanatory power for each spatial model. (GLM=generalized linear model, MAXENT=maximum entropy, TSA=trend surface analysis, AC=autocovariate, SEVM=spatial eigenvector mapping)

Model	**Pearson r**	**Prob>|ρ|**
GLM-TSA	**0.3036**	0.0490
GLM-AC	**0.7608**	0.0410
GLM-SEVM	**0.9137**	0.0034
MAXENT-TSA	-0.037	0.6445
MAXENT-AC	0.0273	0.7506
MAXENT-SEVM	-0.0212	0.6640

Figures

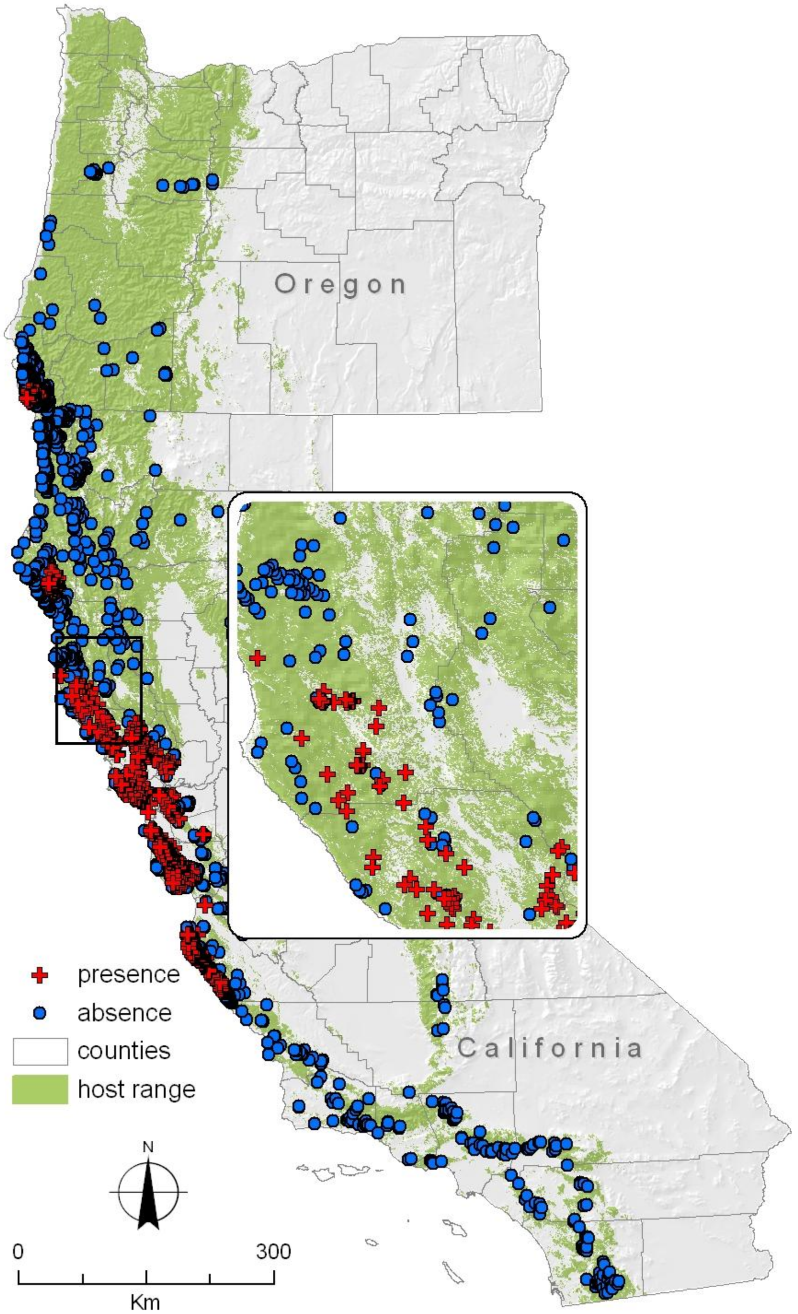


Figure 1: Map of 3787 field plots surveyed for the presence of *Phytophthora ramorum* in susceptible forests across California and Oregon between 2001 and 2009 (presence: n=1673, absence: n=2114)

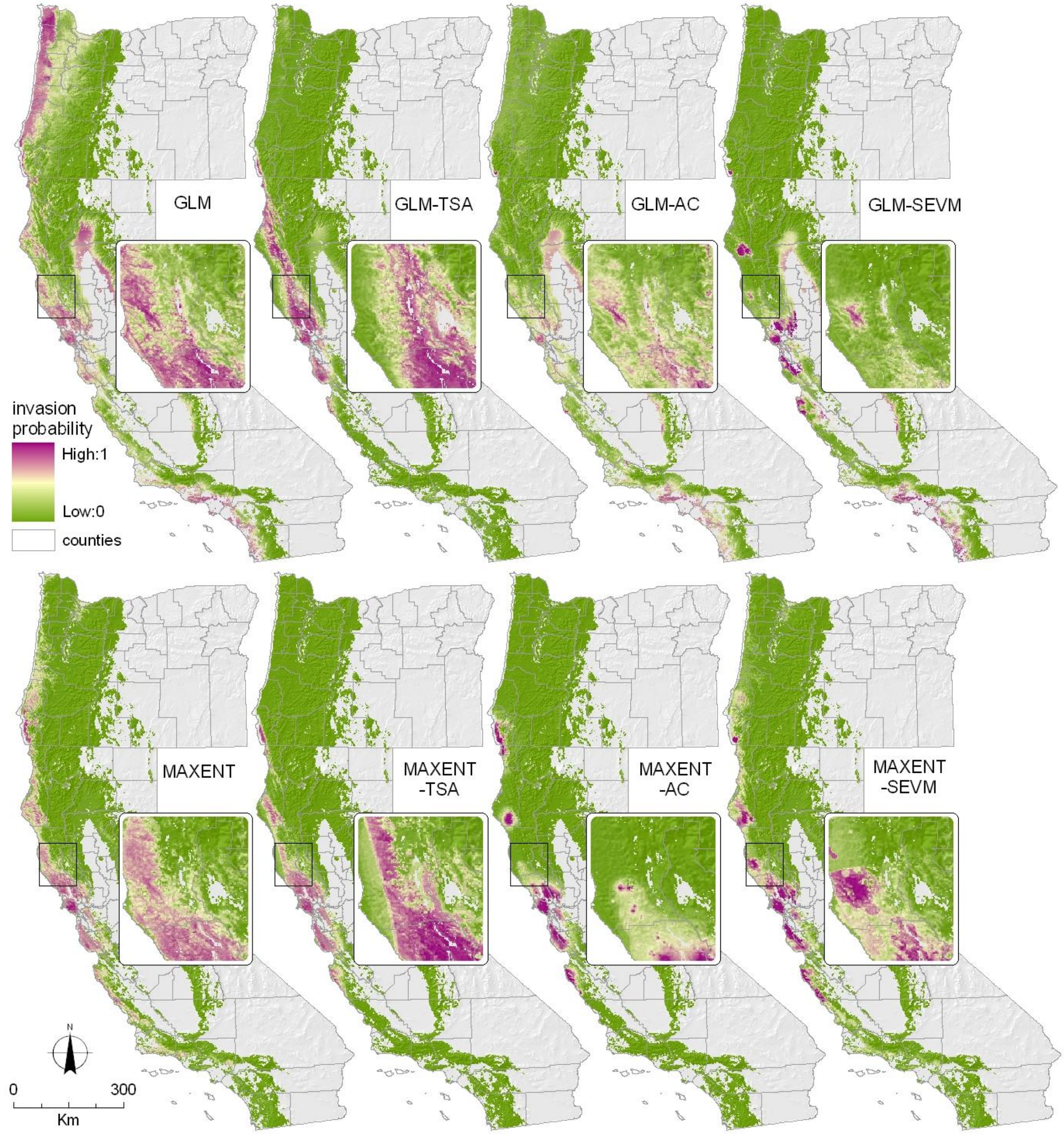
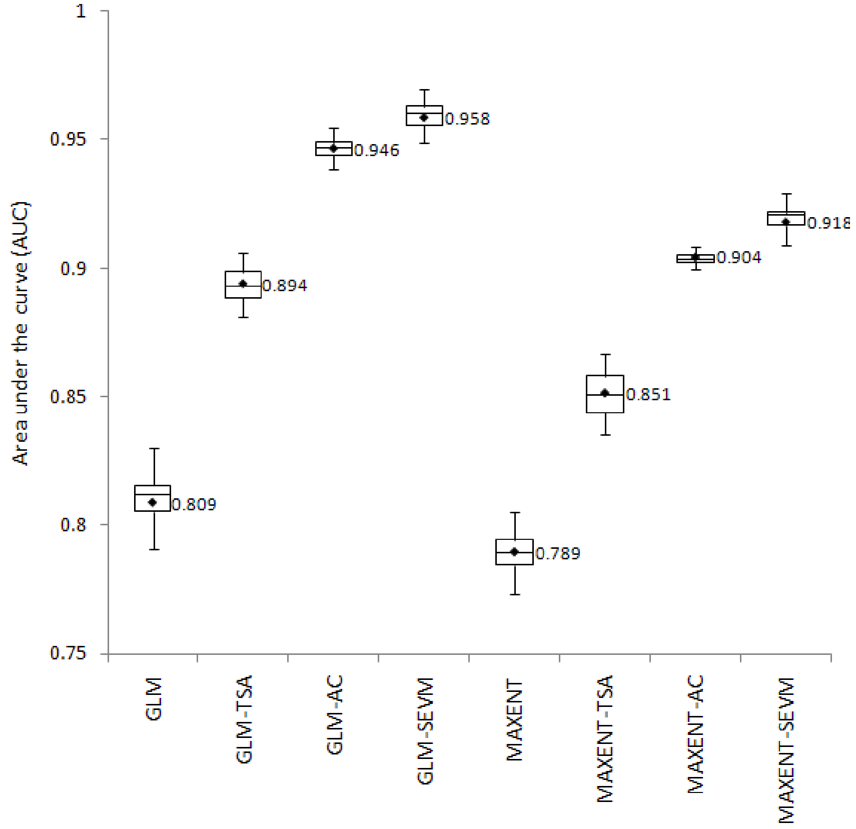
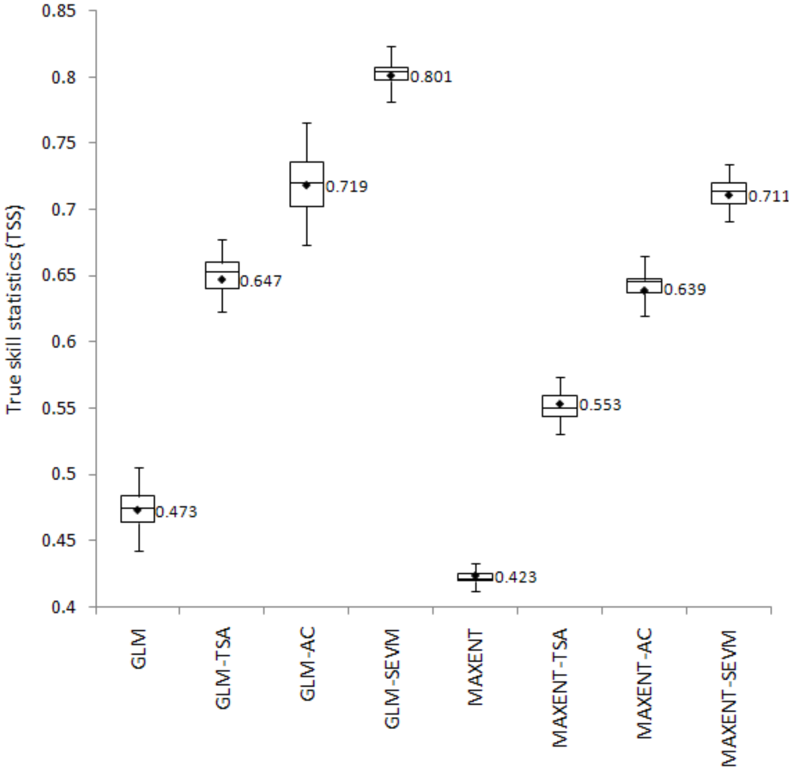


Figure 2: Distribution of *Phytophthora ramorum* invasion probability (0–1 scale) predicted by spatially invariant and spatially structured models. (GLM=generalized linear model, MAXENT=maximum entropy, TSA=trend surface analysis, AC=autocovariate, SEVM=spatial eigenvector mapping)

(a)


Area under the curve (AUC)
1
0.95
0.9
0.85
0.8
0.75
0.809
0.894
0.946
0.958
0.789
0.851
0.904
0.918
GLM
GLM-TSA
GLM-AC
GLM-SEVM
MAXENT
MAXENT-TSA
MAXENT-AC
MAXENT-SEVM

(b)


True skill statistics (TSS)
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0.45
0.4
0.473
0.647
0.719
0.801
0.423
0.553
0.639
0.711
GLM
GLM-TSA
GLM-AC
GLM-SEVM
MAXENT
MAXENT-TSA
MAXENT-AC
MAXENT-SEVM

(c)

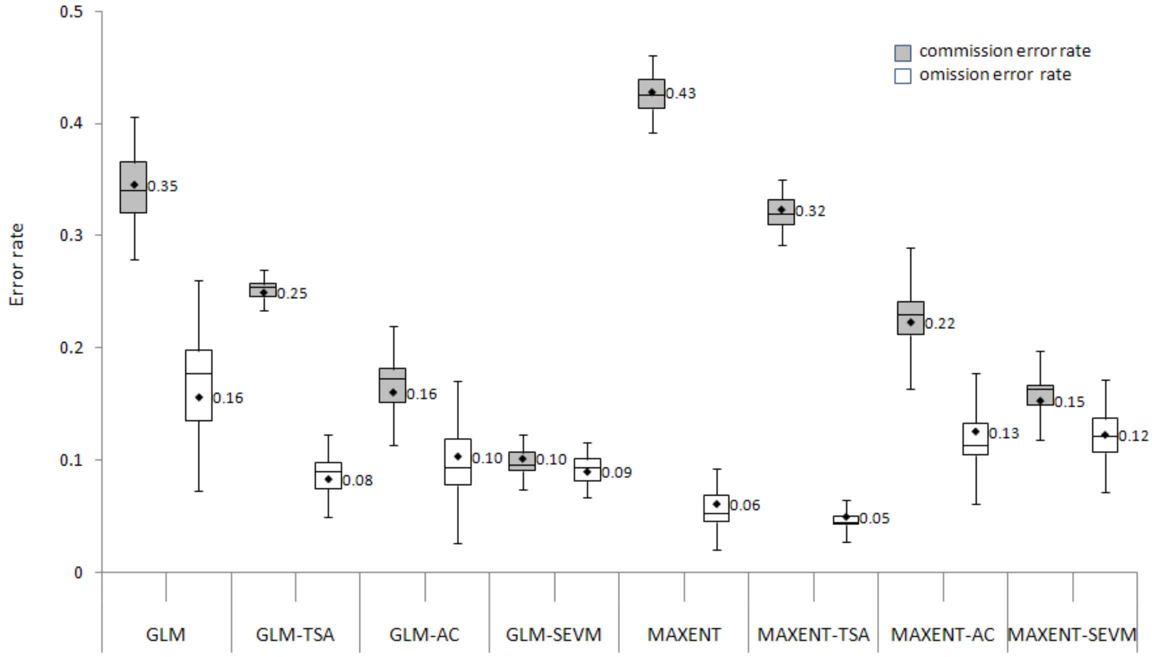


Figure 3: (a) Threshold-independent assessment of model accuracy measured by the area under the curve of the receiver operating characteristic; (b) threshold-dependent assessment of model accuracy maximizing sensitivity and specificity of models measured by the true skill statistic (TSS) and (c) commission and omission error rates. (GLM=generalized linear model, MAXENT=maximum entropy, TSA=trend surface analysis, AC=autocovariate, SEVM=spatial eigenvector mapping)

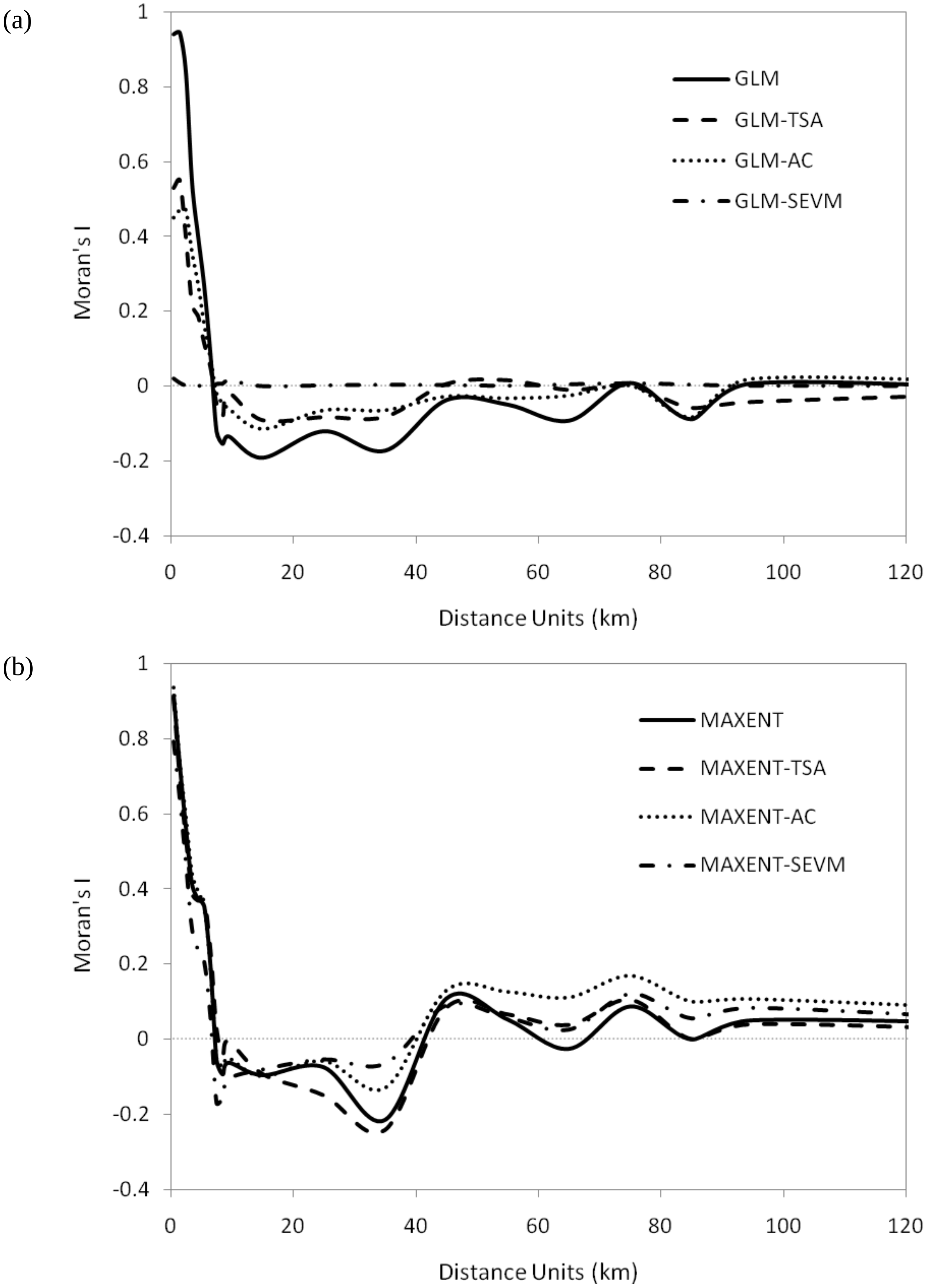


Figure 4: Moran's *I* correlograms of residual spatial autocorrelation for (a) GLM and (b) MAXENT spatial and non-spatial models. (TSA=trend surface analysis, AC=autocovariate, SEVM=spatial eigenvector mapping)

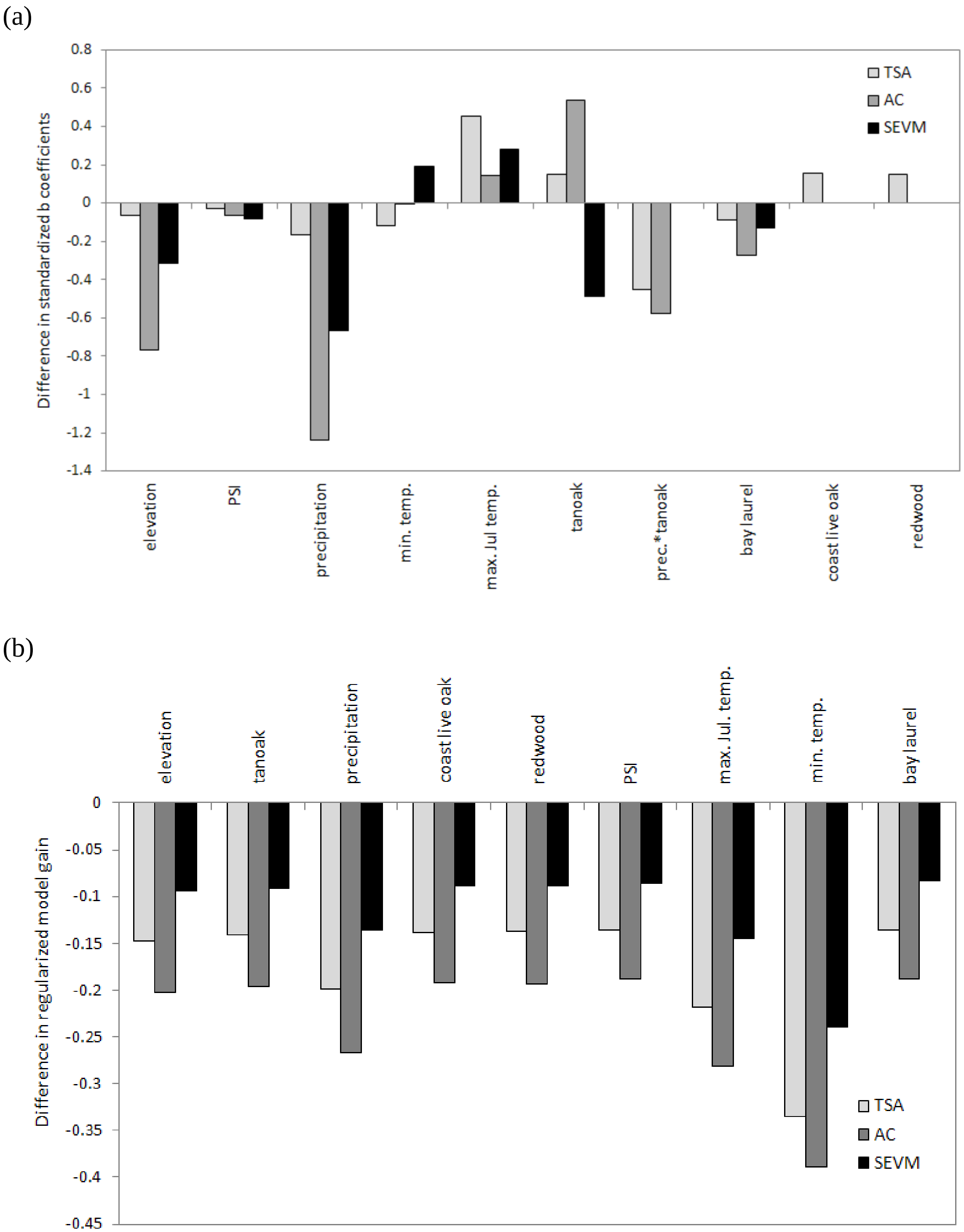


Figure 5: Difference in variable explanatory power among spatial and non-spatial models measured by absolute values of (a) standardized beta coefficients for GLM models and (b) regularized model gain for MAXENT models. Negative values signify a decrease in explanatory power. (TSA=trend surface analysis, AC=autocovariate, SEVM=spatial eigenvector mapping)

CHAPTER 4: PREDICTING POTENTIAL AND ACTUAL DISTRIBUTION OF SUDDEN OAK DEATH IN OREGON. PRIORITIZING LANDSCAPES FOR EARLY DETECTION AND ERADICATION

4.1 Abstract

An isolated outbreak of the emerging forest disease sudden oak death was discovered in Oregon forests in 2001. Despite considerable control efforts, disease continues to spread from the introduction site due to slow and incomplete detection and eradication. Annual field surveys and laboratory tests between 2001 and 2009 confirmed a total of 802 infested locations. Here, we apply two invasive species distribution models (iSDMs) of sudden oak death establishment and spread risk to target early detection and control further disease spread in Oregon forests. The goal was to develop (1) a model of *potential distribution* that estimates the level and spatial variability of disease establishment and spread risk for western Oregon, and (2) a model of *actual distribution* that quantifies the relative likelihood of current invasion in the quarantine area. Our predictions were based on four groups of primary parameters that vary in space and time: climate conditions, topographical factors, abundance and susceptibility of host vegetation, and dispersal pressure. First, we used multi-criteria evaluation to identify large-scale areas at potential risk of infection. We mapped and ranked host abundance and susceptibility using geospatial vegetation data developed with gradient nearest neighbor imputation. The host vegetation and climate variables were parameterized in accordance to their epidemiological importance and the final appraisal scores were

summarized by month to represent a cumulative spread risk index, standardized as five categories from very low to very high risk. Second, using the field data for calibration we applied the machine-learning method, maximum entropy, to predict the actual distribution of the sudden oak death epidemic. The dispersal pressure incorporated in the statistical model estimates the force of invasion at all susceptible locations, allowing us to quantify the relative likelihood of current disease incidence rather than its potential distribution. Our predictions show that 65 km^2 of forested land was invaded by 2009, but further disease spread threatens more than 2100 km^2 of forests across the western region of Oregon (very high and high risk). Areas at greatest risk of disease spread are concentrated in the southwest region of Oregon where the highest densities of susceptible host species exist. This research identifies high priority locations for early detection and invasion control and illustrates how iSDMs can be used to analyze the actual versus potential distribution of emerging infectious disease in a complex, heterogeneous ecosystem.

4.2 Introduction

The rapid spread of invasive organisms and emerging infectious diseases is one of the most important ecological outcomes from the drastic alteration of natural environments by human activities (Vitousek *et al.*, 1996; Foley *et al.*, 2005). In our highly globalized world, only few habitats have been spared from the detrimental impacts of biological invasions on biodiversity, community structure, nutrient cycling, or ecosystem productivity (Mack *et al.*, 2000; Hoffmeister *et al.*, 2005). In managed landscapes, human-induced invasions of exotic organisms and pathogens cause enormous economic losses by threatening our efforts to sustain agricultural production and maintain healthy forest ecosystems (Daszak *et al.*, 2000; Pimentel *et al.*, 2000). Despite intense preventive actions, invaders manage to affect extensive landscapes often due to slow and incomplete discovery of invasion outbreaks. As early detection crucially enhances the efficacy of invasion control and eradication treatments (Simberloff, 2003), there is an increasing need for predictive tools that identify the current geographic extent of invasion spread and the habitats at potential risk of invasion (Thuiller *et al.*, 2005; Franklin, 2010a).

Predicting the spatial distribution of invaders and pathogens is enormously challenging in heterogeneous environments. However, species distribution models (SDMs) that characterize the ecological niche of organisms and relate it to known environmental factors have provided an effective analytical framework for predicting the spread of biological invasions (e.g., Lippitt *et al.*, 2008; Chytry *et al.*, 2009; Strubbe & Matthysen, 2009). To develop invasive species distribution models (iSDMs), two approaches have been generally adopted, although their distinction has often been unclear in the literature. First, researchers predict the *potential distribution* of a biological invasion by identifying locations with environmental conditions potentially suitable for growth and reproduction, in which the invader could exist (Hirzel & Le Lay, 2008; Jeschke & Strayer, 2008). Second, researchers estimate the *actual distribution* of a biological invasion by identifying areas where the invader currently exists, constrained not only by environmental factors but also by colonization time lag and dispersal limitations (Soberon, 2007; Jimenez-Valverde *et al.*, 2008). While the first approach has been used to target various ecosystems potentially threatened by invasive organisms and diseases (Meentemeyer *et al.*, 2004; Lippitt *et al.*, 2008), or to understand the behavior of invaders in novel landscapes (Peterson *et al.*, 2003; Sutherst & Bourne, 2009), the second approach is essential for quantifying the actual range of invasions and predicting their extant consequences in specific environments (Meentemeyer *et al.*, 2008a; Vaclavik & Meentemeyer, 2009). Although knowledge from both types of spatial models can be extremely useful for guiding the management of biological invasions, no studies to date have used both approaches simultaneously to prioritize landscape contexts for early detection surveillance and invasion control.

In this study, we model and map the potential and actual distribution of sudden oak death (SOD) disease in western Oregon. An isolated outbreak of this emerging forest disease, caused by the invasive plant pathogen *Phytophthora ramorum*, was discovered in Oregon forests in 2001 (Hansen *et al.*, 2008), more than 200 km from the closest documented infection in Humboldt County, California. *P. ramorum* causes significant mortality of tanoak (*Notholithocarpus densiflorus*) and oak (*Quercus* spp.) trees and infects a wide range of other plant species, such as Oregon myrtle (*Umbellularia californica*), Pacific rhododendron (*Rhododendron macrophyllum*), and evergreen huckleberry (*Vaccinium ovatum*), considerably altering the composition and structure of forest communities and changing ecosystem processes (Meentemeyer *et al.*, 2008b; Cobb *et al.*, 2010; Davis *et al.*, 2010). The disease symptoms are expressed in two distinct forms, either as lethal infections in canker hosts that serve as epidemiological dead-ends or as non-lethal infections in foliar hosts that produce large amounts of infectious spores on necrotic leaves (Garbelotto *et al.*, 2003; Rizzo & Garbelotto, 2003). These spores are passively transmitted among individual trees and forest patches via rain-splash and wind-driven rain (Davidson *et al.*, 2005), affecting considerable forest area with susceptible host species and favorable environmental conditions.

In contrast to relatively wide distribution throughout California, the pathogen occurs in Oregon only in one small area in Curry County near the town of Brookings (Kanaskie *et al.*, 2009a). Despite substantial control efforts consisting of cutting and burning infected and potentially exposed host plants, and applying herbicide to prevent tanoak sprouting, *P. ramorum* continues to spread from the initial infested sites (Hansen *et al.*, 2008; Kanaskie *et al.*, 2009b). In 2007, SOD quarantine area was extended to current 420 km^2 due to the emergence of six new outbreaks found outside the original 65 km^2 quarantine boundary. The abrupt disease expansion is attributed to several consecutive years of unusually wet and warm weather that promotes long distance dispersal of the pathogen (Davidson *et al.*, 2005; Rizzo *et al.*, 2005). However, it is believed that the major reason why control activities have been only partially successful is the late discovery of disease outbreaks, which propagated across forested landscapes before typical disease symptoms were recognized and infected sites treated (Goheen *et al.*, 2009; Kanaskie *et al.*, 2009c).

Successful containment of SOD depends heavily on early detection, so the pathogen can be destroyed before it can intensify and spread. Aerial surveys searching for dead and dying trees are good detection tools but their effectiveness largely depends on the degree of latency of disease symptoms. Field surveys and stream baiting with subsequent laboratory analyses can detect an infestation in a very early stage but represent labor intensive and costly methods. Predictive risk models thus offer important alternatives for prioritizing areas for early detection and eradication treatments. Although predictive models of *P. ramorum* establishment and spread have been developed and used in California (Meentemeyer *et al.*, 2004; Meentemeyer *et al.*, 2008a), similar modeling has been limited in Oregon due to unavailable vegetation data. This situation now has been remedied by new spatial vegetation data (Ohmann & Gregory, 2002; Ohmann *et al.*, 2007) that allow us to map host susceptibility characteristics across Oregon forests.

Here, we present spatial predictions of *P. ramorum* establishment and spread risk that are being actively used to target early detection and control further disease spread in Oregon forests. The goal of this study was to develop two predictive models: (1) a model of *potential distribution* that estimates the level and spatial variability of *P. ramorum* establishment and spread risk in six ecoregions in Oregon, and (2) a model of *actual distribution* that quantifies the relative likelihood of *P. ramorum* current invasion in the SOD quarantine area. Our predictions are based on GIS analysis of four groups of primary parameters that vary in space and time: climate conditions, topographical factors, abundance and susceptibility of host vegetation, and dispersal pressure. First, we built a heuristic model using multi-criteria evaluation (MCE) method to identify large-scale areas at potential risk of disease infection. Second, using extensive field data for model calibration and calculation of dispersal pressure we applied the machine-learning method, maximum entropy (MAXENT), to predict the actual distribution of the sudden oak death epidemic. Spatially-explicit models of potential and actual distribution of *P. ramorum* invasion in Oregon are urgently needed to provide a better picture of forest resources threatened by this destructive pathogen.

4.3 Methods

4.3.1 Field data collection

To examine factors influencing the spatial distribution of invasion probability of *P. ramorum*, we collected field data over the span of nine years throughout heterogeneous habitat conditions in southwest Oregon. The early detection program was coordinated by the Oregon Department of Forestry and the USDA Forest Service year-round since 2001, using a combination of fixed-wing and helicopter surveys and ground-based checks (Goheen *et al.*, 2006; Kanaskie *et al.*, 2009c). Each year, the forest landscape in southwest Oregon was systematically scanned from a helicopter to look for signs of dead or dying trees, covering the majority of the tanoak host type. Cases, in which apparent crown mortality was discovered, were recorded and mapped using sketch maps and the Global Positioning System (GPS), and followed by thorough field inspections. All mapped sites with tree mortality were visited and evaluated on the ground, although the difficulty in accessing some areas due to rugged terrain and other accessibility obstacles occasionally delayed field visits. Additional ground-based surveys were conducted in areas with known host vegetation because detecting *P. ramorum* from the air is impossible when the symptoms are restricted to necrotic lesions on leaves and twigs or external bleeding on the trunks of infected live trees with healthy-appearing foliage. Transect surveys were used to check for symptomatic vegetation in potential timber sale areas, along roadsides, popular hiking trails, and high-use campgrounds. Extensive watershed-level monitoring was done both inside and outside the quarantine area using stream baiting with tanoak and rhododendron leaves, followed by surveys to locate infected plants when stream baits detected *P. ramorum* (Sutton *et al.*, 2009).

Symptomatic host plants were checked for infection by: (1) isolating and transferring symptomatic tissue directly onto plates with a selective media for *Phytophthora* species, and (2) analyzing samples in Oregon State University and Oregon Department of Agriculture laboratories via traditional culturing and a polymerase chain reaction (PCR)-based molecular assay, using primers designed to amplify *P. ramorum* DNA (Ivors *et al.*, 2004; Goheen *et al.*, 2006). Through these procedures, we obtained a reliable set (n = 802) of confirmed locations for plants infected by *P. ramorum* between 2001 and 2009.

4.3.2 Host species mapping

We restricted our modeling area to six ecoregions in western and central Oregon that have susceptible host species and environmental conditions that can potentially harbor *P. ramorum*: Coast range, Willamette valley, Klamath mountains, Western Cascades, East Cascades – north, and East Cascades – south (Fig. 1). To map and rank susceptibility and distribution of *P. ramorum* hosts, we used geospatial vegetation data developed using gradient nearest neighbor (GNN) imputation (Ohmann & Gregory, 2002; Ohmann *et al.*, 2007; Pierce *et al.*, 2009). The GNN method applies direct gradient analysis (canonical correspondence analysis) and nearest-neighbor imputation to ascribe detailed ground attributes of vegetation to each pixel in a regional landscape. We developed GNN species models for each of the six ecoregions in western and central Oregon in which *P. ramorum* hosts occur. Field plot data consisted of canopy cover of plant species recorded on several thousand field plots installed in regional inventory, ecology, and fuel mapping programs. Spatial explanatory variables were measures of climate, topography, parent material, and geographic location. The resulting GNN models are 30-m-resolution GIS rasters, in which each cell value is associated with codes of individual species and their abundances (percent cover). We extracted abundance data for 14 host species present in the study area (Table 1).

4.3.3 Climate and topography surfaces

We quantified a set of three climate and three topographical variables that play an important role in the establishment and spread of sudden oak death disease. To map weather conditions known to affect foliar plant pathogens (Woods *et al.*, 2005), we derived 30-year monthly averages (1971–2001) of maximum temperature, minimum temperature, and precipitation characteristics from the parameter elevation regression on independent slopes model (PRISM; Daly *et al.*, 2001). PRISM uses point measurements from a large sample of weather base stations and combines them with digital terrain data, coastal proximity, vertical mass layering, and other factors to spatially interpolate climate variability across large landscapes. We used 800 m resolution grids for each month in the rainy season (December to May) that represents the reproductive period for *P. ramorum* in California and Oregon (Davidson *et al.*, 2005). We also derived three topographic variables: elevation, topographic moisture index (TMI), and potential solar irradiation (PSI) from the U.S. Geological Survey 30-m digital terrain model. The TMI describes the effect of topography on local moisture availability and was calculated as the natural log of the ratio between the upslope contributing drainage area and the slope gradient of a grid cell (Moore *et al.*, 1991). The PSI characterizes the potential mean solar irradiation and was calculated for the rainy season using the cosine of illumination angle on slope equation (Dubayah, 1994).

4.3.4 Model of potential distribution

We developed a heuristic (rule-based) iSDM model using multi-criteria evaluation (MCE) method (Malczewski, 1999; Jiang & Eastman, 2000; Mendoza & Martins, 2006) to identify the areas at potential risk of *P. ramorum* establishment and spread in western Oregon. Following methods described in Meentemeyer *et al.* (2004) for California, expert input was used to assign a weight of relative importance to each predictor variable and rank the criterion range to standardize the data and determine the magnitude and direction of their effect on potential disease spread.

4.3.4.1 Ranking host vegetation

We compiled vegetation data to create a host index variable calculated in the GIS by summing the products of the species abundance score and spread score in each 30 m cell. To generate the species abundance score, the percent canopy cover of each species was linearly reclassified into ten abundance classes using equal interval classification scheme. To generate the species spread score, individual host species were scored from 0 to 10 according to their potential to produce inoculum and spread the disease to other hosts (Table 1). With minor changes to account for specific disease behavior in Oregon (Hansen *et al.*, 2008), we followed the scoring scheme previously developed by Meentemeyer (2004) for SOD risk model for California. Tanoak was assigned the highest score of 10 as it is the most affected species (Goheen *et al.*, 2006; Kanaskie *et al.*, 2009c) and predominant sporulating host in Oregon forests (Hansen *et al.*, 2008). Tanoak is susceptible to both foliar and stem infection and is associated with high severity infections in mixed redwood-tanoak and evergreen forest associations (Maloney *et al.*, 2005). Oregon myrtle was scored moderately high (5) because the foliar infection on this host produces significant amounts of inoculum that spreads to other host vegetation in the form of zoospores and sporangia (Davidson *et al.*, 2005; Rizzo *et al.*, 2005). Several landscape epidemiological studies in California consistently observed positive correlation between the presence of Oregon myrtle and *P. ramorum* infection (Kelly & Meentemeyer, 2002; Condeso & Meentemeyer, 2007; Meentemeyer *et al.*, 2008a), but this host appears to play a less important role in the epidemiological system in Oregon (Hansen *et al.*, 2008). *Rhododendron* species were also scored moderately high (5) as they are susceptible to both foliar and branch infection, and are widely distributed in the understory of mixed evergreen and coniferous forests in Oregon (Goheen *et al.*, 2006). Redwood (*Sequoia sempervirens*) was given a score of 3 because the production of sporangia from its foliar infestation is limited but the species is often present in association with more susceptible tanoak (Maloney *et al.*, 2005). The remaining species that are susceptible to foliar infection and provide transmission pathways for the pathogen were assigned a value of 1. Both species of oaks were scored 0, as they represent terminal-hosts in the epidemiological system and their potential to spread inoculum is minimal (Davidson *et al.*, 2005). The final host index values were linearly rescaled into five standard ranks (0–5).

4.3.4.2 Ranking climate factors

We ranked precipitation and temperature conditions using threshold values from Meentemeyer *et al.* (2004) based on published knowledge of *P. ramorum* biophysical properties gained from laboratory tests and field studies (Table 2). Since water must be available on plant surfaces for a substantial period of time (6–12 consecutive hours) before infection is initiated (Garbelotto *et al.*, 2003; Tooley *et al.*, 2009), precipitation represents a significant limiting factor for *P. ramorum*. We assigned the highest score (5) to areas with an average monthly precipitation greater than 125 mm, being the most suitable for disease establishment and inoculum production. Lower scores (4–1) were given to progressively lower rainfall amounts, while areas receiving less than 25 mm of rainfall were given a score of 0. Laboratory experiments demonstrated that *P. ramorum* thrives best at mild temperatures between 18–22 °C, while infection rates decrease to less than 50% at temperatures below 12 °C and above 30 °C (Werres *et al.*, 2001; Garbelotto *et al.*, 2003; Englander *et al.*, 2006; Tooley *et al.*, 2009). Therefore, we assigned the areas with an average maximum temperature between 18–22 °C the highest rank of 5. Areas with maximum temperatures outside the most suitable range were given progressively lower scores. Although little is known about the effect of minimum temperature on infection rates, *P. ramorum* is intolerant to temperatures below freezing (Rizzo & Garbelotto, 2003; Browning *et al.*, 2008). Areas with average minimum temperatures above freezing (0 °C) were assigned a score of 1 and areas with average minimum temperatures below freezing a score of 0.

4.3.4.3 Developing heuristic model

We summarized the final appraisal scores for western Oregon to represent a cumulative spread risk index that was subsequently standardized into five risk categories from very low risk to very high risk. Each predictor variable (criterion), ranked between 0 and 5 to encode the suitability for disease establishment and spread, was assigned a weight according to the estimated relative importance of the variable in the epidemiological system (Table 3). Using the weights and scores of vegetation and climate parameters, the final spread risk was computed for each grid cell by finding the sum of the product of each ranked variable and its weight, divided by the sum of the weights:

$$\bar{S} = \frac{\sum_{i}^{n} W_i R_{ij}}{\sum_{j}^{n} W_i}, \qquad (1)$$

where $\bar{S}$ is the appraisal score (spread risk) for a grid cell, W_i is the weight of the *i*th predictor variable, and R_{ij} is the rank, or score, of the *j*th value of the *i*th variable. We computed the equation for each month in the pathogen's reproductive season (December–May) and averaged the six monthly maps into one cumulative spread risk index. This risk model represents a potential distribution of *P. ramorum* in western Oregon based on site suitability for disease establishment and inoculum production, without considering pathogen's dispersal pressure or human-mediated forms of spread.

4.3.5 Model of actual distribution

We developed a statistical iSDM model using maximum entropy (MAXENT) to estimate the actual distribution of *P. ramorum* infections within the 2008 quarantine area in southwest Oregon. MAXENT is a machine-learning method that predicts the distribution of an organism by finding the probability distribution of maximum entropy (i.e., the closest to uniform) that respects a set of constraints derived from sample locations. The constraints are represented by simple functions of environmental predictor variables, with their means required to be close to the empirical average of occurrence sites (Phillips *et al.*, 2006; Phillips & Dudik, 2008). This method has been shown to perform well in comparison with other algorithms that utilize presence-only data to predict species distributions (Elith *et al.*, 2006; Elith & Graham, 2009; Vaclavik & Meentemeyer, 2009).

4.3.5.1 Developing statistical model

We split the total of 802 field samples of confirmed *P. ramorum* infection chronologically into three datasets. The 2005–2008 dataset (n=482) was used to calibrate the relative likelihood of current invasion based on the relationship between the field observations of disease occurrence and 21 predictor variables including three climate factors (maximum temperature, minimum temperature, precipitation), three topographical factors (elevation, TMI, PSI), abundance of 14 host species (listed in Table 1), and a dispersal pressure variable. The dispersal pressure term was computed with the 2001–2004 dataset (n=218) to quantify the relative force of invasion at all locations in the study area (Hastings *et al.*, 2005; Meentemeyer *et al.*, 2008a) and thus force the MAXENT model to predict the actual or current distribution of the pathogen rather than its potential distribution (Vaclavik & Meentemeyer, 2009). We used a cumulative distance metric that incorporates dispersal limitations in iSDMs without explicitly estimating the dispersal characteristics of the organism (Allouche *et al.*, 2008). The cumulative distance (D_i) summed the inverse of the squared Euclidean distances d_{ik} between each potential source of invasion *k* (confirmed between 2001 and 2004) and target plot *i* (sampled between 2005 and 2008):

$$D_i = \sum_{k=1}^{N} \left(\frac{1}{(d_{ik})^2} \right) \qquad (2)$$

Such a distance-constraining factor is crucial for discriminating the actual distribution from the potential distribution of biological invasions, as it accounts for restrictive forces that prevent invasive species from colonizing habitats environmentally favorable but remote from already invaded locations (Vaclavik & Meentemeyer, 2009; Lobo *et al.*, 2010). If dispersal pressure was omitted, then all sites that are environmentally similar to those already invaded, would be modeled as actual distribution, yielding considerable over-predictions.

Utilizing the MAXENT software version 3.2.1, we iteratively weighted each predictor variable to maximize the likelihood to reach the optimum probability distribution, and used the logistic output to ensure a predicted range between 0 and 1 (Elith & Burgman, 2003; Phillips & Dudik, 2008). We selected 500 iterations for model convergence and employed the regularization procedure that prevents overfitting better than variable-selection methods commonly used in traditional statistical models (Phillips & Dudik, 2008). In addition, we used a jackknife test of the relative contribution to model gain to get insight into the relative importance of individual explanatory variables (Phillips *et al.*, 2006).

4.3.6 Evaluating the models

Since the potential distribution is a hypothetical concept that refers to locations which could be infested by the forest pathogen based on suitable environmental factors, the heuristic model cannot be rigorously assessed with the use of field presence/absence data (Chefaoui & Lobo, 2008; Vaclavik & Meentemeyer, 2009; Lobo *et al.*, 2010). As an alternative, we can examine the correspondence between the predicted risk levels and infested locations and compare it to risk levels at randomly distributed points. We used the GIS to generate the same number of random points as the number of confirmed locations (n=802) and ran a T-test to identify the degree to which predicted potential distribution differs between invaded and random sites.

The actual distribution refers to locations where the pathogen most likely exists at a specific time, as constrained by environmental and dispersal limitations. The performance of the statistical model thus can be rigorously evaluated with field data. The 2009 set of confirmed infections (n=102) was set aside from the model development process to be used as an independent dataset for validation. Samples that were collected and tested in laboratory in 2009 but were negative for infection were used as absences in calculating the accuracy statistics. However, we included only those (n=34) located further than 200 m from known confirmed sites to account for scale on which the disease is known to be clustered (Condeso & Meentemeyer, 2007) and thus avoid potential false negative cases. We compiled these datasets and used the area under the curve (AUC) of the receiver operating characteristics (ROC) to examine the true positive rate as a function of the false positive rate at each possible probability threshold predicted by the model (Fielding & Bell, 1997; Pontius & Schneider, 2001; Hirzel *et al.*, 2006). We also calculated omission and commission error rates at the threshold that maximized specificity and sensitivity of the statistical model (Jimenez-Valverde & Lobo, 2007; Freeman & Moisen, 2008).

4.4 Results

4.4.1 Predicted geographic patterns of potential invasion

The model of potential distribution predicts the level and spatial variability of *P. ramorum* establishment and spread risk in six ecoregions in Oregon (Fig. 2). Nearly 252 km^2 (0.2%) of western and central Oregon’s 111,694 km^2 of land area was predicted as very high risk for disease spread (Table 4). Very high risk habitats occur in the southwest portion of the study area, mostly in the Coast Range ecoregion within 50 km from the Pacific Ocean. They are patchily distributed across the valleys of Chetco River, Wheeler Creek, Pistol River, Rogue River, Elk River, Sixes River, and their tributaries. Very high risk was generally identified over relatively small areas (mean patch size = 0.9 hectares) nested within larger areas of high risk and coinciding with the highest abundances of the most important host species: tanoak, rhododendron, and Oregon myrtle. The very high risk levels occur most frequently in Curry County, in which they encompass a total of 243.3 km^2 (5.8% of county). Three other counties (Coos, Douglas, and Josephine) include very high risk habitats, although these habitats cover only 0.1% of each county area.

Nearly 1,865 km^2 (1.7%) of the study area was mapped high risk. High risk habitats form slightly more continuous stretches (mean patch size = 2.1 hectares) along river valleys of the Coast Range and the western part of the Klamath Mountains ecoregion. Although highly concentrated in the southwest portion of the state, high risk areas extend north in small patches along the coast to the Umpqua River and its tributaries in Douglas County. The majority of continuous areas predicted as high risk occur in Curry County, in which it encompasses a total of 1333.3 km^2 (31.8% of county area). These areas coincide with suitable climate conditions of high moisture availability and relatively warm temperatures in large continuous areas of susceptible forest vegetation. High risk habitats are typically mixed evergreen forests including redwood and Douglas fir but with tanoak as a dominant or co-dominant species. Rhododendron and evergreen huckleberry often occur in the understory of these forest communities. Larger areas of high risk were also identified in Coos County (210.1 km^2; 5.1% of county area) and Josephine County (290.9 km^2; 6.9% of county area).

Over 4,216 km^2 (3.8%) of the study area was mapped moderate risk. Moderate risk habitats are scattered across southern half of the Coast Range ecoregion and western part of the Klamath Mountains but extends in smaller amounts to the Western Cascades ecoregion. In the two counties with the largest moderate risk prediction, Curry County (1084.1 km^2; 25.8% of county area) and Josephine County (1021.7 km^2; 25.8% of county area), moderate risk was mapped mostly in habitats with climatically suitable conditions but with lower values of host vegetation index. In the Klamath region, it forms a buffer-like pattern around high risk areas. In the Western Cascades ecoregion, moderate risk occurs in small patches (mean patch size = 1.5 hectares) and extends to Oregon's northern boundary, following the patchy distribution of rhododendron species in Douglas-fir dominated forest communities.

Over 66,530 km^2 (59.6%) of western Oregon was mapped low risk and 38,708 km^2 (34.7%) of the area was mapped very low risk. Low risk habitats are generally larger in area (mean=65.3 hectares) and extend over a vast portion of the northern part of the Coast Range ecoregion and the entire Willamette Valley ecoregion. Low risk is often associated with moderately suitable temperature and moisture conditions but low abundance and susceptibility of host vegetation. The low risk level was predicted over more than 95% of Benton, Clackamas, Clatsop, Columbia, Lincoln, Marion, Multnomah, Polk, Tillamook, Washington, and Yamhill counties. Very low risk habitats form nearly one large area in both East Cascades ecoregions but extend west in several large patches (mean patch size = 37 km^2) to the Western Cascades and Klamath Mountains ecoregions. Very low risk areas occur further from the coast (>150 km) at higher elevations (> 1200 m) with cold temperatures and low precipitation. There are no hosts species mapped in 77% of the very low risk areas and only species with low abundance and susceptibility (mostly Douglas fir) are mapped in the remaining 23%. The very low risk levels were predicted over nearly 100% of Deschutes, Jefferson, Klamath, Lake, and Wasco counties.

4.4.2 Predicted geographic patterns of actual invasion

The model of actual distribution predicts the relative likelihood of *P. ramorum* current invasion in the 2009 quarantine area in Curry County (Fig. 3a). Using the threshold that maximized model effectiveness, we estimated pathogen's presence across 65.4 km^2 of land area in southwest Curry County, northwest of the town of Brookings in the Chetco River watershed (Fig. 3b). All areas predicted as being infected occur within 15 km of the Pacific coast. Two areas with the highest likelihood of infection occur at the lower section of Chetco River between Joe Hall Creek and Ferry Creek, and across the valley hillsides of North Fork Chetco River and its tributaries Mayfield Creek and Bravo Creek. Outside of the watershed, a large patch of forest predicted by the model occurs north of the town of Brookings between Ram Creek and Shy Creek. Two locations, modeled as being likely infected but in which disease has not been confirmed to date, were identified along Jack Creek and Jordon Creek in the southern portion of the Chetco River watershed and between Houstenade creek and Miller Creek in the northwest part of the quarantine area. These locations are relatively close to known infected sites (~4 km) and coincide with areas mapped as having high abundances of tanoak and evergreen huckleberry.

The jack-knife test of variable importance (Fig. 4) shows that the variable with the highest gain when used in the model in isolation was dispersal pressure, having the most useful information that contributes to final prediction. Similarly, the dispersal pressure term decreased the model gain the most when it was omitted, having the most variability that is not present in other predictors. After the dispersal pressure variable, precipitation, maximum temperature, and elevation followed in their relative importance for model gain. From all 14 host species used for prediction, evergreen huckleberry, tanoak, and Douglas-fir were identified in the respective order as being the most important for model gain.

4.4.3 Model evaluation

The T-test for the model of potential distribution showed that modeled risk is significantly higher at sites identified as infested between 2001 and 2009 (n=802) than at randomly selected locations (P<0.0001) (Table 5a). Most of the 802 infected sites were mapped high risk (48%), followed by low risk (29%), moderate risk (16%), very high risk (7%), and very low risk (0%). Most of the 802 random locations were mapped as low risk (58%), followed by very low risk (36%), moderate risk (3%), high risk (2%), and very high risk (1%).

The ROC test for the model of actual distribution produced the AUC value of 0.91 based on the data from 136 samples analyzed in 2009 (Table 5b). The optimal probability threshold based on maximizing sensitivity and specificity of the model was relatively low (t=0.124) and produced commission and omission error rates of 0.18 and 0.13 respectively.

4.5 Discussion

Mapping the geographic distribution of invasive species and diseases is essential for the examination of their impacts in natural ecosystems and implementation of effective management strategies (Holdenrieder *et al.*, 2004). Predictive, spatial tools that identify current extent of biological invasions and habitats at potential risk of spread are increasingly needed to guide the management of biological invasions (Simberloff, 2003; Plantegenest *et al.*, 2007). In this study, we developed two predictive models of *P. ramorum* potential and actual distribution in western Oregon to prioritize landscape context for early detection surveillance and invasion control.

The heuristic model of *P. ramorum* potential distribution identifies areas that can serve as potential habitats for disease establishment and propagation. Mapped risk is based on combined effects of host species availability and susceptibility, and climate conditions in the pathogen's major reproductive season (December–May). Based on the model criteria, our prediction indicates that numerous forests across the western region of Oregon face considerable risk of sudden oak death invasion. Although concentrated in the southwest part, very high and high risk habitats were mapped across the entire Curry County and identified at smaller-scale in Coos and Josephine counties, more than 150 km away from the currently quarantined areas. This result corroborates findings of previous studies (Meentemeyer *et al.*, 2004; Meentemeyer *et al.*, 2008a) and suggests that *P. ramorum* is in relatively early stage of invasion, occupying only a small portion of its fundamental ecological niche.

The levels of *P. ramorum* establishment and spread risk agree closely with predictions from an equivalent model of potential distribution developed for California (Meentemeyer *et al.*, 2004). Although our estimates for Oregon are based on GNN vegetation data that differ from those used as inputs for modeling in California (CALVEG dataset; USDA Forest Service RSL, 2003), predicted risk levels align considerably well across the border region of north-east California and south-west Oregon. Several discrepancies in risk levels (moderate risk in Del Norte County, high risk across the state border in Curry County) are caused by higher susceptibility ranking of tanoak and lower susceptibility ranking of Oregon myrtle in our model, based on documented differences in the epidemiological role of these species in Oregon (Hansen *et al.*, 2008). Considering the similarity of environmental conditions and a prevalence of redwood-tanoak forests in northern California, we suggest the risk model developed for California may be slightly under-estimating the potential risk of *P. ramorum* invasion in the northernmost region.

The significant T-test suggests the risk model produced plausible predictions; however, 29% of currently infected sites were mapped as low risk. This type of underprediction is likely associated with the accuracy and precision of host vegetation data used as the most important criterion in model building. First, the 30 m spatial resolution of our vegetation data may be coarser than the scale at which the disease occurs. High resolution aerial photographs indicate there are small patches of host species that were mapped as non-host or no forest vegetation because these patches are smaller than the minimal mapping unit of the GNN data. Secondly, the GNN species distributions consist of a single field plot imputed to each pixel, resulting in species maps with small amounts of noise at a local scale. This fine-scale heterogeneity poses no significant problems for regional-scale analyses but may cause some infected locations to overlay with scattered pixels representing no host in the vegetation data and thus low risk in the predicted risk model. Lastly, a non-forest mask was applied to the vegetation model based on maps of ecological systems developed by the USGS Gap Analysis Program. There are 23 infected locations in close vicinity of the town of Brookings that overlap with pixels classified as low density or open space development in the vegetation model.

In addition, “spread risk” in this model is defined as the potential to produce inoculum and further propagate the disease across landscape. Therefore, it places low importance to terminal hosts (e.g., oaks) that may get invaded for a short period of time but serve as epidemiological dead-ends (Rizzo & Garbelotto, 2003; Davidson *et al.*, 2005).

The MAXENT model of *P. ramorum* actual distribution quantifies the relative likelihood of disease infection calculated with data from the 2001–2008 field surveys. Mapped invasion is based on the statistical relationship between known infected sites and a set of climate, topographical, host availability, and dispersal pressure variables. Model predictions suggest that current invasion range covers approximately 15% of the current quarantine area. Locations with the highest relative likelihood of pathogen’s presence occur along the Chetco River and its north fork, matching field observations of disease incidence. However, several places with no field observations were predicted as likely infected and thus indicate felicitous targets for early detection surveys.

The AUC value above 0.9 and low values of commission and omission error rates (<0.2) suggest relatively high prediction accuracy of the statistical model. However, the optimal threshold that maximized model effectiveness was relatively low and reflects the difficulty of predicting invasive organisms far from equilibrium with their environment (Vaclavik & Meentemeyer, 2009). As the motivation was to prioritize landscape contexts for early detection and invasion control, we selected a threshold that gives the same weight to commission and omission errors, in order to balance the importance of detecting the maximum number of infected sites with the relative cost of ground and helicopter surveillance (Meentemeyer *et al.*, 2008a). If commission errors were preferred, the model would produce a conservative scenario, decreasing the probability to detect disease outbreaks. If omission errors were preferred, large areas with marginal likelihood of *P. ramorum* presence would be predicted, increasing the cost of unnecessary field sampling.

Predictions of the statistical model were highly correlated with temperature and moisture conditions, elevation, and abundance of major host species. However, the most important variable based on jack-knife test of model gain was dispersal pressure, representing the force of pathogen's invasion. Considerably stronger effect of dispersal pressure is in contrast with results of Ellis *et al.* (2010) that identified environmental factors to be slightly more important than force of invasion for determining the spatial pattern of *P. ramorum* in northern California. Again, our findings indicate that the pathogen is in an initial stage of invasion, and are consistent with previous studies that recognized incorporation of range-confining variables based on space or distance metrics to be essential for predicting organisms under colonization-lag and non-equilibrium scenarios (Araujo & Pearson, 2005; Allouche *et al.*, 2008; De Marco *et al.*, 2008; Vaclavik & Meentemeyer, 2009). The early stage of invasion is also supported by the fact that dispersal pressure was significant for model prediction, although it was based on simple cumulative distance from initially invaded sites and did not account for potential landscape connectivity (Ellis *et al.*, 2010) or long distance human-mediated forms of spread (Cushman & Meentemeyer, 2008).

In this study, we applied a heuristic and statistical model in the GIS to produce spatial predictions of the potential and actual distribution of *P. ramorum* invasion in Oregon forests. Several studies have mapped potential SOD risk at a state (Meentemeyer *et al.*, 2004) or continental scale (Venette & Cohen, 2006; Kelly *et al.*, 2007) and identified actual patterns of disease spread in California (Meentemeyer *et al.*, 2008a). However, this is the first effort to model *P. ramorum* invasion in Oregon and use both iSDM approaches simultaneously, while clearly distinguishing between their meanings and purposes. While estimates of potential distribution provide a better picture of forests potentially threatened by disease invasion, the model of actual distribution quantifies the current range at unsampled locations. Application of the actual distribution model to on-the-ground management will increase the efficacy of detection and eradication of current outbreaks, especially under conditions of limited resources. Application of the potential distribution model will allow identifying habitats at the highest risk of future disease spread, to which preventive measures can be applied. When resources are available, field monitoring of high risk habitats will increase the chance to detect outbreaks introduced via long-distance dispersal events and minimize future disease impacts. Therefore, complementary knowledge from both types of spatial models is crucially needed to guide monitoring and control activities, especially for organisms in early stages of invasion that have considerable mismatch between their fundamental and realized niche.

iSDMs are getting increasingly popular to tackle early detection and eradication problems. However, models of potential and actual distribution are often confused and have never been used simultaneously for the same organism. In this application, we developed a heuristic model based on current knowledge of *P. ramorum* physiological and epidemiological requirements to represent its fundamental niche (potential distribution). We applied a statistical model trained by field data to portray pathogen's realized niche (actual distribution) and included data on dispersal pressure to constrain predicted range (Allouche *et al.*, 2008; Vaclavik & Meentemeyer, 2009). Although much remains to be learned about possibilities to control further *P. ramorum* spread, these spatially-explicit models provide simple yet effective management tools to prioritize landscape contexts for early detection and eradication of disease outbreaks. As new infected sites are discovered, models should be updated and validated with new data to continually refine management strategies. This research illustrates how the iSDM framework can be used to analyze the actual versus potential distribution of emerging infectious disease in a complex, heterogeneous ecosystem.

4.6 Acknowledgements

We thank the Forest Inventory and Analysis (FIA) of the Pacific Northwest Research Station and the Ecology program of the Pacific Northwest Region, USDA Forest Service, for field plot data used in host species modeling. We also thank members of the Center for Applied GIScience at UNC Charlotte for their helpful comments on the manuscript. This research was supported by grant from USDA Forest Service – Pacific Southwest Research Station (08-JV-11272138-055).

Tables

Table 1: Spread scores of host species based on their potential to spread inoculum of *P. ramorum.*

Hosts	**Score**
Arbutus menzeisii – Pacific madrone	1
Arctostaphylos spp. – manzanita	1
Frangula californica – California buckthorn	1
Frangula purshiana – Pursh's buckthorn	1
Notholithocarpus densiflorus – tanoak	10
Lonicera hispidula – pink honeysuckle	1
Pseudotsuga menziesii – Douglas-fir	1
Quercus chrysolepis – canyon live oak	0
Quercus kelloggii – black oak	0
Rhododendron spp. – rhododendron	5
Rubus spectabilis – salmonberry	1
Sequoia sempervirens – redwood	3
Umbellularia californica – Oregon myrtle	5
Vaccinium ovatum – evergreen huckleberry	1

Table 2: Range of values and assigned scores (*R*), ranked 0–5 from least to most suitable for establishment and spread of *P. ramorum.*

Rank	**Precipitation (mm)**	**Average maximum temperature (°C)**	**Average minimum temperature (°C)**
5	> 125	18-22	-
4	100-125	17-18; 22-23	-
3	75-100	16-17; 23-24	-
2	50-75	15-16; 24-25	-
1	25-50	14-15; 25-26	> 0
0	<25	< 14; > 26	< 0

Table 3: Importance weights (*W*) assigned to predictor variables, ranked 1–6 from lowest to highest importance for *P. ramorum* (according to Meentemeyer *et al.* 2004).

Variable	**Weight**
Host species index	6
Precipitation	2
Maximum temperature	2
Minimum temperature	1

Table 4: Land area of predicted spread risk levels in Oregon counties (in km^2 and percent of total county area).

County	**Area**	**Very low risk**		**Low risk**		**Moderate risk**		**High risk**		**Very high risk**	
	(km2)	**km^2**	**%**	**km^2**	**%**	**km^2**	**%**	**km^2**	**%**	**km^2**	**%**
Benton	1758	0.0	0.0	1738.3	98.9	2.8	0.2	0.0	0.0	0.0	0.0
Clackamas	4866	114.0	2.3	4652.6	95.6	77.4	1.6	0.0	0.0	0.0	0.0
Clatsop	2083	0.0	0.0	2077.0	99.7	0.0	0.0	0.0	0.0	0.0	0.0
Columbia	1693	0.0	0.0	1693.4	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Coos	4135	0.0	0.0	3161.3	76.5	759.5	18.4	210.1	5.1	3.8	0.1
Curry	4195	0.0	0.0	1534.8	36.6	1084.1	25.8	1333.3	31.8	243.3	5.8
Deschutes	4688	4683.8	99.9	4.5	0.1	0.0	0.0	0.0	0.0	0.0	0.0
Douglas	13088	1027.5	7.9	11304.5	86.4	726.8	5.6	28.0	0.2	1.3	0.0
Hood River	1347	504.0	37.4	826.2	61.3	17.1	1.3	0.0	0.0	0.0	0.0
Jackson	7248	2307.1	31.8	4936.3	68.1	4.4	0.1	0.1	0.0	0.0	0.0
Jefferson	1723	1644.8	95.5	71.5	4.1	0.0	0.0	0.0	0.0	0.0	0.0
Josephine	4240	66.1	1.6	2858.4	67.4	1021.7	24.1	290.9	6.9	3.5	0.1
Klamath	15846	15769.2	99.5	62.9	0.4	0.0	0.0	0.0	0.0	0.0	0.0
Lake	8353	8334.2	99.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Lane	11938	913.6	7.7	10651.5	89.2	370.8	3.1	2.5	0.0	0.0	0.0
Lincoln	2520	0.0	0.0	2513.5	99.7	6.4	0.3	0.0	0.0	0.0	0.0
Linn	5980	313.9	5.2	5541.1	92.7	105.0	1.8	0.0	0.0	0.0	0.0
Marion	3092	76.1	2.5	2981.2	96.4	34.4	1.1	0.0	0.0	0.0	0.0
Multnomah	1125	0.0	0.0	1121.3	99.7	3.4	0.3	0.0	0.0	0.0	0.0
Polk	1927	0.0	0.0	1926.6	100.0	0.1	0.0	0.0	0.0	0.0	0.0
Tillamook	2832	0.0	0.0	2832.2	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Wasco	3274	2954.1	90.2	299.7	9.2	2.8	0.1	0.0	0.0	0.0	0.0
Washington	1881	0.0	0.0	1881.4	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Yamhill	1860	0.0	0.0	1860.2	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Total	**111694**	**38708.4**	**34.7**	**66530.4**	**59.6**	**4216.7**	**3.8**	**1864.9**	**1.7**	**251.9**	**0.2**

Table 5: Model evaluation. (a) Potential distribution: T-test of sites infected by *P.ramorum* versus random sites in predicted risk levels. (b) Actual distribution: evaluation statistics for Maxent model calculated with 2009 samples.

(a)

T-test (P<0.0001)				
	infected sites (n=802)		**random sites (n=802)**	
Risk level	**#**	**%**	**#**	**%**
very high	53	6.6	4	0.5
high	386	48.1	19	2.4
moderate	131	16.3	24	3.0
low	232	28.9	463	57.7
very low	0	0.0	292	36.4

(b)

Infected sites in 2009 (n=102) Uninfected sites in 2009 (n=34)		
AUC	0.911	
Threshold	0.124	
Commission error (rate)	6	(0.18)
Omission error (rate)	13	(0.13)

Figures

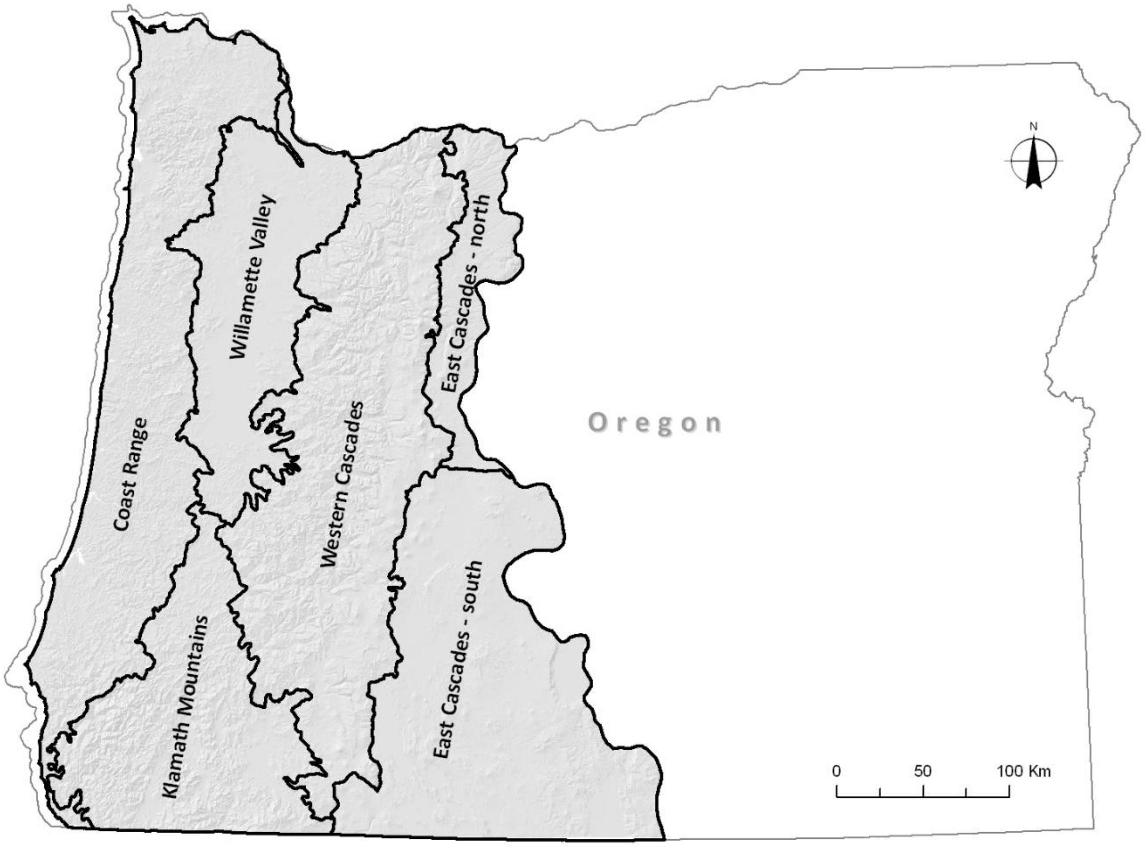


Figure 1: Study area: six ecoregions in western Oregon that have susceptible host species and climate conditions potentially suitable for establishment and spread of *P. ramorum.*

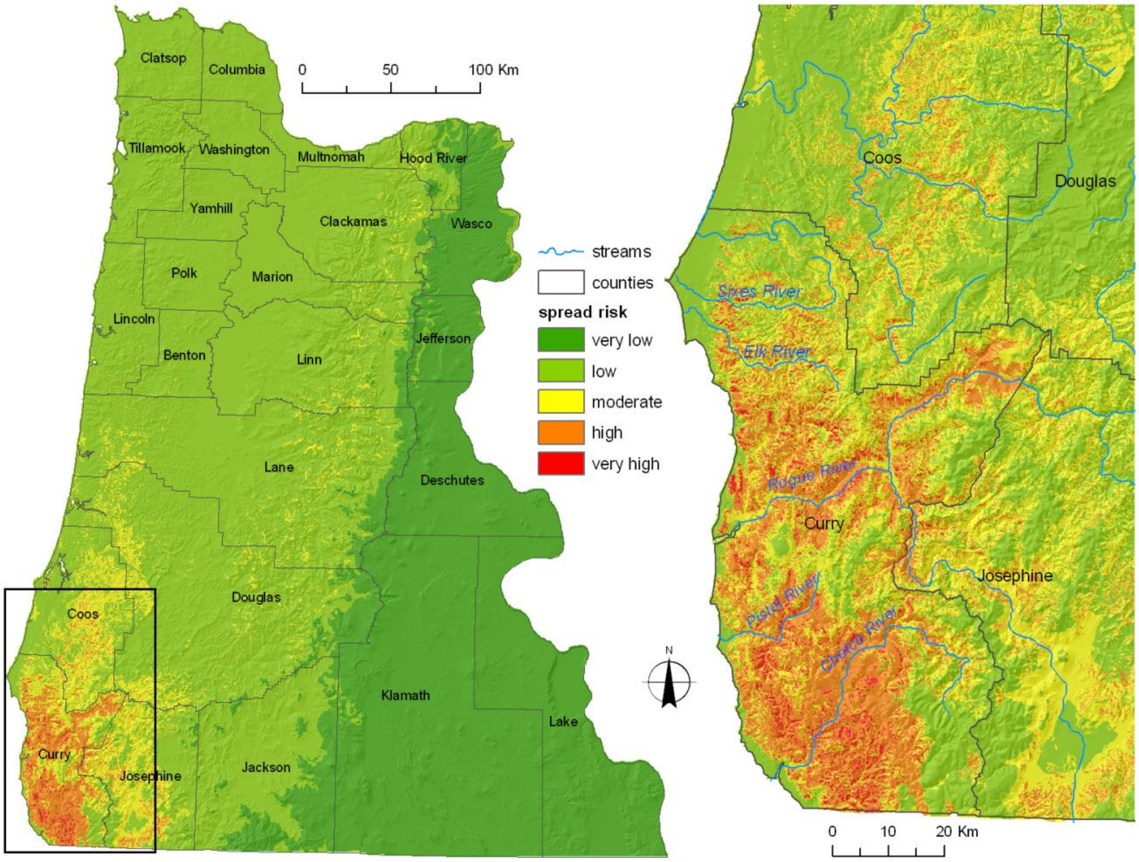


Figure 2: Predicted spread risk map for *P. ramorum* in western Oregon based on heuristic model of potential distribution. The inset shows southwest counties with the highest spread risk levels.

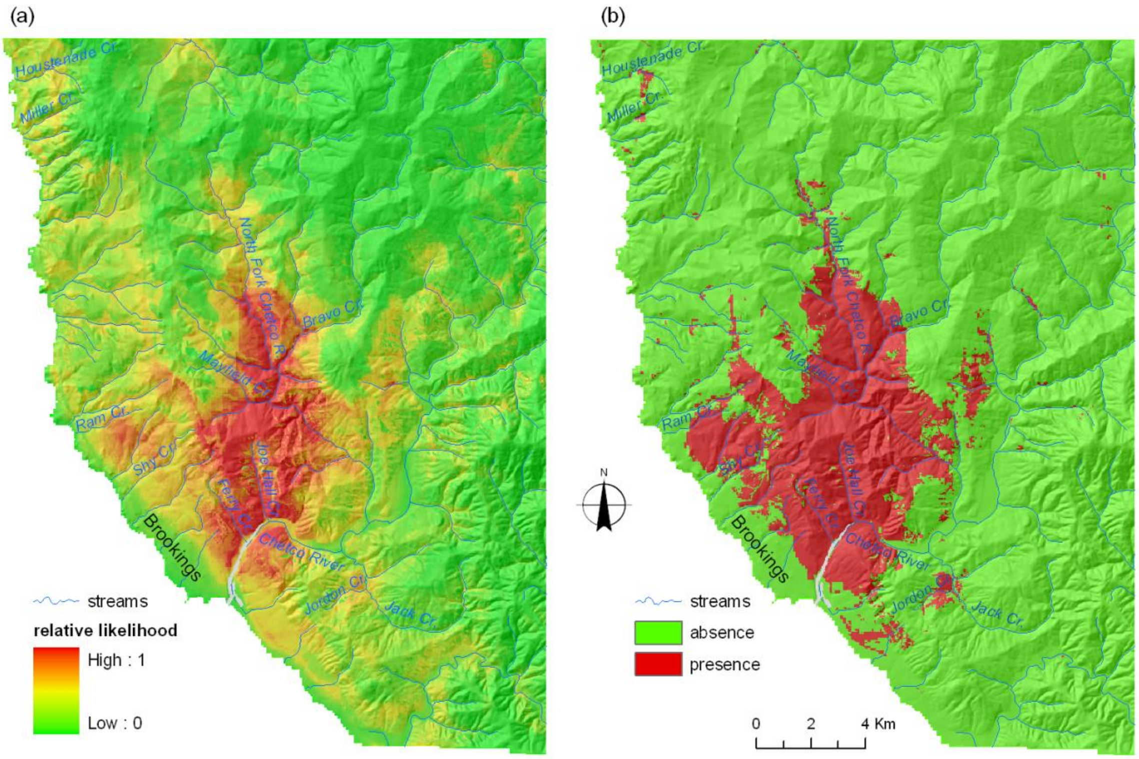


Figure 3: Predicted actual distribution of *P. ramorum* in southwest Curry County based on maximum entropy model. Map (a) shows relative likelihood of pathogen’s presence. Map (b) shows presence/absence realization of actual distribution based on probability threshold that maximized specificity and sensitivity of the model.

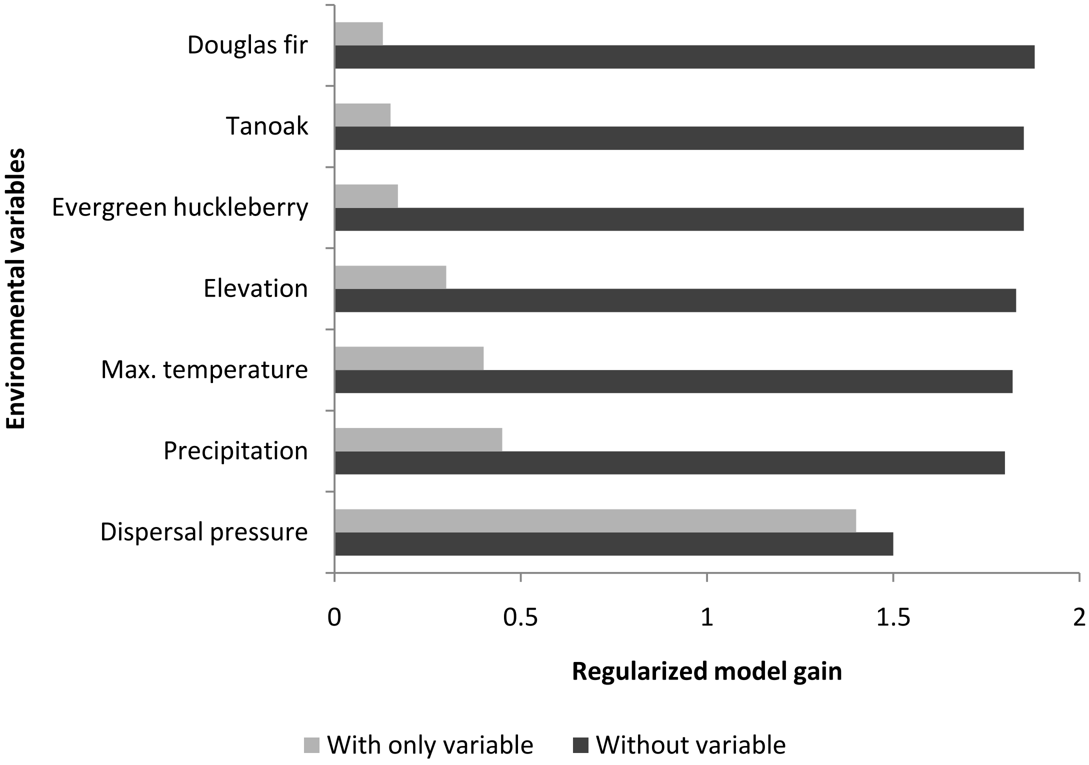


Figure 4: Jack-knife test of variables' relative importance. Graph shows seven most important environmental variables and their influence on regularized model gain when they were used in isolation or omitted.

CONCLUSIONS

The goal of this dissertation was to conduct a comprehensive study that examines three unresolved challenges in invasive species distribution modeling that are important for accurate predictions and ecological understanding of actual and potential distributions of invasive organisms. The advancement of SDM methods and comparisons of novel modeling algorithms experienced an enormous surge in recent ecological and biogeographical literature. However, this dissertation goes back to the heart of the discipline and questions some of the key ecological principles that constitute the foundation of SDM.

Chapter 1 answered the question whether absence data and dispersal constraints are needed to predict actual distributions of invasive species. I evaluated the impact of three types of occurrence data (presence-only, true-absence, pseudo-absence) on model accuracy and assessed the role of dispersal constraints measured by an estimate of propagule pressure. Results show that the prediction of actual distribution is less accurate when absence data and dispersal constraints are ignored. Specifically, presence-only models and models without dispersal constraints had a tendency to over-predict the actual range of invasions. These findings have profound implications for practical conservation management because significant economic resources would be wasted if inefficacious models were used to guide on-the-ground detection and eradication of invasion outbreaks. Moreover, the study shows that the true-absence data are a critical ingredient not only for model calibration but also for ecologically meaningful assessment of iSDMs that focus on predictions of realized distributions. If independent true-absence data are missing, the statistical evaluators can only indicate how well models discriminate data considered in the training process but reveal little about the real prediction capability.

Chapter 2 addressed the issue of modeling potential distributions of invasive species under different degrees of non-equilibrium with environment. Here, it was assumed that an invasive species will be closer to equilibrium the longer it has been present in the new environment because the degree of equilibrium depends largely on species dispersal ability and time since introduction. The results confirm the theoretical hypothesis that a full environmental niche of invasive species cannot be effectively captured with data from the realized distribution that is restricted by processes preventing full occupancy of suitable habitats. iSDMs calibrated under non-equilibrium are less accurate and robust in predicting the habitat potentially prone to invasion than models calibrated under scenarios closer to equilibrium. In addition, iSDMs of species in early stages of invasion had higher tendency to underpredict the potential range than models of species in later stages of invasion. The robustness of findings in this chapter are supported by consistent results for both a real example of a biological invasion in different stages of invasion and a simulated species under three hypothetical scenarios of non-equilibrium. This work also demonstrates how a simulated virtual species, for which the “true” environmental niche is perfectly known, can be used to examine fundamental ecological questions in an environment controlled for complications from natural variation and uncertainties in model assessment.

Chapter 3 focused on a phenomenon that has become a new paradigm in a geographical ecology: spatial dependence, i.e. spatial autocorrelation (SAC) in biogeographical data. In this chapter the implications of SAC for the accuracy and parameterization of iSDMs were examined by comparing spatially invariant models with models that deal with spatial dependence at a broad scale using trend surface analysis, a local scale using autocovariate methods, or multiple spatial scales using spatial eigenvector mapping. The results reveal that accounting for different scales of SAC significantly enhances predictive capability of iSDMs. Predictions improved dramatically when fine-scale SAC was incorporated in the models, suggesting that local range-confining processes are driving the spread of the modelled invasive organism. While the importance of environmental variables was relatively consistent across all models, the explanatory power decreased in spatial models for factors with strong spatial structure. In addition, results show that accounting for SAC reduces the amount of residual autocorrelation for a traditional statistical method based on the probability theory (GLM) but not for a machine learning method (MAXENT). However, this procedure still improved performance of both types of approaches, supporting the hypothesis that dispersal and colonization processes are important factors to consider in distribution models of biological invasions. Furthermore, these findings demonstrate that accounting for SAC is not only vital to avoid common problems in standard statistical approaches but can also be a crucial surrogate for dynamic processes that explain ecological mechanisms of invasion in static iSDMs.

Finally, chapter 4 demonstrated how iSDMs can be used in conservation practice to target areas for early detection and eradication of invasion outbreaks. Following methodological recommendations that stemmed from the work in previous chapters, models of potential and actual distribution of *P. ramorum* were developed for six ecoregions in Oregon. The predictions show that 65 km^2 of forested land was invaded by 2009, but further disease spread threatens more than 2100 km^2 of forests across the western region of Oregon. In addition to the practical goal of quantifying the spatial extent of threatened forest resources, this was the first effort to use predictions of both actual and potential distributions simultaneously, while clearly distinguishing between their meanings, purposes, and methodological nuances. However, it is stressed that complementary knowledge from both types of spatial models is needed to guide monitoring and control activities, especially for organisms in early stages of invasion that have considerable mismatch between their fundamental and realized niche.

Species distribution modeling is essential for both basic and applied research in ecology, biogeography, and conservation biology. The presented research addressing the effects of absence data, dispersal constraints, stage of invasion, and spatial dependence on model performance contributes not only to better spatial predictions of biological invasions but also to the ecological conceptualization of SDM as a discipline. For iSDMs to serve as effective tools for early-detection and management of invasive species in conservation practice, their accuracy and correct interpretation is essential to minimize the ecological impact and economic cost of biological invasions.

REFERENCES

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

Allouche, O., Tsoar, A. & Kadmon, R. (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, 43, 1223-1232

Anacker, B.L., Rank, N.E., Huberli, D., Garbelotto, M., Gordon, S., Harnik, T., Whitkus, R. & Meentemeyer, R. (2008) Susceptibility to *Phytophthora ramorum* in a key infectious host: landscape variation in host genotype, host phenotype, and environmental factors. *New Phytologist*, 177, 756-766

Araujo, M.B. & Guisan, A. (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, 33, 1677-1688

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

Araujo, M.B., Pearson, R.G., Thuiller, W. & Erhard, M. (2005) Validation of species-climate impact models under climate change. *Global Change Biology*, 11, 1504-1513

Augustin, N.H., Mugglestone, M.A. & Buckland, S.T. (1996) An autologistic model for the spatial distribution of wildlife. *Journal of Applied Ecology*, 33, 339-347

Austin, M. (2007) Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling*, 200, 1-19

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

Austin, M.P., Belbin, L., Meyers, J.A., Doherty, M.D. & Luoto, M. (2006) Evaluation of statistical models used for predicting plant species distributions: Role of artificial data and theory. *Ecological Modelling*, 199, 197-216

Ayala, D., Costantini, C., Ose, K., Kamdem, G.C., Antonio-Nkondjio, C., Agbor, J.P., Awono-Ambene, P., Fontenille, D. & Simard, F. (2009) Habitat suitability and ecological niche profile of major malaria vectors in Cameroon. *Malaria Journal*, 8, -

Beale, C.M., Lennon, J.J., Elston, D.A., Brewer, M.J. & Yearsley, J.M. (2007) Red herrings remain in geographical ecology: a reply to Hawkins et al. (2007). *Ecography*, 30, 845-847

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

Berry, P.M., Dawson, T.P., Harrison, P.A. & Pearson, R.G. (2002) Modelling potential impacts of climate change on the bioclimatic envelope of species in Britain and Ireland. *Global Ecology and Biogeography*, 11, 453-462

Bini, L.M., Diniz, J.A.F., Rangel, T.F.L.V.B., Akre, T.S.B., Albaladejo, R.G., Albuquerque, F.S., Aparicio, A., Araujo, M.B., Baselga, A., Beck, J., Bellocq, M.I., Bohning-Gaese, K., Borges, P.A.V., Castro-Parga, I., Chey, V.K., Chown, S.L., de Marco, P., Dobkin, D.S., Ferrer-Castan, D., Field, R., Filloy, J., Fleishman, E., Gomez, J.F., Hortal, J., Iverson, J.B., Kerr, J.T., Kissling, W.D., Kitching, I.J., Leon-Cortes, J.L., Lobo, J.M., Montoya, D., Morales-Castilla, I., Moreno, J.C., Oberdorff, T., Olalla-Tarraga, M.A., Pausas, J.G., Qian, H., Rahbek, C., Rodriguez, M.A., Rueda, M., Ruggiero, A., Sackmann, P., Sanders, N.J., Terribile, L.C., Vetaas, O.R. & Hawkins, B.A. (2009) Coefficient shifts in geographical ecology: an empirical evaluation of spatial and non-spatial regression. *Ecography*, 32, 193-204

Boyce, M.S., Vernier, P.R., Nielsen, S.E. & Schmiegelow, F.K.A. (2002) Evaluating resource selection functions. *Ecological Modelling*, 157, 281-300

Brasier, C. & Webber, J. (2010) Sudden larch death. *Nature*, 466, 824-825

Braunisch, V., Bollmann, K., Graf, R.F. & Hirzel, A.H. (2008) Living on the edge - Modelling habitat suitability for species at the edge of their fundamental niche. *Ecological Modelling*, 214, 153-167

Broennimann, O., Treier, U.A., Muller-Scharer, H., Thuiller, W., Peterson, A.T. & Guisan, A. (2007) Evidence of climatic niche shift during biological invasion. *Ecology Letters*, 10, 701-709

Browning, M., Englander, L., Tooley, P.W. & Berner, D. (2008) Survival of *Phytophthora ramorum* hyphae after exposure to temperature extremes and various. *Mycologia*, 100, 236-245

Burnham, K.P. & Anderson, D.R. (2004) Multimodel inference - understanding AIC and BIC in model selection. *Sociological Methods & Research*, 33, 261-304

Chefaoui, R.M. & Lobo, J.M. (2008) Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological Modelling*, 210, 478-486

Chen, H., Chen, L.J. & Albright, T.P. (2007) Predicting the potential distribution of invasive exotic species using GIS and information-theoretic approaches: A case of ragweed (Ambrosia artemisiifolia L.) distribution in China. *Chinese Science Bulletin*, 52, 1223-1230

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

Cobb, R.C., Meentemeyer, R.K. & Rizzo, D.M. (2010) Apparent competition in canopy trees determined by pathogen transmission rather than susceptibility. *Ecology*, 91, 327-333

Condeso, T.E. & Meentemeyer, R.K. (2007) Effects of landscape heterogeneity on the emerging forest disease Sudden Oak Death. *Journal of Ecology*, 95, 364-375

Cushman, J.H. & Meentemeyer, R.K. (2008) Multi-scale patterns of human activity and the incidence of an exotic forest pathogen. *Journal of Ecology*, 96, 766-776

Daly, C., Taylor, G.H., Gibson, W.P., Parzybok, T.W., Johnson, G.L. & Pasteris, P. (2001) High-quality spatial climate data sets for the United States and beyond. *Transactions of the American Society of Agricultural Engineers*, 43, 1957-1962

Dark, S.J. (2004) The biogeography of invasive alien plants in California: an application of GIS and spatial regression analysis. *Diversity and Distributions*, 10, 1-9

Daszak, P., Cunningham, A.A. & Hyatt, A.D. (2000) Emerging infectious diseases of wildlife - Threats to biodiversity and human health. *Science*, 287, 443-449

Davidson, J.M., Wickland, A.C., Patterson, H.A., Falk, K.R. & Rizzo, D.M. (2005) Transmission of *Phytophthora ramorum* in mixed-evergreen forest in California. *Phytopathology*, 95, 587-596

Davis, E.C. (2004) Predicting potential distributions of invasive land snails via ecological niche modeling. *Integrative and Comparative Biology*, 44, 687-687

Davis, F.W., Borchert, M., Meentemeyer, R.K., Flint, A. & Rizzo, D.M. (2010) Pre-impact forest composition and ongoing tree mortality associated with sudden oak death disease in the Big Sur Region; California. *Forest Ecology and Management* 259, 2342-2354

De'ath, G. & Fabricius, K.E. (2000) Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology*, 81, 3178-3192

de Knegt, H.J., van Langevelde, F., Coughenour, M.B., Skidmore, A.K., de Boer, W.F., Heitkonig, I.M.A., Knox, N.M., Slotow, R., van der Waal, C. & Prins, H.H.T. (2010) Spatial autocorrelation and the scaling of species-environment relationships. *Ecology*, 91, 2455-2465

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

Delmelle, E., Delmelle-Cahill, E., Casas, I. & Barto, T. (2010) H.E.L.P: A GIS-based health exploratory analysis tool for practitioners. *Applied Spatial Analysis and Policy*

Diniz-Filho, J.A. & Bini, L.M. (2005) Modelling geographical patterns in species richness using eigenvector-based spatial filters. *Global Ecology and Biogeography*, 14, 177-185

Diniz-Filho, J.A., Bini, L.M. & Hawkins, B.A. (2003) Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology and Biogeography*, 12, 53-64

Dormann, C.F. (2007a) Assessing the validity of autologistic regression. *Ecological Modelling*, 207, 234-242

Dormann, C.F. (2007b) Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography*, 16, 129-138

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

Dray, S., Legendre, P. & Peres-Neto, P.R. (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, 196, 483-493

Dubayah, R.C. (1994) Modeling a solar radiation topoclimatology for the Rio Grande river basin. *Journal of Vegetation Science*, 5, 627-640

Eastman, J.R. (2006) IDRISI Andes Guide to GIS and Image Processing. In. Clark Labs, Clark Univeristy, IDRISI Productions 1987-2006, Worcester, MA

Elith, J. & Burgman, M.A. (2003) Habitat models for PVA. *Population Viability in Plants. Conservation, Management and Modeling of Rare Plants.* (ed. by C.A. Brigham and M.W. Schwartz), pp. 203-235. Springer-Verlag, New York.

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

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

Elith, J., Kearney, M. & Philips, S. (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

Ellis, A.M., Vaclavik, T. & Meentemeyer, R.K. (2010) When is connectivity important? A case study of the spatial pattern of sudden oak death. *Oikos*, 119, 485-493

Englander, L., Browning, M. & Tooley, P.W. (2006) Growth and sporulation of *Phytophthora ramorum* in vitro in response to temperature and light. *Mycologia*, 98, 365-373

Engler, R. & Guisan, A. (2009) MIGCLIM: Predicting plant distribution and dispersal in a changing climate. *Diversity and Distributions*, 15, 590-601

Engler, R., Guisan, A. & Rechsteiner, L. (2004) An improved approach for predicting the distribution of rare and endangered species from ocurrence and pseudo-absence data. *Journal of Applied Ecology*, 41, 263-274

Engler, R., Randin, C.F., Vittoz, P., Czaka, T., Beniston, M., Zimmermann, N.E. & Guisan, A. (2009) Predicting future distributions of mountain plants under climate change: does dispersal capacity matter? *Ecography*, 32, 34-45

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

Fitt, B.D.L., McCartney, H.A. & Walklate, P.J. (1989) The role of rain in dispersal of pathogen inoculum. *Annual Review of Phytopathology*, 27, 241-270

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

Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K., Helkowski, J.H., Holloway, T., Howard, E.A., Kucharik, C.J., Monfreda, C., Patz, J.A., Prentice, I.C., Ramankutty, N. & Snyder, P.K. (2005) Global consequences of land use. *Science*, 309, 570-574

Fonseca, R.L., Guimaraes, P.R., Morbiolo, S.R., Scachetti-Pereira, R. & Peterson, A.T. (2006) Predicting invasive potential of smooth crotalaria (Crotalaria pallida) in Brazilian national parks based on African records. *Weed Science*, 54, 458-463

Franklin, J. (1995) Predictive vegetation mapping: geographic modeling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography*, 19, 474-499

Franklin, J. (2010a) *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge University Press, Cambridge, UK

Franklin, J. (2010b) Moving beyond static species distribution models in support of conservation biogeography. *Diversity and Distributions*, 16, 321-330

Freeman, E.A. & Moisen, G.G. (2008) A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. *Ecological Modelling*, 217, 48-58

Garbelotto, M., Davidson, J.M. & Ivors, K. (2003) Non-oak native plants are the main hosts for the sudden oak death pathogen in California. *California Agriculture*, 57, 18-23

Giovanelli, J.G.R., Haddad, C.F.B. & Alexandrino, J. (2008) Predicting the potential distribution of the alien invasive American bullfrog (Lithobates catesbeianus) in Brazil. *Biological Invasions*, 10, 585-590

Goheen, E.M., Hansen, E.M., Kanaskie, A., Sutton, W. & Reeser, P. (2009) Persistence of *Phytophthora ramorum* after eradication treatments in Oregon tanoak forests. In: *Phytophthoras in forests and natural ecosystems. Proceedings of the fourth meeting of the international union of forest research (IUFRO) working party*, pp. 173-176. U.S. Department of Agriculture, Forest Service, Pacific Southwest Research Station, General technical report PSW-GTR-221, Albany, California

Goheen, E.M., Kanaskie, A., McWilliams, M., Hansen, E., Sutton, W. & Osterbauer, N. (2006) Surveying and monitoring sudden oak death in southwest Oregon forests. In: *Proceedings of the sudden oak death second science symposium: the state of our knowledge. General technical report PSW-GTR-196* (eds. S.J. Frankel, P.J. Shea and M.I. Haverty), pp. 413-415. Pacific Southwest Research Station, Forest Service, U.S. Department of Agriculture, Albany, CA

Griffith, D.A. & Peres-Neto, P.R. (2006) Spatial modeling in ecology: The flexibility of eigenfunction spatial analyses. *Ecology*, 87, 2603-2613

Grinnell, J. (1917) The niche-relationships of the California Thrasher. *Auk*, 34, 427-433

Guisan, A., Edwards, T.C. & Hastie, T. (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, 157, 89-100

Guisan, A., Lehmann, A., Ferrier, S., Austin, M., Overton, J.M.C., Aspinall, R. & Hastie, T. (2006) Making better biogeographical predictions of species' distributions. *Journal of Applied Ecology*, 43, 386-392

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

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

Guo, Q., Kelly, M. & Graham, C.H. (2005) Support vector machines for predicting distribution of Sudden Oak Death in California. *Ecological Modelling*, 182, 75-90

Hansen, E.M., Kanaskie, A., Prospero, S., McWilliams, M., Goheen, E.M., Osterbauer, N., Reeser, P. & Sutton, W. (2008) Epidemiology of Phytophthora ramorum in Oregon tanoak forests. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, 38, 1133-1143

Hastings, A., Cuddington, K., Davies, K.F., Dugaw, C.J., Elmendorf, S., Freestone, A., Harrison, S., Holland, M., Lambrinos, J., Malvadkar, U., Melbourne, B.A., Moore, K., Taylor, C. & Thomson, D. (2005) The spatial spread of invasions: new developments in theory and evidence. *Ecology Letters*, 8, 91-101

Havel, J.E., Shurin, J.B. & Jones, J.R. (2002) Estimating dispersal from patterns of spread: Spatial and local control of lake invasions. *Ecology*, 83, 3306-3318

Hayden, K.J., Rizzo, D., Tse, J. & Garbelotto, M. (2004) Detection and quantification of Phytophthora ramorum from California forests using a real-time polymerase chain reaction assay. *Phytopathology*, 94, 1075-1083

Higgins, S.I., Richardson, D.M., Cowling, R.M. & Trinder-Smith, T.H. (1999) Predicting the Landscape-Scale Distribution of Alien Plants and Their Threat to Plant Diversity. *Conservation Biology*, 30, 301-313

Hirzel, A.H., Hausser, J., Chessel, D. & Perrin, N. (2002) Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? *Ecology*, 83, 2027–2036

Hirzel, A.H., Hausser, J. & Perrin, N. (2007) Biomapper 4.0. In. Laboratory for Conservation Biology, Department of Ecology and Evolution, University of Lausanne, Switzerland

Hirzel, A.H., Helfer, V. & Metral, F. (2001) Assessing habitat-suitability models with a virtual species. *Ecological Modelling*, 145, 111-121

Hirzel, A.H. & Le Lay, G. (2008) Habitat suitability modelling and niche theory. *Journal of Applied Ecology*, 45, 1372-1381

Hirzel, A.H., Le Lay, G., Helfer, V., Randin, C. & Guisan, A. (2006) Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, 199, 142-152

Hoffmeister, T.S., Vet, L.E.M., Biere, A., Holsinger, K. & Filser, J. (2005) Ecological and evolutionary consequences of biological invasion and habitat fragmentation. *Ecosystems*, 8, 657-667

Holdenrieder, O., Pautasso, M., Weisberg, P.J. & Lonsdale, D. (2004) Tree diseases and landscape processes: the challenge of landscape pathology. *Trends in Ecology & Evolution*, 19, 446-452

Hutchinson, G.E. (1957) Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, 22, 415-427

Ibanez, I., Silander, J.A., Wilson, A.M., Lafleur, N., Tanaka, N. & Tsuyama, I. (2009) Multivariate forecasts of potential distributions of invasive plant species. *Ecological Applications*, 19, 359-375

Iverson, L.R., Schwartz, M.W. & Prasad, A.M. (2004) How fast and far might tree species migrate in the eastern United States due to climate change? *Global Ecology and Biogeography*, 13, 209-219

Ivors, K.L., Hayden, K.J., Bonants, P.J.M., Rizzo, D.M. & Garbelotto, M. (2004) AFLP and phylogenetic analyses of North American and European populations of *Phytophthora ramorum*. *Mycological Research*, 108, 378-392

Jeschke, J.M. & Strayer, D.L. (2008) Usefulness of bioclimatic models for studying climate change and invasive species. *Annals of the New York Academy of Sciences*, 1134, 1-24

Jiang, H. & Eastman, J.R. (2000) Application of fuzzy measures in multi-criteria evaluation in GIS. *International Journal of Geographical Information Science*, 14, 173-184

Jimenez-Valverde, A. & Lobo, J.M. (2006) The ghost of unbalanced species distribution data in geographical model predictions. *Diversity and Distributions*, 12, 521-524

Jimenez-Valverde, A. & Lobo, J.M. (2007) Threshold criteria for conversion of probability of species presence to either-or presence-absence. *Acta Oecologica-International Journal of Ecology*, 31, 361-369

Jimenez-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

Johnson, J.B. & Omland, K.S. (2004) Model selection in ecology and evolution. *Trends in Ecology & Evolution*, 19, 101-108

Kanaskie, A., Goheen, E., Hansen, E., Osterbauer, N., McWilliams, M., Schultz, R., Savona, S., Sutton, W. & Reeser, P. (2009a) Early detection and eradication of

Phytophthora ramorum (sudden oak death) in Oregon forests. *Phytopathology*, 99, S61-S61

Kanaskie, A., Goheen, E.M., Hansen, E.M., Sutton, W., Reeser, P. & Osterbauer, N. (2009b) Monitoring the effectiveness of *Phytophthora ramorum* eradication treatments in southwest Oregon tanoak forests. *Phytopathology*, 99, S61-S61

Kanaskie, A., Hansen, E.M., Goheen, E.M., McWilliams, M., Reeser, P. & Sutton, W. (2009c) *Phytophthora ramorum* in Oregon forests: Six years of detection, eradication, and disease spread. In: *Phytophthoras in forests and natural ecosystems. Proceedings of the fourth meeting of the international union of forest research (IUFRO) working party*, pp. 170-172. U.S. Department of Agriculture, Forest Service, Pacific Southwest Research Station, General technical report PSW-GTR-221, Albany, California

Kearney, M. (2006) Habitat, environment and niche: what are we modelling? *Oikos*, 115, 186-191

Kearney, M., Phillips, B.L., Tracy, C.R., Christian, K.A., Betts, G. & Porter, W.P. (2008) Modelling species distributions without using species distributions: the cane toad in Australia under current and future climates. *Ecography*, 31, 423-434

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

Keith, D.A., Akcakaya, H.R., Thuiller, W., Midgley, G.F., Pearson, R.G., Phillips, S.J., Regan, H.M., Araujo, M.B. & Rebelo, T.G. (2008) Predicting extinction risks under climate change: coupling stochastic population models with dynamic bioclimatic habitat models. *Biology Letters*, 4, 560-563

Kelly, M., Guo, Q., Liu, D. & Shaari, D. (2007) Modeling the risk for a new invasive forest disease in the United States: An evaluation of five environmental niche models. *Computers Environment and Urban Systems*, 31, 689-710

Kelly, M. & Meentemeyer, R.K. (2002) Landscape dynamics of the spread of sudden oak death. *Photogrammetric Engineering and Remote Sensing*, 68, 1001-1009

Kelly, M., Tuxen, K. & Kearns, F. (2004) Geospatial informatics for management of a new forest disease: Sudden oak death. *Photogrammetric Engineering and Remote Sensing*, 70, 1001-1004

Kelly, N.M. & Tuxen, K. (2003) WebGIS for monitoring "sudden oak death" in coastal California. *Computers, Environment and Urban Systems*, 27, 527-547

Kissling, W.D. & Carl, G. (2008) Spatial autocorrelation and the selection of simultaneous autoregressive models. *Global Ecology and Biogeography*, 17, 59-71

Kuhn, I. (2007) Incorporating spatial autocorrelation may invert observed patterns. *Diversity and Distributions*, 13, 66-69

Kupfer, J.A. & Farris, C.A. (2007) Incorporating spatial non-stationarity of regression coefficients into predictive vegetation models. *Landscape Ecology*, 22, 837-852

Legendre, P. (1993) Spatial Autocorrelation - Trouble or New Paradigm. *Ecology*, 74, 1659-1673

Lennon, J.J. (2000) Red-shifts and red herrings in geographical ecology. *Ecography*, 23, 101-113

Lichstein, J.W., Simons, T.R., Shriner, S.A. & Franzreb, K.E. (2002) Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs*, 72, 445-463

Lippitt, C.D., Rogan, J., Toledano, J., Sangermano, F., Eastman, J.R., Mastro, V. & Sawyer, A. (2008) Incorporating anthropogenic variables into a species distribution model to map gypsy moth risk. *Ecological Modelling*, 210, 339-350

Lobo, J.M., Jimenez-Valverde, A. & Hortal, J. (2010) The uncertain nature of absences and their importance in species distribution modelling. *Ecography*, 33, 103-114

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

Lopez-Darias, M., Lobo, J.M. & Gouat, P. (2008) Predicting potential distributions of invasive species: the exotic Barbary ground squirrel in the Canarian archipelago and the west Mediterranean region. *Biological Invasions*, 10, 1027-1040

Lutolf, M., Kienast, F. & Guisan, A. (2006) The ghost of past species occurrence: improving species distribution models for presence-only data. *Journal of Applied Ecology*, 43, 802-815

Mack, R.N., Simberloff, D., Lonsdale, W.M., Evans, H., Clout, M. & Bazzaz, F.A. (2000) Biotic invasions: Causes, epidemiology, global consequences, and control. *Ecological Applications*, 10, 689-710

Maggini, R., Lehmann, A., Zimmermann, N.E. & Guisan, A. (2006) Improving generalized regression analysis for the spatial prediction of forest communities. *Journal of Biogeography*, 33, 1729-1749

Malczewski, J. (1999) *GIS and Multicriteria Decision Analysis*. Wiley and Sons, New York, 177-195 pp.

Maloney, P.E., Lynch, S.C., Kane, S.F., Jensen, C.E. & Rizzo, D.M. (2005) Establishment of an emerging generalist pathogen in redwood forest communities. *Journal of Ecology*, 93, 899-905

Manel, S., Williams, H.C. & Ormerod, S.J. (2001) Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology*, 38, 921-931

McArthur, R.H. (1957) On the relative abundance of bird species. *Proceedings of the National Academy of Science*, 43, 293-295

McCartney, H.A. & Fitt, B.D.L. (1985) Construction of dispersal models. *Advances in Plant Pathology, Vol. 3, Mathematical Modeling of Crop Disease* (ed. by D. Ingram, P. Williams and C.A. Gilligan), pp. 107-143. Academic Press, London.

McCullagh, P. & Nelder, J.A. (1989) *Generalized Linear Models*. Chapmam & Hall, London, UK

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

Meentemeyer, R.K., Anacker, B.L., Mark, W. & Rizzo, D.M. (2008a) Early detection of emerging forest disease using dispersal estimation and ecological niche modeling. *Ecological Applications*, 18, 377-390

Meentemeyer, R.K., Cunniffe, N.J., Cook, A.R., Filipe, J.A.N., Hunter, R.D., Rizzo, D.M. & Gilligan, C.A. (2011) Epidemiological modeling of invasion in heterogeneous landscapes: Spread of sudden oak death in California (1990–2030). *Ecosphere*, 2, art17

Meentemeyer, R.K., Rank, N.E., Shoemaker, D.A., Oneal, C.B., Wickland, A.C., Frangioso, K.M. & Rizzo, D.M. (2008b) Impact of sudden oak death on tree mortality in the Big Sur ecoregion of California. *Biological Invasions*, 10, 1243-1255

Meentemeyer, R.K., Rizzo, D., Mark, W. & Lotz., E. (2004) Mapping the risk of establishment and spread of Sudden Oak Death in California. *Forest Ecology and Management*, 200, 195-214

Mendoza, G.A. & Martins, H. (2006) Multi-criteria decision analysis in natural resource management: A critical review of methods and new modelling paradigms. *Forest Ecology and Management*, 230, 1-22

Meynard, C.N. & Quinn, J.F. (2007) Predicting species distributions: a critical comparison of the most common statistical models using artificial species. *Journal of Biogeography*, 34, 1455-1469

Miller, J. (2005) Incorporating Spatial Dependence in Predictive Vegetation Models: Residual Interpolation Methods. *The Professional Geographer*, 57, 169-184

Miller, J. (2010) Species dsitribution modeling. *Geography Compass*, 4, 490-509

Miller, J. & Franklin, J. (2002) Modeling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial dependence. *Ecological Modelling*, 157, 227-247

Miller, J., Franklin, J. & Aspinall, R. (2007) Incorporating spatial dependence in predictive vegetation models. *Ecological Modelling*, 202, 225-242

Moore, I.D., Grayson, R.B. & Ladson., A.R. (1991) Digital terrain modelling. A review of hydrological, geomorphological, and biological applications. *Hydrological Processes*, 5, 3-30

Ohmann, J.L. & Gregory, M.J. (2002) Predictive mapping of forest composition and structure with direct gradient analysis and nearest-neighbor imputation in coastal Oregon, USA. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, 32, 725-741

Ohmann, J.L., Gregory, M.J. & Spies, T.A. (2007) Influence of environment, disturbance, and ownership on forest vegetation of Coastal Oregon. *Ecological Applications*, 17, 18-33

Pearson, R.G. (2006a) Climate change and the migration capacity of species. *Trends in Ecology & Evolution*, 21, 111-113

Pearson, R.G. (2006b) Model-based uncertainty in species range predicition. *Journal of Biogeography*, 33, 1704-1711

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

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

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

Peterson, A.T., Papes, M. & Kluza, D.A. (2003) Predicting the potential invasive distributions of four alien plant species in North America. *Weed Science*, 51, 863-868

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

Peterson, A.T. & Vieglais, D.A. (2001) Predicting species invasions using ecological niche modeling: New approaches from bioinformatics attack a pressing problem. *Bioscience*, 51, 363-371

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

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

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

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

Phillips, S.J. & Elith, J. (2010) POC plots: calibrating species distribution models with presence-only data. *Ecology*, 91, 2476-2484

Pierce, K.B., Ohmann, J.L., Wimberly, M.C., Gregory, M.J. & Fried, J.S. (2009) Mapping wildland fuels and forest structure for land management: a comparison of nearest neighbor imputation and other methods. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, 39, 1901-1916

Pimentel, D., Lach, L., Zuniga, R. & Morrison, D. (2000) Environmental and economic costs of nonindigenous species in the United States. *Bioscience*, 50, 53-65

Plantegenest, M., Le May, C. & Fabre, F. (2007) Landscape epidemiology of plant diseases. *Journal of the Royal Society Interface*, 4, 963-972

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

Prasad, A.M., Iverson, L.R., Peters, M.P., Bossenbroek, J.M., Matthews, S.N., Sydnor, T.D. & Schwartz, M.W. (2010) Modeling the invasive emerald ash borer risk of spread using a spatially explicit cellular model. *Landscape Ecology*, 25, 353-369

Pulliam, H.R. (2000) On the relationship between niche and distribution. *Ecology Letters*, 3, 349-361

Pysek, P. & Richardson, D.M. (2007) Traits associated with invasiveness in alien plants: Where do we stand? *Biological invasions (Ecological Studies 193)* (ed. by W. Nentwig), pp. 97-126. Springer-Verlag, Berlin & Heidelberg.

Rahbek, C., Gotelli, N.J., Colwell, R.K., Entsminger, G.L., Rangel, T.F.L.V.B. & Graves, G.R. (2007) Predicting continental-scale patterns of bird species richness with spatially explicit models. *Proceedings of the Royal Society B-Biological Sciences*, 274, 165-174

Ripley, B.D. (1976) 2nd-Order Analysis of Stationary Point Processes. *Journal of Applied Probability*, 13, 255-266

Rizzo, D.M. & Garbelotto, M. (2003) Sudden oak death: endangering California and Oregon forest ecosystems. *Frontiers in Ecology and the Environment*, 1, 197-204

Rizzo, D.M., Garbelotto, M. & Hansen, E.A. (2005) *Phytophthora ramorum*: Integrative research and management of an emerging pathogen in California and Oregon forests. *Annual Review of Phytopathology*, 43, 309-335

Robinson, T.P., van Klinken, R.D. & Metternicht, G. (2010) Comparison of alternative strategies for invasive species distribution modeling. *Ecological Modelling*, 221, 2261-2269

Rodder, D., Sole, M. & Bohme, W. (2008) Predicting the potential distributions of two alien invasive Housegeckos (Gekkonidae: Hemidactylus frenatus, Hemidactylus mabouia). *North-Western Journal of Zoology*, 4, 236-246

Rouget, M. & Richardson, D.M. (2003) Inferring process from pattern in plant invasions: A semimechanistic model incorporating propagule pressure and environmental factors. *American Naturalist*, 162, 713-724

Roura-Pascual, N., Bas, J.M., Thuiller, W., Hui, C., Krug, R.M. & Brotons, L. (2009) From introduction to equilibrium: reconstructing the invasive pathways of the Argentine ant in a Mediterranean region. *Global Change Biology*, 15, 2101-2115

Santika, T. & Hutchinson, M.F. (2009) The effect of species response form on species distribution model prediction and inference. *Ecological Modelling*, 220, 2365-2379

Segurado, P. & Araujo, M.B. (2004) An evaluation of methods for modelling species distributions. *Journal of Biogeography*, 31, 1555–1568

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

Sharma, G.P., Singh, J.S. & Raghubanshi, A.S. (2005) Plant invasions: Emerging trends and future implications. *Current Science*, 88, 726-734

Simberloff, D. (2003) How much information on population biology is needed to manage introduced species? *Conservation Biology*, 17, 83-92

Smolik, M.G., Dullinger, S., Essl, F., Kleinbauer, I., Leitner, M., Peterseil, J., Stadler, L.M. & Vogl, G. (2010) Integrating species distribution models and interacting particle systems to predict the spread of an invasive alien plant. *Journal of Biogeography*, 37, 411-422

Soberon, J. (2007) Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters*, 10, 1115-1123

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

Sokal, R.R. & Rohlf, F.J. (1981) *Biometry: The Principles and Practice of Statistics in Biological Research*. W.H. Freeman and Co., New York, 887 pp.

Storch, D., Konvicka, M., Benes, J., Martinkova, J. & Gaston, K.J. (2003) Distribution patterns in butterflies and birds of the Czech Republic: separating effects of habitat and geographical position. *Journal of Biogeography*, 30, 1195-1205

Strubbe, D. & Matthysen, E. (2009) Predicting the potential distribution of invasive ring-necked parakeets *Psittacula krameri* in northern Belgium using an ecological niche modelling approach. *Biological Invasions*, 11, 497-513

Sutherst, R.W. & Bourne, A.S. (2009) Modelling non-equilibrium distributions of invasive species: a tale of two modelling paradigms. *Biological Invasions*, 11, 1231-1237

Sutton, W., Hansen, E.M., Reeser, P.W. & Kanaskie, A. (2009) Stream monitoring for detection of *Phytophthora ramorum* in Oregon tanoak forests. *Plant Disease*, 93, 1182-1186

Svenning, J.C. & Condit, R. (2008) Biodiversity in a warmer world. *Science*, 322, 206-207

Svenning, J.C. & Skov, F. (2004) Limited filling of the potential range in European tree species. *Ecology Letters*, 7, 565-573

Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J. & Collingham, Y.C. (2004) Extinction risk from climate change. *Nature*, 427, 145-147

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

Thuiller, W., Richardson, D.M., Pysek, P., Midgley, G.F., Hughes, G.O. & Rouget, M. (2005) Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. *Global Change Biology*, 11, 2234-2250

Tognelli, M.F. & Kelt, D.A. (2004) Analysis of determinants of mammalian species richness in South America using spatial autoregressive models. *Ecography*, 27, 427-436

Tooley, P.W., Browning, M., Kyde, K.L. & Berner, D. (2009) Effect of Temperature and Moisture Period on Infection of Rhododendron 'Cunningham's White' by *Phytophthora ramorum*. *Phytopathology*, 99, 1045-1052

Tsoar, A., Allouche, O., Steinitz, O., Rotem, D. & Kadmon, R. (2007) A comparative evaluation of presence-only methods for modelling species distribution. *Diversity and Distributions*, 13, 397-405

USDA Forest Service RSL (2003) CALVEG Vegetation Mapping Program. *Sacramento, California, USA.* http://www.fs.fed.us/r5/rsl/projects/mapping/

Vaclavik, T., Kanaskie, A., Hansen, E.M., Ohmann, J.L. & Meentemeyer, R.K. (2010) Predicting potential and actual distribution of sudden oak death in Oregon: Prioritizing landscape contexts for early detection and eradication of disease outbreaks. *Forest Ecology and Management*, 260, 1026-1035

Vaclavik, T. & Meentemeyer, R.K. (2009) Invasive species distribution modeling (iSDM): Are absence data and dispersal constraints needed to predict actual distributions? *Ecological Modelling*, 220, 3248-3258

Vaclavik, T. & Rogan, J. (2009) Identifying trends in land use/land cover changes in the context of post-socialist transformation in Central Europe: A case study of the greater Olomouc region, Czech Republic. *Giscience & Remote Sensing*, 46, 54-76

Venette, R.C. & Cohen, S.D. (2006) Potential climatic suitability for establishment of *Phytophthora ramorum* within the contiguous United States. *Forest Ecology and Management*, 231, 18-26

Vitousek, P.M., DAntonio, C.M., Loope, L.L. & Westbrooks, R. (1996) Biological invasions as global environmental change. *American Scientist*, 84, 468-478

Werres, S., Marwitz, R., Veld, W.A.M.I., De Cock, A.W.A.M., Bonants, P.J.M., De Weerdt, M., Themann, K., Ilieva, E. & Baayen, R.P. (2001) *Phytophthora ramorum* sp nov., a new pathogen on *Rhododendron* and *Viburnum*. *Mycological Research*, 105, 1155-1165

Woods, A., Coates, K.D. & Hamann, A. (2005) Is an unprecedented dothistroma needle blight epidemic related to climate change? *Bioscience*, 55, 761-769

Zaniewski, A.E., Lehmann, A. & Overton, J.M.C. (2002) Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modeling*, 157, 261-280

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

Zimmermann, N.E., Yoccoz, N.G., Edwards, T.C., Meier, E.S., Thuiller, W., Guisan, A., Schmatz, D.R. & Pearman, P.B. (2009) Climatic extremes improve predictions of spatial patterns of tree species. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 19723-19728