MODELING BURN PROBABILITY: A MAXENT APPROACH TO ESTIMATING CALIFORNIA’S WILDFIRE POTENTAL

by

Andrew Charles Thomason

A Thesis Presented to the
FACULTY OF THE USC GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
MASTER OF SCIENCE
(GEOGRAPHIC INFORMATION SCIENCE AND TECHNOLOGY)

August 2015

DEDICATION

For Ricardo.

ACKNOWLEDGMENTS

I would like to extend my gratitude to my thesis committee members, Dr. Darren Ruddell, Dr. Travis Longcore, and Dr. Karen Kemp, for their expertise on various subjects and overall contribution to my thesis work. Special thanks to my writing instructor Mandy for providing invaluable feedback throughout the draft and final stages of the document.

I want to thank my siblings, Ryan, Danielle, and Nicolle, for their support and encouragement throughout the course of the graduate program. I would also like to thank my parents for inspiring me day in and day out. I would not be the person I am today without your support and guidance.

Last but not least, I want to thank my best friend, Carolin – thank you for everything.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Increased wildfire activity throughout California over the past decade demands greater research on wildfire management approaches. Understanding natural, as well as human landscape characteristics that explain spatial patterns of wildfire potential can be used to complement traditional wildfire management approaches, such as fire suppression, by identifying high risk areas. In this study, California's wildfire potential was statistically modeled using wildfire observations from a 30-year period (1984 to 2013) and a wide variety of environmental variables. Locations of burned wildland habitat encountered between 1984 and 2013 were related to ignition sources, climate conditions, topography, and vegetation to estimate the probability of wildfire for regions of California exclusive of past wildfire occurrences. Twenty-nine variables were considered in building the wildfire probability model to determine which factors best indicate environmental susceptibility to wildfires. Two additional models, historic (1984–1988) and recent (2009–2013), were created to assess changes of wildfire probability across California over time.

Results of the long-term wildfire probability model display a heterogeneous distribution of wildfire probability across the state. Comparison between recent and historic wildfire probability values demonstrates fluctuations in wildfire potential near coastal and forested areas. Wildfire probability maps depicting the likelihood of wildfire in California can aid land as well as disaster management activities and can enhance the safety of firefighters and the public, and minimize wildland and property damages.

CHAPTER ONE: INTRODUCTION

With California experiencing one of the most severe droughts in over a decade, the average size and extent of wildfires has increased dramatically in various regions throughout the state, threatening significant wildland habitat, people, and property. In 2014 alone, the California Department of Forestry and Fire Protection (CalFire 2014a) reported 5,620 wildfires, over 630,000 acres burned, and an estimated $184.02 million in damages due to an ambush of large, devastating wildfires across the state. Such increases in wildfire potential across the state are directly related to increases in suitable habitat, or optimal conditions, for wildfire ignitions and spread of wildfire.

Environmental conditions unique to California, such as climate, topography, population density, and vegetation diversity, make regions throughout the state highly susceptible to wildfire occurrence. In recent years, wildfire activity in California and the western United States has increased dramatically, with higher large wildfire frequency, longer wildfire duration, and longer wildfire seasons, as environmental factors become more favorable (Westerling et al. 2006; Parisien and Moritz 2009; Parisien et al. 2012; Burke 2012). Although relationships between wildfire and environmental conditions have been studied extensively, modeling the distribution of wildfire probability, or the probability an area is likely to burn, remains a work in progress (Parisien et al. 2012). The purpose of this study is to use wildfire observations from a 30-year period, a wide variety of environmental variables, and species distribution techniques to model the spatial distribution of long-term (1984 to 2013), historic (1984 to 1988), and recent (2009 to 2013) wildfire probability and identify environmental influences of wildfire occurrence.

1.1. Wildfire in California

Wildfire is virtually inevitable in much of California due to its unique climate, availability of dry fuel, and population density. California's climate is characterized as a Mediterranean climate consisting of dry winds in the fall season, followed by limited winter precipitation and dry, extensive summers (Keeley 2006). Further, a Mediterranean climate is considered "fire adaptive", especially for specific fire regimes, which refer to general patterns of natural fires in a specific ecosystem and are characterized by relationships between plants and fire, the intensity and severity of fire, and the temporal relationship between fire and vegetation. Specifically, current fire regimes across California are heterogeneous; particularly due to differences in forested and non-forested habitats. However, fire regimes also differ from region-to-region due to variations in temperature, seasonal patterns of temperature and precipitation (Keeley and Syphard 2015).

California's historical wildfire regimes have been described as periodic and catastrophic; however, due to changes in land use and climate, wildfires in California have increased dramatically in recent years (Westerling 2006; Barros et al. 2013; Burke 2012). Specifically, nine of the top twenty largest wildfires in California occurred within the last decade and in various counties across the state (2004-Present; Table 1). However, these patterns of increasing wildfire activity are not universal, as some non-forested regions have not experienced increases in fire activity (Baker 2013; Keeley and Syphard 2015). Additionally, while the extent and severity of wildfires have increased in forested regions, recent research suggests there is no increasing trend in the size of mega-fires, or wildfires exceeding 50,000 hectares (Keeley 2014).

Table 1. Details on nine of the state's 20 largest wildfires (by acreage, since 1932); all occurring over the last decade.

Rank	Fire Name (cause)	Date	County	Acres Burned
2	Rush (Lightning)	August 2012	Lassen	271,911 CA
3	Rim (Human related)	August 2013	Tuolumne	257,314
4	Zaca (Human related)	July 2007	Santa Barbara	240,207
6	Witch (Power lines)	October 2007	San Diego	197,990
7	Klamath Theater Complex (Lightning)	June 2008	Siskiyou	192,038
10	Basin Complex (Lightning)	June 2008	Monterey	162,818
11	Day Fire (Human related)	September 2006	Ventura	162,702
12	Station Fire (Human related)	August 2009	Los Angeles	160,557
16	Happy Camp Complex (Lightning)	August 2014	Siskiyou	132,833

Source: Data from CalFire (2014b)

In recent years, both climate and human activity have played an important role in wildfire activity in California. Since 2012, California has been experiencing the most severe drought conditions in the last century (Griffin and Anchukaitis 2014). Based on the most recent Palmer Drought Severity Index measure for 2014, the state of California is currently classified as "extreme drought" (see Figure 1; NOAA 2015a). These more recent changes in drought conditions across the state play an important role in wildfire activity by increasing the amount of dry fuel available for wildfire, length of fire season, and severity of ignited fires in forested habitats (Keeley and Syphard 2015). More importantly, as California's population continues to grow, more people will continue to move toward wildland areas, altering land use, and increasing the risk of both fire ignition and subsequent damage to life, property, and natural resources (Snider et al. 2007; Pincetl et al. 2008; Burke 2012). Due to current drought conditions and human influence on wildland areas, wildfire continues to threaten many regions of California, thus increasing the importance of wildfire management.

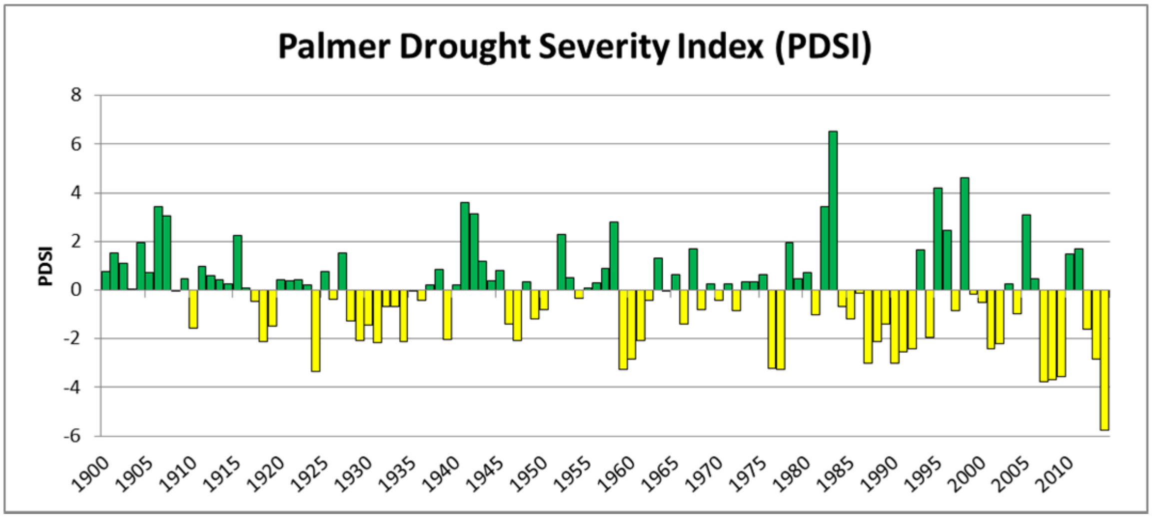

Figure 1. Palmer Drought Severity Index for California between 1900 and 2014.
Source: **NOAA (2015a)**

1.2. Wildfire Management

Wildfire management is a complex process that aims to balance two primary objectives: first, restoring and maintaining fire as an essential natural disturbance and second, minimizing the risk that wildfire poses to people and the surrounding environment (Zaksek and Arvai 2004). To aid in this process, environmental risk and resource managers look toward Geographic Information Systems (GIS), spatial statistics, and habitat suitability modeling to better understand the spatial and temporal distribution of wildfire to support disaster management activities, and minimize the risk to human, property, and the environment. For example, Figure 2 illustrates wildfire threat across California. Such outputs can be used to estimate the potential for impacts on various assets and values susceptible to fire, whereby impacts are more likely to occur at locations with higher threat classes. Understanding the complexity of ignition sources, environmental influences, and characteristics of wildfire can help forecast future wildfire habitat and likelihood of a fire event occurring. Several tools are available to aid in wildfire management, such as species distribution techniques.

Figure 2. Wildfire threat across California. ***Source:*** **Department of Forestry and Fire Prevention (2004)**

Species distribution modeling is an important tool that can provide information pertaining to the potential distribution of species in a given geographic space. Relating known occurrences of species to landscape, climate, and geographic variables, using statistical models can help discover ecological characteristics and predict geographic occurrences at a greater extent (Peterson 2006; Phillips et al. 2006). Much work has focused on modeling the probability of natural wildfire occurrences in a given location using a wide variety of environmental variables (Gedalof et al. 2005; Parisien and Moritz 2009; Parisien et al. 2006, 2012; Krawchuck et al. 2009; Little et al. 2009; Bradstock 2010; Ziesler 2013). More recently, human activity has played a crucial role in wildfire occurrence; therefore, realistic wildfire occurrence estimations require that spatial models incorporate anthropogenic drivers (Cardille et al. 2001; Stocks et al. 2002; Stephens 2005; Syphard et al. 2007, 2008, 2009; Parisien et al. 2012). While species distribution techniques are important in identifying high risk areas, resource maps require ongoing updates.

1.3. Motivation of the Study

While the relationship between environment and wildfire occurrences has been studied in great detail, modeling the distribution of wildfire probability is continuous, mainly due to ongoing wildfire observations across the state, and changes in climate and land use. As climate and land use continues to change in the future, wildfires across California may increase in size and frequency, thus having greater consequences (Westerling et al. 2006, 2011; Westerling and Bryant 2008; Bowman et al. 2009; Krawchuk et al. 2009). Hence, there is a need to continuously update high-resolution maps depicting wildfire probability using recent wildfire occurrence data and environmental variables.

1.4. Research Questions and Objectives

There were three questions in this research. First, what is the spatial distribution of long-term wildfire probability? Second, how do environmental variables affect wildfire probability? Third, where has wildfire distribution and probability changed across California over time?

In order to answer each of these questions, four objectives were set for this thesis. This thesis aimed to: (1) relate random wildfire observations with environmental variables using Maxent software to assess the spatial distribution of long-term (1984 to 2013), recent (2009 to

2013), and historic (1984 to 1988) wildfire probability in California; (2) examine the influence of environmental factors on wildfire probability; (3) assess the utility and robustness of modeling wildfire probability using Maxent software; and (4) compare recent and historic wildfire probability values to evaluate changes in wildfire risk in California over time. Several techniques were utilized in order to accomplish the study's research objectives.

This study utilized wildfire occurrence information, a wide variety of environmental variables, and species distribution techniques in order to model the distribution of long-term wildfire probability in California. Specifically, locations of burned wildland habitat were tied to ignition sources, climate conditions, topography, and vegetation to estimate the probability of wildfire for regions of California exclusive of wildfire occurrence information using Maxent software. Twenty-nine independent variables, representing environmental conditions across the state, were considered for building three wildfire probability models utilized in this study: long-term (1984 to 2013), recent (2009 to 2013), and historic (1984 to 1988).

1.5. Organization of Thesis

This thesis is composed of five chapters, the first being this introductory chapter. Chapter 1 introduced and discussed key information pertinent to the remainder of the thesis, such as background information on wildfire activity in California, wildfire management, and the motivation and overall objectives of the thesis.

Chapter 2 is a review of existing literature examining wildfire processes and wildfire activity in California. In addition, this chapter discusses geospatial technologies for modeling the distribution of wildfire probability and wildfire behavior.

Chapter 3 explains the methodology of the thesis and describes the study area, data, exploratory methodology for choosing input environmental variables for the models, and species distribution modeling techniques. This chapter also details data processing in ArcMap 10.3 and Diva-GIS and requirements for using Maxent software, in addition to methods for analyzing the accuracy and precision of the model outputs.

Chapter 4 presents the long-term, historic and recent wildfire probability maps and additional outputs produced by Maxent software. Details describing the influence of explanatory variables on spatial variability, accuracy and robustness of the models, and comparison between the historic and recent models are discussed in this chapter.

Chapter 5 discusses the overall findings, usefulness of species distribution techniques for determining the patterns of long-term wildfire probability in California, and answers the study's three research questions. This chapter also assesses the relationships between wildfire probability and environmental variables and evaluates changes in wildfire probability in California over time. Results presented in this chapter contribute to the current understanding of long-term, recent and past wildfire probability.

CHAPTER TWO: LITERATURE REVIEW

This chapter discusses wildfire processes and wildfire activity in California in addition to existing geospatial modeling techniques. Explanations of these topics present background information, provide relevant literature, and serve as a basis for the remainder of the thesis.

2.1 Wildfire Processes

Wildfire is highly dependent on the combustion process, ignition source, climate conditions, topographical landscape, and availability of fuels. Wildfire occurs and is maintained as a function of the simultaneous presence of appropriate fuel, ignition agents, and conditions conducive to combustion and spread (i.e. fuel, oxygen, and heat; Cottrell 1989; Fuller 1991; Parisien and Moritz 2009). A fire ignites when fuel, coupled with sufficient oxygen, is exposed to a source of heat above the combustion level while sustaining a rate of rapid oxidation. Aside from ignition sources, appropriate fuel, oxygen, and heat are needed to maintain and spread wildfire across a landscape. Specifically, for fire to develop and spread, heat must be transferred to surrounding fuels in its directional path via convection, radiation, and/or conduction mechanisms. Such mechanisms of heat transfer contribute to the combustion process, depending partially on fuel distribution across a landscape, wind speed and direction, and the slope and aspects of terrain (Viegas 1998). While wildfire is dependent on fuel, oxygen, and heat, wildfire ignitions and wildfire spread is influenced by variations in environmental conditions.

2.2 Wildfire Ignitions and Predicting Variables

Wildfires are ignited by natural or human sources. As shown in Table 2, displaying wildfire ignition statistics for the entire United States between January 2000 and December 2008, lightning is the most common, natural ignition source of wildfires on federal land, causing 45% of reported wildfires and nearly 80% of total area burned in the United States (Prestemon et al. 2013). In contrast, human-caused wildfires, directly or indirectly ignited by a campfire, smoking, fire use, arson, equipment, roads, or juveniles/children, comprise a smaller count and annual area burned. However, ignition sources vary greatly by region, due to variations in environmental variables, such as regional precipitation patterns and dominant vegetation. These climate variables are often used to explain why wildfire ignitions vary across a landscape due to variations in weather and climate, vegetation, geology and topography (Prestemon et al. 2013). For example, dry and warm conditions in southern California promote low fuel moisture and in turn increase suitable conditions for wildfire ignition. Understanding the relationship between wildfire and suitable environmental conditions are important for understanding wildfire probability and behavior in a given area of interest.

Table 2. Fires causes, reported average annual ignitions, reported average annual area burned, and percentage shares of fires by causes between January 2000 and December 2008.

Cause	Average annual ignitions reported	Average annual area burned reported (acres)	Percentage share of reported ignitions (%)	Percentage of share of reported area burned (%)
Natural/Lightning	10,874	5,496,235	45.34	79.90
Campfire	1,964	179,338	8.19	2.61
Smoking	418	22,387	1.74	0.33
Fire Use/Debris Burning	1,538	100,971	6.41	1.47
Incendiary/Arson	2,969	268,962	12.38	3.91
Equipment (Use)	1,338	246,804	5.58	3.59
Railroad	117	14,193	0.49	0.21
Juveniles/Children	1,063	20,464	4.43	0.30
Miscellaneous and Unknown	3,704	529,313	15.44	7.69

Source: Data from Prestemon et al. (2013).

Wildfire is an abiotic physical process that is highly dependent and regulated by its surrounding environment and is therefore a byproduct of suitable environmental conditions, or conditions that are pertinent to wildfire. Due to this, fire frequency and severity fluctuate enormously among different biomes. Recent work has aimed to describe the spatial distribution and environmental requirements of wildfire and explain observed ignition patterns over space and time (Parisien and Moritz 2009; Parisien et al. 2012; Prestemon et al. 2013). As previously noted, an ignition occurs with adequate fuel, oxygen, and heat. Moisture content in available fuel is highly dependent on climate variables such as temperature, solar radiation, humidity, and precipitation, of which precipitation is the most important moisture determinant (Prestemon et al. 2013). Studies have historically related wildfire distribution and ignitions to daily weather conditions, fuel moistures, and fire behavior indices (Haines et al. 1983; Martell et al. 1987; Andrews et al. 2003; Presiler et al. 2004, 2009; Balshi et al. 2009; Finney et al. 2011). Conversely, additional studies have applied monthly or longer-term climate averages of precipitation and temperature, among other weather-derived variables, to estimate the historic distribution of wildfire ignitions. Regardless of differences in temporal scales in previous studies, the frequency of wildfire ignitions is much higher under warmer and drier conditions (Prestemon et al. 2013). Specifically, low precipitation and warmer temperatures are indicative of dry conditions and low fuel moisture, enhancing suitable ignition conditions. Despite climate playing a significant role, wildfire ignitions are also highly dependent on available fuel.

Fuel available for wildfire consists of any substance, or combustible material, that will burn or ignite and is characterized by its moisture contents, size and shape, quantity, and the arrangement across a landscape. Fuel type is categorized as either subsurface, surface, or aerial fuel, while size of fuels can be classified as light, medium, or heavy. Subsurface fuel includes roots, peat, and other decomposed organic matter below ground surface. Fires which burn organic matter in the soil are considered ground fires. On the other hand, surface fuel consists of combustible material up to one meter above of the ground surface and consists of brush, leaves, small trees, among other materials. When ignited, these surface fires allow aerial fuel above to ignited. Aerial fuel includes brush greater than 1 meter above ground surface and once ignited, is referred to as crown fire (Scott and Reinhart 2001).

Fuel can also be categorized by size and helps determine the type of heat transfer, which influences the forward spread of wildfire (Cottrell 1989). Light fuels include short grasses or light brush up to 2 feet, thus burning rapidly. Medium fuels consist of brush up to 6 feet which tend to cause slow but moderate to very high intensity burning, while heavy fuels consist of brush greater than 6 feet that produce low to moderate wildfire spread at a high intensity burn (Randall and Duryea 2003). As described by Prestemon et al. (2013), limited studies have incorporated moisture patterns, coupled with fuel types and vegetation patterns, to describe the spatial distribution of wildfire. More so, studies have focused on exploratory variables that characterize fuel and vegetation types as predictors for understanding wildfire ignitions (Cardille et al. 2001; Syphard et al. 2008, 2009; Littell et al. 2009; Parisien and Moritz 2009; Westerling et al. 2011; Parisien et al. 2012). Vegetation and fuel type variables are better predictors of wildfire ignitions when the temporal distributions of climate variables are monthly or longer term (Prestemon et al. 2013). While climate and fuel variables are direct indicators of ignitions and thus wildfire occurrences, topography indirectly influences wildfires by influencing the moisture content and distribution of fuels, as well as temperature (Carmo et al. 2009).

Wildfire distribution, as well as behavior, is affected by topography of an environment. Specifically, topographic variables such as slope, aspect, and elevation affect incident solar radiation, drying rates of moisture loss from fuels, vegetation type, and climate. Aspect, or direction of the slope, affects solar radiation an area receives, affecting moisture content of fuels and types of vegetation. Based on the positioning of the sun, south- and west- facing slopes tend to have less vegetation and thus less fuel. South-facing slopes receive greater incoming solar radiation and tend to be warmer, allowing vegetation to lose its fuel moisture more quickly, creating suitable conditions for wildfire. Conversely, north-facing slopes tend to be shaded and cooler, delaying the drying process of fuels and reducing suitable conditions for wildfire ignitions (Randall and Duryea 2003).

Elevation in complex terrain indirectly affects wildfire by influencing the amount, timing, and location of precipitation, as well as temperatures and wind direction (Fuller 1991; Fitzgerald). In lower elevations, available fuel tends to be drier and more susceptible to combustion due to lower precipitation and warmer temperatures. The opposite tends to be true for areas of greater elevation (Crimmins and Comrie 2004). Additionally, cloud-to-ground lightning strikes become more prevalent at higher elevations, increasing the risk of ignition (Dissing and Verbyla 2003). Multiple studies to date have included topographic variables, coupled with monthly or long-term climate data and vegetation types, to study the spatial distribution of wildfire (Cardille et al. 2001; Parisien and Moritz 2009; Westerling et al. 2011; Parisien et al. 2012; Paritsis et al. 2013). Specifically, vegetation is known to be correlated with topography, and exclusion of either variable in wildfire empirical models can lead to ambiguity in results. All said, predictor variables, such as climate, vegetation, and topography, in addition to ignition sources are important for understanding wildfire behavior; however, these variables vary by habitat, especially in California.

2.3 Wildfire Patterns in California

Due to California's unique Mediterranean-climate, along with vegetation, topography, and population density, wildfire is inevitable. Differences in these predictors of wildfire vary across the state, thus altering fire regimes. California's climate is characterized by hot, dry summers and cool, moist winters. These conditions enhance fuel accumulation due to the slow decomposition of heavy vegetation, such as forested environments, increasing the severity of wildfires (Westerling et al. 2006). However, in non-forested habitats, such as foothills, variations in wildfire activity are more heavily influenced by effect of higher rainfall, increasing the amount of fuel in subsequent years (Dennison et al. 2008). Westerling et al. (2006) correlated an increase in forest wildfires with warmer spring and summer temperatures, limited precipitation in warmer months, reduced snow pack, or accumulation of snow, early spring snowmelt, and long summer fire seasons at middle and upper elevation ecosystems. Similarly, Keeley and Syphard (2015) concluded that spring and summer temperatures are important drivers of burned area in forest ecosytems. Additionally, multiple regions of California experience strong extremely dry down-sloped winds, known as Santa Ana winds (Keeley 2006; Yue et al. 2014). Such winds are important drivers of wildfire spread in various regions across the state, and particularly southern California (Keeley 2006). From high elevation basins in western North America, cool, dry Santa Ana winds flow downslope toward lower atmospheric pressures off the Pacific Coast, reducing moisture of fuels in its direct path and driving ignited fires (Westering et al. 2004; Moritz et al. 2010; Barros et al 2013). Such climate conditions, such as temperature, precipitation, and winds, significantly influences wildland habitat suitable for fire by affecting moisture content in vegetation, enhancing fuel availability.

Vegetation and availability of fuel greatly affect wildfire regimes in California. Specifically, northern California's landscape is dominated by forests consisting of mixed conifers and mixed evergreen hardwood, which tends to be higher in moisture content due to seasonal precipitation patterns. Conversely, southern California is comprised of drier chaparral and coastal sage scrub shrublands (Barros et al 2013). Differences in fire regimes are directly related to differences in vegetation and fuel type, in addition to climate, among other variables (Sommers, Coloff, and Conard 2011). For example, southern California's shrubland environment offers lower amounts of fuel available for wildfire occurrence; however the dry conditions, coupled with extreme winds, promote optimal conditions for the spread of ignited wildfires. While California's unique climate, topography, and vegetation offers suitable habitat for naturally occurring wildfires, human population, among other variables, have been directly related to increases in wildfire ignitions (Syphard et al. 2007).

Human activities have been linked and are known to promote fire ignitions across much of the state (Barros et al. 2013). Areas with high population densities increase the likelihood of wildfire (Cardille et al. 2001; Syphard et al. 2007, 2009). As the population in California increases in the future, people will move to less populated areas in close proximity to the natural habitat, known as the wildland urban interface (WUI). This increases the density of people and enhances the risk of human-ignited fires (Keeley et al. 1999; Cardille et al. 2001; Snider et al. 2007; Pincetl et al. 2008; Syphard et al. 2007, 2009). Lastly, wildfire is most often managed immediately upon ignition in order to protect human lives and property at risk. Constant suppression of wildfire in forested fire regimes allows fuel to accumulate significantly, thus further increasing the likelihood of future severe fires (Stephens 2009; Barros et al. 2013; Keeley and Syphard 2015). However, recent research suggests that this is not true in chaparral and coastal sage ecosystems. Due to California's changing climate, urban sprawl, and/or abundance of vegetation due to decades of fire suppression, wildfire occurrences in various regions across the state have changed significantly over time. With rising wildfire ignitions over California, there is a greater need to utilize geospatial technologies and statistics for modeling the spatial distribution of wildfire probability to protect wildland habitat, humans, and property.

2.4 Geospatial Techniques for Modeling Wildfire

Geographic data, coupled with geospatial technologies, have been used in previous studies to understand wildfire behavior and determine the distribution of wildfire probability. Specifically, these empirical models use past wildfire occurrence and environmental data to either determine: (1) the growth and distribution of wildfire using fire simulation models; or (2) determine distributions of wildfire probability through species distribution modeling techniques.

Early fire simulation models implemented semi-empirical equations and datasets to determine wildfire characteristics of interest such as fire intensity, rate, and length (Rothermel 1972; FCFDG 1992; Noble et al. 1980; Cheney et al. 1993). More recently, wildfire growth models such as FARSITE and Prometheus have applied environmental and wildfire occurrence relationships to simulate the spread and behavior of wildfire and estimate fire use for resource benefit across a landscape (Finney 1998; Tymstra 2009). FSim, another wildfire behavior model, pairs with FARSITE to simulate fire ignition, growth, and suppression and can be used to model burn probability (Finney 2011). Lastly, FlamMap5, computes fire behavior characteristics such as rate of spread, flame length, and fire line intensity (Finney 2006). As shown in Figure 3, wildfire simulation outputs depict wildfire growth under specific environmental input conditions. Wildfire behavior maps aid in wildfire management activities, such as fire suppression and fuel breaks, increase safety to firefighters and the public, and minimize damage. Although wildfire simulation models are key to understanding the distribution of wildfire behavior, probabilistic modeling techniques, specifically species distribution software, helps to estimate potential wildfire distributions over a landscape.

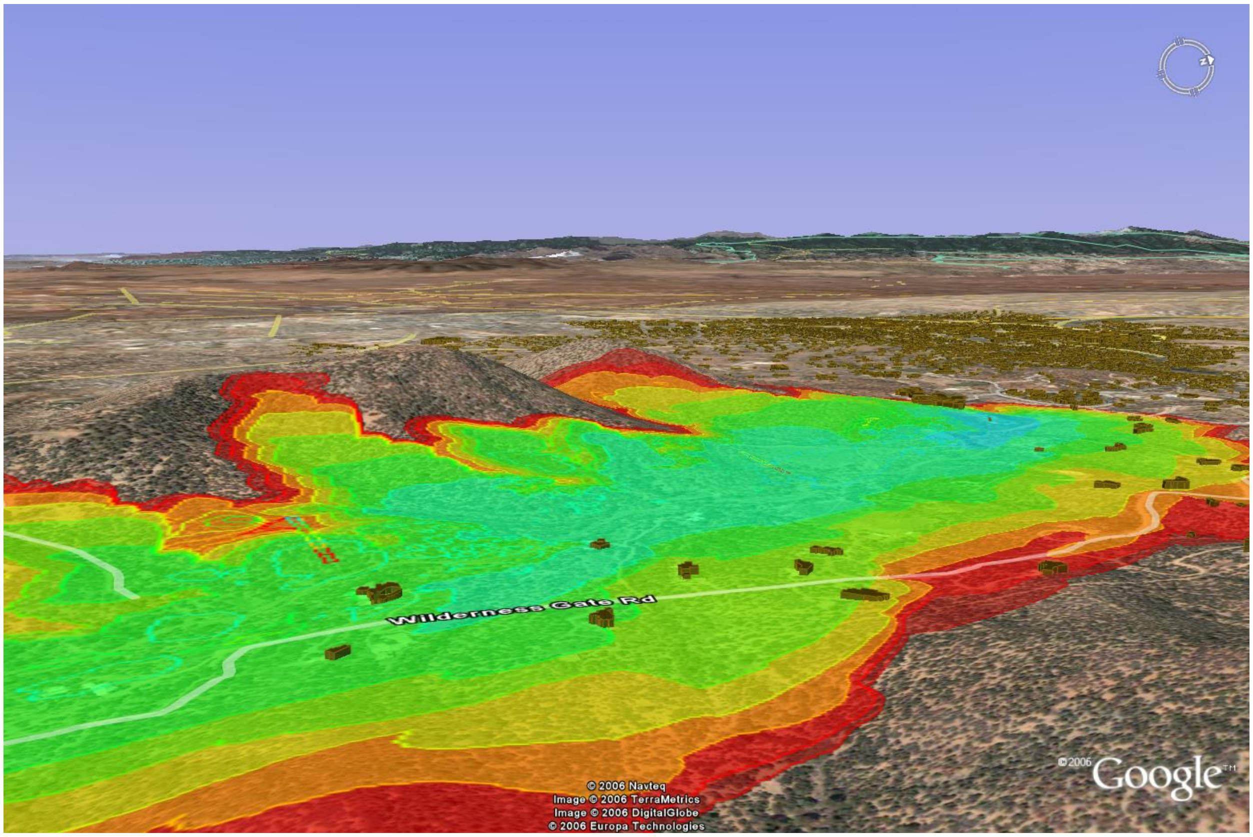

Figure 3. Image depicts a wildfire growth simulation output using FARSITE modeling software. Green area indicates first hour of burn, while red indicates the seventh hour of burn. *Source:* Redfish Group

Species distribution modeling refers to the use of species observations coupled with patterns of biodiversity to predict the potential distribution of a species' habitat. To date, much work has focused on modeling the probability of wildfire occurrences and ignitions in a given location using species distribution techniques (Parisien and Moritz 2009; Syphard et al. 2012, 2013; Parisien et al. 2012; Bar Massada et al. 2013; Paritsis et al. 2013; Peters et al. 2013; Syphard and Keeley 2015). Recently, Maxent software has been utilized in various wildfire applications, such as estimating burn probability, ignition probability, and probability of wildfire risk to property. For example, Parisien et al. (2012) implemented Maxent software for modeling the distribution of wildfire in Western United States. Results of the study (see Figure 4), proved Maxent software to be successful in modeling wildfire probability using wildfire occurrence data (1984-2008) and specific environmental variables (ignitions, climate, vegetation, and topography). Such techniques are an effective method for modeling the distribution of wildfire (Ferrarinil, 2012). Elith et al. (2006) demonstrated that Maxent performed better than other established niche-modeling methods, especially in cases with presence-only data and small population sizes (Syphard and Keeley 2015). Among the software developed and implemented for modeling the likelihood of wildfire ignitions, Maxent has shown to perform better (i.e. AUC) than any other algorithms (Ferrarinil 2012).

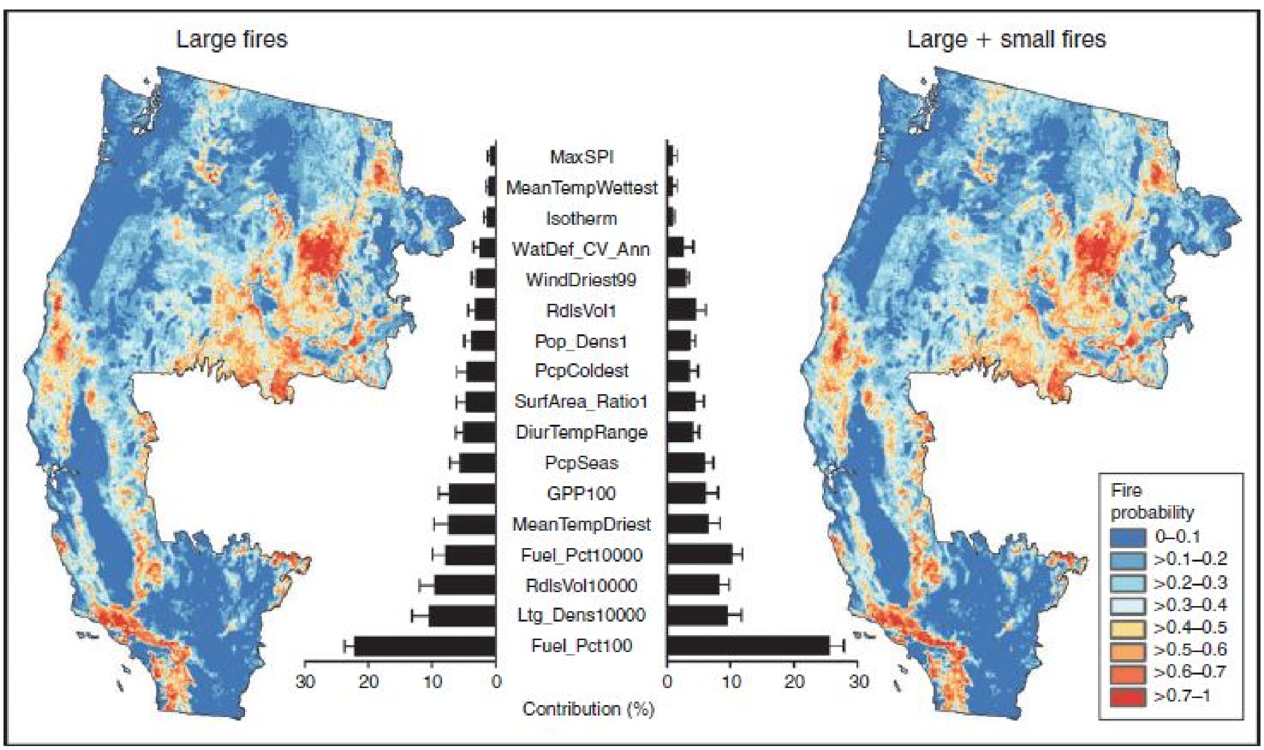

Figure 4. Image depicts wildfire probability output of Maxent software. Warmer colors indicate regions with high probability of suitable habitat, while cooler colors suggest lower probability of suitable habitat. ***Source:*** **Parisien et al. (2012)**

Understanding wildfire processes, wildfire ignitions sources and predicting variables, such as climate, topography, and vegetation, and geospatial technologies are important for modeling the spatial distribution of wildfire potential across California. Chapter 2 provided a review of existing literature and key concepts and serves as a basis for the remainder of the thesis. Following this chapter, methodologies and data sources used in this study are discussed in detail.

CHAPTER THREE: METHODS AND DATA SOURCES

This study aims to utilize explanatory variables, such as ignitions, climate, topography and vegetation, as well as wildfire observations in order to estimate the likelihood of wildfire probability in California. To accomplish this, several key tasks are required: data collection, preparation of environmental layers and wildfire occurrences, exploratory analyses, wildfire probability model design and execution, and results analysis (Figure 5). In this chapter, methods and data sources utilized for modeling the distribution of wildfire at three temporal scales (long-term, recent, and historic) are discussed in detail.

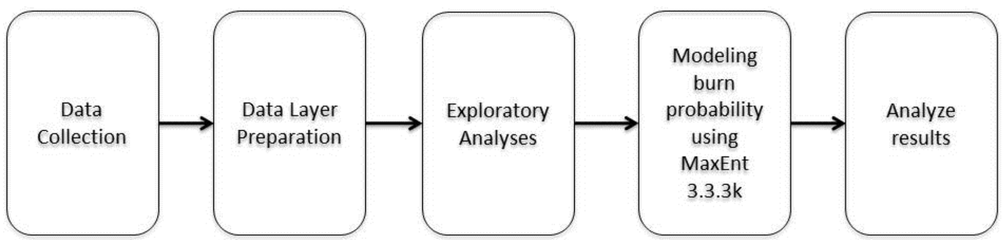

Figure 5. Thesis workflow.

3.1 Study Area

The study area corresponds to the entire state of California, comprising roughly 423,970 km^2 (Figure 6). California is located on the West Coast of the United States and is bordered by Oregon to the north, Nevada to the east, Arizona to the southeast and the US-Mexico border to the south. Based on climate, continental position, elevation, vegetation characteristics, and topographic features, California can be broken down into nine bioregions: North Coast, Central Coast, South Coast, Klamath Mountains, Southern Cascades, Northeastern Plateau, Sierra Nevada, Central Valley and Southeastern Deserts (Barros et al. 2013). Much of the state is characterized by hot and dry summers alternating with cool and wet winters and referred to as a Mediterranean climate. Overall, annual average precipitation is greatest in northern California and is heaviest during winter months. Additionally, topography plays an important role in California climate, as temperature decreases and precipitation increases with elevation. High mountains in the State, such as the Sierra Nevada, are affected by 'alpine climate' with snow in winter and mild to moderate heat in summer. The east side of these high mountains undergoes arid conditions due to the rain shadow effect (i.e. Death Valley). While climate in California varies upon ecoregion, vegetation cover and thus availability of fuel also significantly changes.

Vegetation cover, and thus fuel availability, is highly dependent upon the topography and climate in much of the state. Southern California is dominated by dry chaparral, or shrubland habitat, which experiences high fire frequency between late March and November. Northern California consists of mixed forest, grassland and shrubland vegetation. Fire season runs from mid-May through October in Northern California. Wildfire regime in this portion of the state is highly variable and is reflected by fire return interval and fire severity (Parisien et al. 2012; Barros et al. 2013). Wildfire regimes in California are highly heterogeneous, due mainly to variations in environmental conditions, such as climate and vegetation type, across the state.

California's unique climate, topography, and vegetation cover provide suitable conditions for wildfire occurrence across the entire state, with exception of barren and agricultural regions. As shown in Figure 6, these factors vary significantly across the state and in part cause changes in the distribution of wildfire probability. Wildfire data and environmental variables utilized in this thesis, such as those discussed in this section, are used to model the spatial distribution of wildfire and detailed in subsequent sections.

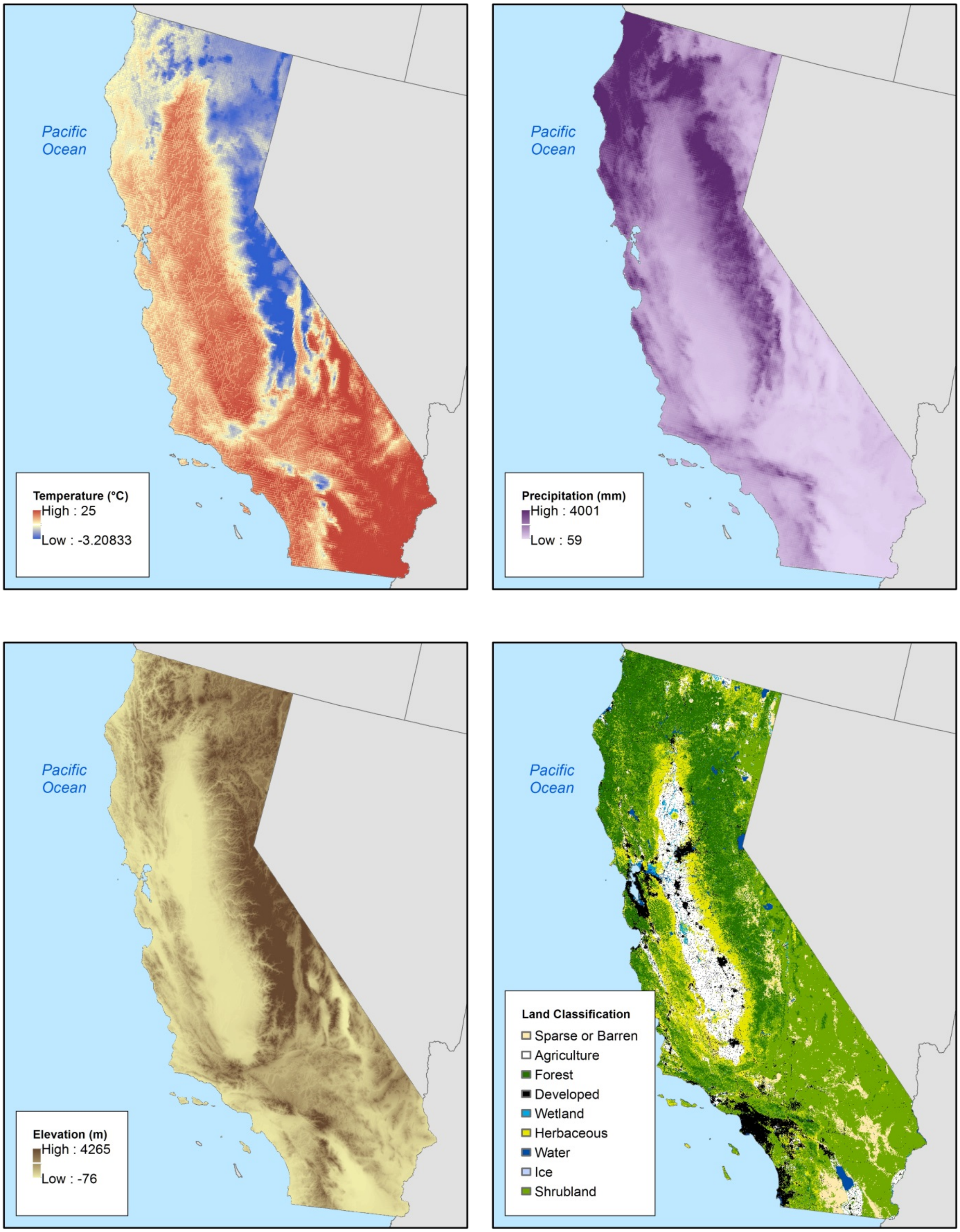

Figure 6. The California study area showing mean annual temperature, mean annual precipitation, elevation, and land cover. ***Sources:*** **Gesch 1996; PRISM 2004; Jin et al. 2013**

3.2 Study Overview

For modeling the distribution of wildfire probability in California, multiple models (long-term [1984-2013], recent [2009-2013], and historic [1984-1988]) were created by correlating wildfire observations between 1984 and 2013 (dependent variable) to explanatory variables, such as ignition sources, climate, topography, and vegetation (independent variables; Table 3). Modeling techniques used in this thesis were influenced by Parisien et al.'s (2012) successful study, effectively modeling the spatial distribution of wildfire probability using Maxent software for the western United States. Maxent software was used to build wildfire probability models.

As detailed in Section 3.3.2.2., mean monthly maximum and minimum temperature and mean monthly precipitation were utilized to extract nineteen bioclimate variables using DIVA-GIS software for use as climate variables in each of the models (long-term [1984-2013], recent [2009-2013], and historic [1984-1988]) (Hijmans et al. 2012). Ignitions, topography, and vegetation variables represent a single dataset (i.e. road density, distance to roads, etc.).

Parisien et al. (2011, 2012) suggest a 'moving-window' approach to represent neighborhood conditions of ignition, topography, and vegetation, or averaged results for spatial scales of interest, in order to model wildfire probability. Specifically, environmental variables at fine spatial scales contain little information about factors precluding to ignitions and spread of fire. Therefore, wildfire occurrence locations can be considered to have conditions somewhat suitable to fire activity within its "neighborhood". While enlarging the spatial scale of environmental variables leads to the inclusion of areas less suitable for wildfire occurrence and helps refine wildfire-environmental relationships, fine-scale information can be lost (Parisien et al. 2011). Therefore, Parisien et al. (2011) recommends evaluating wildfire-environmental relationships at more than one spatial scale. Using *Block Statistics* within the ArcMap 10.3

Spatial Analyst toolbox, three spatial scales were computed for each variable (1, 25, and 100 km^2). All data inputs were processed with ArcMap 10.3. As detailed in Section 3.3.2, all data were converted to use an Albers NAD 1983 equal-area projection and converted to a 1.0 km^2 resolution.

This thesis builds three wildfire probability models to assess the significance of exploratory variables and portrays the spatial distribution of wildfire probability. Specifically, a long-term, or multi-decadal model used wildfire occurrences data for the full temporal scale (1984 to 2013) and utilized selected variables chosen following exploratory analysis. To assess the change in distribution of wildfire probability across California over time, in addition to environmental variables which promote wildfire, two additional models were created. Data collected between 2009 and 2013 was utilized to estimate recent wildfire potential across the state. Conversely, a historic model capturing wildfire observations, climate, and land cover data between 1984 and 1988 was created to assess the California's historic wildfire potential. As shown in Table 3, twenty-nine variables were considered for modeling wildfire probability in California. As discussed in subsequent sections, data sources and yearly averages depended on temporal scale of each model considered.

Table 3. Exploratory variables considered for analysis and their description. Fewer variables were chosen for the final models.

Category	Input Name	Source	Description
Ignitions	Lgt_Dens[s]	NOAA	Annual density of lightning ignited wildfires (ignitions km^{-2} $year^{-1)}$
	Pop_Dens[s]	Gridded population of the world, v.3	Population density (people km^{-2})
	Rd_Dens[s]*	US Census Bureau	Road Density of primary or secondary roads (km km^{-2})
	Distrd_dens[s] *	US Census Bureau	Distance to primary or secondary roads (km^3 $person^{-1}$)
Bioclimate	Bio1	PRISM, DIVA-GIS	Annual mean temperature (°C)
	Bio2	PRISM, DIVA-GIS	Mean diurnal range (°C)
	Bio3	PRISM, DIVA-GIS	Isothermality
	Bio4	PRISM, DIVA-GIS	Temperature seasonality (°C)
	Bio5	PRISM, DIVA-GIS	Max Temperature of Warmest Month (°C)
	Bio6	PRISM, DIVA-GIS	Min Temperature of Coldest Month (°C)
	Bio7	PRISM, DIVA-GIS	Temperature Annual Range (°C)
	Bio8	PRISM, DIVA-GIS	Mean Temperature of Wettest Quarter (°C)
	Bio9	PRISM, DIVA-GIS	Mean Temperature of Driest Quarter (°C)
	Bio10	PRISM, DIVA-GIS	Mean Temperature of Warmest Quarter (°C)
	Bio11	PRISM, DIVA-GIS	Mean Temperature of Coldest Quarter (°C)
	Bio12	PRISM, DIVA-GIS	Annual Precipitation (mm)
	Bio13	PRISM, DIVA-GIS	Precipitation of Wettest Month (mm)
	Bio14	PRISM, DIVA-GIS	Precipitation of Driest Month (mm)
	Bio15	PRISM, DIVA-GIS	Precipitation Seasonality (mm)
	Bio16	PRISM, DIVA-GIS	Precipitation of Wettest Quarter (mm)
	Bio17	PRISM, DIVA-GIS	Precipitation of Driest Quarter (mm)
	Bio18	PRISM, DIVA-GIS	Precipitation of Warmest Quarter (mm)
	Bio19	PRISM, DIVA-GIS	Precipitation of Coldest Quarter (mm)
	Wind_cl[s] *	NREL	Wind class (categorical)
Topography	Elev[s]	DEM, USGS	Elevation above sea level (m)
	Aspect[s]	DEM, USGS	Slope Aspect (degrees)
	Slope[s]	DEM, USGS	Slope Angel (degrees)
Vegetation	Fuel[s]	USA Gap Analysis Land Cover	Fuel vs. Nonfuel (%)
	GPP[s] *	MODIS – MOD17A3	Gross primary productivity (kg C/m2)

[S]Denotes scale dependent variable. These variables were calculated at three spatial scales ([S]): 1, 25, and 100 km^2.

*Dataset not utilized in the historic wildfire probability model.

3.3 Data Sources

A number of data sources and types were utilized in this study. Specifically, to model the distribution of wildfire probability, Maxent software requires presence locations (dependent variable) and environmental variables (independent variables). As detailed in this section and later subsections, wildfire observations were utilized as presence locations, and ignition, climate, topography, and vegetation data as environmental variables.

3.3.1 Wildfire Data

3.3.1.1 Wildfire Presence Data Creation (Dependent Variable)

For modeling the distribution of wildfire potential in California, wildfire observations, or presence locations, are required. Wildfire locations were extracted from burned area perimeter data layers. Specifically, burned area polygons are available as part of the Monitoring Trends in Burn Severity (MTBS) project, conducted by the U.S. Geological Survey National Center for Earth Resources Observation and Science (EROS) and the USDA Forest Service Remote Sensing Applications Center (RSAC; Eidenshink et al. 2007). Wildfire occurrence source data were primarily provided by the Bureau of Land Management (BLM), Bureau of Reclamation (BR), Bureau of Indian Affairs (BIA), United States Fish and Wildlife Service (USFWS), National Park Service (NPS), and United States Forest Service (USFS). The original source data span the 1984-2013 temporal scale and cover the entire United States of America, with natural and human-related fires recorded in the database.

Wildfire polygons in the state of California were retained for the study (Figure 7). Due to inconsistencies in reporting throughout the state, small fires (<300 acres) were omitted from the

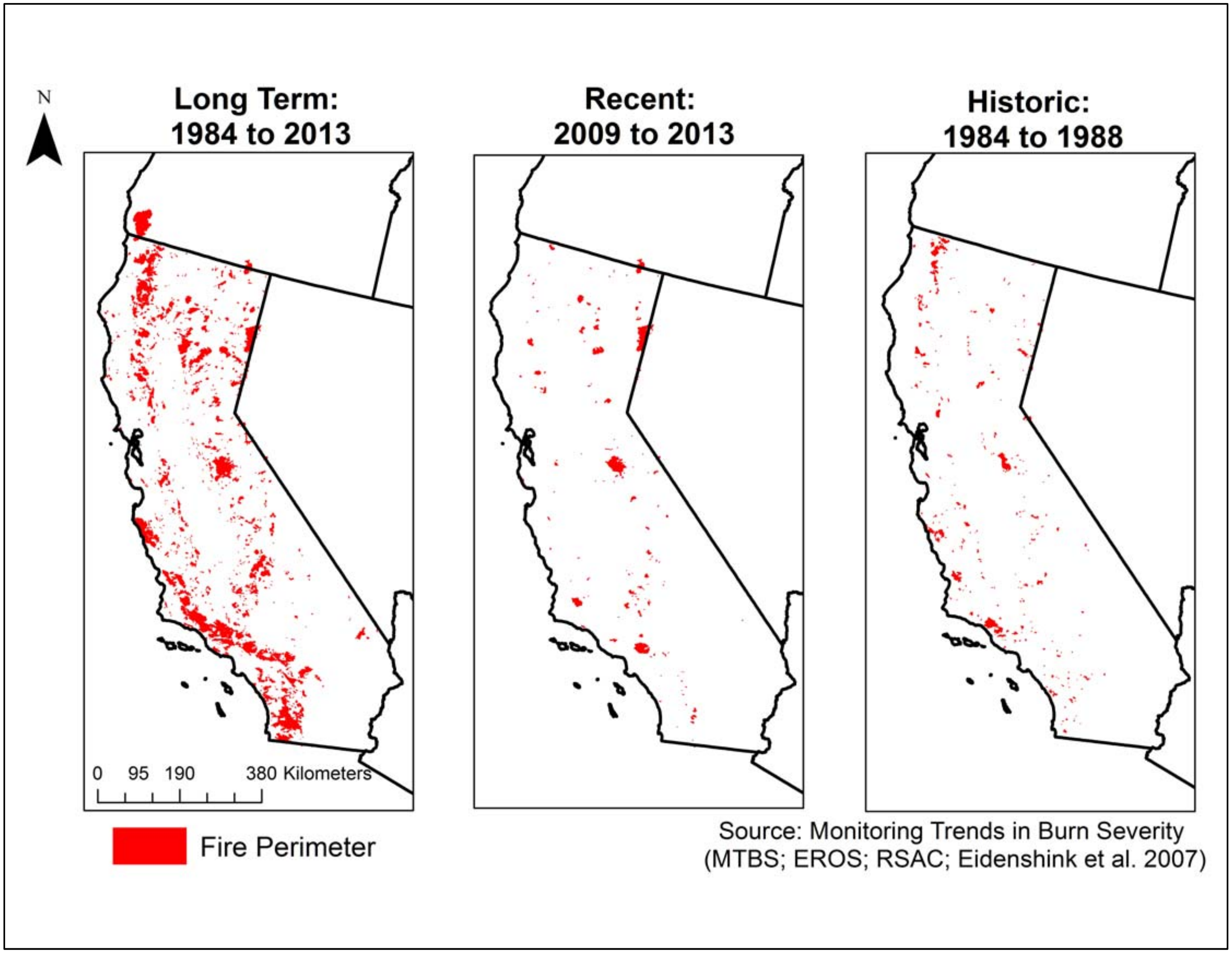

Figure 7. Wildfire polygons the studies three temporal scales: long-term (1984 to 2013), recent (2009-2013), and historic (1984 to 1988).

wildfire database. Using the *Create Random Points* tool in ArcMap 10.3, random points were distributed within the extent of California. The *Select by Location* tool was utilized to obtain random wildfire points which intersect burned area polygons for each period (long-term [1984 to 2013], recent [2009 to 2013], and historic [1984 to 1988]). An example of random wildfire locations from within a burned area polygon is displayed in Figure 8. Each point represents a location of past burned area and wildfire occurrence. Using these techniques, a total of 5,000 random points were obtained and utilized to model the long-term distribution of wildfire probability in California. Due to a more limited temporal scale in the recent and historic analysis, fewer points (n=1,250) were retained for use in these models.

Independent from the wildfire presence locations gathered for modeling (5,000 from long-term, 1,250 from recent, and 1,250 from historic), additional wildfire presence locations were generated for independent model validation outside of Maxent software. Specifically, using the same techniques outlined above, 1,250 wildfire presence locations were extracted from long-term burned area polygons for the long-term wildfire probability model validation. Similarly, 300 recent and 300 historic wildfire presences were gathered for recent and historic wildfire probability model validation, respectively. While wildfire observations were retained as dependent variables for use in Maxent software and for validating Maxent model outputs, additional pseudo-absence locations were needed for model validation (see Section 3.4.5).

3.3.1.2 Wildfire Absence Data Creation

Although pseudo-absences or background locations are not used in presence-only habitat suitability modeling, they are important in validating such models. Similar to gathering wildfire

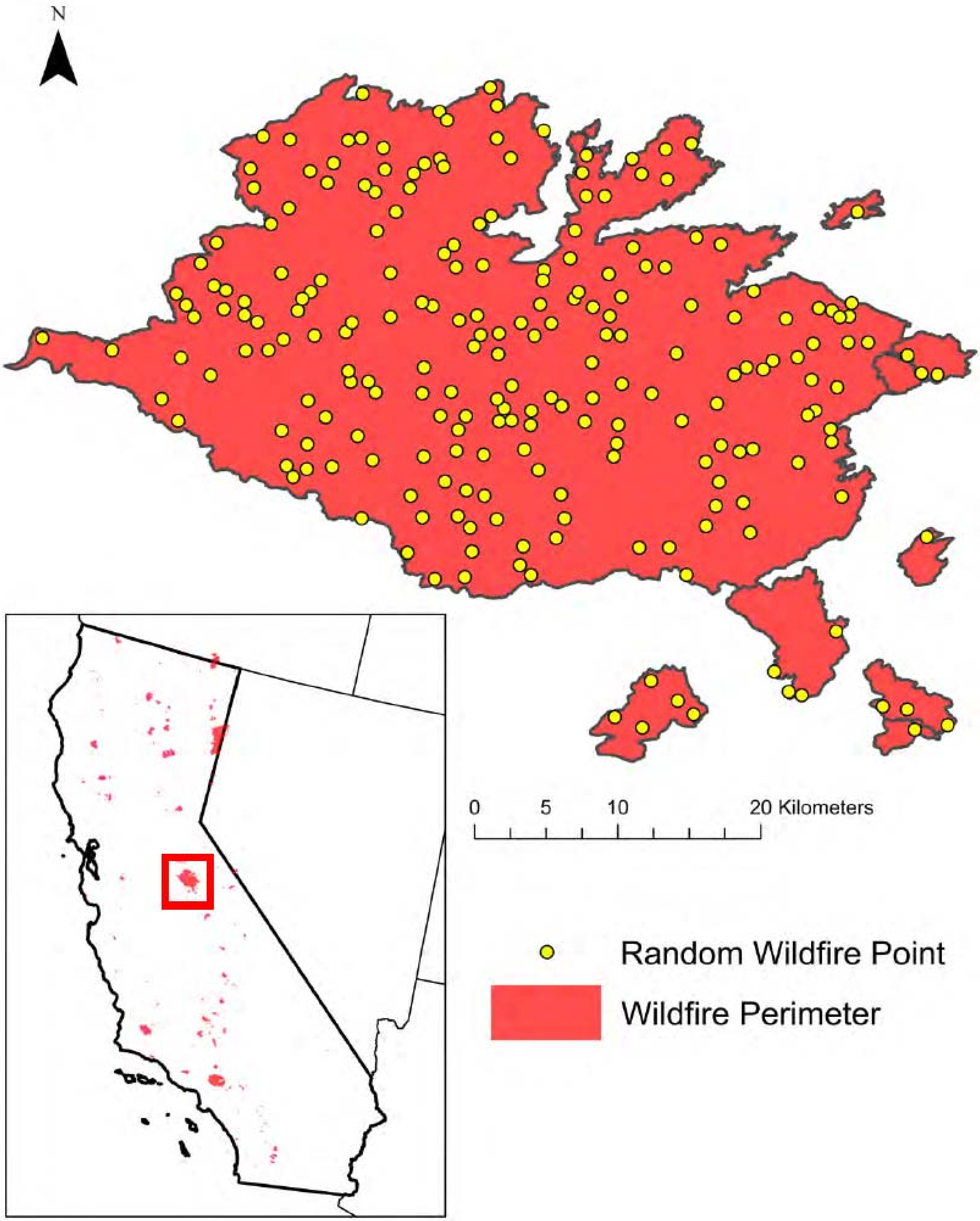

Figure 8. Example of random wildfire presence points extracted from burned area polygons.

presence data, the *Create Random Points* tool in ArcMap was used to distribute random points within the extent of California. The *Select by Location* tool was utilized to select random wildfire points which intersect burned area polygons. Upon wildfire presence selected, the *Switch Selection* tool was utilized to reselected background locations where wildfire is absent at each of the three temporal scales of interest (long-term, recent, and historic). Specifically, a total of 1,250 pseudo-absences were generated for testing the long-term probability model. Similarly, 300 pseudo-absences were obtained for recent and historic long-term model validation. Note that the number of wildfire presence locations generated for model validation matches the number of pseudo-absences. While wildfire presence and pseudo-absence locations were utilized in modeling and testing, exploratory data, or independent variables, were necessary in modeling with Maxent software.

3.3.2 Explanatory Data (Independent Variables)

3.3.2.1 Ignition Sources

Both natural and anthropogenic sources were considered to assess the role of ignitions on wildfire probability, as detailed in Table 3. The only natural ignition considered was lightning, whereas three explanatory variables representing anthropogenic sources were utilized: population density, road density, and distance to nearest road. Gridded summaries of annual density of lightning strikes were obtained from the National Climatic Data Center (NOAA 2015b). These summaries represent the number of cloud-to-ground lightning flashes per year between 1986 and 2012. The summary grids are defined as a 4 km Albers Equal Area grid. For use in wildfire probability models, annual densities of lightning strikes were averaged for the three temporal scales using *Cell Statistics* in ArcMap's Spatial Analyst toolbox, as data permit.

Ignition sources that represent anthropogenic influence are single data layers and therefore are not processed to represent an average of the studies' timeframe. The population density grid (Gridded Population of the WorldVersion 3 [GPWv3]) utilized in this study consists of estimates of human population per unit of area in 2000 (long-term and recent models) and 1990 (historic model; Center for International Earth Science Information Network [CIESIN]; Columbia University, and Centro Internacional de Agricultura Tropical [CIAT] 2005). Road density and distance to roads represent human and vehicle proximity to possible fire prone areas (primary and secondary roads downloaded from U.S. Census Bureau 2014). Based on limited availability of historic roads data, road density and distance to roads were not utilized in the historic model. In all, the anthropogenic ignition variables discussed are considered significant for modeling wildfire probability since human activities are known to alter natural fire regimes (Syphard et al. 2007). While ignition sources are important for modeling wildfire probability, climate is a key factor in wildfire occurrence and behavior.

3.3.2.2 Bioclimate

Bioclimate variables implemented in the study include averages of temperature and precipitation, which represent the effect of variations in climate on fuel moisture and control of vegetation patterns. Bioclimate variables, or climate indices, developed by the USGS, were computed to represent climate variables in each of the wildfire probability models (long-term, recent, and historic). Such bioclimatic variables capture information about annual conditions, seasonal means, and intra-year seasonality (O'Donnell and Ignizio 2012). Variables Bio1 through Bio19 were derived from a single climate data source (PRISM 2004). Mean monthly minimum and maximum temperature and mean monthly precipitation data layers at the 4 km scale were retrieved and further processed in ArcMap 10.3 for use in DIVA-GIS 7.5.0 software. Specifically, using *Cell Statistics* in ArcMap's Spatial Analyst toolbox, monthly climate data was averaged for each period of interest (1984-2013, 2009-2013, and 1984-1988). Twelve datasets (January through December) were derived for each of the three climate variables (mean monthly minimum and maximum temperature and mean monthly precipitation), totaling 36 datasets. The thirty-six datasets were clipped to the study area using the *Extract by Mask* Spatial Analyst tool. Using the "Environments…" tab in the *Extract by Mask* tool window, the output was assigned Albers NAD 1983 equal-area projection and a 1 km spatial resolution. The data layers were imported into DIVA-GIS software to derive nineteen bioclimate variables for use in each of the models. Table 3 details the nineteen bioclimate variables considered in this study.

Wind speed and wind class data, obtainable from National Renewable Energy Laboratory (NREL 2003), is important in determining wildfire spread potential. A polygon feature class consisting of wind power classes (Class 1 through 7) represents the speed of wind for a given area. Class 1 represents zero miles per hour (mph), whereas Class 7 represents an area with wind speeds reaching 21.1 mph. The *Polygon to Raster* tool within ArcMap's Conversion toolbox was used to convert the polygon wind class features to a raster dataset (1 km) for use in Maxent software. Due to availability of information, only the long-term and recent wildfire probability models utilized this dataset. Such indicators of climate are useful for quantifying the effects of climate variables on species distribution (O'Donnell and Ignizio 2012).

3.3.2.3 Topography

Topography is a substantial component of understanding not only wildfire behavior, such as intensity, rate, and direction of fire, but also the spatial distribution of wildfire in a given area. A

30 arc-second (1 km) digital elevation model (DEM) dataset was downloaded and utilized in this study (Gesch 1996). The global 30 arc-second DEM (GTOPO30; tile W140N40) was selected because it matched the 1 km by 1 km cell size that was used for all of the other environmental variables. The original source metadata revealed potential error in the DEM, estimated as the root mean square error (RMSE). Specifically, approximately 30% of the 30 arc-second DEM cells have an absolute vertical accuracy of + or – 30 meters at 90% confidence (USGS 1996). In addition, the metadata gives no information about data accuracy at specific locations within the DEM, adding uncertainty to derivative products (i.e. aspect and slope) (Holmes et al. 2000; Fisher and Tate 2006). However, because this study aims to model relative wildfire estimates, versus an absolute representation, at a small cartographic scale, errors in the 30 arc-second DEM are assumed to be *de minimis*. Errors and uncertainty associated with slope and aspect derivatives are discussed in Section 5.3.

Prior to calculation of topographic derivatives, the 30 arc-second DEM was clipped to the study area using the *Extract by Mask* Spatial Analyst tool. Using the "Environments…" tab in the *Extract by Mask* tool window, the original DEM projection was transformed from WGS 1984 to Albers NAD 1983 equal-area projection and assigned a 1 km spatial resolution. Lastly, small imperfections in the surface raster data were removed using the ArcMap's *Fill* tool. Using the processed DEM, slope and aspect were derived using ArcMap 10.3 Spatial Analyst tools (S*lope* and *Aspect*). Slope represents the ratio of rise over run and is expressed in degrees and provided a measurement of terrain steepness. Aspect is expressed in positive degrees from 0 to 359.9, measured clockwise from north, and refers to the direction of slope. Aspect represents the effect of solar heating, climate, and moisture content in fuels, all important predictors of wildfire potential.

3.3.2.4 Vegetation

Vegetation cover is applied to the wildfire models and represents biomass accumulation and fuel available for burning. A 30 meter categorical land cover class dataset from the National Land Cover Database (2011) was obtained and used to represent vegetation available as fuel for the long-term and recent wildfire probability models (National Gap Analysis Program, USGS; Jin et al. 2013). To represent historic vegetation conditions, a historical land-use and land-cover dataset provided by USGS at a spatial resolution of 30 meters was utilized (Price et al. 2007).

Due to differences in fire regimes among vegetation (i.e. forested vs. non-forested habitat) across the state, the recent and historical categorical land cover class datasets were reclassified to represent fuel versus nonfuel to limit bias and model under- or overfitting. Specifically, areas where wildfire spread is unusual, such as areas of sparse vegetation cover (i.e. deserts, alpine tundra) and permanent wetlands, were considered and reclassified as nonfuel, while all other areas were considered fuel, such as forestlands and rangelands, among others.

In addition to fuel versus nonfuel, gross primary product (GPP) was considered for modeling the spatial distribution of long-term and recent wildfire. GPP represents the rate at which plants store energy as biomass per unit of time, or flammable biomass (Parisien et al. 2012). GPP varies among ecosystems and is highest where temperatures are warm and water and solar energy are abundant (Friedland et al. 2011). A 1 km global Terra/MODIS GPP dataset in HDF-EOS format was downloaded and exported to ESRI GRID format using the National Climatic Data Center Weather and Climate Toolkit (USGS 2003). Four grids were required for this study: h08v04, h08v05, h09v04, and h09v05. The grids were mosaicked into one raster grid using the *Mosaic to New Raster* function in ArcMap 10.3. The projection of the mosaicked raster was transformed from Integerized Sinusoidal (ISIN) to Albers NAD 1983 equal-area projection. GPP is expressed as the amount of organic matter synthesized by producers per unit area in unit time in kg C/m2/year. The GPP dataset for California ranges from 0.03 to 3.6 kg C/m2/year. As discussed in 3.2, all resulting vegetation, as well as ignition and topography variable grids were average to the 25 and 100 km^2 scale using *Block Statistics* prior to modeling burn probability in Maxent software.

3.3.3 Exploratory Analyses

As discussed in Section 3.3.2, twenty-nine variables were considered for modeling the spatial distribution of wildfire probability across California. Of these variables, bioclimate datasets (Bio 1 through Bio 19; Table 3) underwent exploratory analyses in order to determine the correlation amongst variables. This method is utilized to avoid incorporating a large number of variables that overlap information and thus reduce the accuracy and efficiency of the models (i.e. annual precipitation, precipitation of wettest month, precipitation of driest month, precipitation of wettest quarter, and precipitation of driest quarter, precipitation of warmest quarter, precipitation of coldest quarter, etc.). To accomplish this task, bioclimate datasets listed in Table 3 were converted to point features using the *Raster to Point* tool. Following, attribute tables were exported to *CSV* format and combined into a single Microsoft Excel file. The complete file was uploaded into IBM's Statistical Package for the Social Sciences (SPSS) for correlation analysis using Pearson R regression (2013). The fairly uncorrelated yet complementary environmental variables were retained for use in the species distribution model. Specifically, highly correlated variables ($R>0.6$) were excluded in the models. Results of the exploratory analyses are detailed in Table 4 and provided as Appendix A.

While correlation analyses were used to reduce redundant information in modeling wildfire probability, variables were further analyzed through practice model runs. Variables and scales (1, 25, and 100 km^2) which perform the best (i.e. AUC) and contributed the greatest overall percentage were retained in the final model runs. Long-term, recent, and historic models utilized variables that represent accurate environmental conditions during each of the selected temporal scales (Table 4). Difference in variables for each of the models relate to results of the exploratory analyses, practice runs (model performance), and availability of data. Variables detailed in Table 4 were retained as independent inputs for modeling the long-term, historic, and recent distribution of wildfire probability using Maxent software.

Table 4. Variables utilized in long-term, recent, and historic wildfire probability models. All definitions and citations for the sources are provided in Table 3.

Long-term Model	Recent Model	Historic Model
Lgt_dens1	Lgt_dens100	Lgt_dens100
Pop_dens1	Pop_dens1	Pop_dens1
Rd_dens100	Rd_dens100	Bio2
Distrd_dens100	Distrd_dens100	Bio3
Bio2	Bio2	Bio10
Bio4	Bio3	Bio15
Bio9	Bio12	Bio19
Bio14	Bio14	Elev100
Bio19	Bio15	Aspect100
Wind_cl100	Bio18	Slope25
Elev1	Wind_cl100	Fuel1
Aspect100	Elev100	
Slope25	Aspect100	
Fuel25	Slope25	
GPP100	Fuel100	
	GPP100	

Note: The number following each variable (except bioclimate variables) represents the scale utilized (1, 25, or 100 km^2 scale).

3.4 Maxent Software

3.4.1 Overview

Statistical models are used for predicting the behavior of random processes. Maximum entropy (Maxent) is a sophisticated approach to modeling the probability distribution of species habitat from the *n*-dimensional environmental space using presence-only data (Phillips et al. 2006). Maxent software estimates a target probability distribution by fitting the probability distribution of maximum entropy to the environmental variables (independent variables) at each presence-point, or species occurrence (dependent variable). Specifically, Maxent fits sample points to input environmental variables, and estimates the environmental requirements (i.e. suitability) for that species (Phillips et al. 2006; Parisien et al. 2012). The information is used to estimate the species distribution of non-sampled regions using known explanatory variables and produce a habitat suitability map containing “logistic”, “cumulative”, and/or “raw” probabilities. Map values of each cell in outputs represent an estimate of the relative, rather than absolute, probability of presence per grid cell. The software assumes that all sample points were collected unbiased of environmental conditions, or explanatory variables, used in the model (Phillips et al. 2006).

3.4.2 Data Requirement

Wildfire probability models are computed in Maxent 3.3.3k, a free downloadable software from the Internet (Phillips et al. 2006; Elith et al. 2011). Maxent requires presence-only data points to be formatted using comma-separated values (CSV) displayed in three columns: species, longitude, and latitude. As discussed in Section 3.3.1, 5,000 random points intersecting burned area perimeters of wildfires observed between 1984 and 2013 were retained in the long-term model. Fewer points (n=1,250) were selected for use in the recent (2009 to 2013) and historic (1984 to 1988) models. The sample's location inputs were created by converting the dataset to Albers NAD 1983 equal area projection. Upon doing so, the *Calculate Geometry* tool was used to update the X, Y coordinates in the attribute tables. The updated attribute tables were extracted to Excel using the *Table to Excel* tool. In Microsoft Excel, these tables were manipulated and saved as CSV format for use in Maxent software.

Similarly, Maxent requires all environmental variables to be in ASCII raster format and contain the same geographic reference and projection system, geographic extent, and grid cell size in order to execute a model. Using ArcMap 10.3, environmental variables considered for Maxent software were formatted using tools within the Spatial Analyst toolbox. First, raster datasets were clipped to the extent of California using the *Extract by Mask* tool. This tools allows the user to set the parameters needed for each output file; the geographic reference and projection system, geographic extent, and grid cell size for each environmental variable were set exactly the same. All environmental variables implemented Albers NAD 1983 equal area projection and conformed to the grid cell size of 1.0 km^2. The modified environmental variables were converted to ASCII files and stored in a folder labeled "Environmental Variables." The directory file was uploaded into the "Environmental Layers" section in the Maxent software. Upon uploading wildfire presence data and environmental variables in Maxent software, default model parameters were altered.

3.4.3 Model Parameters

Basic and advanced parameters were adjusted before executing the model. The number of model replications was set to 15. This setting runs the model 15 independent times and then averages the results from all models created. Using this feature, in combination with withholding a portion of data for Maxent testing, enables the ability to test the model's performance and provides a way to measure the amount of variability in the model. The random test percentage setting in Maxent is implemented in order to evaluate the model's performance. The random test percent was set to 25 percent, allowing the performance of the resulting model to be tested using a random selection of 25 percent of the presence-only locations. "Subsample" replicated run type was set. This method for evaluating the model's performance is unbiased since no training data need to be employed. Lastly, the maximum number of background points was set to 40,000 for the long-term model and 20,000 for the recent and historic models, increasing the number of points utilized to determine the Maxent distribution.

In the Advanced Settings tab, the number of iterations was set to 5000 (normally set to 500). Increasing the number of iterations allows the model to have adequate time for convergence, thus reducing the uncertainty of over- or under-predicting the relationships. The 10^{th} percentile training presence threshold was implemented for the Maxent model runs. Suggested by Phillips and Dudik (2008), the 10^{th} percentile threshold provides a highly conservative estimate of species' tolerance to each predictor. Models were performed using the logistic function of the Maxent raw values, or an exponential function of the explanatory variables, because they provide the closest estimate of the probability for species presence, given the environment (Elith et al. 2006).

3.4.4 Model Accuracy

To predict the accuracy of the Maxent outputs, several graphs, created by Maxent, were evaluated using metrics that were computed and averaged for each of the 15 model replications.

The estimated fraction of area suitable for wildfire and omission is measured at the wildfire probability threshold, which minimizes the sum of error measurements. These estimations are interpreted together, and dictate the expected rate of false negatives for a given predicted suitable area (Parisien et al. 2012).

Another output graph to measure model performance is the area under the curve (AUC). The AUC graph allows a user to compare performance of one model with another. An AUC value of 0.5 indicates that the performance of the model is no better than random, while 1.0 indicates perfect classification accuracy. However, this study implements a presence-only framework; therefore, it is not possible to achieve unity in AUC because absence locations are unknown. The maximum achievable AUC is equal to $1 - a/2$, where a is the fraction of the study area (California) that the species covers, which is unknown in most cases. For this study, it is appropriate to assume a to be the percentage of pixels where fire was observed (Parisien et al 2012). This method provides an underestimated approximation of prevalence. In addition to assess the modeling accuracy through generated Maxent metrics, the wildfire probability models were validated independent from the Maxent software.

3.4.5 Model Validation

The best performing long-term, recent, and historic wildfire probability models were validated using independent test datasets detailed in Section 3.3.1 and threshold dependent confusion matrices, also known as error matrices, and Cohen's kappa statistic values from the confusion matrices. For the long-term wildfire probability model, a single dataset consisting of 1,250 presence and 1,250 pseudo-absences was created for validating the long-term wildfire probability model. Similarly, recent and historic independent datasets consisting of 300 presence and 300 pseudo-absences were utilized for validating the recent and historic wildfire probability models. Using the *Extract Values to Points* tool within ArcMap's Spatial Analyst toolbox, the independent test points (presence and pseudo absence) were used to extract the pixel values from each of their respective Maxent habitat suitability outputs. The dataset was then exported to generate a spreadsheet containing presence-and pseudo absence information as the ground truth, along with predicted values (as a percentage) by Maxent software. The 10^{th} percentile training presence threshold was used as the bound for determining presence and absence of Maxent's output predictions. Specifically, pixel values above the 10^{th} percentile of training data are determined to have wildfire presence, whereas below this threshold indicates wildfire absence.

The confusion matrix (Table 5) displays the number of false positives (FP), false negatives (FN), true positives (TP), and true negatives (TN) and compares predicted observations with actual observations to yield a percentage of correct observations. Further, the confusion matrix is used to calculate (Equation 1) the following statistical measures of performance: 1) sensitivity or true positive rate; 2) specificity or true negative rate; 3) accuracy; and 4) kappa statistic. The kappa statistic corrects for expected accuracy due to chance and is rated as follows: 0 to 0.2 = slight, 0.21 to 0.4 = fair, 0.41 to 0.6 = moderate, 0.61 to 0.8 = substantial and 0.81 to 1 = near perfected agreement (Landis and Koch 1977; Manel, William, and Ormerod 2001; Allouche, Tsoar, and Kadmon 2006). Upon validating the best performing long-term, recent and historic wildfire probability models using statistical measures of performance, Maxent logistic outputs were further processed in ArcMap for portraying suitable habitat.

Table 5. Confusion matrix for presence/pseudo-absence.

	Recorded			Totals
		Presence (+)	Absence (-)	
Predicted	Presence (+)	True positive (TP)	False positive (FP)	TP + FP
	Absence (-)	False negative (FN)	True negative (TN)	FN + TN
	Totals	TP + FN	FP + TN	Total

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} - \text{TN}}$$

$$\text{Overall Accuracy} = \frac{\text{TP} + \text{TN}}{\text{Total}}$$

$$\text{Kappa} = \frac{\left(\frac{\text{TP} + \text{TN}}{\text{n}}\right) - \frac{(\text{TP} + \text{FP})(\text{TP} + \text{FN}) + (\text{FN} + \text{TN})(\text{TN} + \text{FP})}{\text{n}^2}}{1 - \frac{(\text{TP} + \text{FP})(\text{TP} + \text{FN}) + (\text{FN} + \text{TN})(\text{TN} + \text{FP})}{\text{n}^2}}$$

3.4.6 Mapping Wildfire Burn Probability

Upon running and validating the Maxent long-term, recent, and historic wildfire probability models, logistic output maps were converted into useable format in ArcMap in order to accurately display the probability of burn from wildfire occurrence at each temporal scale of interest. Specifically, Maxent output maps were converted from ASCII format to a floating point raster grid using *ASCII to Raster* conversion tool. Logistic outputs from Maxent software give an estimate between 0 and 1 of probability of presence, whereby 1.0 indicates the best conditions for wildfire occurrence and 0 indicates predictions of unsuitable conditions..

Aside from the effective logistic wildfire probability outputs portraying wildfire likelihood, a wildfire threat map derived from the long-term wildfire probability model was created. Specifically, four risk classes (moderate, high, very high, and extreme) and one fire absence class (non-fuel) was established. A 10th percentile threshold of training data was used to establish the primary threshold bound for moderate wildfire risk (0.327). Comparable to the fire threat map generated by CalFire (2004), the remaining high, very high, and extreme wildfire risk classes were (0.327-0.40), (0.40-0.70), and (0.70-1.0), respectively.

Due to the limited temporal scale for the recent and historic outputs, wildfire threat maps were deemed unnecessary for these models. Instead, the recent and historic logistic wildfire probability maps were reclassified as suitable or unsuitable wildfire habitat using the 10th percentile thresholds of training data (0.262 for recent and 0.272 for historic).

3.4.7 Model Comparison

To analyze the change in wildfire probability in California over time, logistic model outputs (recent and historic) were compared and contrasted using map algebra in ArcMap 10.3 Spatial Analyst Toolbox. Specifically, *Raster Calculator* was used to analyze multiple rasters by subtracting cell values from the historic model from the corresponding cell values of the recent model. This method produces a map output depicting areas where the recent and historic outputs agree or disagree with one another on a cell-by-cell basis. This output shows the change in wildfire probability and helps to elaborate changes in environmental influence and wildfire distribution in California overtime.

As discussed in Chapter 3, data collection, preparation of environmental layers and wildfire occurrences, exploratory analyses, and wildfire probability model execution and validation were performed in this study. Results of the long-term (1984-2013), recent (2009-2013), and historic (1984-1988) wildfire probability outputs are discussed in Chapter 4.

CHAPTER FOUR: RESULTS

Wildfire probability models were created using Maxent software in order to predict California's long-term wildfire potential, assess relationships between environmental variables and wildfire probability, and evaluate the projected change in wildfire distribution in California over time. As discussed in Chapter 3, locations of burned wildland habitat were tied to ignition sources, climate conditions, topography, and vegetation to estimate the probability of wildfire for regions of California exclusive of wildfire occurrence information. Twenty-nine explanatory (independent) variables were considered for building three wildfire probability models utilized in this study: long-term (1984 to 2013), recent (2009 to 2013), and historic (1984 to 1988). The number of variables to be fitted in each model was reduced to between 11 and 17 based on exploratory analyses (see Table 4). Maxent model outputs demonstrate the distribution of wildfire probability across the state of California at each of the three temporal scales. Maxent model outputs, such as response curves, model metrics, and habitat suitability maps, for each of the three models are discussed in detail throughout the remainder of this chapter.

4.1 Long-term Model Results

A long-term wildfire probability model was created to model wildfire potential across the state of California. Specifically, random wildfire presence locations obtained from burned area polygons between 1984 and 2013 and fifteen environmental variables were utilized as dependent and independent input variables in the wildfire probability model.

The relationships between wildfire probability and environmental variables are highly diverse, as shown in the response curves of nine of the fifteen variables (Figure 9). The majority of wildfire responses to environmental variables are non-linear, whereby wildfire probability is maximized over intermediate values (i.e. gross primary productivity, road density, and elevation). On the other hand, the response curve for fuel depicts a positive relationship between fuel and wildfire probability.

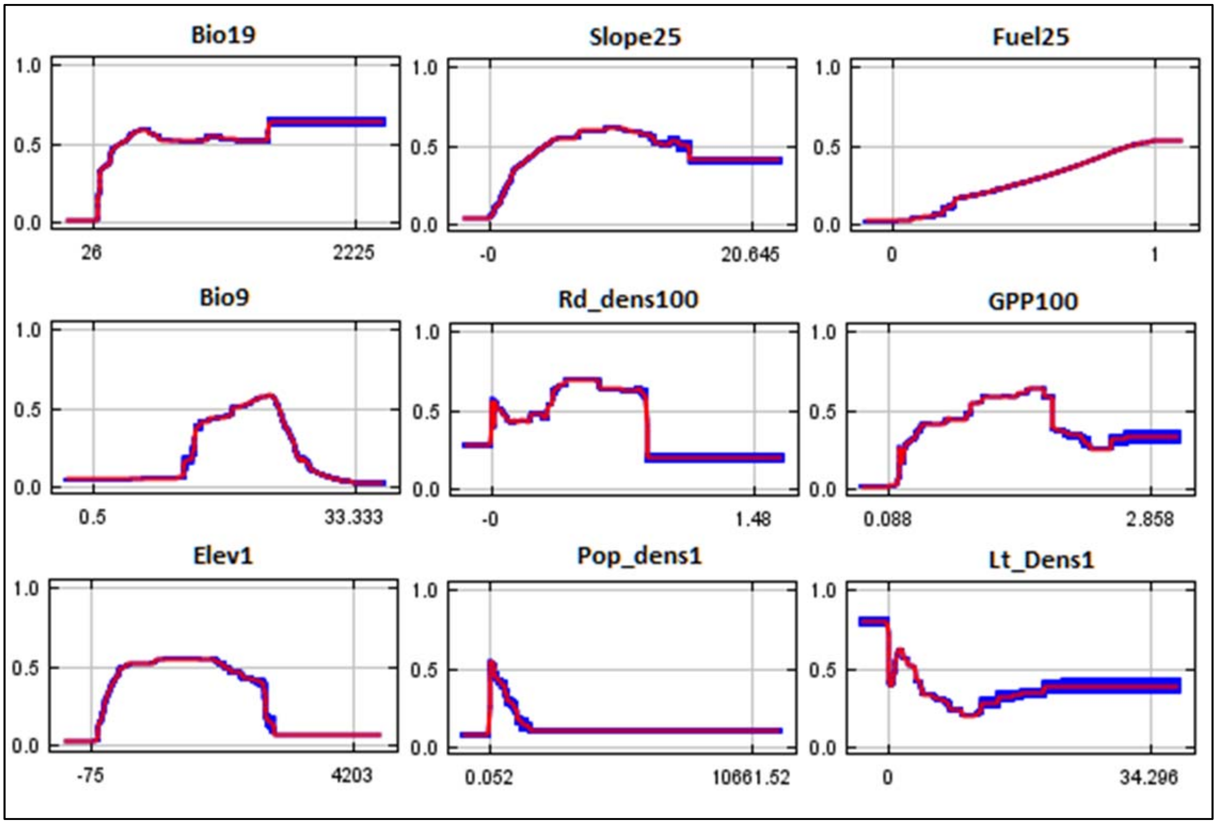

Figure 9. Predicted long-term wildfire probability for nine of the fifteen environmental variables. The red line indicates the mean wildfire probability values, whearas the blue shading represents the standard deviation, as calculated from 15 replicate runs.

Ignition, bioclimate, topography, and vegetation environmental variables appear to all be important predictors of long-term wildfire probability (Table 6). Of these, precipitation of the coldest quarter (Bio19) accounts for the greatest contribution (28.9%). Slope and fuel also significantly contribute to wildfire probability (22 and 13.6%, respectively). Among the ignition sources, road density at 100 km^2 accounts for the great contribution (6.8%).

Table 6. Relative contribution for each environmental variable utilized in the Long-term wildfire probability model given as a percent (%).

Variable	Percent Contribution (%)
Bio19	28.9
Slope25	22
Fuel25	13.6
Bio9	7.8
Rd_dens100	6.8
Gpp100	6.8
Elev1	4.3
Bio4	2.9
Wind_cl100	2.2
Lt_dens1	1.6
Pop_dens1	1.5
Aspect100	0.5
Bio14	0.4
Distrd_dens100	0.3
Bio2	0.3

To predict the accuracy of the models outputs, Maxent software computes evaluation metrics and displays results in two graphs: average omission versus predicted area and sensitivity versus specificity. Model evaluation metrics indicate that the long-term wildfire probability model performed fairly well (Figure 10). Specifically, the AUC for the long-term wildfire probability model indicates a high level of performance (AUC = 0.807). Additionally, the omission rate and predicted area graph displays close conformance between the omission rate for long-term wildfire presence data and the predicted rate of omission, suggesting accuracy of the model (Figure11).

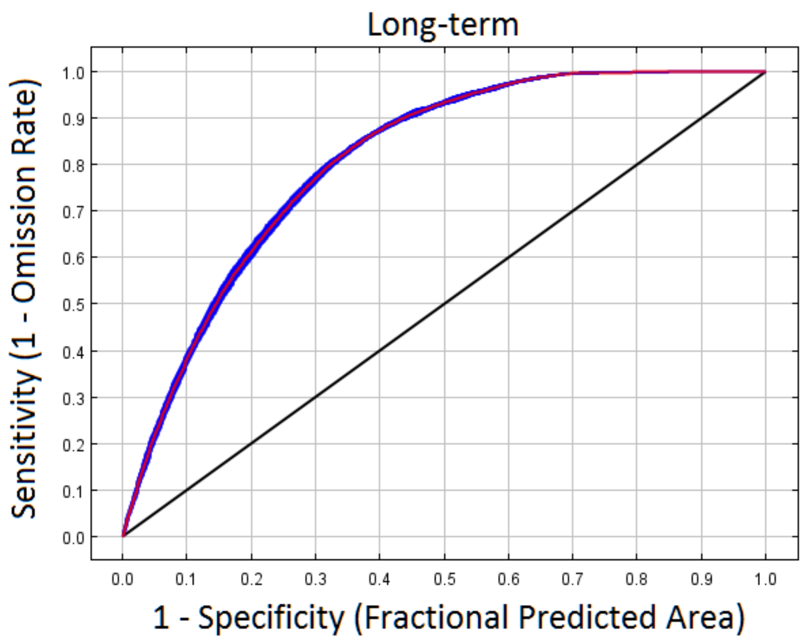

Figure 10. Average model sensitivity vs. specificity obtained by executing the long-term wildfire probability model. The red line indicates the mean AUC, whereas the blue shading represents the mean standard deviation as calculated from 15 replicated runs using random subsets of data.

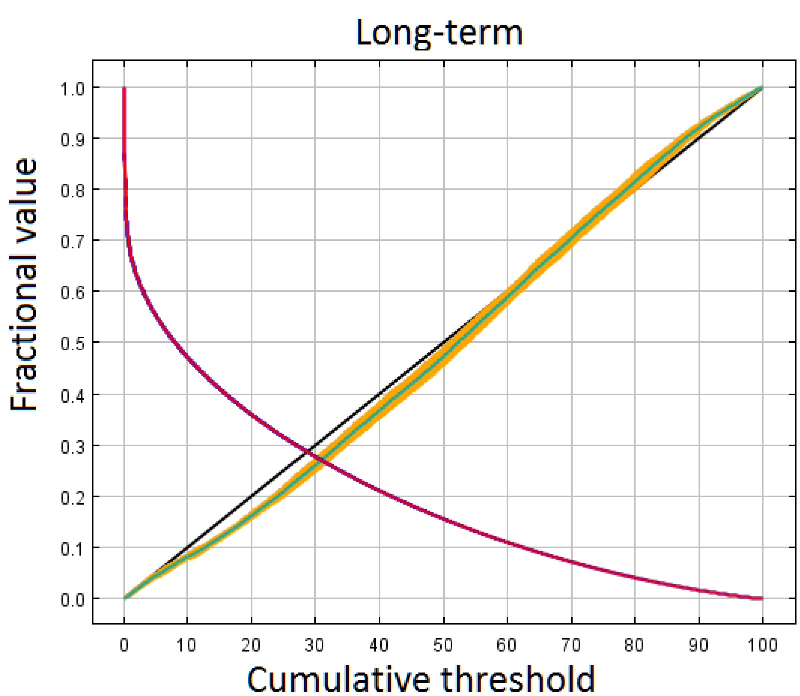

Figure 11. Average omission and predicted area for long-term wildfire occurrence data. The predicted omission (black line; behind yellow) conforms to the mean omission on test data (green line). The orange shading represents the mean standard deviation of omission.

4.1.1. Model Validation

Statistical measures of performance were calculated using the average 10% training presence threshold from the Maxent model runs and error matrix results summarized in Table 7.

Table 7. Error matrix for the long-term wildfire probability model validation using independent test data presences/pseudo-absences (n=2500).

	Recorded			**Totals**
		Presence (+)	Absence (-)	
Predicted	Presence (+)	1100	393	1493
	Absence (-)	150	857	1007
	Totals	1250	1250	2500

As summarized in Table 8, the overall accuracy of the long-term wildfire probability model was 0.783, indicating the model correctly predicted 78.3% of the presence and pseudo-absence point to be included or excluded in predicted wildfire habitat. Further, Maxent performed at a high level predicting wildfire presence where wildfire was observed (sensitivity = 88.0%). However, Maxent performed at a lower level in predicting non-wildfire habitat where pseudo-absences occurred (specificity = 68.3%). This implies the model poorly distinguished between wildfire habitat and non-wildfire habitat by over-predicting suitable habitat. The kappa statistic indicated the long-term wildfire probability model had moderate agreement with the testing dataset (presence and pseudo-absence).

Table 8. Accuracy measures for the long-term wildfire probability model validation using independent test data presences/pseudo-absences (n=2500).

Measures	**Values**
Sensitivity	0.880
Specificity	0.686
Overall Accuracy	0.783
Kappa statistic	0.567

4.1.2. Long-term Wildfire Probability Maps

Maxent model outputs demonstrate the potential distribution of long-term wildfire in California (Figure 12). Mean predicted wildfire probability (based on 15 model replicates), where warmer colors (i.e. red, orange, and yellow) indicate higher probability of suitability and cooler colors (i.e. blues) indicate lower probability of wildfire suitable habitat, are portrayed in Figure 12. Patterns in the modeled wildfire probability output are highly diverse throughout the state of California. Specifically, wildfire probability is moderately high (>0.4) in most areas, however wildfire likelihood is low in desert and agriculture areas.

The long-term wildfire probability map in Figure 12 was reclassified into four wildfire risk classes (Figure 13). Using the class designations, approximately 41% of the total area is classified as moderate, 10% is high risk, 24% is very high risk, and only 1% is extreme wildfire risk; 24% was designated as non-fuel habitat, where wildfire is assumed to be absent (Table 9).

Table 9. Long-term wildfire probability class area and percent of total area.

Wildfire Risk Class	Km^2	Percent of Total Area
Moderate (0 – 0.327)	166,701	41%
High (0.327 – 0.40)	41,441	10%
Very High (0.40 – 0.70)	97,174	24%
Extreme (0.70 – 1.0)	5,141	1%
Non-fuel	97,351	24%

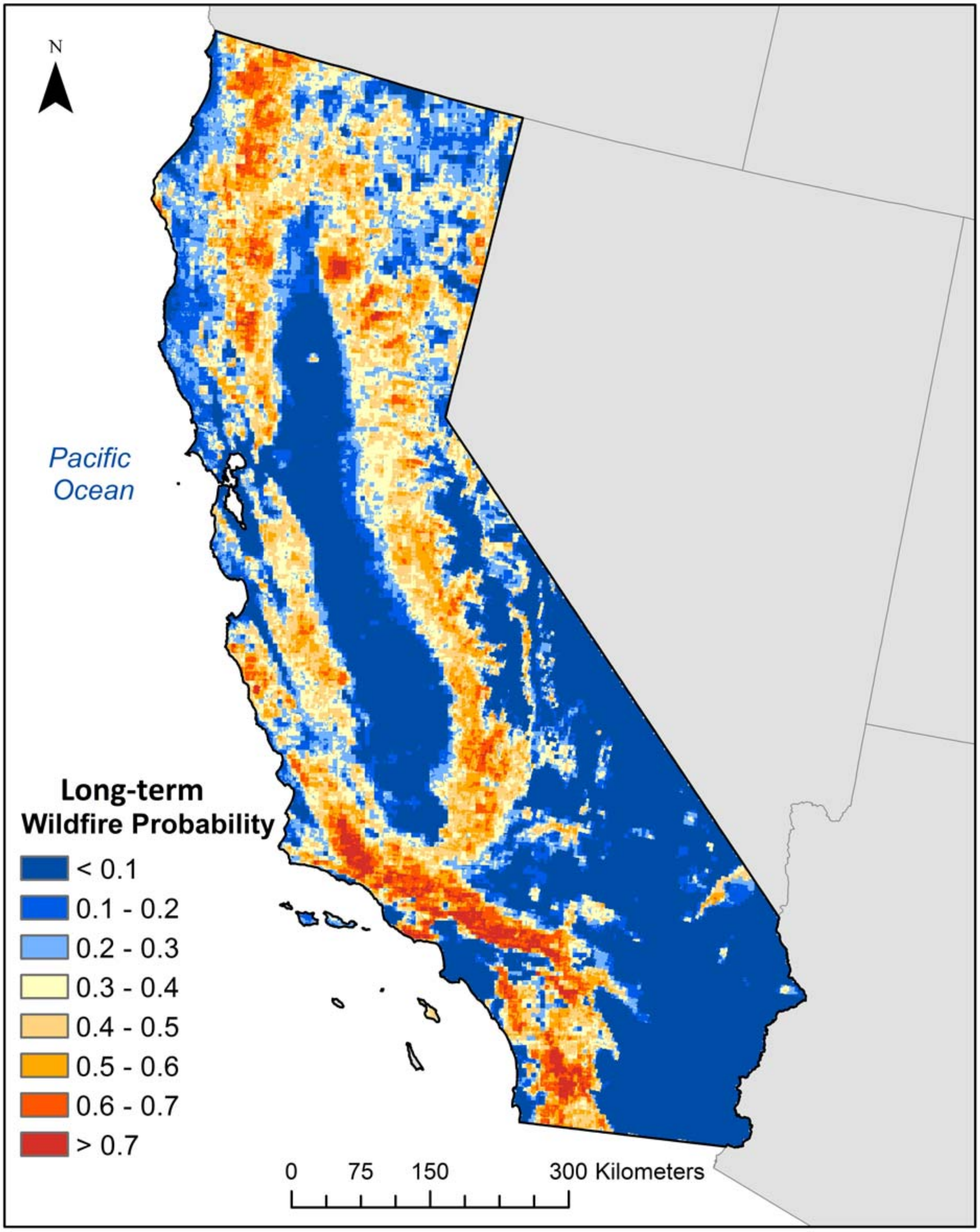

Figure 12. Long-term wildfire probability map using 5,000 random presence-point locations between 1984 and 2013. Warmer colors indicate regions with high probability of suitable habitat, while cooler colors suggest lower probability of suitable habitat (AUC = 0.807).

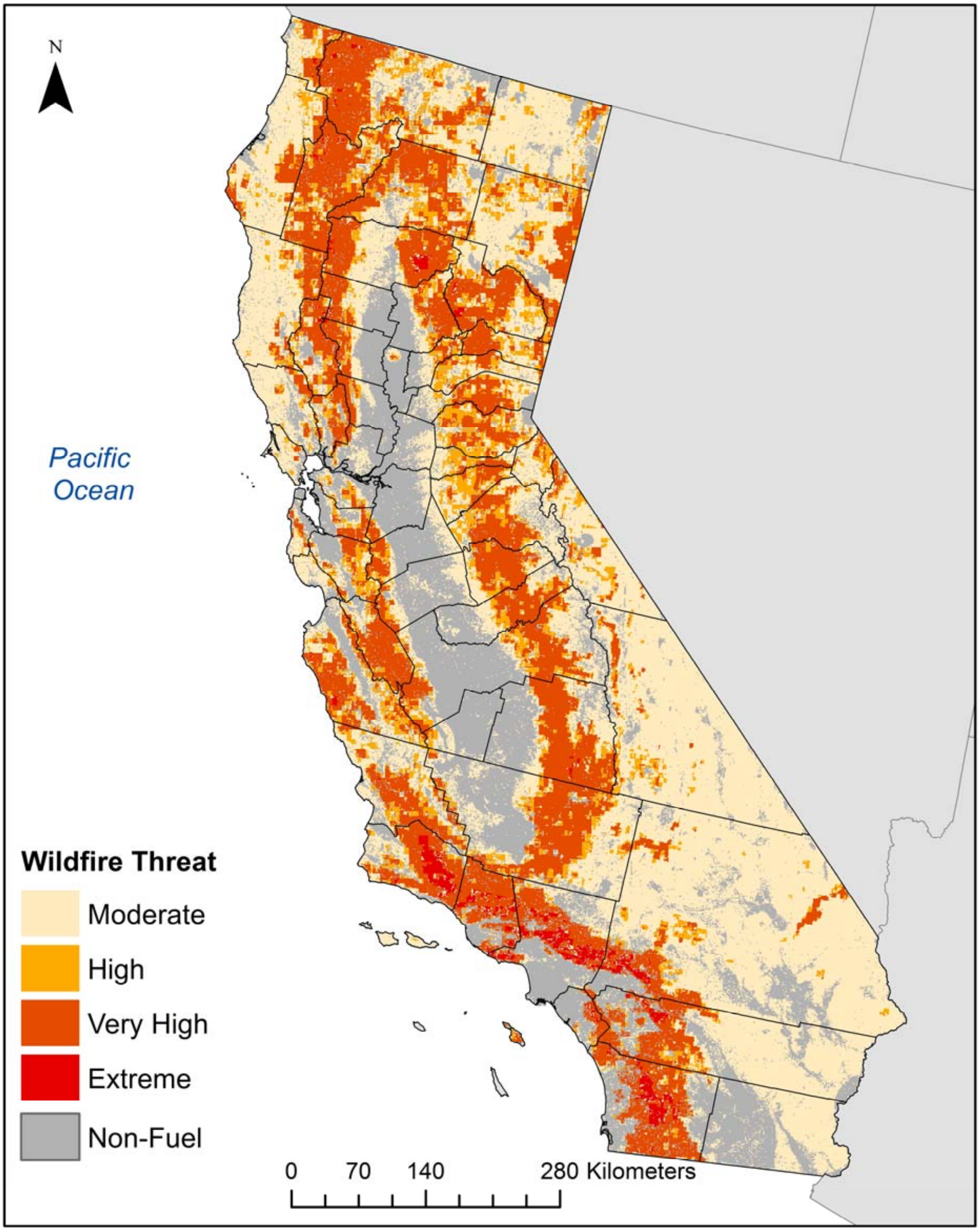

Figure 13. Wildfire threat map derived from the long-term wildfire probability map (Figure 12).

4.2 Recent and Historic Model Results

Short-term wildfire probability models (recent [2009 to 2013] and historic [1984 to 1988]) were created to asses California's recent and past wildfire risks. Specifically, the recent wildfire probability model evaluates recent wildfire risk across the state of California. This output, coupled with the historic wildfire probability model is used to assess the change in wildfire probability and suitable habitat over time. To do so, recent and historic wildfire probability models utilize 1,250 random wildfire presence locations from within burned areas and between 11(historic) and 17 (recent) explanatory variables.

As shown in Table 10, ignitions, climate, topography, and vegetation all play an important role in modeling the recent distribution of wildfire probability. Specifically, elevation (Elev100; 20.9%), annual precipitation (Bio12; 18.8%), and fuel (fuel100; 10.5%) are the three top contributors of the recent model. For the historic model, climate variables contribute 52.6% of the total (100%), with precipitation seasonality (Bio15; 18.6%) and mean temperature of warmest quarter (Bio10; 18%) the greatest contributors. Slope (Slope25; 15%), elevation (Elev100; 12.1%), and fuel (Fuel1; 9.1%) also significantly contribute to the historic model.

As shown in Figure 14, the area under the ROC curve (AUC) for both the recent and historic models indicates high levels of performance (0.923 and 0.871, respectively). The increase in performance with the recent data is most likely the result of the utilization of more environmental variables as independent variables and/or better fit between wildfire locations and independent variables for the models time frame (2009 to 2013). Additionally, for both the historic and recent models, the predicted omission mainly conforms to the mean omission of the test data (Figure 15), with the exception of various cumulative thresholds. This differentiation may be due to the use of a subsample method (25% of presence data). Specifically, the test and training data are not independent of one another because the test data is derived from the same wildfire presence data.

Table 10. Relative contribution for each environmental variable utilized in each short term wildfire probability models (recent and historic) given as a percent (%).

Variable	Recent Percent Contribution (%)	Variable	Historic Percent Contribution (%)
Elev100	20.9	Bio15	18.6
Bio12	18.8	Bio10	18
Fuel100	10.5	Slope25	15
Bio18	9.5	Bio19	13.1
Slope25	7.4	Elev100	12.1
Rd_Dens100	5.5	Fuel1	9.1
Lt_Dens100	5.4	Lt_Dens100	6.8
Pop_Dens1	4.8	Pop_Dens1	3.3
GPP100	4.2	Bio3	2.1
Bio9	3.4	Aspect100	1.1
Bio15	2.6	Bio2	0.8
Distrd_dens100	2.4		
Bio14	1.5		
Aspect100	1.3		
Wind_Cl100	0.9		
Bio3	0.8		
Bio2	0.2		

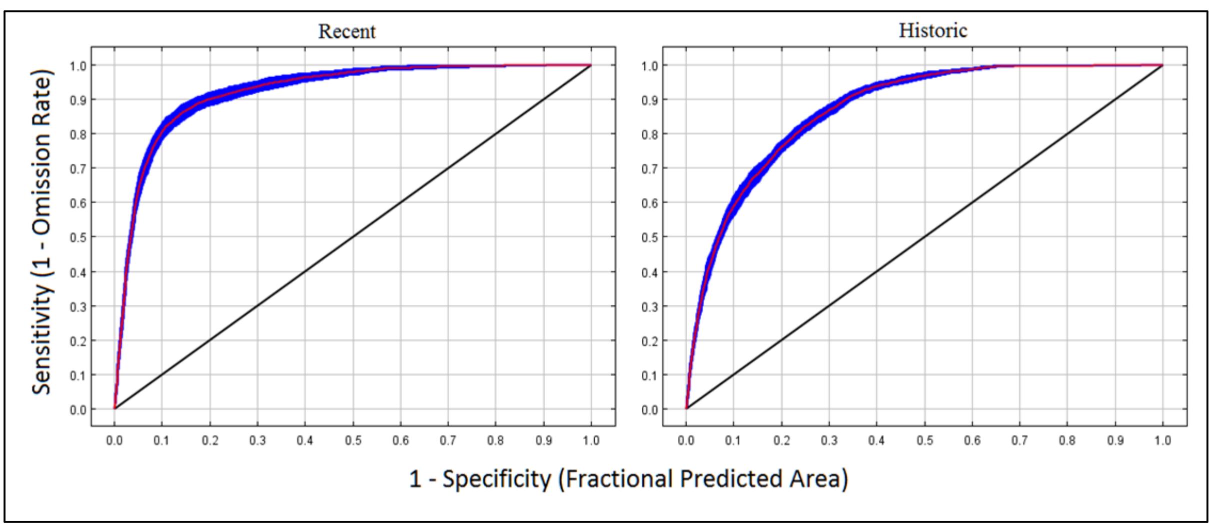

Figure 14. Average model sensitivity vs. specificity for the recent and historic wildfire probability models. The red line indicates the mean AUC, whereas the blue shading represents the mean standard deviation as calculated from 15 replicated runs.

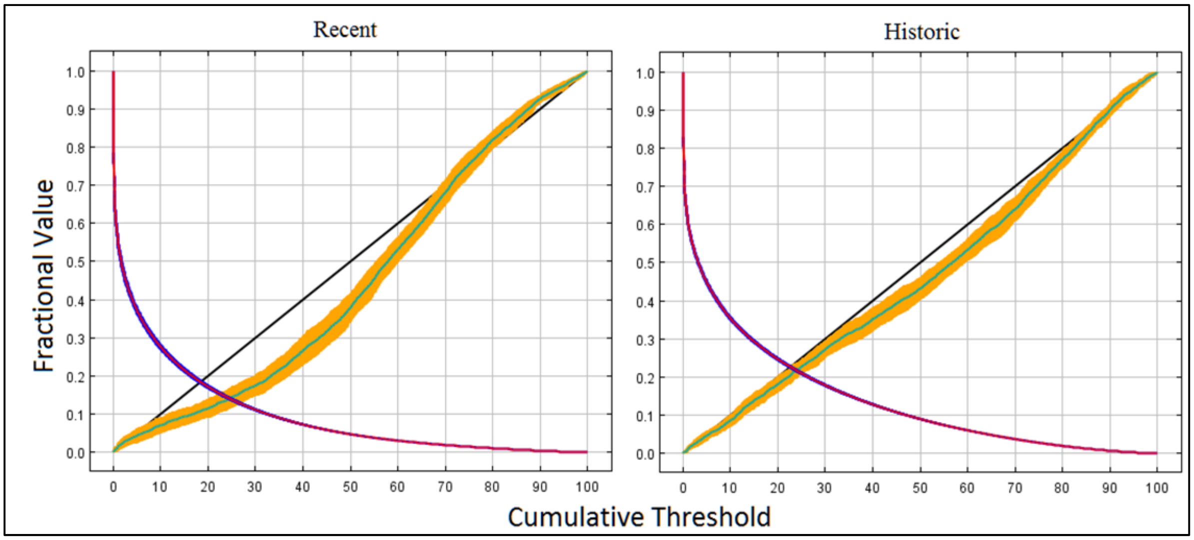

Figure 15. Average omission and predicted area for recent and historic wildfire occurrence data. The predicted omission (black line; behind yellow) mainly conforms to the mean omission on test data (green line) in the recent and historic graphs. The orange shading represents the mean standard deviation of omission.

4.2.1. Model Validation

The accuracy of the recent and historic wildfire probability models were assessed using the presence/pseudo-absence dataset mentioned in Section 3, 10% training presence threshold from the Maxent model runs. Performance measures were calculated for recent and historic Maxent outputs using error matrix tables (Tables 11 and 12).

Table 11. Error matrix for the recent wildfire probability model validation using independent test data presences/pseudo-absences (n=600).

		Recorded		**Totals**
		Presence (+)	Absence (-)	
Predicted	Presence (+)	264	39	303
	Absence (-)	36	261	297
	Totals	300	300	600

Table 12. Error matrix for the historic wildfire probability model validation using independent test data presences/pseudo-absences (n=600).

		Recorded		Totals
		Presence (+)	Absence (-)	
Predicted	Presence (+)	247	56	303
	Absence (-)	53	244	297
	Totals	300	300	600

As summarized in Table 13, the best of the recent and historic wildfire probability models generally performed similarly. The overall accuracy of the recent and historic wildfire probability models was 87.5% and 81.8%, respectively. Unlike the long-term wildfire probability model, recent and historic sensitivity and specificity metrics showed similar results, whereby Maxent nearly performed equally in predicting wildfire presence where wildfire was observed (sensitivity = 88.0% and 82.3%, respectively) and non-wildfire habitat where pseudo-absences occurred (specificity = 87.0% and 81.3%, respectively). The kappa statistic (0.750) indicated the recent wildfire probability model had substantial agreement with the testing dataset (presence and pseudo-absence). For the historic wildfire probability model, the kappa statistic (0.637) indicated the Maxent model also had substantial agreement with its testing dataset.

Table 13. Accuracy measures for the recent and historic wildfire probability models validation using independent test data presences/pseudo-absences (n=600 for recent and n=600 for historic).

	Values	
Measures	**Recent**	**Historic**
Sensitivity	0.880	0.823
Specificity	0.870	0.813
Overall Accuracy	0.875	0.818
Kappa statistic	0.750	0.637

4.2.2. Recent and Historic Wildfire Probability Maps

As with the long-term wildfire probability model, recent and historic wildfire probability model outputs demonstrate the distribution of wildfire in California under each scenario. Based on fifteen model replications, the mean wildfire distribution displays wildfire probability under recent and historic conditions (Figure 16 and 17). Warmer colors (i.e. red, orange, and yellow) indicate higher probability of suitability and cooler colors (i.e. blues) indicate lower probability of wildfire suitable habitat. A visual inspection of Figures 16 and 17 show recent and historic wildfire probability patterns are highly diverse across the state of California.

Figures 16 and 17 were reclassified into suitable and unsuitable habitat for wildfire occurrence. Comparison of suitable and unsuitable habitat for the recent and historic wildfire probability outputs revealed the recent model's prediction of total suitable habitat is nearly half of the historic model's estimation (Table 14). A comparison of recent and historic wildfire probability model outputs are further discussed in Section 4.3.

Table 14. Recent and historic predictions of suitable habitat using the 10th percentile threshold of training data.

Wildfire Model	Suitable Habitat (km^2)	Percent of Total Area
Recent (2009-2013)	55,650	14%
Historic (1984 – 1988)	108,335	27%

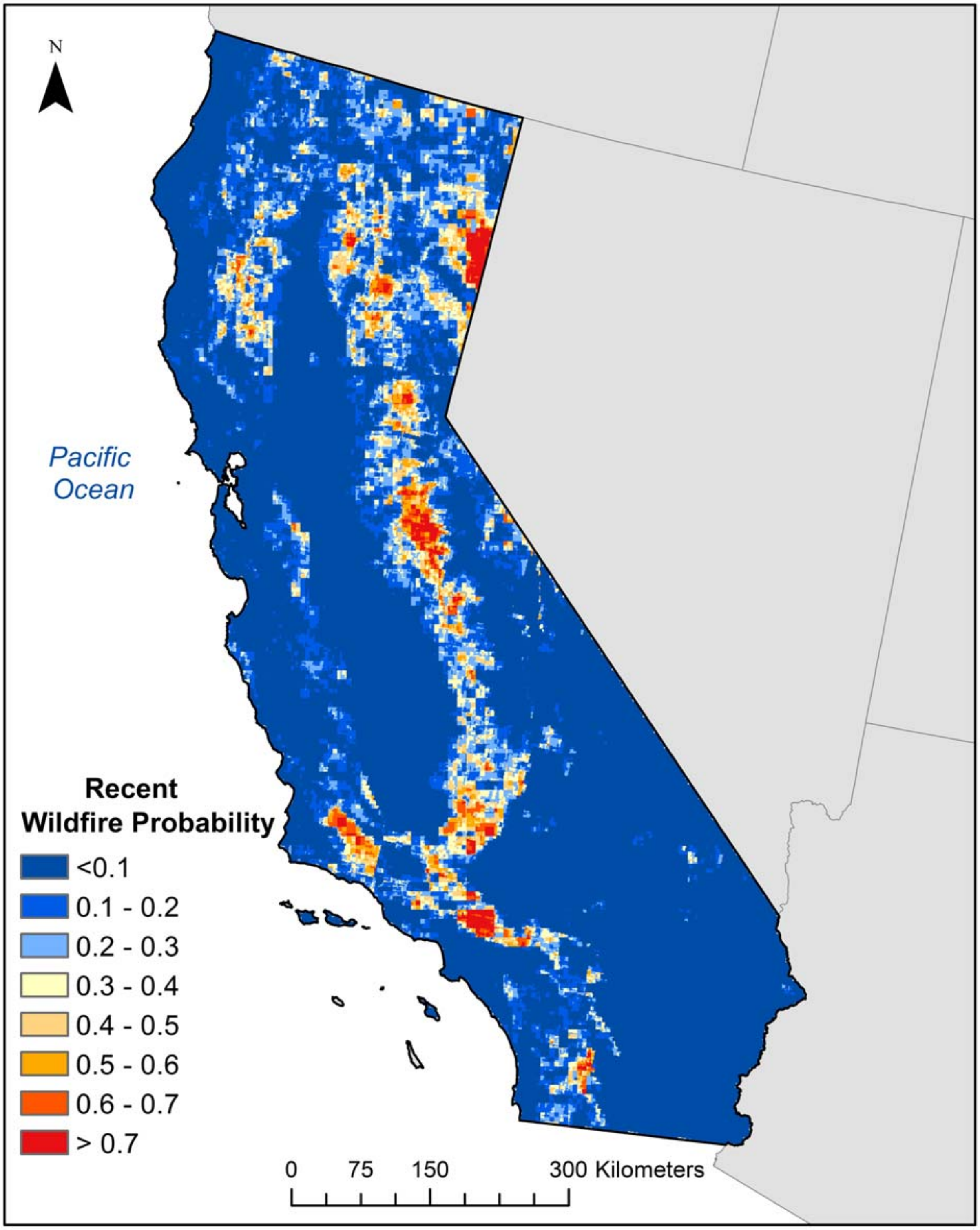

Figure 16. Recent wildfire probability maps using 1,250 random presence-point locations between 2009 and 2013. Warmer colors indicate regions with high probability of suitable habitat, while cooler colors suggest lower probability of suitable habitat (AUC = 0.923).

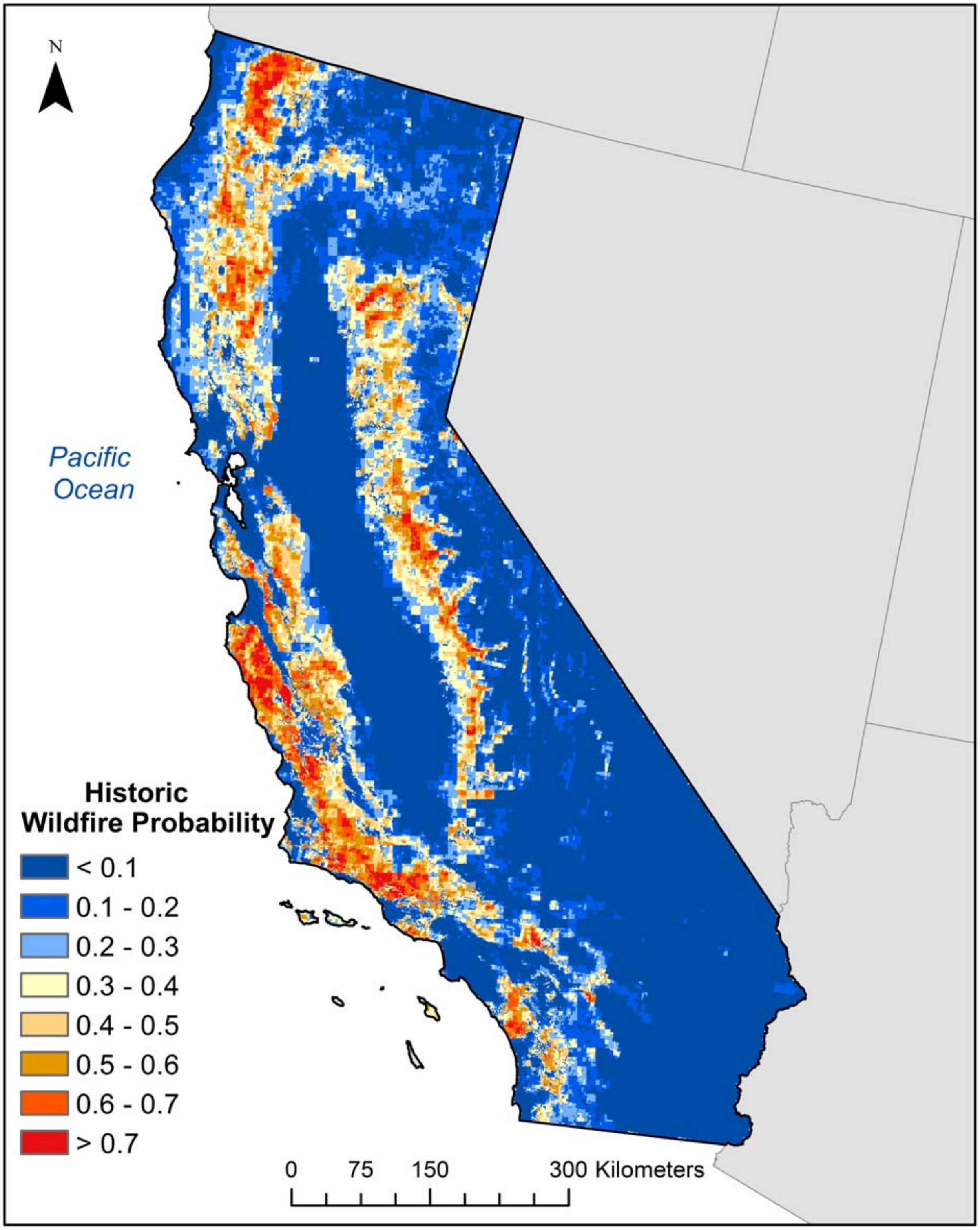

Figure 17. Historic wildfire probability maps using 1,250 random presence-point locations between 1984 and 1988. Warmer colors indicate regions with high probability of suitable habitat, while cooler colors suggest lower probability of suitable habitat (AUC = 0.871).

4.3 Recent and Historic Model Comparison

Outputs of the recent and historic models are compared using an algebraic expression in order to assess the change in wildfire probability across the state of California over time (Figure 18). As shown in Figure 18, red indicates areas of high wildfire probability predicted by the historic model and a lower probability by the recent model. Conversely, green shows high probability areas predicted by the recent model and less suitable areas by the historic model. Yellow indicates areas where the two models are in agreement, and is the majority coverage, as shown in Figure 18. The recent and historic models differentiate in numerous areas across the state. Specifically, historic wildfire probability was greater than recent predictions along the majority of coast, including surrounding forested areas. Wildfire probability also decreased in various areas east of the San Joaquin Basin. Contrary to this, wildfire probability has increased in areas east of the San Joaquin Basin yet west of the California/Nevada border.

Long-term, recent, and historic wildfire probability maps and metric outputs produced by Maxent software were discussed in this Chapter. Chapter 5 discusses the overall findings, evaluates the usefulness of species distribution techniques for determining the patterns of long-term wildfire probability in California, and assesses changes in wildfire probability in California over time.

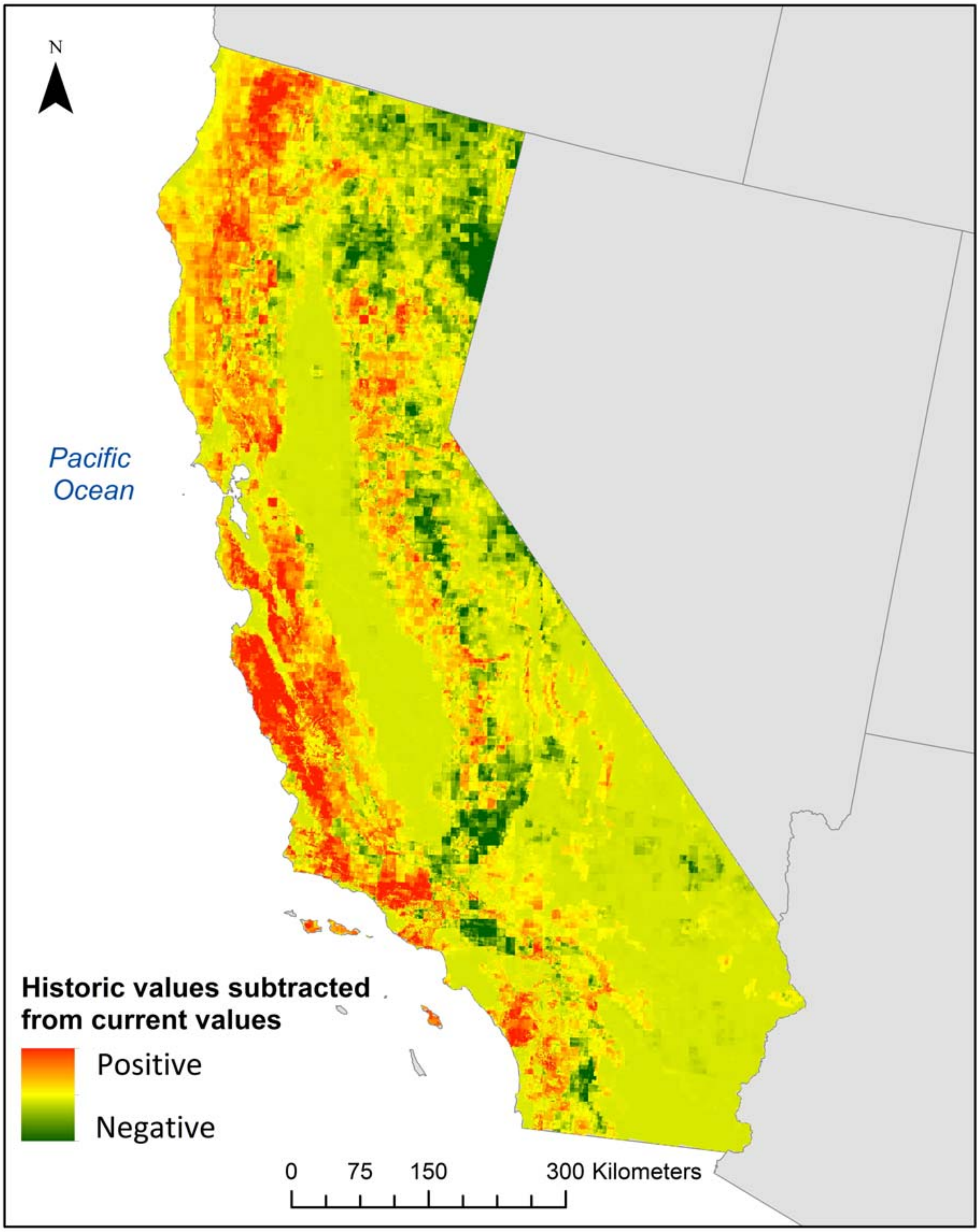

Figure 18. Comparison of recent (2009 to 2013) and historic (1984 to 1988) wildfire probability maps. Cell values from historic model (Figure 17) subtracted from corresponding cell values of the recent model (Figure 16). Red indicates area predicted highly suitable by historic model and less suitable by the recent model. Green indicates areas predicted highly suitable by the recent model and less suitable by the historic model. Yellow shows areas where the two models are in agreement.

CHAPTER FIVE: DISCUSSION AND CONCLUSIONS

The results of this study show habitat suitability modeling techniques, coupled with a wide variety of environmental variables, and wildfire occurrence data can be informative methods for estimating the potential distribution of wildfire in California. Using expansive wildfire occurrence data over a long-term period, habitat suitability models are effective in modeling the overall likelihood of wildfire occurrence. Shorter-term models are successful in modeling recent or past habitat probability for scales of interest. Results from each of these models provide valuable insight to understanding environmental controls, such as climate and vegetation, on state-wide assessments of wildfire likelihood, especially in regions where wildfire occurrence is less frequent.

Results of the long- and short-term wildfire probability models show that wildfire distribution is highly variable across the state of California. Furthermore, resulting distributions of wildfire probability for the long-term, recent, and historic models prove to be dependent on a full range of specific environmental controls, such as ignition sources, climate, topography, and vegetation.

5.1. Long-term Wildfire Probability Model

The long-term probability model was successful in determining the relationship between environmental conditions and wildfire probability and estimating the potential distribution of wildfire likelihood across California (Figure 12). The strongest predictors for estimating the potential distribution of long-term wildfire probability in California are precipitation of the coldest quarter (28.9%), slope (22%), and fuel (13.6%). Results of this study show that fuel has a positive relationship with wildfire probability, whereby increasing fuel increases wildfire potential. While the relationship between fuel and wildfire probability is unsurprising, climate-wildfire relationships are more complex.

Based on results of this study, and similar to results of Parisien et al. (2012), relationships between wildfire and climate are highly complex and vary significantly across the state of California. Specifically, precipitation of the coldest quarter was the most significant contributor of the long-term wildfire model. This variable is directly related to moisture and energy, and thus a primary predictor of geographic distribution of vegetation types (Stephenson 1998; Parisien et al. 2012). The relationship between wildfire and precipitation of the coldest quarter appear to be non-linear, whereby increased fire probability is maximized across intermediate values of the variable. Furthermore, low wildfire probability was estimated for regions where precipitation is nearly non-existent and higher probability once a specific threshold of precipitation is encountered. For example, desert regions in California receive little rainfall (>250 mm) and are comprised of inadequate vegetation for use as fuel. Therefore the distribution of wildfire probability in these regions remains extremely low to absent (see Figure 12). While these complex relationships between wildfire probability and climate are directly related, topography variables indirectly affect suitable habitat and thus wildfire occurrence.

From rigid terrain in the Sierra Nevada to flat and low elevation deserts, California's topographical landscape varies significantly across the state and in turn greatly affects wildfire potential. Slope, elevation, and aspect were derived and utilized as input variables in the model to represent topography. Specifically, slope was the second greatest contributor (22%) of the long-term model. Results of this study depict slope and elevation to have a non-linear relationship with wildfire probability. For instance, wildfire probability increases as slope and elevation increase, up until the maximum threshold is reached. Although slope, elevation and aspect are not directly related to wildfire, these factors act as proxies for other environmental controls which inhibit and/or promote wildfire potential. As shown in Figure 12, wildfire potential across the state of California is high in these areas of rigid terrain (i.e. Sierra Nevada and Klamath Mountains) and low in areas of flatter terrain (i.e. Sacramento and San Joaquin Valley). In areas of rigid topography and high elevation, anthropogenic variables such as population density, road density, and human impact on fuel tends to be low; increasing the potential of wildfire occurrence. However, results show that as elevation increases to a specific threshold, wildfire probability becomes extremely low. This is a result of changes in the limited distribution of vegetation at higher elevations (i.e. sparse vegetation in alpine zones). Similarly, low lands in California are frequently related to agriculture and human development, and thus low fuel content for combustion and wildfire probability.

As for ignitions, both natural and human-related sources generally play a minor role in predicting long-term wildfire probability in California. Lightning contributed only 1.6% to the long-term wildfire probability model. This is not surprising since rainfall is most often related to lightning density, and thus increases moisture content in available fuel. Comparable to Parisien et al. (2012), results of this study showed that wildfire probability have a negative relationship with lightning density. In contrast, results proved human-related sources to have greater influence (8.6% combined), with road density as the greatest contributor (6.8%). The relationship between wildfire probability and road density at the 100 km^2 scale, in addition to distance to nearest road density (100 km^2 scale), appear to be non-linear. These results are consistent with Syphard et al. (2007) and Parisien et al. (2012), whereby human ignitions exhibits an inverse U-shaped, non-linear relationship to wildfire probability. In contrast, population density has a negative relationship to wildfire probability, where increased population results in decreased wildfire likelihood.

Model metrics prepared by Maxent software proved the long-term wildfire probability model to perform at a high and robust level. Similarly, the Maxent outputs were validated using performance measures. These measures revealed that the model performed accurately. Overall, the results of the long-term probability model add to our understanding of wildfire distribution patterns and responses to environmental conditions across the state of California. More importantly, the long-term wildfire results can be further processed to produce up-to-date wildfire threat maps to depict areas of moderate, high, very high, and extreme wildfire risk (Figure 12). Shorter-term models (recent and historic) were used to further assess the change in California's wildfire potential over time.

5.2.Recent and Historical Wildfire Probability Models

Recent and historic wildfire probability models were successful in estimating the distribution of recent and past wildfire probability in California (Figures 16 and 17, respectively). Consistent with the long-term model outputs, climate, topography, and fuel significantly contribute to each of the model runs. Specifically, for the recent wildfire probability model, elevation (20.9%), annual precipitation (18.8%), and fuel (10.5%) were the greatest contributors. For the historic model, precipitation seasonality (19.3%), mean temperature of the warmest quarter (17.1%), and slope (15.6%) significantly added to the output. Unlike the long-term and recent wildfire probability models, fuel was not among the top three contributors of the historic wildfire probability model. However, similar to precipitation of the coldest quarter, precipitation seasonality is a critical environmental factor which affects natural vegetation. Therefore precipitation seasonality acts as a proxy for fuel (Walsh and Lawler 1981).

Relationships between the models' top contributing environmental variables and wildfire probability generally are similar with those discussed for the long-term model (i.e. positive relationship between fuel and wildfire probability). However, population density was the only variable included in the historic model to assess human influence on wildfire probability, of which showed no change in relationship (negative). A detailed analysis of human influence on wildfire probability across the state of California was outside the scope of this study. However, Syphard et al. (2007) were successful in analyzing humans influence on wildfire regimes in California. In brief, they concluded that humans are altering the spatial and temporal pattern of wildfire regimes in California, and is shown through the shift in wildfire location from remote forests to more urbanized environments.

5.2.1. Model Comparison

Wildfire activity and fire severity have increased over the past several decades due to a combination of changes in climate in forest habitats and human influence within lower elevation ecosystems, such as shrublands (Westerling et al. 2006; Littell et al. 2009; Keeley and Syphard 2015). The results of this study show the distribution of suitable wildfire habitat to vary significantly across space over time. Specifically, the study revealed a decrease in suitable burned area habitat from 1984-1888 to 2009-2013. This finding may suggest that although wildfire activity has recently increased in forest ecosystems due to changing climate conditions, humans have played a significant role in influencing the distribution of wildfire occurrence in non-forested environments due to changes in land use (Keeley and Syphard 2015).

In general, changes in the distribution of wildfire probability vary significantly across California. Specifically, wildfire probability has decreased in various regions along the west coast, in addition to various forested areas and regions east of the San Joaquin Basin. This decrease between historic and recent wildfire probability may be directly related to human influence on wildfire probability, whereby increases in human development and population have decreased wildfire at a specific threshold (Syphard et al 2007). For example, in areas of high population and increasing urban development, fuel content tends to be low, reducing suitable habitat for wildfire occurrence.

Although wildfire probability has decreased in areas along the coast of California, multiple regions across the state portray different results. Specifically, wildfire probability has increased in higher elevation forested areas, where roads are less frequent (i.e. remote fires). This may suggest that although land-use has played a crucial factor in wildfire occurrence in lower elevation habitats, recent changes in climate, and particularly drought, have led to an increase in wildfire activity in forested habitats (Westerling et al. 2006; Keeley and Syphard 2015). Although recent research (Westerling et al. 2006) for determining climate impact on wildfire was restricted to large fire events on federally owned land greater than 1370 meters, environmental conditions of increased wildfire probability where generally comparable in this study (i.e. increases in wildfire were seen in areas of higher elevation). Such differences in the wildfire probability models suggest that change in wildfire probability from each of the models (recent and historic) varied significantly across the state of California.

Overall results of this comparative analysis between recent and historic wildfire probability suggest that changes in wildfire potential are highly variable across the state. Understanding such variations within fire regimes (e.g. fire intensity, season, size, and type) was outside the scope of this study. However, recent research suggests that recent changes in wildfire activity vary from region-to-region, such as forested versus non-forested habitats, and may be directly related to recent changes in urban development, population density, and climate.

5.3. Limitation and Future Work

Although results of the wildfire probability models effectively predicted the distribution of wildfire likelihood across California, additional techniques should be evaluated in future work. This study aimed to predict burned area at a single, small cartographic scale, and did not attempt to model variation within nature fire regimes. However, due to differences in fire-climate relationships from one region to the next, spatial context is lost when modeling wildfire probability at small cartographic scales (Little et al. 2009; Parisien and Moritz 2009; Keeley and Syphard 2015). In future work, wildfire probability models created in Maxent should be run on a region-by-region basis to prohibit model under- and overfitting, which in turn may increase the accuracy of the models.

Further, as discussed in Section 3.3.2.3, this study utilized topographic variables (i.e. elevation, slope, and aspect) derived from a 30 arc-second DEM. Based on the large spatial resolution of the dataset (1 km), errors exist and create uncertainty in the analysis. Such errors in the source DEM can greatly affect slope and aspect variables, and subsequently affect overall results of this study (Holmes et al. 2000). However, these errors are assumed to be *de minimis* on the overall results of this study based on the model accuracy and data validation statistics. Future work should address such errors in the DEM dataset in order to reduce uncertainty in the analysis.

Similarly, aspect and slope raster grids averaged for the 25 and 100 km^2 scales contain errors and therefore add uncertainty to the analysis. Specifically, aspect and slope derivatives at the 1 km^2 spatial resolution were portioned into non-overlapping blocks. Values within each block were averaged and the resulting value was assigned to all of the cells in each block to create 25 and 100 km^2 scale outputs. However, averaging 25 and 100 cell values may cause inaccurate representations of aspect and slope. For example, the block average of 25 cells containing seventeen, zero degree cell values (north-facing) and eighteen, 180 degree cell values (south-facing) is 130 degrees (southeast-facing). This method inaccurately misrepresents the seventeen north-facing cells in the raster grid by reassigning the cells a value of 130 degrees (southeast-facing). While these errors create uncertainty due to the use of aspect at the 100 km^2 spatial scale in all final models, practice model runs proved aspect at the source scale (1 km^2) to be a low contributors in each model (long-term, recent, and historic). Although slope is a significant contributor in the models, slope is measured as a continuous value in degrees and therefore the averaged values are assumed to be accurate averaged representations. Consequently, errors in aspect and slope variables at the 25 and 100 km^2 are expected to be minimal and have small effect on the overall results of the study.

To significantly add to our understanding of changes in wildfire probability across California over time, a more in-depth study design and analysis is required. Specifically, the historic wildfire probability model was limited to a single variable to assess human influence on historic wildfire potential (population density). Utilizing additional environmental variables with greater anthropogenic influence, such as housing density and wildland urban interface, among others, would be beneficial additions to the historic and recent models. Further, the recent and historic wildfire probability models were each limited to one, five year time period. Because climate and anthropogenic influences on wildfire have changed significantly overtime, results of this study only present a brief and limited "snapshot" of the changes in wildfire potential over time. Evaluating multiple scales over the past 30 years would provide an elaborate analysis of change in wildfire occurrence across the state of California throughout history. Lastly, limited research to date has aimed to analyze the change in wildfire habitat suitability due to changes in wildfire management practices (Syphard and Keeley 2015). Evaluating the relationship between past and recent wildfire management practices and wildfire activity would greatly add to our understanding of human influence on wildfire habitat suitability.

5.4. Final Thoughts

This study utilized random wildfire presence data, a wide variety of environmental variables, and species distribution modeling techniques to model the spatial distribution of wildfire probability. Specifically, long-term (1984 to 2013), recent (2009 to 2013), and historic (1984 to 1988) wildfire probability models were built using Maxent software to: 1) map the potential distribution of long-term wildfire likelihood across the state of California; 2) investigate the relationship between environmental variables and long-term wildfire probability; and 3) assess the change in the distribution of wildfire probability in California over time.

Based on the results of this analysis, it is evident that species distribution models, and especially Maxent software, is an effective tool for modeling the distribution of wildfire likelihood in California. Although probability outputs are not an absolute representation of wildfire, these provide relative wildfire estimates over long-term periods (Krawchuk et al. 2009; Parisien and Moritz 2009; Parisien et al. 2012). More importantly, these estimations of wildfire likelihood can provide invaluable insight for fire, land, as well as disaster management activities, and in turn can enhance the safety of firefighters and the public, and minimize wildland and property damages.

REFERENCES

Allouche, O., A. Tsoar, and R. Kadmon. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). Journal of Applied Ecology 43 (6): 1223−1232

Andrews, P.L., D. O. Loftsgaarden, L. S. Bradshaw. 2003. Evaluation of fire danger rating indexes using logistic regression and percentile analysis. *International Journal of Wildfire.* 12(2): 213-226.

Baker, W.L. 2013. Is wildland fire increasing in sagebrush landscapes of the western United States? *Association of American Geographers* 103, 5–19.

Balshi M. S., A.D. Mcguire, P. Duffy, M. Flannigan, J. Walsh, J. Melillo. 2009. Assessing the response of area burned to changing climate in western boreal North America using a Multivariate Adaptive Regression Splines (MARS) approach. *Global Change Biology* 15, 578–600.

Bar Massada, A, A.D. Syphard, S.I. Stewart, V.C. Radeloff. 2012. Wildfire ignition-distribution modelling: a comparative study in the Huron-Manistee National Forest, Michigan, USA. *International Journal of Wildfire.* 22(2) 174-183.

Barros, A., J. Pereira, M. Moritz, and S. Stephens. 2013. Spatial characterization of wildfire orientation patterns in California. Forests 4 (1): 197-217.

Bradstock, R.A. 2010. A biogeographic model of fire regimes in Australia: current and future implications. Global Ecology and Biogeography 19, 145–158.

Burke, M. P., T. S. Hogue, M. Ferreira, C. B. Mendez, B. Navarro, S. Lopez, and J. A. Jay. 2010. The effect of wildfire on soil mercury concentrations in southern California watersheds. *Water, Air, & Soil Pollution*. 212 (1): 369-85.

Bowman, D. M., J. K. Balch, P. Artaxo, W. J. Bond, J. M. Carlson, M. A. Cochrane, C. M. D′Antonio, R. S. Defries, J. C. Doyle, S. P. Harrison, F. H. Johnston, J. E. Keeley, M. A. Krawchuk, C. A. Kull, J. B. Marston, M.A. Moritz, I. C. Prentice, C. I. Roos, A. C. Scott, T.W. Swetnam, G. R. van der Werf, and S. J. Pyne. 2009. Fire in the Earth System. *Science* 324(5926): 481–484.

CalFire. 2014a. Incident Information. Available at: http://cdfdata.fire.ca.gov/incidents/incidents_current

CalFire. 2014b. Top 20 Largest California Wildfires. Available at: http://www.fire.ca.gov/communications/downloads/fact_sheets/20lacres.pdf. Published 6, October 2014.

Cardille, J.A., S. J. Ventura; M. G. Turner. 2001. Environmental and social factors influencing wildfires in the upper Midwest, United States. *Ecological Applications.* 11(1): 111-127.

Carmo, M., F. Moreira, R. Casimiro, P. Vaz. 2009. Land use and topography influences on wildfire occurrence in northern Portugal. *Landscape and Urban Planning.* 100 (2011) 169–176.

Cheney, N. P., J. S. Gould, and W. R. Catchpole. 1993. The influence of fuel, weather, and fire shape variables on fire-spread in grasslands. *International Journal of Wildfire*, 3:31-44.

Center for International Earth Science Information Network - CIESIN - Columbia University, and Centro Internacional de Agricultura Tropical - CIAT. 2005. Gridded Population of the World, Version 3 (GPWv3): Population Density Grid. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). Available at: http://sedac.ciesin.columbia.edu/data/set/gpw-v3-population-density.

Cottrell, W. (1989). *The Book of Fire*. Missoula, MT: Mountain Press.

Crimmins, M.A. and A.C. Comrie. 2004. Interactions between antecedent climate and wildfire variability across south-eastern Arizona. *International Journal of Wildfire.* 13: 455-466.

Dennison, P.E., M.A. Moritz, R.S. Taylor. 2008. Evaluating predictive models of critical live fuel moisture in the Santa Monica Mountains, California. *International Journal of Wildland Fire* 17, 18–27.

Dissing, D. and D.L. Verbyla. 2003. Spatial patterns of lightning strikes in interior Alaska and their relationships to elevation and vegetation. *Canadian Journal of Forest Research.* 33.5(2003): 77-782.

Eidenshink, J., B. Schwind, K. Brewer, Z. Zhu, B. Quayle, S. Howard. 2007. A project for monitoring trends in burn severity. *Fire Ecology.* 3(1): 3-21.

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

Elith, J., S.J. Phillips, T. Hastie, M. Dudı´k, Y.E. Chee, C.J. Yates .2011. A statistical explanation of Maxent for ecologists. *Diversity and Distributions*17, 43–57.

Ferrarini, A. 2012. Why not use niche modelling for computing risk of wildfire ignition and spreading? *Environmental Skeptics and Critics.* 1(4):56-60.

Finney, M. A. 1998. FARSITE: Fire Area Simulator—model development and evaluation. Res. Pap. RMRS-RP-4. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 47.

Finney, M. A. 2006. An overview of FlamMap fire modeling capabilities. In: Andrews, P.L., Butler, B.W. (Comps),Fuels Management-How to Measure Success: Conference Proceedings, March 28–30, Portland, OR. USDA Forest Service, Rocky Mountain Research Station Proceedings RMRS-P-41, pp. 213–220.

Finney, M. A., C.W. McHugh, I.C. Grenfell, K.L. Riley, K.C. Short. 2011. A simulation of probabilistic wildfire risk components for the continental United States. *Stochastic Environmental Research and Risk Assessment*. 25(7):973-1000.

Fisher, P. F. and N.J. Tate. 2006. Causes and consequences of error in digital elevation models. *Progress in Physical Geography*. 30(4):467-489

Fitzgerald, S. Oregon State University. *Appendix B. The Effects of Topography, Weather, and Fuel on Fire Behavior.* Available at: https://ir.library.oregonstate.edu/xmlui/bitstream/handle/1957/19402/pnw618zappendixb.pdf?sequence=30.

Fuller, M. 1991. *Forest Fires: An Introduction to Wildfire Behavior, Management, Firefighting, and Prevention.* New York, NY: John Wiley & Sons, Inc.

Forestry Canada Fire Danger Group (FCFDG). Development and structure of the Canadian forest fire behavior prediction system. Forestry Canada, Science and Sustainable Development Directorate, Ottawa, ON, Information Report ST-X-3, 1992.

Friedland, A., R. Relyea, D. Courard-Hauri. 2011. "Environmental Science: Foundations and Applications". New York: W.H. Freeman.

Gedalof Z., D. L. Peterson, N. J. Mantua. 2005. Atmospheric, climatic, and ecological controls on extreme wildfire years in the north-western United States. Ecological Applications 15, 154–174.

Gesch, D.B., and K.S. Larson. 1996. Techniques for development of global 1-kilometer digital elevation models, in Pecora Thirteen, Human Interactions with the Environment--Perspectives from Space, 13th, Sioux Falls, South Dakota, August 20-22, 1996, Proceedings: Bethesda, Maryland, American Society of Photogrammetry and Remote Sensing. Available at: https://lta.cr.usgs.gov/GTOPO30

Griffin D. and K. J. Anchukaitis. 2014. How unusual is the 2012-2014 California drought? American Geophysical Union 41, 9017-9023.

Haines, D.A., W. A. Main, J. S. Frost, A. J. Simard. 1983. Fire-danger rating and wildfire occurrence in the northeastern United-States. *Forest Science.* 29(4): 679-696.

Hijmans, R.J., L. Guarino, P. Mathur. 2012. Diva-GIS Version 7.5 Manual. Available at: http://www.diva-gis.org/docs/DIVA-GIS_manual_7.pdf

Holmes, K.W., O.A. Chadwich, and P.C. Kyriakidis. 2000. Error in a USGS 30-meter digital elevation model and its impact on terrain modeling. *Journal of Hydrology.* 233: 154-173.

IBM Corp. 2013. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.

Jin, S., L. Yang, P. Danielson., C. Homer, J. Fry, G. Xian. 2013. A comprehensive change detection method for updating the National Land Cover Database to circa 2011. Remote Sensing of Environment, 132: 159 – 175.

Keeley, J.E., C.J. Fotheringham, M. Morais. 1999. Reexamining fire suppression impacts on brushland fire regimes. Science 284, 1829–1832.

Keeley, J.E. 2006. South coast bioregion. *In Fire California's Ecosystems*. (Eds N.G. Sugihari, J.W. van Wagtendonk, K.E. Shaffer, J. Fites-Kaufman, A.E. Thode) pp. 350–390. (University of California Press: Berkeley, CA).

Keeley, J.E. 2009. Fire intensity, fire severity and burn severity: A brief review and suggested usage. *International Journal of Wildfire* 18:116-126.

Keeley, J.E. 2014. The association of megafires and extreme droughts in California. Presented at the 99th ESA Annual Meeting, August 12, Sacramento, CA. Available at: http://eco.confex.com/eco/2014/webprogram/Paper45575.html

Keeley, J.E. and A.D. Syphard. 2015. Different fire-climate relationships on forested and non-forested landscapes in the Sierra Nevada ecoregion. *International Journal of Wildland Fire*. 24:27-36.

Krawchuk, M. A., M. A. Moritz, M-A. Parisien, J. Van Dorn, and K. Hayhoe. 2009. Global pyrogeography: the current and future distribution of wildfire. *PLoS ONE* 4, e5102.

Landis, J.R. and G.G. Koch. 1977. The measurements of observer agreement for categorical data. Biometrics 33 (1): 159–174.

Littell, J.S., D. McKenzie, D. L. Peterson, A. L. Westerling. 2009. Climate and wildfire area burned in western US ecoprovinces, 1916–2003. *Ecological Applications* 19, 1003–1021.

Manel, S., H. C. Williams, S. J. Ormerod. 2001. Evaluating presence–absence models in ecology: the need to account for prevalence. Journal of Applied Ecology 38 (5): 921–931.

Martell, D.L., S. Otukol, B. J. Stocks. 1987. A logistic model for predicting daily people-caused forest fire occurrence in Ontario. *Canadian Journal of Forest Research.* 17(5): 394-401.

Miller J.D., H.D. Safford, M. Crimmins, A.E. Thode. 2009. Quantitative evidence for increasing forest fire severity in the Sierra Nevada and southern Cascade mountains, California and Nevada, USA. *Ecosystems.* 12, 16–32.

Moritz, M.A., T.J. Moody, M.A. Krawchuck, M. Highes, A. Hall. 2010. Spatial variation in extreme winds predicts large wildfire locations in chaparral ecosystems. *Geophysical Research Letters.*

National Oceanic and Atmospheric Administration (NOAA). 2015b. National Climatic Data Center. Lightning products and services. Available at: http://www.ncdc.noaa.gov/data-access/severe-weather/lightning-products-and-services.

National Renewable Energy Laboratory (NREL) and True Winds Solutions. 2003. Wind Data. Available at: http://www.nrel.gov/gis/data_wind.html.

Noble, I. R., G. A. V. Bary, and A. M. Gill. McArthur's fire danger meters expressed as equations. *Australian Journal of Ecology,* 5:201--203, 1980.

Parisien M-A, V. S. Peters, Y. Wang, J. M. Little, E. M. Bosch, B. J. Stocks. 2006. Spatial patterns of forest fires in Canada, 1980–1999. *International Journal of Wildfire* 15, 361–374.

Price, C.V., N. Nakagaki, K.J. Hitt, R.C. Clawges. 2007. Enhanced Historical Land-Use and Land-Cover Data Sets of the U.S. Geological Survey: polygon format files, U.S. Geological Survey Digital Data Series 240. Available at: http://water.usgs.gov/GIS/metadata/usgswrd/XML/ds240_landuse_poly.xml.

O'Donnell, M.S. and D.A. Ignizio. 2012. Bioclimatic predictors for supporting ecological applications in the conterminous United States. U.S. Geological Survey Data Series 691.

Parisien, M-A. and M. A. Moritz. 2009. Environmental controls on the distribution of wildfire at multiple spatial scales. *Ecological Monographs.* 79(1): 127-154.

Parisien M-A, S. A. Parks, M. A. Krawchuk , M. D. Flannigan, L. M. Bowman, M. A. Moritz. 2011. Scale-dependent factors controlling area burned in boreal Canada. *Ecological Applications* 21, 789–805.

Parisien, M-A., S. Snetsinger, J. Greenberg, C. Nelson, T. Schoennagel, S. Dobrowski, and M. Maritz. 2012. Spatial variability in wildfire probability across the western United States. *International Journal of Wildfire.* 21: 313-327.

Paritsis, J., H. Andres, T.T. Veblen, T. Kitzberger. 2013. Habitat distribution modeling reveals vegetation flammability and land use as drivers of wildfire in SW Patagonia. *Ecosphere.* 4(5):53.

Peterson, A.T. 2006. Uses and requirements of ecological niche models and related distributional models. *Biodiversity Informantics*. 3: 59-72.

Phillips, S. J., R. P. Anderson, and R. E. Shapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190, 231–259.

Pincetl S., P. W. Rundel, J. C. Blasio, D. Silver, T. Scott, J. E. Keeley, R. Halsey. 2008. It's the Land Use, Not the Fuels: Fires and Land Development in Southern California. *Real Estate Rev.*; 37(1): 25-42.

Preisler, H. K., D. R. Brillinger, R. E. Burgan, J. W. Benoit. 2004. Probability based models for estimation of wildfire risk. *International Journal of Wildfire*. 13(2): 133-142.

Preisler, H. K., R. E. Burgan, J. C. Eidenshink, J. M. Klaver, R. W. Klaver. 2009. Forecasting distributions of large federal-lands fires utilizing satellite and gridded weather information. *International Journal of Wildfire.* 18(5): 508-516.

Prestemon, J. P., T. J. Hawbaker, M. Bowden, J. Carpenter, M. T. Brooks, K. L. Abt, R. Sutphen, S. Scranton. 2013. Wildfire Ignitions: A Review of the Science and Recommendation for Empirical Modeling. United States Department of Agriculture: Forest Service. General Technical Report SRS-171.

PRISM Group. 2004. Climate Data for Recent Years, 1984–2013. (Oregon State University: Corvallis, OR). Available at http://www.prismclimate.org.

Randall C. K. and M. Duryea. 2003. Fire in the Wildland-Urban Interface: Understand Fire Behavior. Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, University of Florida. Retrieved from: https://edis.ifas.ufl.edu/pdffiles/FR/FR13800.pdf.

Redfish Group. Farsite Image: St. John's Wilderness Gate Scenario. Available at: http://www.redfish.com/wildfire/fireScenarios/wildernessGateHomes.jpg.

Rothermel, R.C.. A mathematical model for predicting fire spread in wildfires. USDA Forest Service Research Paper INT-115, 1972.

Scott J. H. and E. D. Reinhardt. 2001. Assessing crown fire potential by linking models of surface and crown fire behavior. Research Paper. RMRS-RP-29. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. p. 59.

Snider G., P. J. Daugherty, D. Wood. 2007. The irrationality of continued fire suppression: an avoided cost analysis of fire hazard reduction treatments versus no treatment. *Journal Of Forestry*; 104: 431-437.

Sommers, W.T., S.G. Coloff, S.G. Conard (2011). "Synthesis of Knowledge: Fire History and Climate Change" (2011). JFSP Synthesis Reports. Paper 19. Available at: http://digitalcommons.unl.edu/jfspsynthesis/19

Stephens, S. L. 2005. Forest fire causes and extent on United States Forest Service lands. *International Journal of Wildfire* 14, 213–222.

Stephens, S., J. Moghaddas, C. Edminster, C. Fiedler, S. Haase, M. Harrington, J. Keeley, E. Knapp, J. McIver, K.K. Metlen, C.N. Skinner, A. Youngblood. 2009. Fire treatment effects on vegetation structure, fuels, and potential fire severity in western U.S. forests. *Ecological Applications*. 19, 305–320.

Stephenson, N.L. 1998. Actual evapotranspiration and deficit: biologically meaningful correlates of vegetation distribution across spatial scales. *Journal of Biogeography* 25, 855–870.

Stocks B.J., J. A. Mason, J. B. Todd, E. M. Bosch, B. M. Wotton, B. D. Amiro, M. D. Flannigan, K. G. Hirsch, K. A. Logan, D. L. Martell, W. R. Skinner. 2002. Large forest fires in Canada, 1959–1997. *Journal of Geophysical Research—Atmospheres* 108(D1), FFR5-1–FFR5-12.

Sugihara, N. G., J.W. van Wagtendonk, K.E. Shaffer. 2006. Fire and California Vegetation. In *Fire in California's Ecosystems*. University of California Press.

Syphard A. D., V. C. Radelof, J. E. Keeley, T. J. Hawbaker, M. K. Clayton, S. I. Stewart, R. B. Hammer. 2007. Human influence on California fire regimes. *Ecological Applications* 17, 1388–1402.

Syphard, A. D., V. C. Radeloff, N. S. Keuler, R. S. Taylor, T. J. Hawbaker, S. I. Stewart, M. K. Clayton. 2008. Predicting spatial patterns of fire on a southern California landscape. *International Journal of Wildfire* 17, 602–613.

Syphard, A. D., V. C. Radeloff, T. J. Hawbaker, S. I. Stewart. 2009. Conservation threats due to human-caused increases in fire frequency in Mediterranean-climate ecosystems. *Conservation Biology* 23(3): 758-769.

Syphard, A.D., J.E.Keeley, A. Bar Massada, T.J. Brenna, V.C. Radeloff. 2012. Housing arrangement and location determine the likelihood of housing loss due to wildfire. *PLoS ONE* 7(3), e33954.

Syphard, A.D. and J.E.Keeley. 2015. Location, timing, and extent of wildfire vary by cause of ignition. *International Journal of Wildfire*. 24, 27-47.

TIGER/Line Shapefiles. 2014. Primary and Secondary Roads/ prepared by the U.S. Census Bureau. 2013. Available at: https://www.census.gov/geo/maps-data/data/tiger-line.html.

Tymstra, C.; R.W. Bryce; B.M. Wotton; O.B. Armitage. 2009. Development and structure of Prometheus: the Canadian wildfire growth simulation Model. Inf. Rep. NOR-X-417. Nat. Resour. Can., Can. For. Serv., North. For. Cent., Edmonton, AB.. Prometheus.

US Geological Survey (USGS). 1996. Global 30 Arc-Second Elevation Data Set. (EROS Data Center Distributed Active Archive Center: Sioux Falls, SD). Available at http://eros.usgs.gov/elevation-products.

US Geological Survey (USGS). 2003. MODIS 17A3 Gross Primary Productivity. Maintained by the NASA Land Processes Distributed Active Archive Center (LP DAAC) at the USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota, 2003. Available at: https://lpdaac.usgs.gov.

Walsh R.P.D. and D.M. Lawler. 1981. Rainfall seasonality: description, spatial patterns and change through time. *Weather*. 36(7): 201-208.

Westerling, A. L., , D.R. Cayan, T.J. Brown, B.L. Hall and L.G. Riddle 2004: Climate, Santa Ana Winds and Autumn Wildfires in Southern California. *EOS, Transactions, American Geophysical Union,* 85(31): 289,296. Available at: http://ulmo.ucmerced.edu/pdffiles/04EOS_Westerling.pdf.

Westerling, A. L., H. G. Hidalgo, D. R. Cayan, T. W. Swetnam. 2006. Warming and earlier spring increases western US forest wildfire activity. Science 313, 940–943.

Westerling A. L. and B. P. Bryant. 2008. Climate change and wildfire in California. *Climate Change*. 87 (Suppl 1):S231-S249.

Westerling, A. L., B. P. Bryant, H. K. Preisler, T. P. Holmes, H. G. Hidalgo, T. Das, and S. R. Shrestha. 2011. Climate change and growth scenarios for California wildfire. *Climatic Change*. 109 (Suppl 1):S445–S463.

Viegas, D.X. 1998. Weather, fuel status and fire occurrence: predicting large fires. In: J.M. Moreno (Ed.) Large Forest Fires, Backhuys, Leiden, pp. 31-49.

Yue, X., L.J. Mickley, J.A. Logan. 2014. Projection of wildfire activity in southern California in the mid-twenty-first century. *Climate Dynamics.* 43:1973-1991.

Zaksek M. and J. L. Arvai. 2004. Toward improved communication about wildfire: mental models research to identidy information needs for natural resource management. *Risk Analysis*. 24(6):1503-14.

Ziesler, P. S. 2013. Modeling burn probability patterns for large fires. Colorado State University. ProQuest LLC. Dissertation.

Appendix A: SPSS Outputs

Table A1. Long-term Exploratory Analysis Output

Correlations

		Bio1	Bio2	Bio3	Bio4	Bio5	Bio6	Bio7	Bio8	Bio9	Bio10	Bio11	Bio12	Bio13
Bio1	Pearson Correlation	1	-.075**	.311**	-.263**	.642**	.875**	-.244**	.902**	.893**	.892**	.932**	.380**	.419**
	Sig. (2-tailed)		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535
Bio2	Pearson Correlation	-.075**	1	.107**	.563**	.588**	-.472**	.849**	-.294**	.142**	.144**	-.313**	-.545**	-.527**
	Sig. (2-tailed)	.000		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535
Bio3	Pearson Correlation	.311**	.107**	1	-.741**	-.102**	.413**	-.426**	.502**	-.057**	-.059**	.494**	.203**	.234**
	Sig. (2-tailed)	.000	.000		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535
Bio4	Pearson Correlation	-.263**	.563**	-.741**	1	.500**	-.618**	.904**	-.580**	.191**	.194**	-.585**	-.542**	-.555**
	Sig. (2-tailed)	.000	.000	.000		.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535

	Pearson Correlation	.642**	.588**	-.102**	.500**	1	.231**	.574**	.328**	.868**	.870**	.342**	-.141**	-.106**
Bio 5	Sig. (2-tailed)	.000	.000	.000	.000		.000	.000	.000	.000	.000	.000	.000	.000
	N	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535
	Pearson Correlation	.875**	-.472**	.413**	-.618**	.231**	1	-.665**	.943**	.621**	.620**	.976**	.644**	.677**
Bio 6	Sig. (2-tailed)	.000	.000	.000	.000	.000		.000	.000	.000	.000	.000	.000	.000
	N	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535
	Pearson Correlation	-.244**	.849**	-.426**	.904**	.574**	-.665**	1	-.542**	.144**	.146**	-.559**	-.650**	-.651**
Bio 7	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000		.000	.000	.000	.000	.000	.000
	N	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535
	Pearson Correlation	.902**	-.294**	.502**	-.580**	.328**	.943**	-.542**	1	.652**	.651**	.970**	.525**	.561**
Bio 8	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000		.000	.000	.000	.000	.000
	N	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535
	Pearson Correlation	.893**	.142**	-.057**	.191**	.868**	.621**	.144**	.652**	1	.999**	.682**	.178**	.212**
Bio 9	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000		.000	.000	.000	.000
	N	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535

Bio 10	Pearson Correlation	.892**	.144**	-.059**	.194**	.870**	.620**	.146**	.651**	.999**	1	.681**	.177**	.211**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000		.000	.000	.000
	N	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535
Bio 11	Pearson Correlation	.932**	-.313**	.494**	-.585**	.342**	.976**	-.559**	.970**	.682**	.681**	1	.558**	.596**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000		.000	.000
	N	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535
Bio 12	Pearson Correlation	.380**	-.545**	.203**	-.542**	-.141**	.644**	-.650**	.525**	.178**	.177**	.558**	1	.989**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000		.000
	N	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535
Bio 13	Pearson Correlation	.419**	-.527**	.234**	-.555**	-.106**	.677**	-.651**	.561**	.212**	.211**	.596**	.989**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	
	N	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535
Bio 14	Pearson Correlation	-.421**	-.277**	-.281**	.036**	-.408**	-.236**	-.114**	-.316**	-.383**	-.383**	-.335**	.202**	.156**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535

	Pearson Correlation	.662**	-.398**	.253**	-.463**	.178**	.799**	-.536**	.693**	.491**	.492**	.759**	.714**	.763**
Bio 15	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535
	Pearson Correlation	.432**	-.528**	.219**	-.543**	-.092**	.682**	-.645**	.564**	.231**	.230**	.602**	.994**	.996**
Bio 16	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535
	Pearson Correlation	-.174**	-.464**	-.181**	-.175**	-.385**	.074**	-.357**	-.042**	-.220**	-.218**	-.039**	.614**	.548**
Bio 17	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535
	Pearson Correlation	-.190**	-.397**	-.187**	-.128**	-.353**	.024**	-.291**	-.076**	-.225**	-.221**	-.075**	.561**	.496**
Bio 18	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535
	Pearson Correlation	.439**	-.528**	.223**	-.546**	-.088**	.687**	-.646**	.567**	.236**	.235**	.608**	.992**	.994**
Bio 19	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535	65535

Correlations

		Bio14	Bio15	Bio16	Bio17	Bio18	Bio19
	Pearson Correlation	-.421	.662**	.432**	-.174**	-.190**	.439**
Bio1	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	65535	65535	65535	65535	65535	65535
	Pearson Correlation	-.277**	-.398	-.528**	-.464**	-.397**	-.528**
Bio2	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	65535	65535	65535	65535	65535	65535
	Pearson Correlation	-.281**	.253**	.219	-.181**	-.187**	.223**
Bio3	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	65535	65535	65535	65535	65535	65535
	Pearson Correlation	.036**	-.463**	-.543**	-.175	-.128**	-.546**
Bio4	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	65535	65535	65535	65535	65535	65535
	Pearson Correlation	-.408**	.178**	-.092**	-.385**	-.353	-.088**
Bio5	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	65535	65535	65535	65535	65535	65535
	Pearson Correlation	-.236**	.799**	.682**	.074**	.024**	.687
Bio6	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	65535	65535	65535	65535	65535	65535
	Pearson Correlation	-.114**	-.536**	-.645**	-.357**	-.291**	-.646**
Bio7	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	65535	65535	65535	65535	65535	65535
Bio8	Pearson Correlation	-.316**	.693**	.564**	-.042**	-.076**	.567**

	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	65535	65535	65535	65535	65535	65535
	Pearson Correlation	-.383**	.491**	.231**	-.220**	-.225**	.236**
Bio9	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	65535	65535	65535	65535	65535	65535
	Pearson Correlation	-.383**	.492**	.230**	-.218**	-.221**	.235**
Bio10	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	65535	65535	65535	65535	65535	65535
	Pearson Correlation	-.335**	.759**	.602**	-.039**	-.075**	.608**
Bio11	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	65535	65535	65535	65535	65535	65535
	Pearson Correlation	.202**	.714**	.994**	.614**	.561**	.992**
Bio12	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	65535	65535	65535	65535	65535	65535
	Pearson Correlation	.156**	.763**	.996**	.548**	.496**	.994**
Bio13	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	65535	65535	65535	65535	65535	65535
	Pearson Correlation	1**	-.245**	.141**	.809**	.805**	.124**
Bio14	Sig. (2-tailed)		.000	.000	.000	.000	.000
	N	65535	65535	65535	65535	65535	65535
	Pearson Correlation	-.245**	1**	.769**	.087**	.031**	.777**
Bio15	Sig. (2-tailed)	.000		.000	.000	.000	.000
	N	65535	65535	65535	65535	65535	65535
Bio16	Pearson Correlation	.141**	.769**	1**	.549**	.495**	.999**

	Sig. (2-tailed)	.000	.000		.000	.000	.000
	N	65535	65535	65535	65535	65535	65535
	Pearson Correlation	.809**	.087**	.549**	1**	.976**	.536**
Bio17	Sig. (2-tailed)	.000	.000	.000		.000	.000
	N	65535	65535	65535	65535	65535	65535
	Pearson Correlation	.805**	.031**	.495**	.976**	1**	.482**
Bio18	Sig. (2-tailed)	.000	.000	.000	.000		.000
	N	65535	65535	65535	65535	65535	65535
	Pearson Correlation	.124**	.777**	.999**	.536**	.482**	1**
Bio19	Sig. (2-tailed)	.000	.000	.000	.000	.000	
	N	65535	65535	65535	65535	65535	65535

**. Correlation is significant at the 0.01 level (2-tailed).

Table A2. Recent Exploratory Analysis Output

Correlations

		Bio1	Bio2	Bio3	Bio4	Bio5	Bio6	Bio7	Bio8	Bio9	Bio10	Bio11	Bio12	Bio13
	Pearson Correlation	1	.266**	.088**	.235**	.866**	.858**	.213**	.884**	.810**	.939**	.924**	-.520**	-.513**
Bio1	Sig. (2-tailed)		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806
Bio2	Pearson Correlation	.266**	1	.253**	.382**	.562**	-.099**	.728**	.124**	.290**	.348**	.096**	-.461**	-.461**

	Sig. (2-tailed)	.000		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806
	Pearson Correlation	.088**	.253**	1	-.761**	-.166**	.306**	-.465**	.346**	-.033**	-.202**	.361**	-.046**	-.021**
Bio 3	Sig. (2-tailed)	.000	.000		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806
	Pearson Correlation	.235**	.382**	-.761**	1	.616**	-.219**	.897**	-.124**	.278**	.550**	-.146**	-.329**	-.353**
Bio 4	Sig. (2-tailed)	.000	.000	.000		.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806
	Pearson Correlation	.866**	.562**	-.166**	.616**	1	.539**	.652**	.638**	.799**	.958**	.641**	-.560**	-.571**
Bio 5	Sig. (2-tailed)	.000	.000	.000	.000		.000	.000	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806
	Pearson Correlation	.858**	-.099**	.306**	-.219**	.539**	1	-.287**	.918**	.693**	.667**	.968**	-.227**	-.219**
Bio 6	Sig. (2-tailed)	.000	.000	.000	.000	.000		.000	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806
Bio 7	Pearson Correlation	.213**	.728**	-.465**	.897**	.652**	-.287**	1	-.101**	.285**	.489**	-.142**	-.432**	-.452**

	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000		.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806
	Pearson Correlation	.884**	.124**	.346**	-.124**	.638**	.918**	-.101**	1	.713**	.717**	.948**	-.371**	-.371**
Bio 8	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000		.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806
	Pearson Correlation	.810**	.290**	-.033**	.278**	.799**	.693**	.285**	.713**	1	.800**	.726**	-.323**	-.328**
Bio 9	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000		.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806
	Pearson Correlation	.939**	.348**	-.202**	.550**	.958**	.667**	.489**	.717**	.800**	1	.745**	-.539**	-.541**
Bio 10	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000		.000	.000	.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806
	Pearson Correlation	.924**	.096**	.361**	-.146**	.641**	.968**	-.142**	.948**	.726**	.745**	1	-.363**	-.347**
Bio 11	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000		.000	.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806
Bio 12	Pearson Correlation	-.520**	-.461**	-.046**	-.329**	-.560**	-.227**	-.432**	-.371**	-.323**	-.539**	-.363**	1	.987**

	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000		.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806
	Pearson Correlation	-.513**	-.461**	-.021**	-.353**	-.571**	-.219**	-.452**	-.371**	-.328**	-.541**	-.347**	.987**	1
Bio 13	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806
	Pearson Correlation	-.654**	-.169**	-.259**	.053**	-.503**	-.656**	.019**	-.657**	-.578**	-.541**	-.683**	.321**	.288**
Bio 14	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806
	Pearson Correlation	.414**	.131**	.355**	-.220**	.229**	.452**	-.146**	.389**	.286**	.286**	.510**	-.259**	-.159**
Bio 15	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806
	Pearson Correlation	-.499**	-.464**	.008**	-.382**	-.567**	-.188**	-.476**	-.343**	-.309**	-.538**	-.320**	.990**	.993**
Bio 16	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806
Bio 17	Pearson Correlation	-.621**	-.311**	-.231**	-.055**	-.525**	-.507**	-.140**	-.564**	-.549**	-.542**	-.591**	.697**	.654**

	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806
Bio18	Pearson Correlation	-.314**	-.257**	-.327**	.159**	-.244**	-.312**	.003	-.344**	-.462**	-.208**	-.366**	.439**	.397**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.062	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806
Bio19	Pearson Correlation	-.495**	-.465**	.021**	-.396**	-.571**	-.180**	-.488**	-.340**	-.306**	-.540**	-.311**	.983**	.988**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806

Correlations

		Bio14	Bio15	Bio16	Bio17	Bio18	Bio19
Bio1	Pearson Correlation	-.654	.414**	-.499**	-.621**	-.314**	-.495**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806
Bio2	Pearson Correlation	-.169**	.131	-.464**	-.311**	-.257**	-.465**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806
Bio3	Pearson Correlation	-.259**	.355**	.008	-.231**	-.327**	.021**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000

	N	407806	407806	407806	407806	407806	407806
Bio4	Pearson Correlation	.053**	-.220**	-.382**	-.055	.159**	-.396**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806
Bio5	Pearson Correlation	-.503**	.229**	-.567**	-.525**	-.244	-.571**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806
Bio6	Pearson Correlation	-.656**	.452**	-.188**	-.507**	-.312**	-.180
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806
Bio7	Pearson Correlation	.019**	-.146**	-.476**	-.140**	.003**	-.488**
	Sig. (2-tailed)	.000	.000	.000	.000	.062	.000
	N	407806	407806	407806	407806	407806	407806
Bio8	Pearson Correlation	-.657**	.389**	-.343**	-.564**	-.344**	-.340**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806
Bio9	Pearson Correlation	-.578**	.286**	-.309**	-.549**	-.462**	-.306**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806
Bio10	Pearson Correlation	-.541**	.286**	-.538**	-.542**	-.208**	-.540**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806
Bio11	Pearson Correlation	-.683**	.510**	-.320**	-.591**	-.366**	-.311**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000

	N	407806	407806	407806	407806	407806	407806
Bio12	Pearson Correlation	.321**	-.259**	.990**	.697**	.439**	.983**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806
Bio13	Pearson Correlation	.288**	-.159**	.993**	.654**	.397**	.988**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806
Bio14	Pearson Correlation	1**	-.477**	.269**	.758**	.584**	.260**
	Sig. (2-tailed)		.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806
Bio15	Pearson Correlation	-.477**	1**	-.164**	-.492**	-.469**	-.144**
	Sig. (2-tailed)	.000		.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806
Bio16	Pearson Correlation	.269**	-.164**	1**	.640**	.378**	.997**
	Sig. (2-tailed)	.000	.000		.000	.000	.000
	N	407806	407806	407806	407806	407806	407806
Bio17	Pearson Correlation	.758**	-.492**	.640**	1**	.824**	.622**
	Sig. (2-tailed)	.000	.000	.000		.000	.000
	N	407806	407806	407806	407806	407806	407806
Bio18	Pearson Correlation	.584**	-.469**	.378**	.824**	1**	.359**
	Sig. (2-tailed)	.000	.000	.000	.000		.000
	N	407806	407806	407806	407806	407806	407806
Bio19	Pearson Correlation	.260**	-.144**	.997**	.622**	.359**	1**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	

N		407806	407806	407806	407806	407806	407806

**. Correlation is significant at the 0.01 level (2-tailed).

Table A3. Historic Exploratory Analysis Output

Correlations

		Bio1	Bio2	Bio3	Bio4	Bio5	Bio6	Bio7	Bio8	Bio9	Bio10	Bio11	Bio12	Bio13
Bio1	Pearson Correlation	1	.254**	.085**	.180**	.874**	.868**	.191**	.820**	.862**	.945**	.936**	-.498**	-.498**
	Sig. (2-tailed)		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806
Bio2	Pearson Correlation	.254**	1	.180**	.387**	.543**	-.095**	.723**	.128**	.285**	.348**	.101**	-.422**	-.427**
	Sig. (2-tailed)	.000		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806
Bio3	Pearson Correlation	.085**	.180**	1	-.803**	-.194**	.323**	-.535**	.178**	-.101**	-.192**	.354**	-.019**	.004*
	Sig. (2-tailed)	.000	.000		.000	.000	.000	.000	.000	.000	.000	.000	.000	.012
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806

	Pearson Correlation	.180**	.387**	-.803**	1	.561**	-.263**	.904**	.013**	.311**	.489**	-.171**	-.318**	-.346**
Bio 4	Sig. (2-tailed)	.000	.000	.000		.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806
	Pearson Correlation	.874**	.543**	-.194**	.561**	1	.570**	.621**	.639**	.866**	.959**	.676**	-.515**	-.527**
Bio 5	Sig. (2-tailed)	.000	.000	.000	.000		.000	.000	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806
	Pearson Correlation	.868**	-.095**	.323**	-.263**	.570**	1	-.290**	.766**	.715**	.684**	.968**	-.222**	-.213**
Bio 6	Sig. (2-tailed)	.000	.000	.000	.000	.000		.000	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806
	Pearson Correlation	.191**	.723**	-.535**	.904**	.621**	-.290**	1	.013**	.328**	.465**	-.136**	-.388**	-.411**
Bio 7	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000		.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806
	Pearson Correlation	.820**	.128**	.178**	.013**	.639**	.766**	.013**	1	.624**	.732**	.815**	-.396**	-.392**
Bio 8	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000		.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806

	Pearson Correlation	.862**	.285**	-.101**	.311**	.866**	.715**	.328**	.624**	1	.862**	.752**	-.348**	-.347**
Bio 9	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000		.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806
	Pearson Correlation	.945**	.348**	-.192**	.489**	.959**	.684**	.465**	.732**	.862**	1	.775**	-.539**	-.549**
Bio 10	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000		.000	.000	.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806
	Pearson Correlation	.936**	.101**	.354**	-.171**	.676**	.968**	-.136**	.815**	.752**	.775**	1	-.358**	-.349**
Bio 11	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000		.000	.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806
	Pearson Correlation	-.498**	-.422**	-.019**	-.318**	-.515**	-.222**	-.388**	-.396**	-.348**	-.539**	-.358**	1	.988**
Bio 12	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000		.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806
	Pearson Correlation	-.498**	-.427**	.004*	-.346**	-.527**	-.213**	-.411**	-.392**	-.347**	-.549**	-.349**	.988**	1
Bio 13	Sig. (2-tailed)	.000	.000	.012	.000	.000	.000	.000	.000	.000	.000	.000	.000	
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806

Bio 14	Pearson Correlation	-.724**	-.109**	-.313**	.180**	-.533**	-.781**	.124**	-.624**	-.652**	-.579**	-.788**	.230**	.184**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806
Bio 15	Pearson Correlation	.206**	-.102**	.485**	-.551**	-.022**	.460**	-.465**	.218**	.285**	-.004**	.403**	.134**	.215**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.008	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806
Bio 16	Pearson Correlation	-.465**	-.425**	.017**	-.354**	-.501**	-.174**	-.418**	-.370**	-.316**	-.522**	-.313**	.994**	.993**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806
Bio 17	Pearson Correlation	-.716**	-.268**	-.330**	.088**	-.559**	-.672**	-.011**	-.597**	-.620**	-.598**	-.733**	.639**	.586**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806
Bio 18	Pearson Correlation	-.416**	-.211**	-.357**	.217**	-.304**	-.449**	.074**	-.219**	-.516**	-.286**	-.475**	.412**	.354**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806

Bio 19	Pearson Correlation	-.459**	-.418**	.032**	-.365**	-.498**	-.165**	-.423**	-.371**	-.312**	-.521**	-.304**	.991**	.989**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806	407806

Correlations

		Bio14	Bio15	Bio16	Bio17	Bio18	Bio19
Bio1	Pearson Correlation	-.724	.206**	-.465**	-.716**	-.416**	-.459**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806
Bio2	Pearson Correlation	-.109**	-.102	-.425**	-.268**	-.211**	-.418**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806
Bio3	Pearson Correlation	-.313**	.485**	.017	-.330**	-.357**	.032**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806
Bio4	Pearson Correlation	.180**	-.551**	-.354**	.088	.217**	-.365**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806
Bio5	Pearson Correlation	-.533**	-.022**	-.501**	-.559**	-.304	-.498**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806
Bio6	Pearson Correlation	-.781**	.460**	-.174**	-.672**	-.449**	-.165

	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806
	Pearson Correlation	.124**	-.465**	-.418**	-.011**	.074**	-.423**
Bio7	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806
	Pearson Correlation	-.624**	.218**	-.370**	-.597**	-.219**	-.371**
Bio8	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806
	Pearson Correlation	-.652**	.285**	-.316**	-.620**	-.516**	-.312**
Bio9	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806
	Pearson Correlation	-.579**	-.004**	-.522**	-.598**	-.286**	-.521**
Bio10	Sig. (2-tailed)	.000	.008	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806
	Pearson Correlation	-.788**	.403**	-.313**	-.733**	-.475**	-.304**
Bio11	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806
	Pearson Correlation	.230**	.134**	.994**	.639**	.412**	.991**
Bio12	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806
	Pearson Correlation	.184**	.215**	.993*	.586**	.354**	.989**
Bio13	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806
Bio14	Pearson Correlation	1**	-.606**	.164**	.791**	.595**	.160**

	Sig. (2-tailed)		.000	.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806
	Pearson Correlation	-.606**	1**	.218**	-.525**	-.635**	.226**
Bio15	Sig. (2-tailed)	.000		.000	.000	.000	.000
	N	407806	407806	407806	407806	407806	407806
	Pearson Correlation	.164**	.218**	1**	.567**	.342**	.997**
Bio16	Sig. (2-tailed)	.000	.000		.000	.000	.000
	N	407806	407806	407806	407806	407806	407806
	Pearson Correlation	.791**	-.525**	.567**	1**	.798**	.556**
Bio17	Sig. (2-tailed)	.000	.000	.000		.000	.000
	N	407806	407806	407806	407806	407806	407806
	Pearson Correlation	.595**	-.635**	.342**	.798**	1**	.328**
Bio18	Sig. (2-tailed)	.000	.000	.000	.000		.000
	N	407806	407806	407806	407806	407806	407806
	Pearson Correlation	.160**	.226**	.997**	.556**	.328**	1**
Bio19	Sig. (2-tailed)	.000	.000	.000	.000	.000	
	N	407806	407806	407806	407806	407806	407806

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).