Modeling Bark Beetle Outbreak and Fire Interactions in Western U.S. Forests and the Invasion Potential of an Invasive Puerto Rican Frog in Hawaii Using Remote Sensing Data

Simon A. Bisrat
Utah State University

Follow this and additional works at: https://digitalcommons.usu.edu/etd
Part of the Ecology and Evolutionary Biology Commons, and the Geographic Information Sciences Commons

Recommended Citation
https://digitalcommons.usu.edu/etd/703

This Dissertation is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Theses and Dissertations by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.
MODELING BARK BEETLE OUTBREAK AND FIRE INTERACTIONS IN
WESTERN U.S. FORESTS AND THE INVASION POTENTIAL OF AN
INVASIVE PUERTO RICAN FROG IN HAWAII USING
REMOTE SENSING DATA

by

Simon A. Bisrat

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Ecology

Approved:

Dr. Michael White
Major Professor

Dr. Karen Beard
Committee Member

Dr. Barbara Bentz
Committee Member

Dr. Richard Cutler
Committee Member

Dr. Ron Ryel
Committee Member

Dr. Byron Burnham
Dean of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2010
ABSTRACT

Modeling Bark Beetle Outbreak and Fire Interactions in Western U.S. Forests and the Invasion Potential of an Invasive Puerto Rican Frog in Hawaii Using Remote Sensing Data

by

Simon A. Bisrat, Doctor of Philosophy
Utah State University, 2010

Major Professor: Dr. Michael White
Department: Watershed Sciences

I used Moderate-Resolution Imaging Spectroradiometer (MODIS) imagery to answer two ecological questions. In the first project, I investigated the interactions between bark beetle-caused tree mortality and fire occurrence in western U.S. forests. I used remotely sensed fire data detected by MODIS satellite and bark beetle-caused tree mortality data. I tested the hypothesis that there is an increased probability of fire incidence in bark beetle-damaged forests compared to healthy forests using conditional probability modeling across the national forests of the western U.S. regardless of forest type. My results did not show a consistent pattern (increase or decrease of conditional probability of fire occurrence, ΔCP) across all lag time periods considered. However, when ΔCP is averaged across the 5-year study period (2001-2005) fire probability increased at 2-year (16%) and 3-year (9%) lags with 0, 1, 4, and 5-year lags showing no
positive effect of bark beetle activity on fire probability. Further, when I analyzed fire-
bark beetle-caused tree mortality separately for persistent fires (fires that lasted for at
least two 8-day composite periods per season) and transient fires (fires that lasted for only
one 8-day composite period per season), the ΔCP increased in all lag periods except the
5-year lag for persistent fires. In the second stage of this project, I used a non-parametric
modeling approach to test how important bark beetle-caused tree mortality is in
influencing fire occurrence relative to other climate and topography-derived variables in
spruce-fir, Douglas-fir, lodgepole, and ponderosa pine forests. My results showed that
climate and topography-derived predictors were consistently selected as important
predictors of fire occurrence while bark beetle-caused tree mortality showing the least
importance. In the second project, I predicted the invasive potential of a Puerto Rican
frog species in Hawaii using the following MODIS products: land surface temperature;
normalized difference vegetation index and enhanced vegetation index; and leaf area
index/fraction of photosynthetically active radiation absorbed by plant canopies. My
predicted maps showed that the invasive frog species in Hawaii is likely to expand its
current habitat. My results also showed that MODIS-derived biophysical variables are
able to characterize the suitable habitats of the invasive frog species.

(154 pages)
DEDICATION

I dedicate this dissertation to all my family members, especially my wife, Makda, our lovely children, Millen and Jonathan, my mom, my dad, and last but not least to my cousin Dawit. Imagine Makda how my life would have been so meaningless without you.
ACKNOWLEDGMENTS

I would like to thank first Dr. Michael White for his unparalleled and relentless guidance and advice throughout the course of these projects and also for making every effort possible to support me financially despite the fact that all my projects had not been fully funded. Mike, you truly deserve the lion share of my praise and gratitude. I would like also to thank Dr. Karen Beard for creating the opportunity to know Mike and ultimately be able to work with him. I would like also to extend my gratitude to Dr. Jim MacMahon for encouraging me to stay here at USU and continue my program in spite of the difficulties I encountered in spring of 2006 and also for ultimately providing funding during the first crucial months of my program. Jim, without your support and encouragement it would have been impossible for me to stay at USU. I thank you for that from the bottom of my heart. I would like also to extend my heartfelt thanks to Dr. Chris Leucke for providing financial support in spring semester of 2006 and fall semester of 2009. I would like also to thank Dr. Barbara Bentz, Matt Hansen, and Jim Vandygriff of the Rocky Mountain Forestry Research Station for their technical support and helping me learn more about bark beetles. My heartfelt gratitude also goes to Dr. Richard Cutler for his guidance and crucial advice during the course of these projects. Even though Dr. Cutler was not on my committee from the very beginning, his technical guidance and advice were so instrumental to complete these projects. I would like also to thank Dr. Ron Ryel for providing much-needed vehicle for free during my field work in Hawaii and also for showing interest to be on my revised committee.
I would like also to thank the people at NASA Applied Science Program Graduate Student Fellowship and the Intermountain Digital Data Archive Center for providing me the funding to support my work in Puerto Rico and Hawaii. I also gratefully acknowledge the USFS Forest Health Protection Teams of Region 1 through 6, the Land Processes Distributed Active Archive Center, located at the U.S. Geological Survey Center for Earth Resources Observation and Science for providing data and assistance. Particularly, I acknowledge Jennifer Ross of Region 2, Daniel Ryerson of Region 3, Richard Halsey of Region 4, Zachary Heath of Region 5, and Julie Johnson of Region 6 for providing the bark beetle aerial detection survey flight and damage maps. I thank also Jesse Poulos for field assistance; Will Pitt, Chris McGinty, John Lowry, and Emily Price during data collection; Dr. Adele Cutler, Dr. Janet Franklin, Nicholas Crookston, and Gong Zang during data analysis. My deepest gratitude is also extended to Qiu Hong Tang of the University of Washington for sharing his processed MODIS-derived snow data.

I would like also to thank my wife and our two little kids for their patience and sacrifice. You truly sacrificed a lot for me to reach here. I would like also to extend my deepest gratitude to my mom, my late dad, and my siblings for loving me and believing in me. I like also say thank you to Efeson Mengesha and Mical Woldeselleassie for your wonderful friendship and making yourself available to baby-sit our kids. My special thanks also go to Dawit Fitsame. Dawit, it would have been impossible for me to come this far without your help. Last but not least, I thank God for giving me all the health and strength to complete my program.

Simon A. Bisrat
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.</td>
<td>Classification Methods</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Model accuracy assessment</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td>Map generation</td>
<td>103</td>
</tr>
<tr>
<td></td>
<td>Results</td>
<td>104</td>
</tr>
<tr>
<td></td>
<td>Discussion</td>
<td>116</td>
</tr>
<tr>
<td></td>
<td>Conclusion</td>
<td>122</td>
</tr>
<tr>
<td></td>
<td>Literature cited</td>
<td>123</td>
</tr>
</tbody>
</table>

5. SUMMARY AND CONCLUSIONS | 131  |

6. VITA | 134  |
<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Major bark beetle species and their major host species</td>
<td>14</td>
</tr>
<tr>
<td>2. Number of ADS and MODIS-detected fire pixels (1 km²) for each year considered in the study. Note that the quality of the MODIS fire data for 2000 fire year was poor</td>
<td>17</td>
</tr>
<tr>
<td>3. Predictor variables used to model the occurrence of fire</td>
<td>51</td>
</tr>
<tr>
<td>4. Accuracy metric values for models built for all areas experiencing fire in each forest type</td>
<td>66</td>
</tr>
<tr>
<td>5. Accuracy metric values for models built only for areas with bark beetle activity and fire occurrence</td>
<td>67</td>
</tr>
<tr>
<td>6. Accuracy measures for predictions of coqui presence on the island of Hawaii using monthly climatological values calculated from MODIS-generated biophysical variables as predictors</td>
<td>107</td>
</tr>
<tr>
<td>7. Distribution of extant land cover types as shown in Figure 2, measured coqui presence, and modeled habitat (probability between 0.41 and 1.0). All values are shown in percent, rounded to nearest whole number</td>
<td>109</td>
</tr>
<tr>
<td>8. Accuracy measures for predictions of coqui presence on the island of Hawaii using dry and It season values calculated for each MODIS-generated biophysical variable as predictors</td>
<td>116</td>
</tr>
</tbody>
</table>
META DATA

Figure | Page
--- | ---
1. Study area. Exact boundaries of ADS flight paths (grey) vary slightly by year; 2004 is shown here | 13
2. Hypothetical example showing how the spatial analysis was done: (a) percent bark beetle infestation based on 100 100 m² pixels within each 1 km² pixel (b) MODIS fire imagery with a 1 km pixel size; numbers represent the number of 8-day composite period fires per year (higher numbers indicating frequent fires) | 16
3. Percent change in the conditional probability of fire associated with bark beetle presence, ΔCP, at 0- to 5-year time lags analyzed separately for low and high levels of bark beetle infestation. Low infestation indicates that between 1% and 10% of the area within a 1 km pixel was covered by a bark beetle infestation; high infestation indicates greater than 10% infestation. Positive ΔCP indicates that, when compared to areas with bark beetle absence, bark beetle presence was associated with an increased probability of fire. Negative ΔCP indicates the reverse | 24
4. Percent change in the conditional probability of fire associated with bark beetle presence, ΔCP, at 0- to 5-year time lags calculated for each fire year. Positive ΔCP indicates that, when compared to areas with bark beetle absence, bark beetle presence was associated with an increased probability of fire. Negative ΔCP indicates the reverse | 24
5. Percent change in the conditional probability of fire associated with bark beetle presence, ΔCP, at 0- to 5-year time lags. All ΔCP values are averages over the entire study period. Positive ΔCP indicates that, when compared to areas with bark beetle absence, bark beetle presence was associated with an increased probability of fire. Negative ΔCP indicates the reverse | 25
Percent change in the conditional probability of fire associated with bark beetle presence, \( \Delta CP \), at 0- to 5-year time lags analyzed separately for transient and persistent fires. Transient fires burned for one 8-day MOD14A2 composite period per year; persistent fires burned for two or more composite periods. Positive \( \Delta CP \) indicates that, when compared to areas with bark beetle absence, bark beetle presence was associated with an increased probability of fire. Negative \( \Delta CP \) indicates the reverse .................................................................25

FPAR amplitude (A and B) and annual water balance (C and D) for pixels with bark beetle outbreaks and subsequent fire occurrence. In (A) and (C), probability distributions are shown separately for transient (dashed line) and persistent (solid line) fire categories. Probability distribution differences are negative where the FPAR amplitude (B) or WB (D) probability distributions are lower for persistent than transient fires. In (B) and (D), thick line is a five-year running average......26

The vegetation maps for spruce-fir (a), lodgepole pine (b), Douglas-fir (c), and ponderosa pine (d) forests used in this study are shown in green color. The red color shows the 2001-2005 fire pixels detected by MODIS in each forest type. The black solid boundary line shown around the states of Arizona and New Mexico in ponderosa pine forest map (d) shows the study area for SW ponderosa pine forests.................................................................44

Variable importance plots for spruce-fir (A), lodgepole pine (B), Douglas-fir (C), ponderosa pine (D), and SW ponderosa pine (E) forests for models fitted for all areas that experienced fire (2001-2005). A variable with the highest mean decrease in accuracy is the most important variable...............................................61

Mean variable importance plots for models fitted for all areas that experienced fire in all forest types. A variable with the highest mean decrease in accuracy is the most important variable .................................................................62
11. Variable importance plots for spruce-fir (A), lodgepole pine (B), Douglas-fir (C), ponderosa pine (D), and SW ponderosa pine (E) forests for models only fitted for areas that experienced bark beetles outbreak (2001-2005). A variable with the highest mean decrease in accuracy is the most important variable ..................................64

12. Mean variable importance plots for models only fitted for areas that experienced bark beetle outbreak in all forest types. A variable with the highest mean decrease in accuracy is the most important variable ..........................................................65

13. Maps of predictor variables: bark beetle-caused tree mortality observed in 2004 in all forest types (a), 2001-2005 mean annual precipitation (b), 2001-2005 mean fire season maximum temperature (c), 2001-2005 mean fire season fire season dew point (d), 2001-2005 mean spring and summer average temperature (e), 2001-2005 mean number of snow days (f), 2001-2005 mean PDSI (g), 2001-2005 mean NPP (h), elevation (i), slope (j), and aspect (k)........................................83

14. Sampled coqui presence/absence points in Puerto Rico superimposed on land cover map. Black points show sampled absence points and red points show sampled presence points. The land cover map is reformatted to show only the major cover types (Source: International Institute of Tropical Forestry - http://www.tropicalforestry.net for Puerto Rico.........................................................94

15. Sampled coqui presence/absence points in Hawaii superimposed on land cover map. Black points show sampled absence points, red points show sampled presence points, and blue points show known presence points obtained from HISC. The land cover map is reformatted to show only the major cover types (Source: NOAA’s Coastal Services Center, Land Cover Analysis - http://www.csc.noaa.gov/crs/lca/hawaii.html for Hawaii).............................................95

16. Coqui predicted distribution maps using monthly data from native range generated by CT (a), RF (c), and SVM (e) and using invaded range data generated by CT (b), RF (d), and SVM (f). Known coqui presence points are represented
with black dots. The legend indicates the predicted probabilities of coqui presence and known coqui presence points. Shaded relief of the island of Hawaii is shown as a background to represent the surface features ......................................................... 106

17. Predicted probability map showing the mean classification (with an upper limit of 0.70 and lower limit of 0.32 of 95% confidence interval) generated out of the six classification maps shown in Fig. 4 using monthly values as predictors (a), map showing the prediction uncertainty calculated using standard error (b). Shaded relief of the island of Hawaii is shown as a background to represent the surface features .................................................................................................................. 108

18. Map showing the prediction uncertainty calculated using standard error for the six prediction maps shown in Figure 4. Shaded relief of the island of Hawaii is shown as a background to represent the surface features................................................................. 110

19. Frequency distribution of pixels with moderate to high predicted probability of coqui presence (0.41–1.0) at 11 elevation bins .......................................................... 111

20. Graphs showing the important variables generated using RF. The left hand panel (a) shows variable importance for models built on native range data and the right hand panel (b) shows for models built on invaded range data. The x-axis shows the mean decrease in accuracy. Note that a variable with the least mean decrease in accuracy is the most important variable ................................................................. 113

21. Coqui predicted distribution maps using dry and wet season values as predictors from native range generated by CT (a), RF (c), and SVM (e) and using invaded range data generated by CT (b), RF (d), and SVM (f). Known coqui presence points are represented with black dots. The legend indicates the predicted probabilities of coqui presence and known coqui presence points. Shaded relief of the island of Hawaii is shown as a background to represent the surface feature.................................................................................................................. 114
22. Sensitivity analysis of the test data by progressively removing 20%, 40%, and 60% of the data using CT (a, b), RF (c, d), and SVM (e, f). The right hand panel shows for models built on native range data and the left hand panel shows for models built on the invaded range data.
CHAPTER 1

BACKGROUND AND RESEARCH GOALS

Remote sensing technologies are used to collect synoptic data of terrestrial ecosystems over wide spatial and temporal scales which are difficult to measure by conventional methods. In addition to providing basic descriptive parameters, remote sensing derived variables can be used as model inputs to constrain the functions of statistical or process models (Franklin, et al. 1997). The National Aeronautics and Space Administration (NASA) is making substantial contributions to understanding the dynamics of global change since it launched the Earth Observing System (EOS) in late 1990’s. The first EOS satellite, called Terra (EOS AM) was launched in 1999, carrying five remote sensors. The second EOS satellite, called Aqua (EOS PM) was launched in 2002. Among these sensors, MODIS offers a unique combinations of features: (i) it detects a wide spectral range of electromagnetic energy (36 spectral bands ranging in wavelength from 0.4 µm to 14.4 µm); (ii) it takes measurements at three spatial resolutions (two bands at 250 m, five bands at 500 m and 29 bands at 1 km); and (iii) it has a 1-2-day repeat time of imaging the globe (Justice, et al. 2002). MODIS provides measurements in large-scale global dynamics including changes in Earth's cloud cover, radiation budget and processes occurring in the oceans, on land, and in the lower atmosphere improving our understanding of global dynamics and processes. MODIS is also crucial for the development of validated, global, interactive Earth system models able to predict global change accurately.

I selected MODIS over other imaging systems mainly for its high radiometric resolution (36 spectral bands) and finer temporal resolution as it is able to capture images
across the entire globe at 1 to 2 days repeat cycle (Justice, et al. 2002). This continual and comprehensive coverage by MODIS is frequent enough to capture many rapid biological and meteorological changes. Other imaging systems such as Landsat’s ETM+, for instance, capture the image of the Earth in finer spatial detail (30 m spatial resolution), but with a temporal resolution of 16 days which is not suitable to capture rapid physical and biological processes. The 16 day temporal resolution of Landsat’s ETM+ images can even get worse due to cloud cover contamination (Sano, et al. 2007). The second motivation for selecting MODIS imagery is that the images are freely available to users.

I used several MODIS products to answer two critically important ecological questions. The first ecological question I attempted to answer was the interactions between bark beetle-caused tree mortality and fire occurrence in contiguous western U.S. forests. The prevailing hypothesis on the interactions between these two important natural disturbance agents is that bark beetle-caused tree mortality predisposes forests to increased risk of wildfire. There is, however, little existing literature to support this hypothesis. Besides, the scientific evidence on this important ecological question is either inadequate in some forest types or non-existent in other forest types.

I used MODIS fire product (MOD14A2, 1 km, 8-day composites) and bark beetle-caused tree mortality polygons obtained from United Forest Service (USFS), Forest Health Protection aerial detection survey database to answer this important research question. All of the studies that investigated bark beetle-caused tree mortality and fire interactions in western U.S. to date used ground-based fire data collected by multi-agencies such as the USFS, United States Bureau of Land Management; to the best of my knowledge, no study to date has used satellite-derived fire products to investigate
the interactions between bark beetles-caused tree mortality and fire. The multi-agency
ground-based fire datasets are known to have some positional inaccuracies and also they
may not include all fires ignited (Crimmins and Comrie 2004, Morgan et al. 2008,
Parisien, and Moritz, 2009). Satellite-detected fire data, on the other hand, could be a
more reliable source of fire occurrence data than the data acquired from traditional fire

In the first stage of this project, I used conditional probability modeling to
determine if there is an increased probability of fire incidence given bark beetle-caused
tree mortality across the entire forests of the western United States regardless of forest
type (Chapter 2). In the second stage of my project, I used a multivariate statistical
modeling approach to test how important bark beetle-caused tree mortality is in
influencing subsequent fire occurrence relative to other climate and topography-derived
variables in spruce-fir, lodgepole pine, Douglas-fir, and ponderosa pine forests (Chapter
3). I used climate and topographic layers along with bark beetle-caused tree mortality
layer to identify predictor variables that are likely to influence fire occurrence.

For the second ecological question that I attempted to answer, I used MODIS-
derived biophysical variables to predict the invasion potential of an invasive Puerto Rican
frog species in Hawaii (Chapter 4). The reason I used remote sensing-derived predictors
in this project was mainly due to the lack of reliable gridded climate data interpolated
from well distributed weather stations both in Hawaii (invaded habitat) and Puerto Rico
(native habitat). Remotely-sensed vegetation data are known to provide a better proxy for
climate variables than climate data interpolated from sparse weather stations (Bradley
and Fleishman 2008). I used MODIS-derived land surface temperature (MOD11A2, 1
km, 8-day composites) (Wan et al. 2004), normalized difference vegetation index and enhanced vegetation index (MOD13Q1, 250 m, 16-day composites) (Huete et al. 2002), and leaf area index/fraction of photosynthetically active radiation absorbed by plant canopies (MOD15A2, 1 km, 8-day) (Myneni et al. 2002) for both the native range as well as the invaded range. I assumed here that by incorporating data from the native range, where the frog species has been for thousands of years, I will have a better predictive model than using data solely from the invaded range.

Traditionally, ecologists used models to predict species distribution or identify variables that are important in determining the occurrence of fire or other natural disturbances assuming a linear relationship between species/disturbance occurrence (as response variable) and environmental factors (as explanatory variables) and for error to be normally distributed. However, satisfying such assumptions in ecological data is difficult; as a result new non-parametric modeling approaches such as Classification Trees (CT) and Random Forests (RF) that can model non-linear relationships and also handle discrete response and predictor variables are being increasingly used in ecological research (Guisan and Zimmermann 2000). Furthermore, the classic least square regression techniques do not lend themselves to handle discrete environmental variables as all the explanatory as well as the response variables need to be continuous (Manel et al. 1999). In the second stage of my first project where I investigated the interactions between bark beetle-caused tree mortality and fire, I used RF algorithm to identify variables important in influencing fire occurrence. In the second project where I predicted the invasion potential of an invasive frog species in Hawaii, I used a combination of CT,
RF, and Support Vector Machine (SVM) algorithms to predict the invasion potential of
the frog species in Hawaii.

Literature cited


Crimmins, M. A., and A. C. Comrie. 2004. Interactions between antecedent climate and
wildfire variability across south-eastern Arizona. International Journal of

from two years of MODIS data. International Journal of Wildland Fire 14:117-
130.

Estimation of forest Leaf Area Index using remote sensing and GIS data for
18:3459-3471.


Overview of the radiometric and biophysical performance of the MODIS


CHAPTER 2

INTERACTIONS BETWEEN BARK BEETLE-CAUSED TREE MORTALITY AND FIRE IN WESTERN U.S. FORESTS

Abstract

In western U.S. forests the interaction between bark beetle-caused tree mortality and fire is a poorly understood yet pressing management concern. Although conventional wisdom holds that dead trees will increase fire risk, no comprehensive study has attempted to study bark beetle–fire interactions over the entire western U.S. Here, using 2000 to 2004 bark beetle aerial detection survey (ADS) data and 2001 to 2005 Moderate Resolution Imaging Spectroradiometer (MODIS) fire images (MOD14A2), I calculated the influence of bark beetle-caused tree mortality on 0- to 5-year lagged subsequent fire occurrence. I calculated the conditional probability of fire given bark beetle presence \[P(F|BB\text{pres})\] and absence \[P(F|BB\text{abs})\] and then change (increase or decrease) in the conditional probability of fire associated with bark beetle presence, termed \(\Delta CP\), as

\[\frac{P(F|BB\text{pres}) - P(F|BB\text{abs})}{P(F|BB\text{abs})} \times 100.\]

Bark beetle-caused tree mortality increased the probability of fire, but only at the 2-year (\(\Delta CP 16\%\)) and 3-year (\(\Delta CP 9\%\)) lags. At 0- and 1-year lagged \(\Delta CP\) approached zero; at longer lags, bark beetle-caused tree mortality appears to have a negative effect on fire probability (\(\Delta CP -20\%\) at 4-year lag; \(\Delta CP -8\%\) at 5-year lag). When analyzed separately as persistent fires, (lasting two or more MOD14A2 8-day composite periods per season, a 12\% subset of the full dataset), \(\Delta CP\) was positive.
at all lags except the 5-year lag, with peaks of 39% at 2-year and 41% at 3-year lags. My results suggest that bark beetle-caused tree mortality can increase fire probability in contiguous western U.S. forests, especially at the 2- and 3-year time lags and for persistent fire events.

Introduction

Natural disturbances are ecologically important because they can create long-term ecological legacies and determine vegetation patterns in forested landscapes (Turner and Dale 1998, Veblen 2000, Bebi et al. 2003,). The interaction between bark beetles and fire in western U.S. forests is a pressing management concern. Fire is often considered the most important ecological disturbance in western United States (U.S.) forests and as a result is well studied (Veblen 2000). However, insects and diseases also influence age structure and species composition of most forests (Amman 1991). Researchers have historically focused on understanding the distribution, frequency, size and magnitude of disturbances (e.g., Turner et al. 1994, McCullough et al. 1998,) with only a few studies investigating synergistic disturbance interactions (e.g., Kulakowski et al. 2003) and/or long-lasting disturbance legacies influencing ecosystems processes, population dynamics and subsequent disturbances (Knight 1987, McCullough et al. 1998).

One such disturbance interaction, and the subject of this paper, is the interaction between bark beetle-caused tree mortality and fire disturbances. Many bark beetle species have large ranges and can be found throughout the majority of forest ecosystems in the western U.S. (Wood 1981) (Table 1). Although fire is more visually dramatic, insect outbreaks, particularly bark beetle outbreaks, are widespread in western North American
coniferous forests and in some cases may be as ecologically important as fire (Dale et al. 2001, Bebi et al. 2003). Aggressive bark beetles assume this importance because they typically kill host trees prior to successful reproduction, influence tree mortality via phloem girdling which interrupts photosynthate flow, and influence forest ecosystem structure by modifying forest succession, nutrient cycling, and the species, size and distribution of trees. In some pine forests, fire suppression can alter stand structure and composition, making forests more prone to broad scale insect outbreaks (McCullough et al. 1998). For example, the large scale mountain pine beetle (*Dendroctonus ponderosae* Hopkins) outbreaks currently active in lodgepole pine (*Pinus contorta* Dougl) stands in British Columbia are in part related to fire management policy the past 100 years (Taylor et al. 2006).

Bark beetle outbreaks have been widely perceived to increase forest fire ignition and severity in western U.S. forests (Hopkins 1909, Schmidt 1988). Some authors argue that beetle-caused tree mortality increases the incidence of subsequent fires as a result of: snags acting as lightning rods (Barrows 1951, Schmid and Hinds 1974); dead needles increasing the total amount of fuel load and vertical continuity of crown fires (Brown 1975, Knight 1987); or increased radiation transmission through the defoliated canopy increasing fuel desiccation (Brown 1975, Furayev et al. 1983).

Indeed, some researchers reported an increased rate of fire incidence in bark beetle attacked forests. In a study of Colorado subalpine forests following the 2002 Hayman fire, a 1940s beetle outbreak slightly increased the rate of fire incidence, particularly in high severity fires (Bigler et al. 2005). However, pre-burn stand structure and forest type were found to be more important than the beetle-caused tree mortality for
explaining fire severity. Similarly, in the 1988 Yellowstone fires, forests that were killed by mountain pine beetles in the period 1972 to 1975 showed an 11% increase in rate of fire occurrence, whereas areas with trees killed by beetles in the early 1980s did not correlate well with the fire pattern in 1988 (Lynch et al. 2006). Both studies suggest that the secondary effects of bark beetle-caused mortality on stand structure exceed the primary effects of beetle-induced fuel accumulation on increasing fire incidence. These results also suggest that the effect of bark beetle killed trees may exhibit time lag dependencies.

On the other hand, other researchers found that severe episodes of bark beetle-caused tree mortality do not increase the rate of fire incidence. In support of this argument, at least three studies in subalpine forests of Colorado have shown that spruce bark beetle outbreaks did not increase susceptibility to subsequent fires. First, in a study that tested the effect of 1998 to 2001 beetle outbreaks on the 2002 fire in northwestern Colorado, no conclusive evidence was found to support a hypothesis of increased fire extent or severity following bark beetle outbreaks (Kulakowski and Veblen 2007). Second, in a ~2800 km² study area in the White River National Forest in northwestern Colorado, a severe spruce bark beetle outbreak of the 1940s was not correlated with increased fire frequency per area during the subsequent 50 years (Bebi et al. 2003). The authors found no empirical evidence to support the three principal arguments indicated above for increased fire hazard in pine forests following bark beetle outbreaks. Third, in another White River National Forest study, the 1940s spruce bark beetle outbreaks had no effect on increasing the rate of fire incidence in the 1950s (Kulakowski et al. 2003).
Thus, based on results from the available geographically limited studies, there is evidence that bark beetle-caused tree mortality may influence subsequent fire probability in some ecosystems and regions, but not in others. There is, therefore, an urgent need to investigate bark beetle and fire disturbance interactions at broad scales relevant to policy decisions controlling salvage logging and fire prevention strategies. To address this need, I assembled complementary geospatial datasets of bark beetle outbreaks and fire occurrence at 1 km spatial resolution and assessed the following key questions across western U.S. forests: 1) as a preliminary step for framing subsequent questions, do within-pixel levels of beetle infestation influence subsequent fire probability? 2) does bark beetle-caused tree mortality affect subsequent fire probability and if so, are there time lags at which the effect is stronger? 3) does the effect of bark beetle-caused tree mortality on subsequent fire probability vary by fire persistence levels?

Methods

Study area

My study area covered the contiguous western U.S. not including Alaska and consisted primarily of National Forests within Regions 1 through 6 of the United States Forest Service (USFS, Fig. 1). However, some forests managed by the Bureau of Land Management, individual states and the National Park Service that were surveyed for bark beetle detection are also included in the study.
**Bark beetles: data and processing**

I obtained the 2000 to 2004 USFS-Forest Health Protection aerial detection estimates of bark beetle-caused tree mortality. The bark beetle aerial detection survey (hereafter ADS data) are collected using low-level flights, typically 300 to 600 m above ground level, and manual mapping of tree mortality by the main species of tree killing bark beetles in all western U.S. forests types (Table 1). ADS data include numerous fields, such as polygon boundaries for groups of beetle-killed trees by species, polygon area, number of trees killed, trees killed per area, host species, and forest type, but not all fields are available for all polygons. Consequently I conducted analysis with only the geographical polygon boundaries that included bark beetle activity.
<table>
<thead>
<tr>
<th>Common Name</th>
<th>Scientific Name</th>
<th>Major Host Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Douglas fir beetle</td>
<td><em>Dendroctonus pseudotsugae</em> Hopkins</td>
<td>Douglas-fir (Pseuotsuga menziesi Mirb.)</td>
</tr>
<tr>
<td>Douglas fir pole beetle</td>
<td><em>Pseudohylesinus nebulosus</em> LeConte</td>
<td>Douglas-fir (Pseuotsuga menziesi Mirb.)</td>
</tr>
<tr>
<td>Jeffrey pine beetle</td>
<td><em>D. jeffreyi</em> Hopkins</td>
<td>Jeffrey pine (<em>Pinus jeffreyi</em> Balf.)</td>
</tr>
<tr>
<td>Mountain pine beetle</td>
<td><em>D. ponderosae</em> Hopkins</td>
<td>Lodgepole (<em>Pinus contorta</em> Dougl.) and ponderosa pines (<em>P. ponderosa</em> Laws.)</td>
</tr>
<tr>
<td>Western pine beetle</td>
<td><em>D. brevicomis</em> LeConte</td>
<td>Ponderosa pines (<em>Pinus ponderosa</em> Laws.)</td>
</tr>
<tr>
<td>Spruce beetle</td>
<td><em>D. rufipennis</em> Kirby</td>
<td>Engelmann spruce (<em>Picea engelmanni</em> Parry) and subalpine fir (<em>Abies lasicorpa</em> Nutt.)</td>
</tr>
<tr>
<td>Western basalm bark beetle</td>
<td><em>Dryocoetes confuses</em> Swaine</td>
<td>Engelmann spruce (<em>Picea engelmanni</em> Parry) and subalpine fir (<em>Abies lasicorpa</em> Nutt.)</td>
</tr>
<tr>
<td>True fir bark beetles</td>
<td><em>Scolytus spp.</em> Geoffrey</td>
<td>Engelmann spruce (<em>Picea engelmanni</em> Parry) and subalpine fir (<em>Abies lasicorpa</em> Nutt.)</td>
</tr>
<tr>
<td>Pine engraver</td>
<td><em>Ips pini</em> Say</td>
<td>Lodgepole pine (<em>Pinus contorta</em> Dougl.), Ponderosa pines (<em>Pinus ponderosa</em> Laws.)</td>
</tr>
<tr>
<td>Ips engraver</td>
<td><em>Ips spp.</em> De Geer</td>
<td>Lodgepole pine (<em>Pinus contorta</em> Dougl.), Ponderosa pines (<em>Pinus ponderosa</em> Laws.)</td>
</tr>
</tbody>
</table>

I implemented a four-step processing scheme of the ADS data. First, I established a geographically consistent study area. ADS flight paths, which were provided by USFS staff, vary by USFS regional office and by year within regions. In order to establish a consistent study area for comparison with the MODIS fire data (see below), I extracted the annual geographical bounding longitude (most western and eastern) and latitude (most southern and northern). I then bounded the final study area by the maximum and minimum latitudes and longitudes from the 5-year ADS dataset. Second, within each year, I individually assessed Regions 1 through 6 and defined my analysis region only as those areas covered by the ADS flight paths to exclude areas that are not flown.
Consistent digital information on flight paths from all regions was unavailable prior to 2000, thus fixing the study period (USFS staff, personal communications). In support of delineating study regions, I also used ancillary information from the MODIS fire images (see below) to exclude water bodies. Third, I converted the ADS polygons to a 100 m resolution raster layer (ArcGIS version 9.1; ESRI) such that each pixel contained a 0 (bark beetle-caused tree mortality absent) or 1 (bark beetle-caused tree mortality present). Fourth, for comparison to the 1 km MODIS fire data, I created two 1 km ADS maps from the 100 m maps for each fire year (Table 2) to indicate: (1) bark beetle activity classified as either low (between 1% and 10% of 100 m pixels with bark beetle activity) or high (greater than 10% of 100 m pixels with bark beetle activity, Fig. 2); and (2) a simple binary classification of bark beetle activity absence (no 100 m pixels with bark beetles activity present) or presence (at least one 100 m pixel with bark beetle activity present). The majority of 1 km pixels with bark beetle activity (60%) were classified as low.

Fire: data and processing

I obtained 2001 to 2005 1 km MODIS Terra 8-day composite thermal anomalies/fire product (MOD14A2, Collection 4) from the Land Processes Distributed Active Archive Center (LP-DAAC) Data Gateway. Launched in 1999 on the Terra platform and in 2002 on the Aqua platform, MODIS began capturing images as of February 2000 but 2000 fire season data are considered to be unreliable (Giglio 2005). MOD14A2 (Justice et al. 2002) is based on MODIS brightness temperatures derived in the 4µm ($T_4$) and 11µm channels ($T_{11}$). When possible, $T_4$ is derived from the MODIS band with saturation at 311K, which has low noise and small quantization error, rather
than the band saturating at 500K. $T_{11}$ is computed from the 11µm channel, which saturates at nearly 400 K (Justice et al. 2002). The thermal anomaly/fire product is then processed as a daily Level 2 product (MOD14) describing active fires and thermal anomalies. I used the Level 3 MOD14A2 product, which represents the 8-day maximum value of the individual Level 2 values (Masuoka et al. 1998). Each MOD14A2 1 km fire is assigned one of three detection confidence levels: low-confidence fire, nominal-confidence fire, or high-confidence fire (Justice et al. 2002). In this study, I accepted all three classes as valid fire pixels as the low-confidence fire comprise only 1.3 % of the total fires detected.

![Hypothetical example showing how the spatial analysis was done: (a) percent bark beetle infestation based on 100 100 m² pixels within each 1 km² pixel (b) MODIS fire imagery with a 1 km pixel size; numbers represent the number of 8-day composite period fires per year (higher numbers indicating frequent fires).](image)

I implemented a four-step processing scheme for the MODIS fire data. First, for each 8-day composite, I mosaicked the individual MODIS 1200 km×1200 km tiles, reprojected and co-registered to the ADS Albers Equal Area projection, and then...
subsetted to the ADS study area described above (ERDAS Imagine version 9.0, Leica Geosystems). Second, for each 8-day composite period, I assigned a 0 to pixels without fires and a 1 to pixels with fires. Third, I summed all 46 8-day composite periods images within individual years, producing an annual fire value for each 1 km pixel such that higher values indicated more persistent fires (Fig. 2 and Table 2). Fourth, I categorized burned pixels as transient or persistent fires. I identified the 88% of pixels with fires burning in only one 8-day composite period per year as transient fires. The remainder, burning for two or more 8-day composite periods per year, I termed persistent fires. Further subdivision into multiple persistence categories was unjustified due to limited sample sizes. MOD14A2 does not directly provide burn intensity or burn area within each 1 km pixel; my calculation of persistence is a generic proxy indicator in which I subjectively assume that longer burning fires are more severe.

**TABLE 2.** Number of ADS and MODIS-detected fire pixels (1 km$^2$) for each year considered in the study. Note that the quality of the MODIS fire data for 2000 fire year was poor.

<table>
<thead>
<tr>
<th>Year</th>
<th>ADS</th>
<th>Fire</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>50009</td>
<td>No data</td>
</tr>
<tr>
<td>2001</td>
<td>63964</td>
<td>3319</td>
</tr>
<tr>
<td>2002</td>
<td>96546</td>
<td>6869</td>
</tr>
<tr>
<td>2003</td>
<td>158710</td>
<td>6077</td>
</tr>
<tr>
<td>2004</td>
<td>155283</td>
<td>2671</td>
</tr>
</tbody>
</table>
Ancillary data and processing

I assembled two additional dataset in support of analysis of bioclimatic interactions with persistent versus transient fires following bark beetle outbreaks. First, as an indicator of overall moisture-related climate conditions, I calculated 1990 to 2003 average annual water balance (WB):

\[
WB = \frac{\sum_{\text{years}=1990}^{2003} \sum_{\text{doy}=0}^{364} \text{precipitation} - \text{PET}}{n}
\]

where \( n \) is the number of years, \( \text{precipitation} \) is daily precipitation, \( \text{PET} \) is daily potential evapotranspiration calculated with a Priestley-Taylor approach (Priestley and Taylor 1972), and meteorological inputs are from the Daymet dataset (Thornton et al. 1997). High WB indicates precipitation either exceeds or nearly meets evaporative potential. Although long-term WB averages are insensitive to the period of calculation, I included the decade prior to the study period to incorporate historical moisture conditions likely to influence system drought and fuel load. Second, as a synoptic indicator of biological activity, I obtained 2000 to 2005 1 km resolution monthly Fraction of Photosynthetically Active Radiation (FPAR) composites from the University of Boston (http://cliveg.bu.edu/modismisr, standard MODIS MOD15 product) (Myneni et al. 2002). To reduce atmospheric and cloud effects, I processed maximum value monthly FPAR composites (i.e. one value per month representing the highest monthly FPAR from 2000 to 2005) and then calculated the annual FPAR maximum, minimum, average, and amplitude.
Modeling and analysis

I used conditional probability modeling (Papoulis, A., and Pillai, S.U. 2002) to investigate the effect of bark beetle outbreaks on subsequent fire probability at six time lags as a function of three conditions: (1) bark beetle presence at low and high infestation levels combined with any fire persistence; (2) bark beetle presence at any infestation level combined with any fire persistence; and (3) bark beetle presence at any infestation level combined with transient and persistent fires. I first calculated the probability of beetle presence, $P(B\text{B}_{\text{pres}})$, for each individual year as:

$$P(B\text{B}_{\text{pres}}) = \frac{n_{BB_{\text{pres}}}}{n} \quad (2)$$

where $n$ is the number of study pixels (specific to annually-defined ADS flight paths for Forest Service Regions 1 to 6) and $n_{BB_{\text{pres}}}$ is the number of study pixels with ADS bark beetle presence. Note that depending on which of the three above conditions is under analysis $n_{BB_{\text{pres}}}$ refers to bark beetle presence at any infestation level or bark beetle presence at high or low infestation levels. I then calculated the joint probability of both fire and bark beetle occurrence, $P(F \cap B\text{B}_{\text{pres}})$, for each individual year as:

$$P(F \cap B\text{B}_{\text{pres}}) = \frac{n_{F_{t+\text{lag}} \cap BB_{\text{pres},t}}}{n} \quad (3)$$
where \( n F_{t+\text{lag}} \cap BB_{\text{pres}} \) is the number of study pixels with both MODIS fire detection at time lag \( t+\text{lag} \) and ADS bark beetle presence at time \( t \). Again depending on the condition being assessed, \( F \) refers to fire detection for any class of fire or fire detection for transient and persistent fires. Time lags for my study ranged from zero to 5 years. Finally, the conditional probability of fire given bark beetle presence, \( P(F|BB_{\text{pres}}) \), is:

\[
P(F | BB_{\text{pres}}) = \frac{P(F \cap BB_{\text{pres}})}{P(BB_{\text{pres}})}
\]  

Using the same sequence of equations but substituting bark beetle absence for bark beetle presence, I calculated the conditional probability of fire given bark beetle absence, \( P(F|BB_{\text{abs}}) \) and assigned the percent change in the conditional probability of fire associated with bark beetle presence, \( \Delta CP \), as

\[
\Delta CP = \frac{P(F | BB_{\text{pres}}) - P(F | BB_{\text{abs}})}{P(F | BB_{\text{abs}})} \times 100
\]  

\( \Delta CP \) thus represents the impact of bark beetle presence on subsequent fire probability relative to the probability of fire in forests unaffected by bark beetles. I averaged the annual \( \Delta CP \) values over the entire 2000 to 2005 study period.

Following the \( \Delta CP \) analysis, I pursued a related fire persistence analysis. Instead of assessing fire probability as a function of bark beetle presence versus bark beetle absence, I analyzed only those pixels containing bark beetle presence. For each year from
2000 to 2004, I identified all pixels containing either transient or persistent fires at any
time lags. For these pixels, I then extracted WB and FPAR minimum, maximum, mean,
and amplitude and calculated the probability densities for the transient and persistent fire
categories.

Results

Lagged bark beetle – fire interactions at
different infestation levels

Assessed at 1 km resolution for the western U.S. using ADS data and MODIS fire
data, I did not find conclusive evidence to support an affirmative answer to question 1,
“Do within-pixel levels of beetle infestation influence subsequent fire probability?” (Fig.
3). I base my response on the central metric of ΔCP, the percent change in the conditional
probability of fire associated with bark beetle presence relative to bark beetle absence,
averaged across the 2000 to 2005 study period. Averaged across the 0- to 5-year time
lags, high bark beetle infestation levels relative to low infestation levels were associated
with only a 5% increase in ΔCP largely driven by a peak ΔCP of -20% at the 4-year lag
(Fig. 3; note that as our results are presented for a census, rather than a sampling, error
bars are not included). High infestation ΔCP exceeded low infestation ΔCP by 6% or less
at the 0-, 1-, and 3-year lags but was less than low infestation ΔCP at the 2- and 5-year
lags. Essentially, while four of six lags showed evidence that greater amounts of bark
beetle-killed trees within a 1 km pixel were associated with higher subsequent fire
probability, the magnitude of the effect was usually minor and conflicting evidence
expected at two time lags. Based on these inconclusive results, I conducted all remaining
analysis using a simple binary classification of each 1 km pixel as either containing or not containing ADS-derived bark beetle-caused tree mortality.

**Lagged bark beetle – fire interactions**

Based on the analysis of a simple binary classification of each 1 km pixel as either containing or not containing bark beetle-caused tree mortality for each fire year, my results did not show a consistent increase or decrease in percent change of conditional probability of fire ($\Delta CP$) associated with bark beetle presence relative to bark beetle absence across the different time lags (Fig. 4). For instance, there was a decrease in $\Delta CP$ at 2-year lag in the 2002 fire year but the 2-year lag showed an increase in $\Delta CP$ in the 2003 fire year. However, when the $\Delta CP$ for each lag time is averaged for the entire study period (2001-2005), I found either no effect, a positive effect, or a negative effect of bark beetle presence on subsequent fire probability (Fig. 5). My response to question 2, “Does bark beetle-caused tree mortality affect subsequent fire probability and if so, are there time lags at which the effect is stronger”, is therefore affirmative but qualified. For the 0- and 1-year lags, $\Delta CP$ was near zero, indicating that the conditional probability of fire given bark beetle presence, $P(F|BB_{pres})$, was essentially the same as the conditional probability of fire given bark beetle absence, $P(F|BB_{abs})$. Bark beetle-caused tree mortality produced a peak $\Delta CP$ of 16% at the 2-year lag followed by 9% at the 3-year lag (Fig. 5) and decreased fire probability at 4-year ($\Delta CP$ -20%) and 5-year ($\Delta CP$ -8%) time lags. In summary, my results suggest that when information on MODIS fire persistence is not included, i.e. pixels in any fire season are considered to have burned or not burned; bark beetle activity does not affect short-term (0- to 1-year time lag) fire probability.
However, two to three years after a mortality event fire probability increases by approximately 13%, and thereafter decreases.

*Lagged bark beetle – fire interactions for transient and persistent fires*

When I categorized MODIS fire pixels as burned by transient fires detected in only one MODIS 8-day composite period per year or burned by persistent fires detected in two or more composite periods per year, ΔCP showed marked difference between the two fire types at all time lags, often by more than 10% (Fig. 6). ΔCP was higher for persistent fires at all time lags except 5-year (Fig. 6), strongly suggesting that long-lived fires are more affected by bark beetle activity than are transient fires. My answer to question 3, “*Does the effect of bark beetle-caused tree mortality on subsequent fire probability vary by fire persistence levels*” is consistently affirmative. Similar to the overall analysis (Fig. 5), ΔCP was highest at the 3-year (41%) and 2-year (39%) time lags and lowest at the 4-year lag. Persistent fire ΔCP was positive and greater than 10% at all time lags except at five years, where ΔCP was -5 (Fig. 6). The effect of bark beetles activity on transient fires (Fig. 6), which comprised 88% of all fire-affected pixels in my study, was within ± 4% of the ΔCP for all fires lumped into one category (Fig. 5) at all time lags.
FIG. 3. Percent change in the conditional probability of fire associated with bark beetle presence, \( \Delta CP \), at 0- to 5-year time lags analyzed separately for low and high levels of bark beetle infestation. Low infestation indicates that between 1\% and 10\% of the area within a 1 km pixel was covered by a bark beetle infestation; high infestation indicates greater than 10\% infestation. Positive \( \Delta CP \) indicates that, when compared to areas with bark beetle absence, bark beetle presence was associated with an increased probability of fire. Negative \( \Delta CP \) indicates the reverse.

FIG. 4. Percent change in the conditional probability of fire associated with bark beetle presence, \( \Delta CP \), at 0- to 5-year time lags calculated for each fire year. Positive \( \Delta CP \) indicates that, when compared to areas with bark beetle absence, bark beetle presence was associated with an increased probability of fire. Negative \( \Delta CP \) indicates the reverse.
FIG. 5. Percent change in the conditional probability of fire associated with bark beetle presence, ΔCP, at 0- to 5-year time lags. All ΔCP values are averages over the entire study period. Positive ΔCP indicates that, when compared to areas with bark beetle absence, bark beetle presence was associated with an increased probability of fire. Negative ΔCP indicates the reverse.

FIG. 6. Percent change in the conditional probability of fire associated with bark beetle presence, ΔCP, at 0- to 5-year time lags analyzed separately for transient and persistent fires. Transient fires burned for one 8-day MOD14A2 composite period per year; persistent fires burned for two or more composite periods. Positive ΔCP indicates that, when compared to areas with bark beetle absence, bark beetle presence was associated with an increased probability of fire. Negative ΔCP indicates the reverse.
Compared to transient fires, persistent fires in forests with bark beetle-caused tree mortality tended to occur in areas characterized by high annual amplitude in FPAR (ranging from 0% – no plant absorption, to 100% – complete plant absorption; a proxy indicator of photosynthetic activity) and positive water balance (WB) (Fig. 7). Transient fires occurred in pixels with peak FPAR amplitude probability of 17% while persistent fires peaked at 44% (Fig. 7A). Expressed as a probability distribution, FPAR amplitudes greater than about 40 were consistently more probable in persistent than in transient fires (Fig. 7B). WB probability difference (Fig. 7D) switched from negative to positive at about 0.5 m year\(^{-1}\). Note that I also analyzed annual minimum, maximum, and average FPAR but found no significant patterns.

**FIG. 7.** FPAR amplitude (A and B) and annual water balance (C and D) for pixels with bark beetle outbreaks and subsequent fire occurrence. In (A) and (C), probability distributions are shown separately for transient (dashed line) and persistent (solid line) fire categories. Probability distribution differences are negative where the FPAR amplitude (B) or WB (D) probability distributions are lower for persistent than transient fires. In (B) and (D), thick line is a five-year running average.
Discussion

Results from my study suggest that in western forests of the contiguous U.S., bark beetles create a moderate and apparently temporary increase in the conditional probability of fire occurrence, especially two to three years after bark beetle activity. I hypothesize that bark beetle-caused tree mortality may change fuel quality and quantity through an increase in the abundance of dead, dry fuels within the canopy in the years immediately following beetle activity, which may result in an increased risk of crown fire (Knight 1987). My findings support this argument. Because MODIS fire data was only available beginning in 2001 (2000 excluded due to quality issues), other processes, such as effects related to change in stand structure and cover type in the longer-term (Bigler et al. 2005, Lynch et al. 2006) cannot be assessed here.

I theorize that my inconclusive findings relating bark beetle infestation levels to ΔCP (Fig. 1) may be related to the 1 km spatial resolution of the MODIS fire data: no information is available within the 1 km pixel regarding the location of the fire. Given that 60% of pixels in the 1 km percent bark beetle infestation layer were between 1% and 10% infested, most bark beetle-caused tree mortality occurred at much finer spatial resolution than could be resolved with MODIS fire data. My aspatial comparison at the 1 km level was therefore unlikely to be the most appropriate method with which to assess fine resolution infestation level – fire linkages. See conclusions for specific recommendations.

Use of the 8-day composite MODIS data did provide the opportunity to assess the effects of bark beetles on transient versus persistent fires (Fig. 6), a distinction that would be difficult to discern with maps of fire presence and absence only. I submit that even
though transient fires (n=12,510) were far more frequent than persistent fires (n=1,786), persistent fires are likely to be more relevant to long-term fire fighting efforts, air quality, erosion, and other fire-related consequences (Woindzell and King 2003, Phuleria et al. 2005). The effect of bark beetle activity on fire, as measured by ΔCP, was often more than three times higher for persistent versus transient fires, suggesting that bark beetle activity has only a minor effect on short-lived fires but a measurable and positive effect on persistent fires, at least within a 0- to 4-year time lag.

My results also imply that bioclimatic conditions influence the probability of transient versus persistent fires in areas with bark beetle-caused tree mortality (Fig. 7). I found transient fires to be more probable in dry and low FPAR amplitude systems, while persistent fires are more probable in moist and high FPAR amplitude systems. High WB indicates that precipitation exceeds or nearly meets evaporative potential; high FPAR amplitude in evergreen needle leaf western forests is most likely related to winter snow cover, again indicating the potential for extended high soil moisture and subsequently higher biomass. Possibly confounded by climate change (Breshears et al. 2005), high WB/FPAR amplitude systems may thus be poised to function as generators of persistent fires following bark beetle outbreaks – this distinction, if confirmed by higher resolution studies (see Conclusion), may serve as a useful management tool for prioritizing fire danger reduction and/or fire control.

Despite appearances to the contrary, my central finding of increased ΔCP for persistent fires at the 0- to 4-year time lags does not directly contradict other researchers reporting no effect of bark beetle activity on subsequent fires (Bebi et al. 2003, Kulakowski et al. 2003, Kulakowski and Veblen 2007). I attribute the seeming
discrepancy to different study durations: I assessed short-term lagged effects (0 to 5 years) while other studies investigated time lags within the range of 5 to 50 years. My results showing decreasing or negative ΔCP at longer lags (Fig. 2 and 3) are in fact consistent with the shortest of the longer-term studies, which at a 5-year lag showed no effect of beetle outbreaks on subsequent fire probability (Kulakowski and Veblen 2007).

The following sequence of events describing bark beetle activity and subsequent fire probability is consistent with a linkage between my short-term studies and existing long-term studies. First, during the initial and following year of bark beetle-caused mortality, the availability of fine fuels in the canopy is increased but reductions in understory moisture content caused by increased direct beam radiation are offset by reductions in canopy evaporation and sublimation. Second, the availability of dry, flammable fuels in tree crowns as a result of the beetle-caused tree mortality peaks 2 and 3 years after beetle outbreaks, prior to extensive understory regrowth. Third, at the 4-5 year lag, most trees are standing dead sticks, which will not contribute to crown fire. After 5 years, the dead snag may fall to the ground and thereby contribute to ground fire; however, at this stage the regrowth of moist understory vegetation may reduce fuel availability and flammability (Kulakowski et al. 2003).

My results also differ in scope from other work reporting no effect of bark beetle activity on fire probability as most of the studies used only the cool, mesic, high-elevation subalpine forests of the Rocky Mountains (Bebi et al. 2003, Kulakowski et al. 2003, Bigler et al. 2005, Kulakowski and Veblen 2007). My data considered all western U.S. forest types including the dry, hot, low-elevation Pinus ponderosa pine forests and the mid-elevation mixed conifer forests in addition to the cool, mesic, high-elevation...
subalpine forests. In subalpine forests, fire is regulated more by climate than by fuel
type and quantity: therefore the addition of dry and flammable fuel as a result of beetle-
caus ed mortality may not increase fire probability unless combined with suitable and
infrequent climate (Bebi et al. 2003, Kulakowski et al. 2003, Schoennagel et al. 2004,
Kulakowski and Veblen 2007.). In the dry, hot low-elevation P. ponderosa forests of the
Southwest, which I included, fire occurrence and spread is more controlled by fuel type
and quantity than by climate (Swetnam and Betancourt 1998). Mid-altitude mixed conifer
forests are intermediate, such that fire frequency and regime is a function of both fuel and
climate (Schoennagel et al. 2004). In summary, I argue that the use of shorter time lags
and the inclusion of fuel-sensitive forest ecosystems in my study influenced the apparent
discrepancy between existing literature and my results.

Although the MODIS fire data is well validated using USFS in situ fire
observations, high resolution remote sensing, and other regional validation sites
(http://modis-fire.umd.edu/validation.asp), the ADS data contain significant uncertainties.
Primarily, as the ADS bark beetle polygons are manually drawn during flights, the
precise boundaries contain uncertainties of approximately 250 m (Lynch et al. 2006),
likely influencing my finding of no influence of infestation level on fire probability.
Analyses of the ADS data at resolutions finer than my 1 km raster would therefore be
subject to additional uncertainties. Biologically, uncertainties exist in time lag detection
and, to a lesser extent, host species misidentification. I assumed that ADS detection
indicates that the beetle attack occurred in the prior year followed by canopy desiccation
and detectable canopy coloring in the current year. However, for some beetle species
and/or host species combinations and particularly in dry years, visual canopy coloring
may occur during the year of initial attack. Such spatial and biological uncertainties, combined with incomplete metadata for many polygons, translated directly to difficulties in exploring causality in ΔCP interannual variability. The area of bark beetles attacks is increasing (Table 2), possibly leading to more continuous fuel sources. In future work, I will expand the present analysis to track this effect through 2009.

Conclusion

My study is the first to explore short-term lagged bark beetle–fire interactions within the contiguous western U.S. forests. I found that at 2- to 3-year time lags, bark beetle outbreaks increased subsequent fire probability by approximately 13% with negligible or negative effects at shorter and longer time lags. When I analyzed persistent fires as a separate class, though, bark beetle outbreaks increased fire probability at 0- to 4-year lags with a peak of 41% at a 3-year lag. Despite this marked increase, the persistent fires comprise only 12% of the fires leaving the majority of fires either unaffected by bark beetle activity or having a small effect. Given intense management and public health interest in managing fires in the western U.S., my results suggest that bark beetle–fire interactions become a significant consideration only for persistent fires. Disturbance interactions are highly complex: given the coarse resolution of both the ADS and MODIS data and related spatial and biological uncertainties, my study cannot disentangle factors such as insect and host species, disturbance history, fuel load and moisture conditions, and short- and long-term climate variability likely to influence variable beetle–fire interactions. To address this shortcoming and to provide information of more relevance for local decision makers, I advocate for forest type or ecoregion-
specific longitudinal or chronosequence studies combined with high resolution remote sensing, biogeochemical modeling, and climatic analysis along a bioclimatic gradient in western North America. All forest ecosystems in contiguous western U.S. forests lack information on the short-term effect (0-5 years) following bark beetle activity and I, therefore, will attempt to investigate further the effect of the short-term interactions between these two important natural disturbances in all major forest types in the next chapter.

**Literature cited**


Barrows, J.S. 1951. Forest fires in the northern Rocky Mountains. USDA Forest Service, Northern Rocky Mountain Forest and Range Experimental Station, Missoula, Montana, USA.


CHAPTER 3

MODELING THE EFFECTS OF BARK BEETLES, CLIMATE, AND
TOPOGRAPHY ON FIRE OCCURRENCE
IN WESTERN U.S. FORESTS

Abstract

The complex interaction between bark beetles and fire in western U.S. forests is not yet well understood. Even though the conventional wisdom suggests that bark beetle-caused tree mortality predisposes forests for an increased fire risk, the existing scientific evidence to date is either inconclusive in some forest types or non-existent in other forest types. There is, therefore, an urgent demand for more investigation of this critical research question. In order to bridge this knowledge gap, I conducted a study using data from spruce-fir, lodgepole pine, Douglas-fir, and ponderosa pine forests in contiguous western U.S. forests. I employed a retrospective research approach by comparing previously observed fire occurrences for 2001 to 2005 fire years and bark beetle outbreaks for 2000 to 2005. I used the random forest (RF) classification algorithm to evaluate the role of bark beetle-caused tree mortality in influencing fire occurrence in the four forest types relative to other climate and topography-derived predictors. When my models – with an accuracy within the range of well-performing models – are fitted for all fires that occurred in each forest type, climate and topography-derived predictors were overwhelmingly the most selected predictors to influence fire occurrence in all forest types and fire years while bark beetle-caused tree mortality showing no or little

---

2 Coauthored by S. A. Bisrat, M. A. White, B. Bentz
importance to influence fire occurrence varying by fire years. When only areas with bark beetle activity included, the results were largely unchanged. My results generally do not support the hypothesis that bark beetle-caused tree mortality increases the likelihood of fire occurrence in western U.S. forests.

1. Introduction

Fire and insect outbreaks are key components of the natural disturbance regimes that determine the type and stand structure of forested landscapes in the contiguous western United States (U.S.). Fire is often considered to be the most important disturbance agent in western U.S. forests but insect outbreaks, particularly bark beetle (Coleoptera: Curculionidae, Scolytinae) outbreaks that can result in large areas of dead trees, may affect larger areas than fire and are also as ecologically important as fire (Dale et al. 2001; Raffa et al., 2008). While warmer temperatures, earlier snowmelt, and more severe summer droughts have exacerbated fire activity in western U.S. forests (Westerling et al., 2006), the current extent and severity of bark beetle activity in western U.S. forests is also at levels not seen in recent history (Raffa et al., 2008). Climate change is likely to further enhance both bark beetles and fire activity (Hicke et al., 2006; Logan et al., 2003; Tymstra et al., 2007).

Despite our good knowledge of these two natural individual disturbance agents, we know comparatively little about the interactions between bark beetle-caused tree mortality and fire and their changing disturbance regimes. Evidence shows that bark beetle-caused tree mortality changes the composition, amount, and arrangement of living and dead materials in various fuel complexes (Jenkins et al., 2008; Kulakowski et al.,
2003; Page and Jenkins, 2007b). But the effect of this changing fuel complex on the incidence of fire is far from conclusive. Some authors have suggested that bark beetle-caused tree mortality increases the rate of fire incidence in western U.S. forests (Brown, 1975; Furayev et al., 1983; Hopkins, 1909; Knight, 1987; Schmidt, 1988). Theoretical treatment of this subject suggests the critical importance of time since bark beetle outbreak in influencing crown fire behavior (Romme et al., 2006). This theory asserts that crown fire risk may increase immediately following outbreak (in the first one or two years) because of the dead and dry needles remaining on the trees; risk subsequently decreases because of the reduced canopy bulk density (total amount of live and dead fuels in the canopy) once the dead needles make it to the ground. This theory further asserts that the risk of crown fire once again increases after several decades when the beetle-caused tree mortality increases the amounts of standing dead trees (snags). To date, no empirical evidence strongly supports this conceptual model with the exception of the small increase in probability of fire occurrence observed at 2 and 3-year lags in Chapter 2.

The bulk of the studies to date that investigated the effect of bark beetle-caused tree mortality on subsequent fire occurrence used a retrospective research approach where previously observed bark beetle outbreak pattern is compared with expected patterns of fire activity or fire behavior models to look into potential fire behavior. Most of the recent retrospective studies in Engelmann spruce (*Picea engelmannii* Parry)/subalpine fir (*Abies lasiocarpa* Nutt.) forests (hereafter spruce-fir) in Colorado do not support the hypothesis that bark beetle-caused tree mortality increases the incidence of subsequent fire decades after bark beetles killed the trees (Bebi et al., 2003; Bigler et
al., 2005; Kulakowski and Veblen, 2007; Kulakowski et al., 2003). One of these studies in subalpine forests (Bigler et al., 2005) observed a slight increase in the probability of fire incidence; however, pre-burn stand structure and forest type were more important predictors for explaining fire severity than bark beetle-caused tree mortality. Subalpine spruce-fir forests are characterized by infrequent, high severity fire regimes that require extreme and usually less frequent dry conditions for fire to occur (Schoennagel et al., 2004; Sibold and Veblen, 2006). When the extreme dry conditions conducive to initiate high severity fires do occur, the effects of dead fuels as a result of insect outbreaks become minor as there is always adequate fuel load to sustain fire in these high biomass forests (Romme et al., 2006).

Results from similar retrospective studies in lodgepole pine (Pinus contorta Dougl.) forests are also inconclusive because the two studies in this forest type (Lynch et al., 2006; Turner et al., 1999) showed mixed results. Turner et al., (1999) found that the probability of crown fire in lodgepole pine forests in the 1988 Yellowstone National Park was increased in stands with severe bark beetle-caused tree mortality but the likelihood of fire incidence decreased in stands with light or moderate bark beetle-caused tree mortality. Lynch et al., (2006) showed that during the 1988 Yellowstone fires, there was an 11% increase in the probability of fire incidence in lodgepole pine forests that were affected by the 1972 bark beetle-caused tree mortality (16 years following bark beetle activity), but there was no change in the likelihood of fire incidence in forests that were killed during the early 1980s bark beetle outbreaks (approximately eight years following bark beetle activity). Retrospective ground-based studies in Douglas-fir (Pseudotsuga
menziesi Mirb.) and ponderosa pine (Pinus ponderosa Laws.) forests do not exist, resulting in a critical knowledge gap in these important forest types.

Modeling studies have shown a nuanced interaction between bark beetle-caused tree mortality and subsequent fire activity. Two modeling studies to date showed increases in surface fire intensity, fireline intensity, and rate of surface fire spread in the first five years following bark beetles outbreak (0-5 years post-outbreak) in Douglas-fir, lodgepole pine, and spruce forests and increased fireline intensity and higher probability of crowning 5-60 years after bark beetle outbreaks (Jenkins et al., 2008; Page and Jenkins, 2007). The rate of crown fire spread in these studies, however, decreased 5-60 years after outbreaks as a result of the reduced canopy density associated with the defoliating effect of the bark beetles. A more recent modeling study conducted in the central Rocky Mountain spruce forests have concluded that extreme fire behavior is not an inevitable consequence of spruce beetle-caused tree mortality (Derose and Long, 2009).

To date, most studies that used a retrospective approach used a time lag within the range of 5 to 50 years since bark beetle outbreak; only one study (Kulakowski and Veblen, 2007) has used a 5-year time lag. Intercomparisons of retrospective studies are also hampered by the different research methodologies used to investigate the interactions between bark beetle-caused tree mortality and fire. To the best of my knowledge, no retrospective study has systematically explored lagged effect at the sub-five year time scale across multiple forest types.

Climatic variability has similarly variable effects on fire occurrence. Forest fire occurrence in the western U.S. is fundamentally controlled by climate and subsequent
effects on moisture availability and fuel availability and flammability, but these effects are spatially and temporally discontinuous (Westerling, 2008). Fuel flammability effects—controlled by interannual climatic variations—are more pronounced in moist, densely vegetated forest types such as spruce-fir, lodgepole pine, and Douglas-fir forests; on the other hand, fuel availability effects—controlled by long-term above-ground biomass accumulation—are more evident in arid, sparsely vegetated forest types such as ponderosa pine (Westerling and Bryant, 2008). Because fire regimes, stand structure, and the regions occupied by each forest type vary along latitudinal and elevational gradients (mainly caused by variations in fuel flammability versus fuel availability) and because the interactions between bark beetle-caused tree mortality and subsequent fire behavior seems to vary by forest type and to demonstrate important but potentially variable time lags, I hypothesized that there will be differences in the effect of bark beetle-caused tree mortality on occurrence of fire across major forest types.

To explore this bark beetle-fire disturbance complex, I used a consistent retrospective approach and tested the relative importance of 0- to 5-year lagged bark beetle outbreaks, topographic conditions, and interannual climate variability on fire activity across lodgepole pine, spruce-fir, Douglas-fir, and ponderosa pine forest ecosystems in the contiguous western U.S. I sought to answer the following two questions. First, assessed for all areas experiencing fire, did bark beetle-caused tree mortality play an important role in influencing the occurrence of fire relative to other climate and topography derived predictor variables? Second, when restricted to those areas experiencing bark beetle outbreaks, did the impact of bark beetle-caused tree mortality on fire occurrence change?
2. Methods

2.1. Study area

I modeled fire occurrence for each fire year from 2001-2005. The beginning of my study period is dictated by the lack of good satellite-derived fire data for 2000 (Giglio, L. 2005). I chose 2005 as the end of the study period mainly because theoretical treatment of this ecological problem suggests that the first few years after bark beetle outbreak may experience the highest increase in fire risk (Romme et al., 2006).

The study area includes four major forest types in the western U.S. The spruce-fir type (Fig. 8a) extends from 2500 m to more than 3000 m in elevation and is characterized by low frequency and high severity crown fire regimes with fire return interval greater than 100 years (Veblen et al., 1994). Spruce beetle (*D. rufipennis* Kirby) is the main cause of the death in spruce fire forests.

Lodgepole pine forest ecosystems (Fig. 8b) range from 2500 m to 3000 m in elevation and exhibit both high frequency, low intensity surface fires and low frequency, high intensity crown fire regimes with fire return interval ranging from 20 years in lower elevation forests up to 200 years in higher elevation forests (Arno, 1980; Lotan et al., 1985). Trees in lodgepole pine forests are killed mainly by mountain pine beetles (*Dendroctonus. ponderosae* Hopkins).

Douglas-fir forest ecosystems (Fig. 8c) occur at mid elevations between the ponderosa pine and mixed conifer forests and exhibits both high-frequency, low intensity surface fires and low frequency, high intensity crown fire regimes with fire return intervals ranging between 15 years at lower elevation forests and 60 years at higher
elevation forests (Arno, 1976; Wright and Bailey, 1982). Douglas-fir forests are killed mainly by Douglas-fir beetles (*D. pseudotsugae* Hopkins).

![Vegetation maps](image)

**Fig. 8.** The vegetation maps for spruce-fir (a), lodgepole pine (b), Douglas-fir (c), and ponderosa pine (d) forests used in this study are shown in green color (Source: USFS Remote Sensing Application Center - [http://fsgeodata.fs.fed.us/rastergateway/forest_type](http://fsgeodata.fs.fed.us/rastergateway/forest_type)). The red color shows the 2001-2005 fire pixels detected by MODIS in each forest type. The black solid boundary line shown around the states of Arizona and New Mexico in ponderosa pine forest map (d) shows the study area for SW ponderosa pine forests.
Ponderosa pine forest ecosystems (Fig. 8d) occur in drier areas with an elevation range between sea level and 2800 m. Fire in ponderosa pine forests is characterized by high-frequency and low-intensity fire regimes with return intervals within the range of 5 to 30 years (Arno, 2000). Ponderosa pine forests are killed mainly by mountain pine beetles (*Dendroctonus. ponderosae* Hopkins). The fire season in western U.S. forests generally shows a strong seasonality occurring between May and October (Westerling et al., 2003), but in southwestern (SW) ponderosa pine forests the fire season is concurrent with dry spring and early summer conditions bounded by winter snowpack and July and August monsoonal rains (Arno, 2000, Westerling et al., 2003). I therefore analyzed bark beetle-fire interactions in SW ponderosa pine forests of Arizona and New Mexico separately (thick black solid line in Fig. 8d).

2.2. Methods of Analysis

I used a machine learning classification algorithm to predict the occurrence of fire. I chose this approach over traditional statistical modeling methods such as multiple linear regression because machine learning methods are fully nonparametric and thus are not subject to distributional assumptions for the data and do not require *a priori* specification of a data model (Guisan and Zimmermann, 2000). When the nature of the natural process in question is complex or largely unknown, as in the case of modeling fire behavior, machine learning approaches are highly suitable (Parisien and Moritz, 2009). In this study, I used the RF classification algorithm (Breiman, 2001) because – in addition to the usual advantages of machine learning techniques – it exhibits high classification accuracy in a wide range of ecological applications (Cutler et al., 2007; Lawler. et al.,
does not over-fit models, generates a reliable internal estimate of classification accuracy; and – critically for this study – produces a useful assessment of the relative importance of explanatory variables.

2.2.1. Random Forests Algorithm

A classification tree – upon which RF is conceptually based – is a decision tree fit to data by recursive partitioning of the data into increasingly homogeneous subgroups with respect to the classification variable (Breiman, et al. 1984). In contrast, RF fits many classification trees to the data and combines the predictions of the individual trees to create a more accurate prediction. The RF algorithm begins with the random selection of many (the default is 500) bootstrap samples from the original data. In each bootstrap sample approximately 63% of the original data is selected (with replacement) for tree development. The observations in the original dataset that do not occur in the bootstrap sample are said to be out-of-bag for that bootstrap sample. To each bootstrap sample a classification tree is fit with the restriction that at each partitioning point only a small number of variables (the default is the square root of the total number of variables) is available for splitting the data. Each fitted tree is then used to predict the withheld out-of-bag observations (in this case, fire presence or absence). A final prediction for each observation is then generated by majority vote from the individual out-of-bag predictions, with ties in the voting decided randomly. Separation of observations used in model fitting from those used in model testing – combined with several hundred trees – helps to minimize over-fitting and to ensure an unbiased estimator of the true error rate (Breiman et al., 1984). Use of a small number of variables at each partitioning point in each tree
ensures that trees fit to different bootstrap samples are minimally correlated, which tends to produce classification accuracies that are higher than those developed from a single classification tree (Breiman, 1996).

One feature that sets apart RF from other “black box” machine learning classifiers, such as artificial neural network and support vector machine, is the permutation algorithm used to identify the variables that are most important to the classification. This variable importance measure in RF has proven useful for topics such as forest succession (Falkowski et al., 2009), wetland vegetation distribution (Peters, J. et al. 2008), microarray gene expression (Archer and Kirnes, 2008), and road safety (Harb et al., 2009). In situations where there are several hundreds of predictor variables – as in microarray gene expression studies – variable importance measures in RF can also be used to reduce the number of predictor variables and produce a more parsimonious model without compromising prediction accuracy (Jiang et al., 2004).

The RF permutation variable importance algorithm is as follows. RF starts by considering a single tree and a single explanatory variable of interest. Each out-of-bag observation for that tree has a predicted class. Next, the values of the variable of interest are permuted for the out-of-bag observations and new predicted classes obtained. If the variable is not important to the classification then there should be little difference between the two sets of predicted classes, but if the variable is important, one would expect the original predictions to be more accurate than those obtained by using the permuted variable values. For each observation, the difference in accuracies using raw and permuted values of the variable of interest are computed over all bootstrap samples and classification trees for which the observation is out-of-bag, and then aggregated over
all observations and divided by standard error. The resulting z-scores may be tabulated or plotted (Liam and Wiener, 2002). I used this RF permutation approach to identify which variables are important in predicting fire occurrence. I implemented the RF algorithm in R computing environment using ‘randomForest’ package (Liam and Wiener, 2002).

2.3. Data
2.3.1. Forest type data

I used forest type data sets generated by the United States Forest Service (USFS) Forest Inventory and Analysis Program and the USFS Remote Sensing Application Center (http://fsgeodata.fs.fed.us/rastergateway/forest_type) to define my analysis region by selecting spruce-fir, lodgepole pine, Douglas-fir, and ponderosa pine forest types (Fig. 8). This dataset shows 141 forest types at 250 m spatial resolution for the contiguous U.S. The forest types are defined as logical ecological groupings of species mixes (Eyre, 1980). The classification was generated in 2008 by combining FIA plot data, Moderate Resolution Imaging Spectroradiometer (MODIS) images from the 2002 and 2003 growing seasons, MODIS vegetation indices [Enhanced Vegetation Index and Normalized Difference Vegetation Index (Huete, A. et al., 2002)], MODIS Vegetation Continuous Fields (Hansen et al., 2002) in combination with nearly 100 other geospatial data layers that included Parameter-elevation Regressions on Independent Slopes Model climate data (Daly, C. et al. 2002), National Land Cover Database layers (Vogelmann et al., 2001), a Digital Elevation Model (DEM), and ecological regions layers (Bailey, 1994).
2.3.2. *Fire data as response variable*

Most fire modeling studies to date have used fire data compiled by federal agencies including the USFS, Bureau of Land Management, and National Park Service. It is common to observe problems associated with underreporting and inaccuracy of fire locations in such multi-agency fire data sets; as a result some fire data reported by the different agencies underestimate the actual fire occurrence (Brown et al., 2002; Crimmins and Comrie, 2004; Parisien and Moritz, 2009; Westerling et al., 2003). Fire atlases (Gibson, 2006), also known as digital polygon fire histories, are also being used but may not include small fires and may also have some inaccuracies in fire locations (Morgan et al., 2008). Satellite-detected fire data, on the other hand, could be a more reliable source of fire occurrence data than the data acquired from traditional fire monitoring programs (Csiszar et al., 2005; Korontzi, et al., 2006). Because the thermal emissive power from fires is orders of magnitude higher than that from the surrounding non-burning land surface, even fires as small as 1 ha are detectable with a 1 km pixel (Robinson, 1991). In this study, I used fire occurrence data estimated from MODIS (Moderate Resolution Spectroradiometer) data. MODIS fire data is well validated using simulation models, independently collected high resolution satellite data, USFS *in situ* fire observations, and other regional validation sites (Giglio et al., 2003; Morisette et al., 2005). The detection rate of fires by MODIS active fire products in U.S. was high (82%) with the detection rate being greater in the western U.S. than in the eastern U.S. (Hawbaker et al., 2008).

I obtained the 2001 to 2005 1 km MODIS Terra 8-day composite thermal anomalies/fire product (MOD14A2, Collection 4) from the Land Processes Distributed Active Archive Center (LP-DAAC) Data Gateway. Launched in 1999 on the Terra
platform, MODIS began capturing images as of February 2000 but the 2000 fire season data are considered to be unreliable (Giglio, 2005). MOD14A2 (Justice et al., 2002) is based on MODIS brightness temperatures derived in the 4µm ($T_4$) and 11µm channels ($T_{11}$). When possible, $T_4$ is derived from the MODIS band with saturation at 311K, which has low noise and small quantization error, rather than the band saturating at 500K. $T_{11}$ is computed from the 11µm channel, which saturates at nearly 400 K (Justice et al., 2002).

The thermal anomaly/fire product is then processed as a daily Level 2 product (MOD14) describing active fires and thermal anomalies. I used the Level 3 MOD14A2 product, which represents the 8-day maximum value of the individual Level 2 values (Masuoka et al., 1998). Each MOD14A2 1 km fire is assigned one of three detection confidence levels: low-confidence fire, nominal-confidence fire, or high-confidence fire (Justice et al., 2002). In Chapter 2, I included the low-confidence fire as they comprise only small fraction (1.3 %) of the total fires detected by MODIS. Since this study is conducted at a higher spatial detail (related to spatial extent as this study is conducted by species type as opposed to the entire contiguous western U.S. forests in Chapter 2), I chose to accept only the nominal and high-confidence fire as valid fire pixels even though the influence of the low-confidence fires is minor.

I implemented a three-step processing scheme for the MODIS fire data. First, for each 8-day composite, I mosaicked the individual MODIS 1200 km × 1200 km tiles, reprojected and co-registered them to the Albers Equal Area projection (ERDAS Imagine version 9.0, Leica Geosystems). Second, for each 8-day composite period, I assigned a 0 to pixels without fires and a 1 to pixels with valid fires (i.e. excluding pixels with low-confidence fires). Third, I summed all 46 8-day composite period images within
individual years, producing an annual fire value for each 1 km pixel. I considered a pixel to be a fire pixel if, in any given calendar year, fire was present in at least in one 8-day composite period.

2.3.3 Predictor variables

Fire occurs on a given landscape as a function of climate, topography and fuels (Pyne et al., 1996). As a result, any research effort to model fire needs to involve data inputs from climate, topography, and vegetation (Anderson, 1982). In addition to the bark beetle layers, I used precipitation, temperature, dew point, number of snow days, Palmer Drought Severity Index (PDSI), topography-derived layers (elevation, slope, aspect), and net primary production (NPP), as predictors (described in more detail below, see also Table 3 for description and Fig. 13 (the Appendix) for maps).

<table>
<thead>
<tr>
<th>Predictor variables</th>
<th>Data source</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bark beetles</td>
<td>USFS</td>
<td>Annual bark beetle-caused tree mortality in percent at 1 km spatial resolution</td>
</tr>
<tr>
<td>Precip</td>
<td>PRISM</td>
<td>Annual total precipitation in mm at 4 km spatial resolution</td>
</tr>
<tr>
<td>FSTmax</td>
<td>PRISM</td>
<td>Fire season maximum temperature in °C at 4 km spatial resolution</td>
</tr>
<tr>
<td>FSDewpt</td>
<td>PRISM</td>
<td>Fire season dew point in °C at 4 km spatial resolution</td>
</tr>
<tr>
<td>SpSuTemp</td>
<td>PRISM</td>
<td>Annual average spring and summer temperature in °C at 4 km spatial resolution</td>
</tr>
<tr>
<td>SnowDays</td>
<td>MODIS</td>
<td>Annual number of snow days at 1 km spatial resolution</td>
</tr>
<tr>
<td>PDSI</td>
<td>NCDC</td>
<td>Annual Palmer Drought Severity Index (-6 to +6) at 2.5 degree spatial resolution</td>
</tr>
<tr>
<td>NPP</td>
<td>MODIS</td>
<td>Annual Net Primary Production in kg C/m² at 1 km spatial resolution</td>
</tr>
<tr>
<td>Elevation</td>
<td>USGS</td>
<td>Elevation in m at 1 km spatial resolution</td>
</tr>
<tr>
<td>Slope</td>
<td>USGS</td>
<td>Slope in percent at 1 km spatial resolution</td>
</tr>
<tr>
<td>Aspect</td>
<td>USGS</td>
<td>Transformed aspect values ranging from 0 to 1 at 1 km spatial resolution</td>
</tr>
</tbody>
</table>
2.3.3.1 Bark beetles data

I obtained the 2000 to 2005 USFS-Forest Health Protection aerial detection estimates of bark beetle-caused tree mortality. The bark beetle aerial detection survey (ADS) is collected using low-level flights, typically 300 to 600 m above ground level, and manual mapping of tree mortality by the main species of tree killing bark beetles in all western U.S. forests types. The bark beetle ADS data include numerous fields, such as polygon boundaries for groups of beetle-killed trees, polygon area, number of trees killed, trees killed per area, host species, and forest type, but not all fields are available for all polygons. Consequently, I only used the geographical polygon boundaries of bark beetle killed trees and related metadata describing bark beetles species.

I extracted the bark beetle ADS polygons for the four forest types by their respective bark beetles codes as documented on the ADS metadata. I used the beetle codes for spruce beetle (*D. rufipennis* Kirby), western balsam bark beetle (*Dryocoetes confuses* Swaine), and true-fir bark beetles (*Scolytus spp.* Geoffrey) to extract the ADS polygons in spruce-fir forests. In lodgepole pine forests, I used the beetle codes for mountain pine beetle (*D. ponderosae* Hopkins), lodgepole pine beetle (*D. murrayanae* Hopkins), pine engraver (*Ips pini* Say), and ips engraver (*I. spp.* De Geer) to extract the ADS polygons. I used the codes for Douglas-fir beetle (*D. pseudotsugae* Hopkins) and Douglas-fir pole beetle (*Pseudohylesinus nebulosus* LeConte) to extract polygons from Douglas-fir forests. For ponderosa pine forests, I used beetle codes for mountain pine beetle (*D. ponderosae* Hopkins), western pine beetle (*D. brevicomis* LeConte), roundhead pine beetle (*D. adjunctus* Blandford), pine engraver (*I. pini* Say), and ips engraver (*I. spp.* De Geer).
I then converted the ADS polygons in each forest type to a 100 m resolution raster layer (ArcGIS version 9.1; Environment Systems Research, Inc.) such that each pixel contained a 0 (bark beetle-caused tree mortality absent) or 1 (bark beetle-caused tree mortality present). I then created a 1 km ADS maps from the 100 m maps to indicate the level of bark beetle activity by summing the 100 m pixels within each 1 km pixel. For instance, if a 1 km pixel contains only one 100 m pixel with beetle activity that 1 km pixel is assigned 1% value for beetle activity. Similarly, if 1 km pixel contains 50 100 m pixels with beetle activity that 1 km pixel is assigned a value of 50% for beetle activity.

2.3.3.2 Climate data

Fire occurrence in western U.S. forests is highly influenced by climate, particularly precipitation and temperature, through its effect on fuel availability and fuel flammability (Westerling, 2008). Fire season maximum temperature and fire season relative humidity indicate the suitability of conditions for fire ignition as they influence moisture available in fuels (Swetnam and Betancourt, 1998; Veblen et al., 2000; Westerling et al., 2003). Fire season is defined as May through October, when 94% of fire and 98% of burned area creation in western U.S. forests occurs (Westerling et al., 2003). The current and preceding year annual precipitation indicate the effect of moisture available for fuel production (Swetnam and Betancourt, 1998). Spring and summer average temperature indicates the timing of spring, the length of the dry season and consequently the length and severity of the fire season (Westerling et al., 2006). The variability in spring and summer temperature through its effect on the duration and severity of summer drought is an important driver of interannual variability in forest
wildfire in the western U.S. (Morgan et al., 2008; Westerling, 2008; Westerling et al., 2006). Warmer temperatures can decrease moisture available for fuel production through increased evapotranspiration, reduced snowpack, and earlier snowmelt (Westerling, 2008). Reduced snowpack and earlier snowmelt in their turn results in longer and drier summer fire season in many high elevation areas (Westerling et al., 2006). Therefore, the number of days with snow in any given year is an important indicator of the length of the dry season and the severity of the fire season (Westerling and Bryant, 2008).

PDSI represents cumulative precipitation and temperature anomalies by combining precipitation, evapotranspiration, and soil moisture as a measure of the severity and spatial extent of drought with values ranging from -6 (extreme drought) to +6 (extreme It) (Alley, 1985). The current year PDSI indicates the moisture content of fuel, consequently indicating the conditions for fire ignition. PDSI for the preceding year indicates the availability of moisture for fuel production. Prior work has shown links between antecedent moisture available for fuel production, defined by PDSI, and wildfire activity in western U.S., albeit to a varying degree regionally and temporally (Westerling et al., 2003). Burned area in the interior western U.S. shows significant positive relationships with preceding summer PDSI averages (one year lag) and significant negative relationships with current year summer PDSI averages (Collins et al., 2006). I used summer (June, July, August) average instrumental PDSI layer at 2.5° resolution produced for conterminous U.S. (Cook et al., 2004). I used summer average PDSI mainly because it captures the moisture available during the fire season in western U.S. forests (Baisan and Swetnam, 1990). I obtained 2001-2005 summer PDSI point data for the contiguous western U.S. (86 grid points) from National Climate Data Center.
I then created an annual PDSI raster layer from the point data. In addition to climate factors, local-scale variation in topography through its effect on microclimate conditions can influence the occurrence of fire by changing the moisture content of the fuel and the availability of continuous fuel for fire to spread across the landscape (Taylor and Skinner, 1998).

I used 2001-2005 annual and monthly mean maximum temperature and dew point and annual and monthly total precipitation from the PRISM dataset (Daly et al., 2002). Dew point is a measurement of absolute humidity and is defined as the temperature to which air must be cooled in order for condensation to occur. I calculated fire season precipitation, fire season maximum temperature and fire season dew point from the monthly values for each fire year. I calculated spring and summer average temperature from the monthly values of the PRISM maximum and minimum temperature data.

The number of snow days was calculated from MODIS daily snow cover data obtained from the National Snow and Ice Data Center Distributed Data Archive (http://nsidc.org/daac). The MODIS/Terra daily snow cover product (MOD10A1, Collection 5) at 500 m resolution (Hall and Riggs, 2007) produced for the data period from 1 January 2001 to 31 December 2005 was used. MOD10A1 is the most reliable currently available snow cover dataset (Hall et al., 2002) and provides classifications of snow presence, snow absence, and cloud cover. It also provides a class that represents missing or erroneous data. An 8-day compositing method was used to generate daily snow-cover dynamics based on the MOD10A1 product (Q. Tang and D. P. Lettenmaier, in preparation). When the product provides classification of cloud cover or missing data, the closest available day within the 8-day composite was used to determine whether snow
is present or absent. This way, the compositing method effectively minimized cloud contamination and provided a provisional estimate of daily snow cover dynamics.

2.3.3.3 *Topographic data*

I obtained topographic variables from the GTOPO30 global DEM (http://edc.usgs.gov/products/elevation/gtopo30/gtopo30.html) produced at approximately 1 km resolution (30 arc second). I transformed the values of aspect from its circular scale into values that range from 0 to 1 using a trigonometric transformation:

\[
\text{Transformed Aspect} = \frac{1 - \cos(2\pi(\text{Aspect} - 30\degree)/360)}{2}
\]

(Roberts and Cooper 1989).

2.3.3.4 *Vegetation data*

To characterize biomass production status in each pixel I used MODIS-derived primary production product (MOD17). MOD17 is produced as the accumulated carbon fixed by vegetation through photosynthesis at 1 km resolution and at 8-day interval. Since the Collection 4 MOD17 had some errors associated with cloud and aerosol contaminations, I used the improved Collection 5 MOD17 product to obtain reliable estimates of accumulated NPP (Zhao et al., 2005). I obtained the improved annual NPP product from 2001-2005 provided by Numerical Terradynamic Simulation Group at the University of Montana (http://ntsg.umt.edu).

2.4. *Model fitting*

I reprojected the fire, forest type, and predictor layers to the Albers Equal Area projection (ArcGIS 9.1, Environmental Systems Research Institute, Inc.). I then extracted the MODIS fire pixels for each forest type and each fire year (Fig. 8). RF classification
algorithm requires both fire presence and absence points. When absence points are not available, it is possible to randomly generate absence points from areas where fire is actually absent (Parisien and Moritz, 2009). Subsequently, I masked all fire pixels and randomly generated equal number of non-fire pixels for each forest type and each fire year to represent fire absence points, commonly known as pseudoabsences. The fire pixels are assigned a value of 1 to denote fire presence and the non-fire pixels are assigned a value of 0 to denote fire absence. The fire and non-fire pixels are then converted into points at the centroid of the MODIS pixel to represent fire presence and absence locations. The fire presence-absence data set was then used as binary response variable and the respective values for each predictor that corresponds with all the fire presence-absence points were extracted.

For each fire year, I used the respective predictor layers for that particular fire year. In addition, I used the preceding year PDSI and annual precipitation in each fire year to investigate the effect of antecedent moisture on fire occurrence. With respect to bark beetles, I used bark beetle-caused tree mortality layer from the preceding year(s) as well as the current year since there is evidence that suggests bark beetle-caused tree mortality and fire interactions show time-lag dependencies as fuel dynamics and stand structure change over time.

I first fitted models for all areas that experienced fire in each forest type. To investigate further the influence of bark beetles-caused tree mortality on fire occurrence, I fitted models only to those areas that experienced bark beetle outbreaks. I then compared the results obtained from this two-stage modeling approach.
2.5 Model accuracy assessment

I used 10-fold cross-validation to estimate the accuracy of prediction. In most cases, the out-of-bag estimates automatically produced by RF are nearly identical to 10-fold cross-validated estimates. I assessed classification accuracies using percentage of correctly classified (PCC), commission error (false positive), omission error (false negative), kappa, and area under the Receiver Operating Characteristic curve commonly known as AUC (each described in more detail below). All these five measures of model accuracy have values that range from zero (very poor accuracy) to one (a perfect agreement between the observed and the predicted).

PCC, commission error, and omission error are calculated from an error matrix commonly known as the confusion matrix (Fielding and Bell, 1997). PCC is calculated by dividing the number of correctly classified presences and absences by the total number of points in the error matrix. Commission errors (defined as incorrect prediction of presences) are calculated from sensitivity (as $1 - \text{sensitivity}$) where sensitivity measures the percentages of presences correctly predicted. Omission error (defined as incorrect prediction of absences), on the other hand, are calculated from specificity (as $1 - \text{specificity}$) where specificity measures the percentages of absences correctly predicted. PCC, sensitivity, and specificity use an arbitrary threshold (commonly 0.5) to convert the probabilistic estimates of fire presence into dichotomous presence-absence predictions making them threshold-dependent accuracy measures (Fielding and Bell, 1997).

Kappa, like PCC, sensitivity, and specificity, is also calculated from the confusion matrix (Fielding and Bell, 1997). Kappa measures the agreement between predicted presences and absences with the actual observed presences and absences corrected for
agreement that might be due to chance alone. Kappa is essentially a measure of randomness with more accurate models having a value close to 1 and a random predictor having a value of 0 (Cohen, 1960). Kappa also uses arbitrary threshold to convert the probability of species occurrence into presence or absence (Fielding and Bell, 1997).

AUC evaluates the model’s ability to discriminate between sites where the fire is present or absent by plotting the sensitivity (true positives) against 1-specificity (false positives) (Fielding and Bell, 1997). An AUC value of 1 indicates perfect discrimination between presence and absence sites and a value of 0.5 indicates no discrimination ability. AUC evaluates the performance of a model across all possible threshold values to convert the probabilistic estimates of fire occurrence into presence/absence making it one of the most reliable threshold-independent measure of accuracy (Broennimann et al., 2007). Model predictions with AUC values between 0.7 and 0.9 indicate moderate or useful performance; whereas AUC values greater than 0.9 are considered excellent predictions (Pearce and Ferrier, 2000; Swets, 1988). Models with AUC values less than 0.5 are considered not better than random models with omission and commission error rates being higher than correct predictions (Broennimann et al., 2007). I used the ‘PresenceAbsence’ package (Freeman, 2007) in R to calculate all the accuracy metric.

3.0. Results

Bark beetle-caused tree mortality was generally not identified as being an important influence on fire occurrence in the forest ecosystems analyzed (Fig. 9). For models fitted for all areas that experienced fire in all the forest types and fire years, bark-
beetle caused mortality was selected as one of the top three predictors only in the 2002 fire year in ponderosa pine forests (Fig. 9D).

Climate and topography (mainly elevation) derived predictors were consistently the most selected predictors to influence fire occurrence in all forest types across the five fire years. Mean variable importance measure computed for each forest type across the five fire years showed that elevation, PDSI for both current and preceding years, fire season dew point, annual precipitation for current year, spring and summer average temperature, the number of snow days, aspect, and NPP, were the predictors most selected to influence fire occurrence (Fig. 10A-E).

I was not able to get sufficient number of fire pixels in some of the forest types in 2001 and 2002 fire years to fit models for areas that had bark beetle activity. For instance, I had no fire presence points in 2001 fire year and only five fire presence points were available in 2002 fire year in spruce-fir forests. Also, there were only 8 fire presence points in lodgepole pine in 2001 fire year. As a result, I fitted models for areas that experienced bark beetles outbreak only for 2003, 2004 and 2005 fire years to have consistent number of fire years modeled in each forest type.
Fig. 9. Variable importance plots for spruce-fir (A), lodgepole pine (B), Douglas-fir (C), ponderosa pine (D), and SW ponderosa pine (E) forests for models fitted for all areas that experienced fire (2001-2005). A variable with the highest mean decrease in accuracy is the most important variable.
Fig. 9 (cont.)

Fig. 10. Mean variable importance plots for models fitted for all areas that experienced fire in all forest types. A variable with the highest mean decrease in accuracy is the most important variable.
My results show that when models were fitted only for those areas with recorded bark beetle activity, the importance of bark beetle-caused tree mortality to fire occurrence was somewhat higher in some forest types and fire years compared to the models fitted for all areas that experienced fire (Fig. 11). For instance, in spruce-fir forests 2004 bark beetle-caused tree mortality was selected as the third most important variable in the 2005 fire year (1-year time lag) (Fig. 11A). Similarly, the 2003 bark beetle-caused tree mortality was the 4th important predictor in lodgepole pine forests in 2004 fire year and the 2001 and 2002 bark beetle-caused tree mortality were the 3rd and 4th important predictors in Douglas-fir forests in 2003 fire year (Fig. 11B and 11C). Also, the 2002 bark beetle-caused tree mortality was the 2nd important predictor in SW ponderosa pine forests in 2003 fire year (Fig. 11E).

In spite of the relatively higher importance exhibited by bark beetle-caused tree mortality in influencing fire occurrence in some forest types and fire years, when the variable importance measure is averaged across the three fire years, bark beetle-caused tree mortality was not consistently selected as important predictors to influence fire occurrence in all forest types (Fig. 12A-E). Similar to the results I obtained when I fitted models for all areas that experienced fire, climate and topography-derived predictors were consistently the most important predictors to influence fire occurrence across the five fire years.
Fig. 11. Variable importance plots for spruce-fir (A), lodgepole pine (B), Douglas-fir (C), ponderosa pine (D), and SW ponderosa pine (E) forests for models only fitted for areas with bark beetles activity (2001-2005). A variable with the highest mean decrease in accuracy is the most important variable.
Fig. 11 (cont.)

Fig. 12. Mean variable importance plots for models only fitted for areas with bark beetle activity in all forest types. A variable with the highest mean decrease in accuracy is the most important variable.
Table 4

Accuracy metric values for models built for all areas that experienced fire in each forest types.

<table>
<thead>
<tr>
<th>Forest types</th>
<th>Year</th>
<th>No. of observations</th>
<th>PCC (%)</th>
<th>Commission error (%)</th>
<th>Omission error (%)</th>
<th>Kappa</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Spruce/fir:</td>
<td>2001</td>
<td>718</td>
<td>0.73</td>
<td>0.26</td>
<td>0.22</td>
<td>0.38</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>2002</td>
<td>760</td>
<td>0.77</td>
<td>0.27</td>
<td>0.19</td>
<td>0.53</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>2003</td>
<td>2166</td>
<td>0.70</td>
<td>0.24</td>
<td>0.18</td>
<td>0.47</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>2004</td>
<td>182</td>
<td>0.60</td>
<td>0.36</td>
<td>0.42</td>
<td>0.21</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>2005</td>
<td>546</td>
<td>0.73</td>
<td>0.15</td>
<td>0.33</td>
<td>0.46</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td></td>
<td>0.71</td>
<td>0.26</td>
<td>0.27</td>
<td>0.41</td>
<td>0.76</td>
</tr>
<tr>
<td>(b) Lodgepole pines:</td>
<td>2001</td>
<td>766</td>
<td>0.72</td>
<td>0.30</td>
<td>0.24</td>
<td>0.44</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>2002</td>
<td>952</td>
<td>0.77</td>
<td>0.25</td>
<td>0.20</td>
<td>0.54</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>2003</td>
<td>2400</td>
<td>0.80</td>
<td>0.25</td>
<td>0.19</td>
<td>0.53</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td>2004</td>
<td>364</td>
<td>0.67</td>
<td>0.26</td>
<td>0.37</td>
<td>0.34</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>2005</td>
<td>758</td>
<td>0.75</td>
<td>0.24</td>
<td>0.17</td>
<td>0.43</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td></td>
<td>0.74</td>
<td>0.26</td>
<td>0.23</td>
<td>0.46</td>
<td>0.78</td>
</tr>
<tr>
<td>(c) Douglas-fir:</td>
<td>2001</td>
<td>3344</td>
<td>0.61</td>
<td>0.36</td>
<td>0.41</td>
<td>0.22</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>2002</td>
<td>7478</td>
<td>0.77</td>
<td>0.27</td>
<td>0.18</td>
<td>0.54</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>2003</td>
<td>4982</td>
<td>0.63</td>
<td>0.36</td>
<td>0.38</td>
<td>0.26</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>2004</td>
<td>3178</td>
<td>0.61</td>
<td>0.35</td>
<td>0.41</td>
<td>0.23</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>2005</td>
<td>3312</td>
<td>0.63</td>
<td>0.35</td>
<td>0.38</td>
<td>0.27</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td></td>
<td>0.65</td>
<td>0.34</td>
<td>0.35</td>
<td>0.30</td>
<td>0.72</td>
</tr>
<tr>
<td>(d) Ponderosa pines:</td>
<td>2001</td>
<td>1724</td>
<td>0.62</td>
<td>0.36</td>
<td>0.40</td>
<td>0.24</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>2002</td>
<td>4372</td>
<td>0.68</td>
<td>0.33</td>
<td>0.31</td>
<td>0.36</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>2003</td>
<td>2450</td>
<td>0.70</td>
<td>0.31</td>
<td>0.28</td>
<td>0.41</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>2004</td>
<td>1550</td>
<td>0.72</td>
<td>0.29</td>
<td>0.27</td>
<td>0.43</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>2005</td>
<td>1558</td>
<td>0.71</td>
<td>0.27</td>
<td>0.31</td>
<td>0.42</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td></td>
<td>0.69</td>
<td>0.31</td>
<td>0.31</td>
<td>0.37</td>
<td>0.73</td>
</tr>
<tr>
<td>(e) SW ponderosa pines:</td>
<td>2001</td>
<td>416</td>
<td>0.58</td>
<td>0.39</td>
<td>0.44</td>
<td>0.25</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>2002</td>
<td>1610</td>
<td>0.74</td>
<td>0.20</td>
<td>0.30</td>
<td>0.49</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>2003</td>
<td>1358</td>
<td>0.68</td>
<td>0.12</td>
<td>0.39</td>
<td>0.35</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>2004</td>
<td>564</td>
<td>0.71</td>
<td>0.15</td>
<td>0.35</td>
<td>0.43</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>2005</td>
<td>520</td>
<td>0.71</td>
<td>0.11</td>
<td>0.35</td>
<td>0.40</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td></td>
<td>0.68</td>
<td>0.19</td>
<td>0.37</td>
<td>0.38</td>
<td>0.71</td>
</tr>
</tbody>
</table>

*Number of observations is twice the number of fire pixels because they include equal number of both fire presence and absence points.
For models fitted for all areas that experienced fire, mean PCC and AUC values across the five fire years were between 0.65 and 0.78 for all forest types; while mean commission and omission errors ranged between 0.19 and 0.36 and mean kappa ranged between 0.30 and 0.46 (Table 4). Models that were fitted only for areas that experienced bark beetles outbreaks showed mean PCC and AUC values across the five fire years between 0.59 and 0.74 for all forest types; while mean commission and omission errors ranged between 0.15 and 0.45 and mean kappa ranged between 0.20 and 0.39 (Table 5).

### Table 5

Accuracy metric values for models built only for areas with bark beetle activity and fire occurrence.

<table>
<thead>
<tr>
<th>Forest types</th>
<th>Year</th>
<th>No. of observations</th>
<th>PCC (%)</th>
<th>Commission error (%)</th>
<th>Omission error (%)</th>
<th>Kappa</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Spruce/fir:</td>
<td>2003</td>
<td>1038</td>
<td>0.72</td>
<td>0.24</td>
<td>0.35</td>
<td>0.30</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>2004</td>
<td>70</td>
<td>0.60</td>
<td>0.28</td>
<td>0.30</td>
<td>0.23</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>2005</td>
<td>265</td>
<td>0.64</td>
<td>0.22</td>
<td>0.40</td>
<td>0.25</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td></td>
<td>0.65</td>
<td>0.25</td>
<td>0.35</td>
<td>0.26</td>
<td>0.62</td>
</tr>
<tr>
<td>(b) Lodgepole pines:</td>
<td>2003</td>
<td>1147</td>
<td>0.77</td>
<td>0.22</td>
<td>0.31</td>
<td>0.23</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>2004</td>
<td>107</td>
<td>0.66</td>
<td>0.33</td>
<td>0.35</td>
<td>0.29</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>2005</td>
<td>304</td>
<td>0.80</td>
<td>0.34</td>
<td>0.20</td>
<td>0.33</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td></td>
<td>0.74</td>
<td>0.30</td>
<td>0.29</td>
<td>0.28</td>
<td>0.59</td>
</tr>
<tr>
<td>(c) Douglas-fir:</td>
<td>2003</td>
<td>931</td>
<td>0.52</td>
<td>0.48</td>
<td>0.46</td>
<td>0.15</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>2004</td>
<td>1047</td>
<td>0.53</td>
<td>0.46</td>
<td>0.41</td>
<td>0.18</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>2005</td>
<td>1116</td>
<td>0.57</td>
<td>0.40</td>
<td>0.45</td>
<td>0.20</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td></td>
<td>0.54</td>
<td>0.45</td>
<td>0.44</td>
<td>0.18</td>
<td>0.55</td>
</tr>
<tr>
<td>(d) Ponderosa pines:</td>
<td>2003</td>
<td>1017</td>
<td>0.66</td>
<td>0.37</td>
<td>0.30</td>
<td>0.33</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>2004</td>
<td>784</td>
<td>0.74</td>
<td>0.30</td>
<td>0.23</td>
<td>0.47</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>2005</td>
<td>900</td>
<td>0.71</td>
<td>0.28</td>
<td>0.29</td>
<td>0.37</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td></td>
<td>0.70</td>
<td>0.32</td>
<td>0.27</td>
<td>0.39</td>
<td>0.72</td>
</tr>
<tr>
<td>(e) SW Ponderosa pines:</td>
<td>2003</td>
<td>637</td>
<td>0.63</td>
<td>0.10</td>
<td>0.45</td>
<td>0.30</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>2004</td>
<td>319</td>
<td>0.65</td>
<td>0.16</td>
<td>0.40</td>
<td>0.29</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>2005</td>
<td>294</td>
<td>0.67</td>
<td>0.19</td>
<td>0.36</td>
<td>0.32</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td></td>
<td>0.65</td>
<td>0.15</td>
<td>0.40</td>
<td>0.30</td>
<td>0.63</td>
</tr>
</tbody>
</table>
4.0. Discussion

Direct comparison of my results with the few retrospective studies that investigated the effect of bark beetle-caused tree mortality on subsequent fire occurrence in spruce-fir and lodgepole pine forests is not possible as these studies (Bebi et al., 2003; Bigler et al., 2005; Kulakowski and Veblen, 2007; Kulakowski et al., 2003; Lynch et al., 2006; Turner et al., 1999) used longer time lags that ranged between 5 to 50 years than the shorter time lags (0-5 years) I used in this study. However, my results achieved in spruce-fir forests at 5-year time lag agree with the results achieved from the only study that used a 5-year time lag in spruce-fir forests. In that study, Kulakowski et al., (2007) found no effect of bark beetle-caused tree mortality on fire incidence five years after the spruce beetle outbreak in spruce-fir forests in Colorado. My results, therefore, generally support the argument made by other studies (Bebi et al., 2003; Bigler et al., 2005; Kulakowski and Veblen 2007; Kulakowski et al., 2003) that bark beetle-caused tree mortality in spruce-fir forests has no role in influencing fire occurrence.

The two retrospective studies conducted in lodgepole pine forests to date indicated a mixed effect of bark beetle-caused tree mortality on subsequent fire occurrence. In one of these two studies, Turner et al., (1999) found an increased probability of crown fire occurrence during the 1988 fires in Yellowstone National Park in stands with severe bark beetle-caused tree mortality affected between 1969 and the mid 1980s but the probability of crown fire decreased in stands with moderate beetle activity. The second study (Lynch et al., 2006) showed an 11% increase in probability of fire incidence in forests that were affected by mountain pine beetle 16 years before the incidence of the 1988 fire in Yellowstone National Park, but their study also showed no
change in the probability of fire incidence in forests that were affected by bark beetles eight years before the 1988 fire. Despite the fact that my study in lodgepole pine forests vary in time lags with the two studies to date, my results reinforces the argument that bark beetle-caused tree mortality in lodgepole pine forests do not cause increased fire risk during the first 1-5 years following bark beetle activity.

My result in ponderosa pine forests for the 2002 fire year is the only evidence that I have seen in this study that suggested the importance of bark beetle-caused tree mortality in influencing subsequent fire occurrence when I fitted the model to all fires. Bark beetle activity in ponderosa pine forests in 2002 was not the highest as the peak bark beetle activity in this study period was observed in 2005. This exceptional result I observed in ponderosa pine forests in 2002 fire year may be related to the fact that the 2002 fire year in ponderosa pine forests was the year where higher incidence of fire detected by MODIS satellite as compared to the other fire years (Table 4). There are no other retrospective studies conducted in ponderosa pine forests to compare my results with. Similarly, there are no also retrospective studies carried out in Douglas-fir forests as well except the one simulation modeling study conducted by Jenkins et al., (2008).

Simulation modeling studies that looked into potential fire behavior in Douglas-fir, lodgepole pine and Engelmann spruce forests reported a consistent increase in fireline intensity and rate of surface fire spread a few years following beetle outbreaks (0-5 years post outbreak) by comparing forests that are affected and not affected by bark beetles (Jenkins et al., 2008, Page. and Jenkins, 2007). Even though my study used the same time lag as these simulation studies, my results differ. However, another simulation study conducted in Engelmann spruce forests in central Rocky Mountain showed no clear
increase in extreme fire behavior following spruce beetle-caused tree mortality (Derose and Long, 2009). This contrasting evidence from the two simulation modeling studies in Engelmann spruce forests underscores the complexity of the interactions between bark beetle-caused mortality and fire. Such contrasting evidence also signifies the importance of pre-beetle outbreak forest composition, stand structure, and fuel condition in influencing fire behavior following beetle attack (Derose and Long, 2009).

My results showing the predominant role of climate as the dominant factors influencing fire occurrence in western U.S. forests is consistent with results obtained from other studies (Heyerdahl et al., 2008; Westerling et al., 2006). There is clear evidence that fire occurrence in western U.S. forests is controlled by climate, primarily through its effect on fuel availability and fuel flammability (Westerling and Bryant, 2008). A study conducted in spruce-fir and lodgepole pine forests in the Northern Rockies underscored the importance of spring and summer average temperature to influence fire occurrence as fires were more extensive during warm and dry summers following warm springs resulting in earlier melting of the snowpack (Heyerdahl et al., 2008). Earlier snowmelt predisposes forests to greater opportunity to wildfire because in early snowmelt years spring and summer temperature are higher than average, winter precipitation is below average, soil moisture is low and intense, and vegetation is drier than normal (Westerling et al., 2006). Similarly, in my study I also found spring and summer average temperature among the top five most selected predictors in spruce fir, ponderosa pine, and SW ponderosa pine forests while in lodgepole pine forests it was the 6th important predictor as shown by the mean variable importance measure computed for each forest type (Fig. 10).
The strong but inconsistent influence of elevation on fire occurrence across all forest types and fire years (Fig. 10) can be attributed to its effect on temperature. Generally, the potential for suitable fire weather decreases as we go higher along elevational gradients (Bessie and Johnson, 1995). The effects of elevation can also be attributed to its influence on fuel moisture and snowmelt as fuel moisture increases with elevation during the fire season and also snow melts more quickly at lower elevation than at higher elevations (Heyerdahl et al., 2001).

Models that were fitted for all areas that experienced fire in all forest types did not exhibit AUC values of greater than 0.90 to be described models with excellent accuracy; however, the AUC values I achieved are within the range of 0.71 and 0.78 and may be described as well-performing and useful models (Swets, 1988).

5.0 Conclusion

Bark beetles-caused tree mortality did not show strong and consistent influence in predicting fire occurrence in spruce-fir, lodgepole, ponderosa and SW ponderosa pine forests in contiguous western U.S. forests in 2000-2005 fire years and at 0-5 year lag times in comparison to climate and topography-derived predictors. My hypothesis that stated the influence of bark beetle-caused tree mortality would differ across the four major forest types is, therefore, not clearly supported by my results. When bark beetle-caused tree mortality was considered alone, it had a small but transient effect on the probability of subsequent fire occurrence (Chapter 2); however, when compared with climate and topographic variables the effect of bark beetles on fire is negligible. The interaction between bark beetles and fire is highly complex as evidenced by the
contrasting results achieved between some of the studies conducted so far. The evidence gathered to date on this important interaction is far from conclusive and there is still the need for additional studies mainly on the short-term effect of bark beetle-caused tree mortality on fire. We particularly need more studies in ponderosa pine and Douglas-fir forests as the interactions in these forest types is less clear than spruce-fir and lodgepole pine forests. Nonetheless, despite the dramatic impact of bark beetles across the landscape of the contiguous western U.S. forests, generalized forestry management actions to reduce fire risks associated with beetle defoliation within 1-5 years following bark beetle activity do not appear to be justified at present based on my findings.

**Literature cited**


Arno, S.F., 1976. The historical role of fire on the Bitterroot National Forest USDA Forest Service, Intermountain Forest and Range Experiment Station, Res. Paper INT-187, Ogden, UT, USA.


conterminous United States from Landsat Thematic Mapper data and Ancillary data sources. Photogrammetric Engineering and Remote Sensing 67, 650-+


Fig. 13 (Appendix). Maps of predictor variables: bark beetle-caused tree mortality observed in 2004 in all forest types (a), 2001-2005 mean annual precipitation (b), 2001-2005 mean fire season maximum temperature (c), 2001-2005 mean fire season fire season dew point (d), 2001-2005 mean spring and summer average temperature (e), 2001-2005 mean number of snow days (f), 2001-2005 mean PDSI (g), 2001-2005 mean NPP (h), elevation (i), slope (j), and aspect (k).
Fig. 13 (Appendix) (cont.)
CHAPTER 4

PREDICTING THE DISTRIBUTION POTENTIAL OF AN INVASIVE PUERTO RICAN FROG (*Eleutherodactylus coqui*) IN HAWAII USING REMOTE SENSING DATA

Abstract

*Eleutherodactylus coqui* (commonly known as coqui) is a frog species native to Puerto Rico and non-native in Hawaii, where it predate upon Hawaii's endemic fauna and reduces real estate values because of its loud mating calls. Despite widespread problems, its potential range in Hawaii is currently unknown, making control and management efforts difficult. Here I predicted the distribution potential of the coqui on the island of Hawaii at 1 km spatial resolution using five biophysical variables (land surface temperature, leaf area index, fraction of photosynthetically active radiation, normalized difference vegetation index, enhanced vegetation index) derived from the Moderate Resolution Imaging Spectroradiometer (MODIS) as predictors, presence/absence data collected from Puerto Rico and Hawaii, and three classification methods – Classification Trees (CT), Random Forests (RF), and Support Vector Machines (SVM). Models developed using data from the native range and the invaded range predicted potential coqui habitats in Hawaii within the range of well-performing models. Across the three classification methods, mean area under the Receiver Operating Characteristic curve commonly known as AUC was 0.75 for models trained using the native range data and 0.88 for models trained using the invaded range data. I achieved the

3 Coauthored by S. A. Bisrat, M. A. White, K. H. Beard, D.R. Cutler
highest AUC value of 0.90 using RF for models trained with invaded range data. My predictions show that coquis are likely to expand well beyond their current distributions in Hawaii with one of my models predicting as high as 49 % of the island to be suitable coqui habitats. My predictions also show that most of the evergreen forests with an elevation range between 0 and 2000 m which receive higher amount of precipitation are suitable coqui habitats whereas the cool and dry high elevation areas beyond 2000 m elevation are predicted as unsuitable. While my results show that coquis in Hawaii tend to occupy areas with high rainfall and abundant evergreen vegetation similar to their habitats in Puerto Rico, the models do not control for biotic factors, such as competition and predation, in predicting potential habitats, and therefore can not demonstrate niche conservation. I recommend that future work be conducted at finer spatial resolutions with consideration of other biotic factors in limiting coqui distribution. My results also show that MODIS-derived biophysical variables are capable of characterizing coqui habitats in Hawaii.

Introduction

Non-native species have become a global problem because of the potential threat they pose to biodiversity and ecosystems (Lockwood et al. 2007, Lodge et al. 2006). Non-native species also have the potential to generate extensive economic losses (Pimentel et al. 2000). Early detection of non-native species and the ability to predict their distribution patterns in regions outside the native range is fundamental in minimizing the ecological and economic impacts (Zhu et al. 2007). One approach to predict their distribution is to use empirical species distribution models by quantifying the
relationships between field observations and environmental predictor variables (Guisan and Zimmermann 2000). For this practical purpose, species distribution modeling is being widely used to predict the spread of invasive species ranging from small regional scales (Giovanelli et al. 2008, Guo et al. 2005) to continental scales (Broennimann et al. 2007, Loo et al. 2007).

The theoretical underpinning behind species distribution modeling is that the species is able to maintain a viable population without immigration in a multidimensional coordinate system whose axes are ecological and environmental variables (Grinnell, J. 1917). Such models are commonly built using data on the geographical distribution of the species in its native range and projected onto the invaded range with the assumption that the spread of the species in the invaded range will be determined by the same factors as in the native range (Peterson 2003). But this assumption may not hold for all species and may depend on whether the non-native species is introduced intentionally or unintentionally, whether the species is introduced a single time or multiple times in a successive introduction events (related to propagule size), and other factors that determine invasion potentials (Lockwood et al. 2005). In situations where acquiring species data from its native range is difficult or assuming that the invader species will face the same factors that limit its distribution in its native range is not practical, models can be built using species data from its invaded range (Mau-Crimmins et al. 2006). Models developed using only invaded range data, however, can underestimate the potential distribution of the invasive species as the species is not yet at equilibrium with the full range of ecological conditions (Loo et al. 2007). Despite the fact that there is a strong compelling reason to compare species distribution models using both the native
range and the invaded range to achieve accurate prediction (Broennimann and Guisan 2008), there are few studies that compare prediction accuracies between the native and invaded ranges (e.g., Mau-Crimmins et al. 2006).

There is a growing list of classification methods used in species distribution modeling that range from traditional methods that use linear separators among classes (e.g., Generalized Linear Models) to the more modern machine learning based classifiers that use non-linear separators among classes, such as Classification Trees (CT), Random Forests (RF), Support Vector Machines (SVM), and Artificial Neural Networks (Jeschke and Strayer 2008, Olden et al. 2008). The main factor that made these modern classifiers appealing to many ecologists is that there are no requirements for distributional assumptions and for the observed data to be independent (Drake et al. 2006, Guisan and Zimmermann 2000). These different classification methods, however, can yield varying prediction accuracies when applied even to same data sets (Guisan and Zimmermann 2000, Jeschke and Strayer 2008). As a result, there are increasing numbers of studies in species distribution modeling literature that compare different models on identical data sets (Cutler et al. 2007, Franklin et al. 2009).

Some authors argued that when species distribution models built on native range data predict the invaded range well, it can indicate that the species is conserving its niches (Peterson 2003, Peterson and Vieglais 2001). However, not all invading species are able to conserve their niches. Recent evidence of niche shift between native and invaded ranges is observed with herbaceous spotted knapweed (Centaurea maculosa L.) in western North America and Europe where the species showed a clear climatic shift in its invaded range (Broennimann et al. 2007). Other researchers have argued species
distribution models in their current form are not able to answer the question of niche conservation unless they incorporate biotic factors such as competition in addition to abiotic factors (Fitzpatrick et al. 2007, Guisan and Thuiller 2005). For example, Fitzpatrick and Weltzin (2005) argue that species distribution models based only on environmental predictors cannot fully explain whether the species is showing niche conservation because the success of invading species in their new habitat is also mediated by other biotic factors, such as predation and competition. These authors further argue that species distribution models based on environmental explanatory variables alone can only answer the question of whether the niches invading species occupy generally differ from their niches in their native habitat but not niche conservation per se (Fitzpatrick and Weltzin 2005).

*Eleutherodactylus coqui* Thomas (hereafter coqui) is nocturnal, terrestrial frog endemic to Puerto Rico. It was accidentally introduced into Hawaii around 1988 via commercial nurseries and in the past 10 years has spread across the island of Hawaii and onto all four main islands (Kraus and Campbell 2002). The coqui invasion threatens Hawaii’s unique ecological communities because coquis predate upon endemic invertebrates (Beard 2007) and achieve extremely high population densities (up to 90,000 frogs/ha), consuming an estimated 700,000 prey items/ha/night (Beard et al. 2008). The coqui, with extremely loud mating calls (80-90 dB at 0.5 m), has also been determined to negatively impact real estate values (Kaiser and Burnett 2006) and consequently is reducing sales in floriculture and nursery industries, because it often spreads through the sale of nursery plants (Beard et al. 2009). Despite its potential impacts, the future range
of the coqui and its distributional limits in Hawaii are unknown, making efforts to control and manage the species more difficult.

The objective of this study was to predict the potential distribution of coqui frogs on the island of Hawaii using different classification methods, and data from its native and invaded range. My hypothesis was that the environmental niche this invasive frog species occupies in its invaded range is not different to its environmental niche in the native range. In this study, I compare CT, RF, and SVM models for their ability to predict the invasive frog habitat in the island of Hawaii. Although there are studies that compared CT and RF, to my knowledge, no study has yet compared SVM with CT and RF on their use for species distribution modeling.

Methods

Study area

The study area covers the island of Hawaii (19° 41’ 1” N, 155° 23’ 35” W at its center location) where coqui invasion is widespread (Fig. 14). Data from Hawaii Invasive Species Council (HISC) shows that coquis on the island of Hawaii are actively spreading with almost one new population being detected every month (HISC 2007). Coqui invasion on the remaining smaller islands of Hawaii is not widespread. For instance, as of 2008, there were 100s of wild populations (outside of nurseries) on the island of Hawaii (hereafter Hawaii) whereas there were only nine populations on Maui, one population on Kauai, and one on Oahu. Eradication efforts have controlled most of the coqui populations in Maui, Oahu, and Kauai (Beard et al. 2009). As a result, I did not include the other islands in this study.
Species

Anurans have major challenges to cope with fluctuating environmental temperature and water availability because of their permeable skin (i.e., little cutaneous resistance to water loss) (Preest and Pough 1989). Consequently, the control of body temperature and the regulation of evaporative water loss are closely linked in limiting anuran species distribution (Rogowitz et al. 1999). Coquis have direct development, and therefore while they require moisture, do not require the use of bodies of water for any life stage. They have been found to inhabit most parts of Puerto Rico including montane rain forest, urban areas, and dry forests (Beuchat et al. 1984).

In Hawaii, the coqui has spread quickly at low elevations (<500 m) of the eastern side (windward side) where precipitation is high as a result of the orographic uplifting of the tradewinds, but slowly at high elevations (>1000 m). The spread on the western side (leeward side) of the island is also slow compared to the eastern side of the island where mean annual precipitation is lower because the rain shadow effect (Chu and Chen 2005). Coqui in Puerto Rico are found from sea level to the top of the highest peak (1200 m), suggesting that coqui may extend to suitable habitats at higher elevations in Hawaii. Invasion into high elevation forests is of concern because many endemic Hawaiian species are restricted to these habitats (Beard and Pitt 2005). There is, therefore, an urgent need to understand the geographic distribution potential of this invasive frog.

Species distribution data

I collected presence/absence points of coqui frogs from Puerto Rico (native range) and Hawaii (invaded range). I overlaid a 1 km grid over the two islands and then selected
every other pixel of the grid intersecting with the major road network for sampling. I obtained the road layers for Hawaii from Hawaii Data Clearinghouse website (http://hawaii.wr.usgs.gov/hawaii/) and from GIS data depot maintained by an online geospatial community (http://data.geocomm.com/) for Puerto Rico. I used points overlapping the road network for two primary reasons: 1) logistically I would be able to sample more points in a shorter period of time, and 2) because I am studying an invasive species that can spread through vehicular movement (Peacock et al. 2009), areas along roads that coquis could inhabit are more likely to be occupied. The datasets for the predictor variables are collected at 250 m and 1 km spatial resolution (see the section on environmental variables below). I therefore selected the 1 km spatial resolution for sampling to correspond with the coarsest spatial resolution of the predictor variables (Guisan and Thuiller 2005). Furthermore, individual frogs can be easily heard at about 200 m and a chorus can be heard up to 1 km.

At each road stop, I collected data by listening for the mating calls of the male coqui frog at night during peak calling, between 1900 and 100 h, to determine whether the species was present or absent. The mating calls of male coquis are loud and easily identified. Previous studies have found that coquis have higher period of reproductive activity in Puerto Rico measured by high number of calling during the month of May and June (Townsend and Stewart 1994). Accordingly, I collected 757 presence/absence points in Puerto Rico during a two-week period in May 2007 (Fig. 14).
Most of the rainfall in Hawaii occurs in winter between November and March (Chu and Chen 2005) with its heaviest rains occurring by winter storms between October and April (Price 1983); but the notably dry leeward side of the island experiences its maximum rainfall during summer (May to October) (Kolivras and Comrie 2007). I collected 495 presence/absence points in Hawaii using the same sampling protocols as in Puerto Rico during a one week period in late October 2007 to coincide with the beginning of the rainy season at the windward side and at the same time capture the peak rainfall in the leeward side.
rainfall patterns in the leeward side of the island (Fig. 14). I used Magellan SporTrack Map Handheld GPS receiver (Magellan Navigation, Inc, Santa Clara, CA) to geolocate the points. I also obtained 859 known coqui presence points from HISC database (Fig. 15), which I used to validate predicted maps.

Figure 15. Sampled coqui presence/absence points in Hawaii superimposed on land cover map. Black points show sampled absence points, red points show sampled presence points, and the light blue points show known presence points obtained from HISC. The land cover map is reformatted to show only the major cover types (Source: NOAA’s Coastal Services Center, Land Cover Analysis - http://www.csc.noaa.gov/crs/lca/hawaii.html).
Because my sampling protocol was not truly probabilistic or design-based (Edwards et al. 2006), I determined whether my sampling intensity differed from what could have been expected by chance alone, by conducting a $\chi^2$-test on the land cover frequencies from land cover maps for the two regions and the land cover frequencies associated with my sampling points. To calculate the land cover frequencies, I used land cover maps produced by International Institute of Tropical Forestry for Puerto Rico (http://www.tropicalforestry.net) (Helmer et al. 2002) and by National Oceanic and Atmospheric Administration’s Coastal Services Center (http://www.csc.noaa.gov/crs/lca/hawaii.html) for Hawaii (see Fig. 13 and 14 but note that the figures show only the major land cover types).

Environmental data

The interplay between thermoregulation and hydroregulation determines the distribution of coquis (Preest and Pough 1989). As a result, coqui distribution modeling requires climate data on temperature and precipitation. Reliable climate data interpolated from geographically well-distributed weather stations, however, were not available for both the native and invaded ranges. Instead, I used remotely sensed temperature variable and vegetation data derived from remote sensing as a proxy for precipitation (Hashimoto et al. 2008). Note that the remotely-sensed data used here do not produce an estimate of precipitation; rather I assume that the remote sensing metrics described below are related to precipitation (i.e. in tropical climates, high precipitation is accompanied by dense vegetation, which is detectable via remote sensing). Remotely-sensed vegetation data are known to provide a useful alternative for climate variables, especially when unreliable or
sparse climate data must be interpolated from sparse weather stations (Bradley and Fleishman 2008). Topography is often used in species distribution modeling but is an indirect variable with no direct physiological relevance to the species as opposed to direct variables (such as temperature) and resource variables (such as water and food for animals) (Guisan and Zimmermann 2000). Models that include indirect variables, such as elevation, slope, and aspect, cannot be applied outside the geographical extent where the models are built, as species tend to look for microsites in their new habitat to compensate for regional topographic differences (Austin and Smith 1989). For this reason and to maximize the compatibility of models developed in Puerto Rico and Hawaii (Randin et al. 2006), I did not include topographic variables.

I used four Moderate Resolution Imaging Spectroradiometer (MODIS) biomass/greenness products as surrogates for precipitation (Bradley and Fleishman 2008) and one MODIS temperature dataset. I compiled MODIS data (Collection 4) from 2000 to 2005 for Hawaii and Puerto Rico (from both Aqua and Terra platforms). I obtained the MODIS data sets from the Land Processes Distributed Active Archive Center located at United States Geological Service Center for Earth Resources Observation and Science. I selected MODIS over other sensors mainly for its high radiometric resolution (36 spectral bands) and its finer temporal resolution as it is able to capture images across the entire globe at 1 to 2 days repeat cycle (Justice et al. 2002). This continual and comprehensive coverage by MODIS is frequent enough to capture many rapid biological and meteorological changes. MODIS products are generated from daily surface reflectance data by compositing them at 8-day or 16-day period. The compositing is mainly done to reduce atmospheric contamination (Justice et al. 2002).
I obtained temperature data from the 8-day composite, 1 km spatial resolution Land Surface Temperature (LST) product (MOD11A2). LST is retrieved from MODIS thermal infrared bands by combining the results of surface-atmosphere interactions and energy fluxes between the atmosphere and the ground (Wan et al. 2004). I obtained Leaf Area Index (LAI) and Fraction of Photosynthetically Active Radiation (FPAR) data from the 8-day composite, 1 km spatial resolution product (MOD15A2). LAI defines the structural property of a plant canopy, which is the one-sided leaf area per unit ground area. FPAR, on the other hand, measures the proportion of available radiation in the photosynthetically active wavelengths (400-700 nm) that a canopy absorbs (Myneni et al. 2002). I obtained Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) data from the 16-day composite, 250 m spatial resolution product (MOD13Q1). NDVI and EVI are the two MODIS vegetation index products that represent measurements of vegetation biophysical properties (greenness) that indicate photosynthetic activity (Huete et al. 2002). While NDVI is chlorophyll sensitive, the EVI is modified NDVI to make it more responsive to canopy structural changes, to improve its sensitivity in high biomass regions, to decouple the canopy background signal, and also to reduce atmospheric influences (Huete et al. 2002).

**Pre-processing of the data**

I produced composite period (8-day or 16-day based on the type of products) and monthly climatological means from both Aqua and Terra satellites by averaging values over the 6 year period (2000-2005). Before I generated the climatological means, I consulted the Quality Control and Quality Assurance (QC and QA) layers for LST,
LAI/FPAR and NDVI/EVI products that are produced at pixel level. The algorithms that produce these products are executed regardless of input quality and the QC and QA layers need to be applied to retrieve reliable values. Based on the QC and QA definitions for both products, I controlled data quality by eliminating pixel values where snow/ice, atmospheric aerosol, cloud, cloud shadow, and cirrus clouds are detected. Where no good data is available as a result of the QA and QC procedures, I interpolated the missing values using the overall mean for each variable. Based on the examination of preliminary results on prediction accuracies generated between the composite period and monthly climatological means, I found that monthly climatological means gave better prediction accuracy. Hence, I ultimately used the monthly climatological means as predictor variables in this study.

**Modeling approach**

I used three machine learning classification methods in my analyses: CT, RF, SVM. First, I fitted the classifiers using data from Puerto Rico, the native range, and then predicted the invaded range in Hawaii. The data from Hawaii was used as an independent data set to test the fitted classifications. My second approach was to fit the classifiers to the Hawaii data and then to predict the Hawaiian range. Because I did not have a separate, independent data set collected in the invaded range to test the predictions, I used 10-fold cross-validation rather than the overly optimistic resubstitution accuracies to estimate the generalization accuracy (Edwards et al. 2006). I also conducted a sensitivity analysis on my test data set by progressively removing data to test the sensitivity of my results to sampling intensity. I removed 20%, 40%, and 60% of the data randomly and
conducted the sensitivity analysis iteratively 30 times and averaged the corresponding accuracy metrics. I then tested whether there is a significant change in prediction accuracy at each drop in the size of the test data using Wilcoxon’s signed rank test.

Classification Methods

CT work by recursive binary partitioning of the data space into increasingly homogenous regions, known as nodes, with regard to the classification variable (Breiman et al. 1984). At each step in the fitting process, the algorithm selects a node, a predictor variable, and a cut-off value to maximize the decrease in the Gini index (Breiman et al. 1984). The partitioning is continued until further subdivisions no longer reduce the Gini index. Such a tree is said to be fully grown and may have as many terminal nodes as there are distinct combinations of values of predictor variables in the dataset. The lower branches of a fully grown classification tree are typically modeling noise in the data and must be pruned off. The pruning is accomplished by selecting an appropriate value of the complexity parameter, cp, which is the smallest decrease in the value of the Gini index for which a split can take place. Once the fully grown tree has been constructed, a plot of cross-validated error versus cp may be obtained and the value of cp that minimizes the cross-validated error is chosen (Breiman et al. 1984b). I used ‘rpart’ package (Therneau and Atkinson 2007), which is the R implementation of CT, for all my analyses using this method.

RF fits many classification trees to the data and combines the predictions of the individual trees to create more accurate predictions (Breiman 2001). The RF algorithm begins with the random selection of many (the default is 500) bootstrap samples from the
original data. In each bootstrap sample approximately 63% of the original data points occur one or more times. The observations in the original dataset that do not occur in the bootstrap sample are said to be \textit{out-of-bag} for that bootstrap sample. To each bootstrap sample a classification tree is fit with the restriction that at each node only a small number (the default is the square root of the total number of variables) of variables is available to be split on. The fitted trees are used to predict the class for each observation that is out-of-bag with respect to the bootstrap sample the tree is fit to. Finally, the predicted class of an observation is determined by majority vote of the predicted classes for all trees for which the observation is out-of-bag, with ties in the voting decided randomly. The fact that each tree is only used to predict for the observations in the original data set that were not used in the fitting of the tree ensures that RF does not overfit and that the out-of-bag accuracy estimates are unbiased estimator of the true generalization accuracy. For this study, I used ‘randomForest’ package (Liam and Wiener 2002), which is the R implementation of RF.

SVM is a learning algorithm that is designed for two-class problems, such as species presence/absence modeling (Vapnik 1998). SVM projects the original data into a very high dimensional \textit{feature space}. SVM then seeks for the optimal separating hyperplane that separates the two classes (presences and absences in my application) in the high dimensional space by maximizing the margin between the closest training points of the two classes (Hastie 2001). The boundaries between the classes in high dimensional space are then projected back into the original data space, where they may be highly non-linear. SVM is designed to provide high flexibility for approximating class boundaries while at the same time avoiding over-fitting problems (Guo et al. 2005). I used the svm
function in the R package ‘e1071’ (Dimitriadou et al. 2008), which is an interface to the LIBSVM library (Chang and Lin 2008).

Model accuracy assessment

I assessed prediction performance of each classifier in the two modeling approaches using percentage of correctly classified (PCC), sensitivity, specificity, kappa, and area under the Receiver Operating Characteristic curve, commonly known as AUC (each described in more detail below). All these five measures of model accuracy have a value that range from zero (very poor accuracy) to one (a perfect agreement between the observed and the predicted). I used the ‘PresenceAbsence’ package (Freeman 2007) in R to calculate all the accuracy metric. I compared model performance across the three classification methods using Kruskal-Wallis rank sum test in R. I used the Wilcoxon’s signed rank test in R to compare model performances between the two modeling approaches.

PCC, sensitivity, and specificity are calculated from the matrix of correct and incorrect classifications commonly known as the confusion matrix (Fielding and Bell 1997). PCC is calculated by dividing the number of correctly classified presences and absences by the total number of points in the error matrix. Sensitivity measures the percentages of presences correctly predicted, whereas specificity measures the percentages of absences correctly predicted. I get high sensitivity values when omission errors (defined as incorrect prediction of absences) are less common. Similarly, I get high specificity when commission errors (defined as incorrect prediction of presences) are infrequent (Fielding and Bell 1997). PCC, sensitivity, and specificity use an arbitrary
threshold (commonly 0.5) to convert the probabilistic estimates of species presence into dichotomous presence-absence predictions making them threshold-dependent accuracy measures (Fielding and Bell 1997).

Kappa, like PCC, sensitivity, and specificity, is also calculated from the confusion matrix (Fielding and Bell 1997). Kappa measures the agreement between predicted presences and absences with the actual observed presences and absences corrected for agreement that might be due to chance alone. Kappa is essentially a measure of randomness with more accurate models having a value close to 1 and a random predictor having a value of 0 (Cohen 1960). Kappa also uses arbitrary threshold to convert the probability of species occurrence into presence or absence (Fielding and Bell 1997).

AUC evaluates the model’s ability to discriminate between sites where the species is present or absent by plotting the sensitivity (true positives) against 1-specificity (false positives) (Fielding and Bell 1997). An AUC value of 1 indicates perfect discrimination between presence and absence sites and a value of 0.5 indicates no discrimination ability. AUC evaluates the performance of a model across all possible threshold values to convert the probabilistic estimates of species occurrence into presence/absence making it one of the most reliable threshold-independent measures of accuracy (Broennimann et al. 2007).

**Map generation**

To visualize the predictions from the three classification methods, I generated maps of the predicted probability distribution of coqui by applying the selected models to the entire grid of the invaded range using ‘yalmpute’ package (Crookston and Finley 2008) in R. Once I generated the maps for each classification method and the two
modeling approaches, I created a mean classification map by averaging the six maps. I also generated a map that depicts the uncertainty associated with my predictions by calculating standard error around the mean of the six classification maps (Foody and Atkinson 2002).

**Results**

Out of the 757 presence/absence points I collected from Puerto Rico, 439 were presences and 318 were absences (Fig. 14). Out of the 495 presence/absence points I collected in Hawaii, 152 were presences and 343 were absences (Fig. 15). A $\chi^2$-test revealed that the land cover frequencies of Puerto Rico were not different from land cover frequencies associated with my sampling points ($\chi^2=5.839$, df=32, p-value=0.999). The difference between the land cover frequencies and the land cover frequencies associated with my sampling points in Hawaii, however, were significant ($\chi^2=34.925$, df=12, p-value<0.001). Grasslands, scrub/shrub, and cultivated land cover types (Figure 15) are well represented during sampled presence/absence data collection. On the other hand, evergreen forest and bare land cover types (Fig. 15) are underrepresented; whereas high intensity developed and low intensity developed cover types (Fig. 15) are overly represented.

Out of the three classification methods trained on Puerto Rico and applied to Hawaii, RF showed higher performance in specificity and Kappa while SVM showed higher accuracy in PCC and sensitivity (Table 6). CT had the lowest prediction accuracy (Table 6a). RF performed better than the other two methods developed from native range
data. However, the differences in all measures of accuracy across the three classification methods on native range data were not significant (H=1.30, df=2, p-value=0.522).

For models developed from and applied to Hawaii, the three classification methods had a better accuracy than models developed on the native range (W=0, P<0.001). RF had slightly higher accuracy in all the accuracy metrics considered, followed by SVM, with CT showing the least accuracy (Table 6b). Once again, the accuracy measures across the three classification methods on invaded range data were not different (H=2.56, df=2, p-value=0.278).

The predicted maps generated by the three classification methods indicate that coquis are likely to expand their Hawaiian range well beyond their current distribution (Fig. 16a-f). RF predicted a larger area of the island as potential suitable coqui habitats (Fig. 16c, 16d) than CT (Fig. 16a, 16b) and SVM (Fig. 16e, 16f) predicted. Note that SVM predicted few suitable coqui habitats in the southeast and western side of the island compared to CT and RF. For models built on native range data, RF predicted that coquis would increase their distribution to cover about 49% of the island, whereas CT predicted it would cover about 32% and SVM predicted it would cover about 20% considering pixels with moderate to high probabilities of coqui presence (between 0.41 and 1.0). Similarly, RF predicted about 48%, CT about 20%, and SVM about 15% of the island will be covered by coquis using invaded range data.
Figure 16. Coqui predicted distribution maps using monthly data from native range generated by CT (a), RF (c), and SVM (e) and using invaded range data generated by CT (b), RF (d), and SVM (f). Known coqui presence points are represented with black dots. The legend indicates the predicted probabilities of coqui presence and known coqui presence points. Shaded relief of the island of Hawaii is shown as a background to represent the surface features.
Table 6. Accuracy measures for predictions of coqui presence on the island of Hawaii using monthly climatological values calculated from MODIS-generated biophysical variables as predictors.

| Classification method | Accuracy metric |  |
|-----------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                       | PCC (%)         | Sensitivity (%) | Specificity (%) | Kappa           | AUC             |
| (a) Models developed in native range: |                 |                 |                 |                 |                 |
| Classification Trees   | 0.67            | 0.66            | 0.71            | 0.32            | 0.74            |
| Random Forests         | 0.71            | 0.70            | 0.75            | 0.40            | 0.79            |
| Support Vector Machine | 0.72            | 0.73            | 0.69            | 0.38            | 0.72            |
| (b) Models developed in invaded range: |                 |                 |                 |                 |                 |
| Classification Trees   | 0.80            | 0.84            | 0.69            | 0.52            | 0.85            |
| Random Forests         | 0.86            | 0.87            | 0.83            | 0.66            | 0.90            |
| Support Vector Machine | 0.85            | 0.86            | 0.82            | 0.63            | 0.88            |

The mean prediction map (Fig. 17) generated by averaging the six prediction maps (Fig. 16a-f) assigned moderate to high probabilities of coqui presence (between 0.41 and 1.0) to most of the moist, evergreen lowland forests (compare Fig. 17 and the land cover map in Fig. 15). On the other hand, the mean prediction map assigned low probabilities of coqui presence (between 0.00 and 0.30) for the dry, cool, and high elevation areas that are mostly barren lands and areas that are dry at lower elevation areas (compare Fig. 17 and the land cover map in Fig. 15). The contour maps shown in Figure 17 for the mean prediction map indicate that the elevation range for the moderate to high probabilities of coqui presence (between 0.41 and 1.0) is between 0 m and 2000 m with the number of pixels that are suitable coqui habitats in each elevation bins decreasing with increasing elevations (Fig. 19). Elevation ranges beyond about 2000 m are predicted as unsuitable areas for coquis. Slightly more than half of the pixels that are assigned with moderate to high probability of coqui presence were in evergreen forest land cover type followed by grassland cover type with about 22% of the predicted suitable coqui habitats (Table 7). By contrast, bare land cover type received only 0.44% of the moderate to high
probability of presence pixels (Table 7). The 95% confidence interval calculated for the mean prediction map shown in Fig. 16 has a 0.70 probability of coqui presence at its upper limit and a 0.32 of probability of coqui presence at its lower limit. The mean prediction map also shows that areas with moderate to high probability of coqui presence (pixels with values with in the range of 0.41 and 1.0) match with about 90% of the 859 known coqui presence points obtained from HISC (Fig. 17). However, the mean prediction map also showed that areas with the three known coqui presence points on the southern part and two known coqui presence points on the northwest part of the island were not predicted to have moderate to high probabilities of coqui presence.

Figure 17. Predicted probability map showing contour maps at 200 m spacing and the mean classification (with an upper limit of 0.70 and lower limit of 0.32 of 95% confidence interval) generated out of the six prediction maps shown in Fig. 15 using monthly values as predictors. Shaded relief of the island of Hawaii is shown as a background to represent the surface features.
The prediction uncertainty map showed standard errors that ranged from 0.007 to 0.140 (Fig. 18). The uncertainty map showed that there is medium to high prediction uncertainty in the north, northeast, south, southeast, southwest, and west region of the island with standard error ranging from 0.061 to 0.140. The eastern side of the island showed a mixed prediction uncertainty with some pockets of the eastern side receiving a standard error of as low as 0.007, while some pockets receiving medium to high prediction uncertainty (standard error of 0.061-0.140). On the other hand, most of the high elevation and dry areas of the central and northwest part of the island received a prediction uncertainty between low and medium with standard error that ranged from 0.035 to 0.060 (light green color in Fig. 18).

Table 7. Distribution of extant land cover types as shown in Figure 14, measured coqui presence, and modeled habitat (probability between 0.41 and 1.0) All values are shown in percent, rounded to nearest whole number.

<table>
<thead>
<tr>
<th>Land cover type</th>
<th>Areal coverage</th>
<th>Field-sampled presence points</th>
<th>Modeled potential coqui habitat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unclassified</td>
<td>1</td>
<td>14</td>
<td>4</td>
</tr>
<tr>
<td>High Intensity Developed</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Low Intensity Developed</td>
<td>1</td>
<td>19</td>
<td>1</td>
</tr>
<tr>
<td>Cultivated Land</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Grassland</td>
<td>24</td>
<td>20</td>
<td>22</td>
</tr>
<tr>
<td>Evergreen Forest</td>
<td>30</td>
<td>19</td>
<td>55</td>
</tr>
<tr>
<td>Scrub/Shrub</td>
<td>19</td>
<td>22</td>
<td>15</td>
</tr>
<tr>
<td>Palustrine Scrub/Shrub Wetland</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Palustrine Emergent Wetland</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Unconsolidated Shore</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Bare Land</td>
<td>24</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 18. Map showing the prediction uncertainty calculated using standard error for the six prediction maps shown in Figure 16. Shaded relief of the island of Hawaii is shown as a background to represent the surface features.

The variable importance plots generated using RF (Fig. 20) reveal that for models built on native range data, monthly values derived only from LST, NDVI, and EVI were among the most important variables with NDVI occurring in winter/spring seasons, EVI occurring in most of the year, and LST occurring in all months (Fig. 20a). For models built on invaded range data, on the other hand, monthly values derived from all the MODIS variables (i.e., LAI, FPAR, LST, NDVI, EVI) were among the most important predictor variables with NDVI and EVI concentrating in summer/fall, LST in spring and fall, and LAI mostly in fall seasons (Fig. 20b). These variable importance plots revealed two important patterns. First, monthly LST values occurred in most of the months as
important variables for both regions suggesting that monthly LST values alone might seem to be suitable for explaining coqui distribution. Predictions using only LST, however, resulted in poor accuracy measures and poor spatial predictions (results not presented). Second, the variable importance plots revealed that there is an apparent clustering of NDVI and EVI into different months in the two regions, suggesting that the important months in Puerto Rico and Hawaii are different.

Figure 19. Frequency distribution of pixels with moderate to high predicted probability (0.41–1.0) of coqui presence at 11 elevation bins.

The results gathered from the variable importance plots led to the investigation of the seasonality of precipitation by identifying the two driest and two wettest months in the two regions and calculating the mean wet season and mean dry season values for each predictor variable and the difference between the mean wet and dry seasons values. When
I reran the models using these new 15 predictor variables, I achieved comparable accuracy metric values when compared with the accuracy metrics obtained from monthly values (compare Table 6 and 7). The new 15 predictor variables also generated similar prediction maps to the one produced using monthly values in each of the classification methods and the two modeling approaches with areas that have moist, evergreen vegetation receiving high probability value of coqui presence and the dry, high altitude areas receiving low probability values (compare Fig. 16 and 21).

Results on sensitivity analysis of the test data indicate that the accuracy metric for all the three classification methods remained generally consistent when I progressively remove more and more of the test data (Fig. 22). For instance, AUC in RF for models built on native range data slightly changed from 0.79 to 0.78 when only 40% of the data is used relative to the entire test dataset (Fig. 22c). Similarly, AUC changed from 0.88 to 0.89 in SVM for models built on invaded range data when I used only 40% of the test data (Fig. 22f). CT showed overall higher variability than RF and SVM when the test data is progressively reduced. For instance, AUC value reduced from 0.85 to 0.76 in CT when 40% of the data is used relative to the entire dataset for models built on invaded range data (Fig. 22b). Models built on invaded range data showed more variability in accuracy metric values as a result of the sensitivity analysis in comparison to models built on native range data. However, the change in accuracy metric values as a result of the progressive removal of the test data was not significant \( (P>0.05) \) for all models built on native range data as well as invaded range data.
Figure 20. Graphs showing the important variables generated using RF. The left hand panel (a) shows variable importance for models built on native range data and the right hand panel (b) shows for models built on invaded range data. The x-axis shows the mean decrease in accuracy. Note that a variable with the highest mean decrease in accuracy is the most important variable.
Figure 21. Coqui predicted distribution maps using dry and wet season values as predictors from native range generated by CT (a), RF (c), and SVM (e) and using invaded range data generated by CT (b), RF (d), and SVM (f). Known coqui presence points are represented with black dots. The legend indicates the predicted probabilities of coqui presence and known coqui presence points. Shaded relief of the island of Hawaii is shown as a background to represent the surface features.
Figure 22. Sensitivity analysis of the test data by progressively removing 20%, 40%, and 60% of the data using CT (a, b), RF (c, d), and SVM (e, f). The right hand panel shows for models built on native range data and the left hand panel shows for models built on the invaded range data.
Table 7. Accuracy measures for predictions of coqui presence on the island of Hawaii using dry and wet season values calculated for each MODIS-generated biophysical variable as predictors.

<table>
<thead>
<tr>
<th>Classification method</th>
<th>Accuracy metric</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PCC (%)</td>
</tr>
<tr>
<td>(a) Models developed in native range:</td>
<td></td>
</tr>
<tr>
<td>Classification Trees</td>
<td>0.69</td>
</tr>
<tr>
<td>Random Forests</td>
<td>0.73</td>
</tr>
<tr>
<td>Support Vector Machine</td>
<td>0.72</td>
</tr>
<tr>
<td>(b) Models developed in invaded range:</td>
<td></td>
</tr>
<tr>
<td>Classification Trees</td>
<td>0.82</td>
</tr>
<tr>
<td>Random Forests</td>
<td>0.87</td>
</tr>
<tr>
<td>Support Vector Machine</td>
<td>0.84</td>
</tr>
</tbody>
</table>

Discussion

One of the objectives of this study was to test which modeling approach (modeling based on native range data vs. modeling based on invaded range data) is best for predicting potential coqui habitats. Given highest AUC values achieved in each modeling approach (0.79 for models on Puerto Rico data and 0.90 for models on Hawaii data), I conclude that my models in the two approaches predicted the invaded range within the accepted range for well-performing models (Pearce and Ferrier 2000, Swets 1988) with the models built using the invaded range data performing better than the models from the native range. Also, due to the fact that my model predictions for moderate to high probability of coqui presence match with about 90% of the known coqui presence points acquired from HISC (Fig. 17) supports my argument that my models using both approaches predicted the invaded range well.

My results, which showed potential areas that coquis can invade in Hawaii to exceed their current distribution, is of great management concern. However, my results have shown that coquis may have natural limitations to establish in hot spot areas such as
around Volcanoes National Park where most endemic invertebrates are located. This was evident in my results that showed a drastic decrease in the number of pixels predicted as potential coqui habitats in elevation ranges beyond 1000-1200 m. It is estimated that coquis covered about 8,000 ha of the island by 2007 (Beard 2009) which is less than 1 % of island’s total land area. My prediction maps, however, revealed that up to 49 % of (using RF as a classifier) the island’s land area could be suitable habitat for coquis. I recommend, therefore, for control and management policies to be geared towards those areas with higher likelihood of coqui presence (such as the evergreen forests and pocket wet areas in grasslands and shrub/scrub land cover types) and with low prediction uncertainty. Furthermore, areas with low likelihood of coqui presence should receive the least control and management policy considerations. Also, this type of information can be used to search for incipient populations and can direct resources towards areas of higher invasion probability and reduces resources spent looking for populations in low probability areas. Although my modeling effort in this study was specifically focused on the big island, the modeling approaches I used here were also applied to predict suitable coqui habitats in the remaining smaller Hawaiian Islands (results not presented here).

Generally, when the performance of a species distribution model is tested using training data either via cross-validation or data splitting as is the case in my invaded range data, AUC values greater than 0.9 can be achieved; however, species distribution models perform less well when the test data is spatially independent (Jeschke and Strayer, 2008). When my best model built on native range was projected onto spatially independent invaded range, I achieved the highest AUC value of 0.79 which is, therefore,
far better than random model indicating satisfactory model fit (Pearce and Ferrier 2000, Swets 1988). Loo et al. (2007) reported AUC value of 0.73 for models built on native range data to predict the native range of a freshwater snail (*Potamopyrgus antipodarum*) in New Zealand. But when the New Zealand model was projected onto North America, where the species is invasive, AUC declined to 0.53. It can be argued though that models developed on invaded range data can underestimate the potential distribution of invasive species (Loo et al. 2007). Because I achieved an AUC value of 0.90 using only invaded range data as opposed to 0.79 on native range data, my results can suggest that it is not impossible to obtain reasonable range predictions using only invaded range data in situations when acquiring data from native range is difficult or too expensive. One of the reasons the native range data may not have predicted the invaded range as well as the invaded range data can also be attributed to the fact that coquis in Hawaii are currently still spreading. In the future when the coqui populations fully occupy all available suitable coqui habitats, it would be interesting to retest these models and determine if predictions from the native range increase.

Remote sensing data are increasingly used in species distribution modeling. There is, however, a concern with remote sensing data as to whether the existing sensors are capable of detecting environmental variables and biophysical properties that influence species distribution. Current sensors like MODIS are capable of generating relevant data that can be used as input variables for species distribution modeling (Bradley and Fleishman 2008). MODIS-derived variables such as NDVI and LAI have been effectively applied in species distribution modeling (Buermann et al. 2008, Saatchi et al. 2008). My results also support the argument that MODIS-derived LAI, FPAR, LST, NDVI, and EVI
values can be used as proxy measurements for climate and were able to detect relevant environmental variables and characterize coqui species habitat. My results also showed that more than one MODIS-derived biophysical variable was needed to characterize coqui habitats as shown in the poor prediction accuracy I observed when I used only monthly LST values as predictors despite the fact that monthly values derived from LST occurred in almost all months of the year as important variables (see Fig. 7).

When species distribution models based on climate variables predict the invaded range well at biogeographic scale, it can be argued that the species is primarily limited by climate (Wiens and Graham 2005). Pearson and Dawson (2003) suggested that it is when species distribution is modeled at less than 1 km spatial resolution that we expect for biotic interactions to become more important. My results suggest that at the 1 km² spatial scale at which my models are built, coqui distribution in Hawaii is probably primarily limited by climate. However, future local-scale studies should test the role of biotic factors in limiting coqui distributions to eventually validate my biogeographic scale predictions (Pearson and Dawson 2003, Wiens and Graham 2005).

CT and RF are receiving increasing application in species distribution modeling (Crossman and Bass 2008, Cutler et al. 2007, De'ath and Fabricius2000, Franklin et al. 2009, Marmion et al. 2009, Peters et al. 2007). CT, however, suffer from over-fitting the training data and are unstable as small changes in the data can result in marked change in the results obtained by CT (Breiman 2001, Cutler et al. 2007). RF, on the other hand, is proving to be a more robust and stable classifier due to its superior prediction accuracy and its ability in avoiding the over-fitting and instability problem associated with other classification tree methods (Cutler et al. 2007, Lawler et al. 2006, Marmion et al. 2009).
The higher variability exhibited by CT than RF and SVM during my sensitivity analysis (Fig. 21) can be attributed to the instability of CT. My results, therefore, support the argument that RF and SVM outperforms CT in stability as well as in prediction accuracy. SVM is rarely used in species distribution modeling (Drake et al. 2006, Guo et al. 2005, Kelly et al. 2007). My results suggest that it has a great promise as one additional tool in the ecologist’s modeling tool box by showing a competing performance with RF generating comparable prediction accuracies but predicting smaller coqui suitable areas than RF. Compared to other optimization methods such as Artificial Neural Networks, SVM has fewer parameters and requires less tuning (Drake et al. 2006). One way of testing how good a modeling procedure is, is to find out whether the results change when the predictor variables are rearranged or the input variables are changed. The fact that I obtained comparable accuracy measures and spatial predictions when I both used monthly data as well as dry and wet season data underscores the robustness of my modeling procedures.

On the question of whether my results proved niche conservation or not, I concur with Fitzpatrick and Weltzin’s (2005) assertion by concluding that my results can not conclusively show that coquis in Hawaii are showing niche conservation; however, my results show that the niches coquis occupy in Hawaii are not different from the niches they occupy in their native ranges in Puerto Rico. I argue that my results do not show conclusive evidence of niche conservation for three main reasons. First, my models relied mainly on environmental variables without including biological processes such as interactions among species (e.g., competition and predation) and dispersal limitations (Fitzpatrick et al. 2007, Pearson and Dawson 2003). For instance, in areas where my
models predicted suitable coqui habitats, predators, competing species, or dispersal limitations could prevent the coqui from occupying the suitable habitats (Guisan and Thuiller 2005), although predators and competing species are less likely to be an issue for coqui in this invaded range because there are few such species (Beard and Pitt 2005). Second, coqui populations in Hawaii are not yet at equilibrium with their available suitable habitats and are still spreading with new populations being discovered almost every month (HISC 2007). In situations where the species is not fully colonizing the available suitable habitats, models based on current distribution show only a snapshot of the distribution of the species and may not capture the full distribution potential of the species (Fitzpatrick and Weltzin 2005). Third, ecological niches can actually evolve albeit slowly and under constrained circumstances, while this would not influence whether there is niche conservation today, it could influence whether coquis have conserved their niche into the future, and thus their predicted ranges could change; consequently, coquis can evolve their niches to occupy habitats that my models predicted as unsuitable (Holt 1992). This scenario could pose a serious problem in predicting future invasion of coquis due to the fact that coquis are recently introduced to Hawaii and their potential to evolve their niches is not yet known. None of the models correctly predicted known coqui presences in the southern and northwestern regions, possibly indicating some sort of niche evolution. On the other hand, this mismatch between the few coqui presence points and my predictions could also be due to suitable microhabitats undetectable from MODIS. In particular, the highly fractured volcanic landscape in the southern tip of Hawaii may support mesic microhabitats capable of supporting coqui populations.
Conclusion

Predicting the potential distribution of invasive species, both as a proactive or retroactive mitigating strategy, is key to control and manage invasive species. My results have shown the potential areas that coquis can invade in Hawaii are by far larger than their current distribution with one of my models predicting up to 49% of the island as suitable coqui habitats. Hence, the predicted coqui distribution map and the uncertainty map I generated under this study have important management implications. Management and control of coqui populations in the island of Hawaii should, therefore, be geared towards those areas that have higher probabilities of coqui presence and low uncertainty of prediction. Besides, this type of information can be used to search for incipient populations and can direct resources towards areas of higher invasion probability and reduces resources spent looking for populations in low probability areas.

My null hypothesis that stated the habitats coquis occupy in Hawaii are similar to their habitats in Puerto Rico is accepted as my results suggested similarities between the two coqui habitats. From a methodological point of view, my study underscored three main points. First, coqui habitats in Hawaii can be effectively predicted using species distribution modeling approach and data from both the native as well as the invaded range. Second, remotely sensed data can be effectively used to detect important predictor variables that characterize coqui habitats. Third, because species distribution models are correlative methods, I recommend more local-scale mechanistic phylogenetic, physiological, and ecological studies to further understand the role of both biotic and abiotic factors in limiting coqui distributions and, thereby, validate my coarse-scale predictions.
Literature cited


Dimitriadou, E. et al. 2008. e1071. R Development Core Team.


Romme, W. H. et al. 2006. Recent Forest Insect Outbreaks and Fire Risk in Colorado
Forests: A Brief Synthesis of Relevant Research. Colorado State University,
Colorado Forest Restoration Institute, pp. 1-24.


1285-1293.

Core Team.


Zhu, L. et al. 2007. Predicting the spatial distribution of an invasive plant species
CHAPTER 5
SUMMARY AND CONCLUSION

My studies have shown the successful implementation of MODIS satellite imagery to address two ecologically important research questions. In situations where extensive collection of high quality field data is impracticable, remote sensing can be a good alternative. Most of the studies that investigated the interactions between bark beetles and fire or studies that attempted to model fire behavior in western forests have used datasets collected by conventional means. However, these datasets are known to have some errors associated with incomplete reporting or inaccurate geolocations. Remotely sensed fire products such as the MODIS fire product I used in this study, can be a good alternative sources. To my best knowledge, no study has yet attempted to explore the interactions between bark beetle and fire using remotely sensed fire datasets.

Using conditional probability modeling, I showed bark beetle-caused tree mortality can modestly increase the probability of fire occurrence 2- to 3- years following beetle outbreak at an average of 13% for the two time lags without taking forest type into account. There was even marked increase in the probability of fire occurrence when persistent fires are considered relative to transient fires. However, these persistent fires, although dramatic and of high intensity, are only 12% of the total fire datasets considered in this study leaving the majority of the fires ignited in western U.S. forests unaffected by bark beetle-caused tree mortality. Other studies in lodgepole pine forests similarly showed a modest increase in probability of fire occurrence following beetle outbreak but these studies looked at a longer term lagged effects of beetle outbreaks. Further, when I
tested the importance of bark beetle-caused tree mortality to influence fire occurrence relative to other climate and topography-derived predictors, bark beetle-caused tree mortality was consistently the least selected predictor to influence fire occurrence in spruce-fir, lodgepole pine, Douglas-fir, and ponderosa pine forests.

My results from these two studies generally reinforce the existing scientific evidence that indicate bark beetle and fire interactions are highly complex and the former does not necessarily increase the probability of occurring of the latter. The studies to date has proved that it is very difficult to draw a simple and general conclusion as the effects of the change in fuel loading that can result from bark beetle-caused tree mortality can be mediated by several other factors such as stand structure following bark beetle outbreak. Time since the last bark beetle outbreak also plays a crucial role in influencing the risk of fire hazard as fuel dynamics and stand structure tend to change over time. Therefore, a one size fits all approach to manage bark beetle affected forests with the ultimate goal of minimizing fire hazard may not be warranted at this stage.

I believe the two studies I conducted on the question of whether bark beetle-caused tree mortality can increase subsequent fire occurrence are the first retrospective studies that explored the short term lagged effects of bark beetle activity in contiguous western U.S. forests. Besides, my study is the first one to explore effects of bark beetle activity on fire occurrence in ponderosa pine forests and the first study that used retrospective research approach in Douglas-fir forests as most of the studies to date are highly localized being conducted only either in subalpine forests of Colorado or lodgepole pine forests of Yellowstone National Park. There is, therefore, an urgent need
for more studies on this important management concern but yet poorly understood research question especially in Douglas-fir and ponderosa pine forests.

Traditionally, species distribution models are implemented using climate data as explanatory variables. My results underscored the crucial role remotely-sensed data can play in situations where reliable climate data are not available. This was evident because my species distribution models have shown the relevance of MODIS-derived biophysical variables to effectively characterize coqui habitats in Hawaii. This could be another evidence on the utility of remotely sensed data sets in species distribution modeling in addition to the growing number of studies which used remotely sensed data sets successfully when modeling species distributions. Further, my study has confirmed the use of satellite-derived vegetation data such as vegetation indices, LAI, and FPAR as useful surrogates for precipitation when reliable gridded precipitation data is not available. From a methodological point of view, my study has also explored the implementation of SVM for species distribution modeling and compared its prediction accuracies with CT and RF for the first time. Despite being one of the ‘black box’ classification algorithms, SVM showed comparable prediction accuracies in comparison to RF which is one of the most robust classification algorithms. However, when the two classification algorithms are used to predict spatially, they showed marked difference with SVM predicting fewer pixels as suitable coqui habitats in comparison with RF.
VITA

EXECUTIVE SUMMARY:

- Academic and professional experience in landscape ecology, remote sensing and GIS
- Excellent technical skills including digital image processing, GIS spatial analysis, spatial modeling, species distribution modeling, and habitat mapping
- Demonstrated training/teaching and communication skills
- Outstanding work ethic
- Multi-cultural background gained through international study and travel

EDUCATION:

PhD, Ecology (Applications of Remote Sensing and GIS), December 2009, Utah State University, Graduate Advisor: Dr. Michael White

MPhil, Rangeland Ecology, 2002, The University of Queensland, Australia, Graduate Advisor: Dr. Max Shelton

BS, Animal Production and Rangeland Management, 1997, Debub University, Ethiopia (with Distinction)

RELEVANT SKILLS AND EXPERIENCE:

- Developed species distribution model of an invasive frog species in Hawaii using machine-learning algorithms (Random Forests, Support Vector
Machine, Classification Trees) in a project funded by NASA through its 
Applied Science program

- Developed models that predicted the occurrence of wildfire in western United 
  States in a project funded by Ecology Center at Utah State University
- Used image processing software (ENVI, ERDAS Imagine) to build species 
  distribution models from MODIS satellite imagery
- Used GIS software (ESRI ArcGIS, ArcInfo and other open source GIS 
  libraries) to map suitable habitats of organisms
- Developed scripts in IDL and Python to extract data from satellite imagery 
  and GIS layers necessary to build spatial models
- Used several data-mining algorithms (Random Forests, Boosted Classification 
  Trees, Classification Trees, Support Vector Machine) to extract data necessary 
  to build spatial models
- Used traditional statistical models (Generalized Additive Models, Generalized 
  Linear Models, Linear Discriminant Analysis) to predict and map species 
  distributions in R statistical computing environment
- Used R, SAS, SPSS to analyze data and produce reports
- Determined the primary productivity and rainfall use efficiency of 
  rehabilitated grasslands following coal mining in central Queensland, 
  Australia
- Demonstrated teaching/training skills
WORK/LEADERSHIP HISTORY:

- Research Assistant at Utah State University (August 2004-October 2009)
- Research Assistant at The University of Queensland, Australia (September 1999 – September 2001)
- Teaching Assistant at Utah State University, Department of Watershed Sciences (Spring semester 2006)
- Lecturer at University of Asmara, Eritrea (September 2001-July 2004)
- Graduate Teaching Assistant at University of Asmara, Eritrea (November 1997- August 1999)
- Served as student volunteer at the 2009 US-IALE 24th Annual Landscape Ecology Symposium at Snowbird Utah
- Served as secretary for the Academic Commission of the Freshman Program of the University of Asmara, Eritrea (January 2002- July 2004)
- Served as member of Curriculum Review Committee for College of Agriculture at University of Asmara, Eritrea, 2002
- Served as member of the organizing committee for the International Symposium on the Role of Research on Nation Building held in Asmara, Eritrea, 2002
- Served as editor of the Newsletter for the College of Agriculture, University of Asmara, 2002
PRESENTATIONS:

- **Poster**: ‘Predicting the Invasion Potential of a Puerto Rican Frog in Hawaii using Remote Sensing Data’ at the 2009 US-IALE 24th Annual Landscape Ecology Symposium, Snowbird, Utah

- **Poster**: ‘Predicting the Invasion Potential of a Puerto Rican Frog in Hawaii’ at American Geophysical Union (AGU) 2008 Fall Meeting, San Francisco, California

- **Talk**: ‘Interactions Between Bark beetle Outbreak and Fire in Western United States Forests’ at American Geophysical Union (AGU) 2006 Fall Meeting, San Francisco, California

HONORS AND PROFESSIONAL MEMBERSHIP:

- Competitive Graduate Student Research Award Fellowship from NASA’s Applied Sciences Program through Intermountain Region Digital Image Archive Centre, 2007 ($20,000.00)

- Ecology Center Graduate Student Competitive Research and Stipend Award, Utah State University, 2006 ($8,200.00)

- Member of Intentional Association for Landscape Ecology

- Member of American Geophysical Union (AGU)

- Member of American Society for Photogrammetry and Remote Sensing (ASPRS)
PUBLISHED PAPERS:


