Ecological Monographs, 79(1), 2009, pp. 127–154

Environmental controls on the distribution of wildfire at multiple spatial scales

Marc-André Parisien[1,2,3] and Max A. Moritz[1]

[1]*Department of Environmental Sciences, Policy and Management, University of California, 137 Mulford Hall, Berkeley, California 94720 USA*
[2]*Natural Resources Canada, Canadian Forest Service, Northern Forestry Centre, 5320-122nd St., Edmonton, Alberta T5H 3S5 Canada*

Abstract. Despite its widespread occurrence globally, wildfire preferentially occupies an environmental middle ground and is significantly less prevalent in biomes characterized by environmental extremes (e.g., tundra, rain forests, and deserts). We evaluated the biophysical "environmental space" of wildfire from regional to subcontinental extents, with methods widely used for modeling habitat distributions. This approach is particularly suitable for the biogeographic study of wildfire, because it simultaneously considers patterns in multiple factors controlling wildfire suitability over large areas. We used the Maxent and boosted regression tree algorithms to assess wildfire–environment relationships for three levels of complexity (in terms of inclusion of variables) at three spatial scales: the conterminous United States, the state of California, and five wildfire-prone ecoregions of California. The resulting models were projected geographically to obtain spatial predictions of wildfire suitability and were also applied to other regions to assess their generality and spatial transferability. Predictions of the potential range of wildfire had high classification accuracy; they also highlighted areas where wildfires had not recently been observed, indicating the potential (or past) suitability of these areas. The models identified several key variables that were not suspected to be important in the large-scale control of wildfires, but which might indirectly affect control by influencing the presence of flammable vegetation. Models transferred to different areas were useful only when they overlapped appreciably with the target area's environmental space. This approach should allow exploration of the potential shifts in wildfire range in a changing climate, the potential for restoration of wildfire where it has been "extirpated," and, conversely, the "invasiveness" of wildfire after changes in plant species composition. Our study demonstrates that habitat distribution models and related concepts can be used to characterize environmental controls on a natural disturbance process, but also that future work is needed to refine our understanding of the direct causal factors controlling wildfire at multiple spatial scales.

Key words: boosted regression trees; disturbance ecology; environmental space; fire regime controls; habitat distribution models; Maxent algorithm; spatial fire prediction; wildfire.

Introduction

One pillar of ecology and conservation biology is the study of the distribution of biota as a function of their basic habitat requirements (Hutchinson 1959, MacArthur 1972, Brown 1984, Schemske et al. 1994); however, analogous characterization of abiotic patterns and processes is rare. Wildfire is an abiotic ecological process that is strongly regulated by its environment. Despite abundant research quantifying the local response, behavior, and landscape regimes of fire as a function of environmental variables, we lack studies that describe the spatial distribution and envrironmental requirements of fire over broad scales as a function of multiple environmental gradients. Such assessments are needed to refine understanding of controls on fire and fire regimes at broad spatial scales (regional to continental) while allowing mapping of the likelihood and potential spatial extent of wildfire under both current and future climate scenarios.

Fire occurs as a function of suitable environmental conditions—the co-occurrence of adequate fuels, conditions conducive to combustion and spread, and ignition agents. Clearly, more frequent or more numerous consecutive days of hot, dry, and windy weather will translate into more numerous and often larger wildfires when these conditions coincide with the presence of ignitions. Ignitions can be a limiting factor; therefore, patterns of lightning and human land use can have a substantial influence on the occurrence of fire. Humans ignite most of the fires in North America, typically in highly clustered spatial patterns, but humans also extinguish ignitions that could have become large wildfires. Time since the most recent fire and accumulation of biomass can also drive fire probabilities in many ecosystems. It is crucial, however, that suitable conditions coincide in both space and time. For

Manuscript received 6 August 2007; revised 1 May 2008; accepted 6 May 2008. Corresponding Editor: F. He.
[3] E-mail: parisien@nature.berkeley.edu

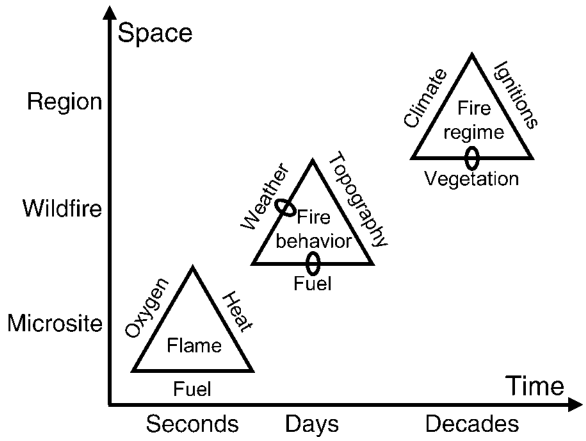

FIG. 1. Dominant factors affecting fire at multiple spatial and temporal scales. A loop on the side of a triangle indicates potential feedbacks. This study was primarily concerned with the fire regime triangle, as all analyses were performed for large study areas (>20 000 km^2) and multidecadal time periods. Each leg of the fire regime triangle is a function of environmental gradients that reflect suitability for fire occurrence.

instance, if fire likelihood is correlated with the annual number of days during which weather conditions would support wildfires, the deserts of the world would be classified as the most fire-prone areas. Yet, the areas experiencing extremely fire-conducive weather conditions are not necessarily those with the most active fire regimes, because their climates typically do not support continuous flammable vegetation. In contrast, some areas with very high rates of growth for aboveground biomass, such as tropical and temperate rain forests, only rarely (if ever) experience fire-conducive conditions. Therefore, the diverse terrestrial biomes of the world that are suitable for wildfire represent a complex ecological envelope or "middle ground" between various climatic extremes (Bond and van Wilgen 1996, van der Werf et al. 2003, Spessa et al. 2005).

In places where recurring fires are possible, the relative influence of environmental controls on fire activity (Fig. 1) can vary spatially and temporally (Schoennagel et al. 2004, Moritz et al. 2005). At finer scales, physics and a mechanistic understanding of fire behavior can be used to describe fire patterns (Rothermel 1972, Van Wagner 1977). Knowledge of wildfires at the landscape or subregional scale (10^4–10^6 ha) is growing (Krawchuk et al. 2006), and studies have successfully linked historical fire observations to weather, topography, and fuels (e.g., Heyerdahl et al. 2001, Stephens 2001, Rollins et al. 2002, Moritz 2003, Mermoz et al. 2005, Collins et al. 2007). For example, Rollins et al. (2002) showed that in a moisture-limited area of the western United States, wildfires preferentially occurred on the northern slopes, where the fuels were more continuous; in contrast, more wildfires on southern slopes were reported for a colder and wetter area, where fire-conducive weather was a limiting factor. A similar contrast between fuel moisture and fuel amount has been made by Meyn et al. (2007) in many parts of North America, in a review of literature concerning large and infrequent fires. Far fewer studies of fire–environment relationships exist at the regional to continental scale (>10^6 ha; e.g., Vasquez et al. 2002, Taylor and Skinner 2003, Russell-Smith et al. 2007, Syphard et al. 2008), and there is a need for an assessment of how well relationships may hold across large spatial scales (but see Cardille et al. [2001], Schulte et al. [2005], Cyr et al. [2007] for sub-regional to regional comparisons). At such coarse scales, associations between wildfire and environmental factors exerting a local and quantifiable influence, such as topography, become obscured (Cary et al. 2006). As a result, there are few quantitative descriptions of global fire patterns. Satellite data (Giglio et al. 2006) offer a global perspective of current (i.e., the past decade) fire activity but without the temporal extent required to allow an understanding of fire regimes. Dynamic global vegetation models have recently been used to identify general relationships between fire and vegetation patterns (e.g., Bond and Keeley 2005, Thonicke et al. 2005), and these models offer a promising set of tools with which to study fire regimes and fire probability patterns. At present, however, we still lack a quantitative framework for characterizing and understanding the geography of fire at different scales across the planet.

The complex suite of environmental factors that allows wildfires to occur can be viewed as fulfilling its habitat requirements. A set of meaningful environmental variables should therefore capture both the drivers of and the constraints on wildfire occurrence, as direct and indirect influences, across spatiotemporal scales (Fig. 1). The multivariate set of environmental conditions that defines the suitability of a given area for wildfire represents an *n*-dimensional environmental variable space. As for any process or organism, evaluating the true potential range of wildfire and thus quantifying the entire set of possible conditions under which fire might be observed, is challenging, if not impossible. This is because fire may not always be recorded where it has the potential to occur through a failure to detect occurrence (i.e., insufficient sampling or reporting), competition for space (i.e., land-use conversion), or other extrinsic factors that prevent its occurrence (i.e., fire suppression). It is therefore more feasible to evaluate the extent to which fire habitat is occupied under current conditions. This "realized" portion of suitable habitat represents the current potential geographic range of wildfire under certain environmental and anthropogenic constraints and, from a practical standpoint, may be more useful in parts of the world where human influence on fire regimes is pervasive or irreversible. Within this geographic range where fire is at least possible, the quality of fire habitat will vary and thus the suitability distributions for fire will also vary spatially (Larsen 1997, Rollins et al. 2002, Parisien and Sirois 2003, Collins and Stephens 2007).

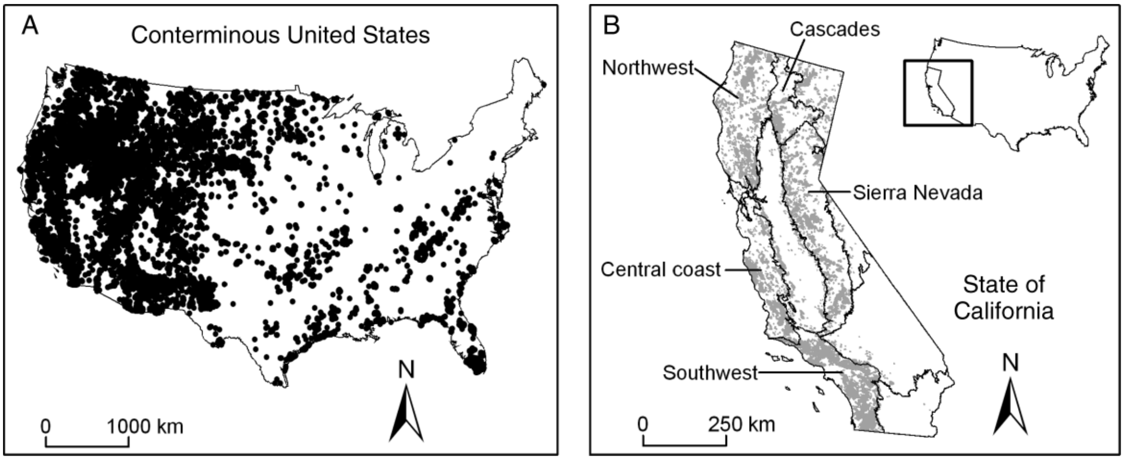

FIG. 2. Study areas representing three spatial scales: (A) the conterminous United States and (B) the state of California, with its five fire-prone ecoregions. Data from the two databases of large (>121 ha) fires used in this study are shown: (A) presumed points of ignition of fires reported for U.S. federal lands, 1980–2003 (circles) and (B) fire perimeters for California, 1950–2003 (in gray).

Recent interest in modeling habitat suitability with distribution models has propelled the refinement of models and the development of several new techniques (Franklin 1995, Austin 2002, Graham et al. 2004, Guisan and Thuiller 2005, Elith et al. 2006). Although a vast array of habitat distribution models is available, these techniques have the following features in common: they can incorporate a large number of variables, their output consists of an explicit probability of occurrence, and they provide some form of model evaluation. This paper documents a new application of habitat distribution models for evaluating controls on the distribution of wildfire. We focus on the distribution of wildfire at broad spatial scales in areas experiencing a wide range of environmental conditions. In this sense, we assume that wildfire behaves like an organism for which site-specific factors, such as forest structure and topography, are more important locally and climate-related factors prevail regionally (Pearson et al. 2004). This assertion is strongly supported in a continent-wide assessment of fire–environment relationships in Australia (Russell-Smith et al. 2007).

A primary goal of this study was to quantify the influence of environmental variables on the distribution of wildfire at three spatial scales: the conterminous United States, the state of California, and five fire-prone ecoregions within California. Our central hypothesis was that, in light of the strong codependence between fire regimes and the ecosystems that they "inhabit," distinct combinations of environmental variables may emerge as important at different spatial scales and in different locations. A related expectation is that fire response functions to key environmental gradients would be convex and unimodal, reflecting certain conditions that are more optimal for fire and others that are less suitable (i.e., marginal habitats). In constructing wildfire suitability models, we used the Maxent algorithm (Phillips et al. 2006) and boosted regression trees (BRT; Friedman 2002), two of the best-performing methods in a recent extensive comparison of modeling techniques (Elith et al. 2006). Observational data from existing databases of wildfire and mapped surfaces of many relevant environmental variables were used as model inputs. Because little is known about the application of habitat distribution models to spatial wildfire patterns, we assessed the sensitivity of the models to increasing numbers of variables, the transferability of models to other regions (e.g., Fielding and Haworth 1995, Bonn and Schröder 2001, Vanreusel et al. 2007), and the effect of sample size on model performance. In addition to demonstrating a new modeling framework for estimating spatial fire probabilities across large areas, we hope to stimulate discussion about the possible application of habitat distribution modeling methods and concepts in disturbance ecology.

STUDY AREA

Conterminous United States

The conterminous United States (Fig. 2) encompasses a wide range of climates controlled by latitudinal, altitudinal, and continental gradients, all of which are strongly influenced by the position of the eastward-flowing jet stream. The eastern United States experiences significant summer precipitation, but in the west, most precipitation falls during the winter months, except for some inland areas of the southwest, which are dry areas where about half of the precipitation falls during the summer months. The coastal northwest, which has mild winters and warm summers, experiences the most precipitation in the conterminous United States. The southwest, which encompasses the deserts, is dry but has areas of significant winter precipitation near the coast. Inland, the Rocky Mountain regions of the western United States are characterized by warm summers and

TABLE 1. The elevation range, temperature, and precipitation normals, major vegetation types, and number of fires and area burned for the studied California ecoregions.

Variable	Cascades	Central coast	Northwest
Area (km^2)	20 633	36 982	55 812
Elevation range (m)	85–3885	0–1954	0–2520
Maximum July temperature (°C) (range)†	29.7 (21.9–37.4)	29.8 (18.0–36.9)	29.8 (17.8–37.4)
Minimum January temperature (°C) (range)†	−3.0 (−8.8 to 3.4)	2.8 (−3.9 to 7.4)	0.7 (−5.8 to 6.0)
Total annual precipitation (mm) (range)‡	1101 (277–2711)	552 (190–1793)	1320 (335–3947)
Vegetation types (% of total cover)§	mixed conifer (59.7), oak woods (9.8), chaparral (6.5), pine forest (6.4), red fir (5.8)	oak woods (44.3), chaparral (28.8), annual grassland (14.7)	mixed conifer (30.0), redwood (11.9), cedar–hemlock–Douglas-fir (24.2), oak woods (13.7), California mixed evergreen (9.3), chaparral (7.2)
Nonfuel cover (%)‖	5.1	21.0	4.3
Mean annual no. fires $\geq$ 121 ha $\pm$ SD¶	3.8 $\pm$ 3.3	8.6 $\pm$ 5.4	13.4 $\pm$ 11.7
Mean annual area burned by fires $\geq$ 121 ha $\pm$ SD (ha)¶	6715 $\pm$ 13 742	14 769 $\pm$ 19 963	22 590 $\pm$ 39 040

† The averaged maximum/minimum monthly temperatures from the PRISM climate normal grids, 1970–2000. The range represents all areas within the ecoregion, not the interannual variability.
‡ The mean total precipitation from the PRISM climate normal grids, 1970–2000. The range represents all areas within the ecoregion, not the interannual variability.
§ From the Potential Natural Vegetation Groups, version 2000 (Schmidt et al. 2002).
‖ From the California Gap Analysis.
¶ From the California Fire and Resource Assessment Program.

severe winters with significant snowfall. To the east, in the rain shadow of the Rockies, lies the Great Plains area, typified by a dry continental climate. Most of the areas near the east coast are also continental, but they experience high moisture in the summer season through the influence of the Atlantic Ocean and other major water bodies (e.g., the Great Lakes).

The numerous climates of the United States are reflected in its vegetation cover, which ranges from sparse xeric vegetation to lush forests, as well as in its fire regimes, which range from regimes dominated by frequent low-intensity fires to those dominated by infrequent high-intensity fires (Pyne et al. 1996, Hardy et al. 2001). The coastal areas of the northwest support temperate rain forests that experience very long fire return intervals, whereas those of the southwest, which are covered by a mosaic of grasslands, shrublands, and forests, experience some of the highest burning rates in North America. Despite significant forested areas at higher elevations, much of the inland area between the coast and the Rocky Mountains is covered with xeric vegetation, such as deserts in the south and pinyon–juniper woodland in the north. Wildfire is rare in many of these areas because of the sparseness of fuels. By contrast, most of the Rocky Mountains region can experience large fires, ranging from low-intensity surface fires in the more open, drier forests to active crown fires in closed-canopy forests. Although large wildfires are occasionally reported in the Great Plains, the natural fire regimes of this region are largely uncertain, as this biome has undergone massive conversion to agriculture. The vegetation cover of the southeast is highly variable, consisting mainly of shrubland, deserts, dry forests, and subtropical forests. Much of its coastal area (i.e., Florida and the Gulf of Mexico coast) experiences short fire cycles, despite significant moisture year-round. The vegetation of the northeast is characterized by deciduous forests, and the natural fire regimes typically have a very long fire cycle, except for the areas surrounding the Great Lakes, which are dominated by coniferous forests that would support a fairly active fire regime such as the boreal forest. Humans have greatly modified natural fire regimes in most of the continental United States either through direct action, such as fire suppression and accidental or deliberate (i.e., prescribed) burning, or indirectly through the alteration of natural vegetation cover.

California

The climate of California has three gradients: a west–east continental gradient, a north–south gradient of decreasing winter precipitation and increasing summer temperature, and an elevation gradient of increasing precipitation and decreasing temperature. The summer position of a massive Pacific high-pressure system promotes a mediterranean climate with well-defined dry and wet seasons (Minnich 2006). Except for the inland deserts of the southeastern part of the state, almost all of the precipitation in California falls from November to April. However, the North American monsoons that occur from July to early September may bring occasional summer precipitation to southern and eastern California. There is also considerable local variation in weather because of its interaction with the rugged terrain in most of the state. As a result of numerous steep climatic, topographic, and geologic

TABLE 1. Extended.

Sierra Nevada	Southwest
63 117	33 767
36–4144	0–3237
28.1 (17.8–37.4)	31.0 (22.8–39.1)
−2.9 (−17.1 to 5.0)	3.3 (−8.1 to 9.5)
938 (118–2369)	472 (158–1457)
mixed conifer (37.8), oak woods (17.5), lodgepole–subalpine (12.4), chaparral (8.8), red fir (8.2)	chaparral (79.2), juniper–pinyon (5.8), annual grassland (6.2)
11.6	34.2
16.1 ± 8.8	25.6 ± 14.9
19 172 ± 19 557	41 720 ± 49 941

gradients, California represents an area of extraordinary vegetation diversity (Barbour and Major 1988; see Plate 1), which translates into a diversity of fire regimes (Stephens et al. 2007).

One commonality of wildfires in parts of California is their association with foehn-type winds out of the east and/or north, which promote extreme fire behavior in the late summer and autumn. These intense winds, called the "Santa Anas" in southern California, are generally deflected by the Sierra Nevada and are channeled though lower mountain passes in the southern part of the state. This type of wind also occurs elsewhere throughout the state, such as the Mono winds of the Sierra Nevada (Schroeder et al. 1964), but are generally less frequent in northern California. Ignition patterns show great variation in California, with lightning strikes occurring primarily in the deserts, the Sierra Nevada, and the Modoc plateau (in the northeast corner of the state) and much less frequently along the coast (van Wagtendonk and Cayan 2007). This general pattern has led many to hypothesize that prehistoric Native American burning was a frequent and dominant source of ignitions for several millennia throughout most of coastal California and the central valley (Anderson 2006). Today, these practices have been eliminated, but humans still ignite most fires reported in California, significantly influencing the modern fire regimes (Syphard et al. 2007). Much of the state has undergone intense fire suppression for about a century, which has altered the structure and composition of vegetation that co-evolved with more frequent fires. This effect is especially pronounced in the state's low- and mid-elevation forest ecosystems, where logging and subsequent fuel build-up may have had a substantial impact on recent wildfires. To better describe regional characteristics in the California wildfire environment, we subdivided fire patterns according to bioclimatic ecoregions (Hickman 1993; Fig. 2). The five most fire-prone ecoregions were then analyzed and characterized in terms of climatic conditions, vegetation, and wildfire activity, as provided in Table 1.

Cascades ecoregion

The Cascades ecoregion is characterized by prominent volcanic peaks (e.g., Mount Shasta), but the topography is gentler than that of the neighboring Klamath Mountains (to the west) and the Sierra Nevada. The dominant vegetation types are coniferous forests, woodlands, and shrublands, which vary widely in terms of precipitation, topography, and substrate. Like the other four fire-prone ecoregions, this ecoregion has a highly variable climate, with annual precipitation differing by a factor of 10 from the driest to the wettest part of the region (Table 1). In general, precipitation is lower on the northwest side of the range because it lies in the rain shadow of the Klamath Mountains. Unlike other mainly forested regions, some areas of the Cascades ecoregion may have been susceptible historically to periodic crown fires, as evidenced by the presence of serotinous species that rely on severe crown fires for regeneration (Bekker and Taylor 2001, Skinner and Taylor 2006).

Central Coast ecoregion

The topography of the Central Coast ecoregion is defined by the Santa Cruz and Santa Lucia mountains on the northern and southern coasts, respectively, and by the Interior Coast Ranges inland. The vegetation usually consists of a progression from coastal prairie and coastal shrubland through mixed hardwoods to coniferous forest at the highest elevations, whereas grasslands, oak woodlands, and shrubland are prevalent further inland. There are climatic differences among these sub-ecoregions that are important for fire, notably the northward decrease in the frequency of Santa Ana conditions. Given the similarity in vegetation throughout the area (i.e., large tracts of chaparral), the lack of large fires in the northern part of the Central Coast ecoregion (with few large fires in recent decades) is surprising (Davis and Borchert 2006).

Northwest ecoregion

The diverse topography, geologic types, and climate patterns of the Northwest ecoregion result in exceptional floristic diversity and complexity (Skinner et al. 2006). The ecoregion ranges from flat marine terraces to the extremely rugged Klamath Mountains. The coastal part of the ecoregion is characterized by a thin strip of coastal shrubland and prairie, whereas the inland areas are covered by coniferous forests, a mix of coniferous and hardwood forests, woodlands, and shrublands. Overall, the fire regime has become less active since implementation of a fire-suppression policy, but fuel load and structural changes may be responsible for the occurrence of larger fires in some parts of the ecoregion (Stuart and Stephens 2006) and not others (Odion et al. 2004). Historically, there was less frequent fire activity near the coast, which experiences moister and cooler weather than inland. Apart from the immediate coast, this may no longer be the case, as wildfires have been reported in a variety of climates within this ecoregion.

Sierra Nevada ecoregion

The Sierra Nevada ecoregion consists of a block mountain range with a steep eastern escarpment. Although most of the precipitation falls during the winter as snow, subtropical monsoonal moisture produces some summer precipitation. From west to east, the main vegetation types are shrublands, woodlands, mixed-conifer forests, subalpine forests, and alpine meadows and montane shrublands. Except for the sparsely vegetated alpine zone, almost all parts of the Sierra Nevada ecoregion have experienced substantial wildfire activity. Like those of the Cascades and Northwest ecoregions, the presettlement surface-dominated fire regimes of the Sierra Nevada ecoregion may have been significantly affected by extensive logging and the exclusion of fire (van Wagtendonk and Fites-Kaufman 2006).

Southwest ecoregion

The Southwest ecoregion is bordered on the north by the Transverse Ranges and on the east by the Peninsular Ranges. It accounts for 8% of the land area of California and, being home to over half of the state's population, has undergone considerable anthropogenic change (Davis et al. 1995). As elsewhere on the coast, summer precipitation is rare, but occasional Mexican monsoonal moisture affects the interior mountains, particularly in the east. The ecoregion is a complex mosaic of grasslands, shrublands, forests, and woodlands. It contains large tracts of chaparral shrublands, where numerous large fires have been recorded over the past century. Most of these large fires have been driven by Santa Ana winds in the fall (Keeley 2006). Although almost all nonurban areas of this ecoregion burn relatively frequently, some regions are affected by Santa Ana winds to a greater extent than others (Moritz et al. 2004).

Other California ecoregions

Although this study focused on the most fire-prone ecoregions of California, almost every ecoregion of the state may experience large wildfires. The Modoc ecoregion lies in the rain shadow of the Cascades and is characterized by a relatively dry, cold continental climate. The dominant vegetation is a mixture of sagebrush steppe and conifer forests. Although most of this ecoregion is fire-prone, it was not considered for analysis here because most of its extent is located outside California. The Central Valley ecoregion, located between the Sierra Nevada and the Central Coast and Northwest ecoregions, is by far the flattest part of California. It was once a vast mixture of prairies, oak savannas, grasslands, freshwater marshes, and riparian woodland, but today most (>90%) of its natural habitats have been converted to agriculture or urban uses. There is a strong north–south gradient, with the north receiving a moderate amount of annual precipitation and the south experiencing rainfall levels similar to those of deserts. The Southeastern Deserts ecoregion consists of the Mojave, Colorado, and Sonoran deserts, along with part of the Great Basin. Virtually all of this ecoregion is arid because it is in the rain shadows of the Sierra Nevada, Transverse, and Peninsular ranges, causing vegetation to be highly discontinuous. The region experiences 10–25 days with afternoon thunderstorms (monsoons) between July and early September (Brooks and Minnich 2006). Major vegetation types include arid and semiarid shrublands, grasslands, and woodlands. In terms of recent human modifications to these ecoregions, the fire regimes of the Modoc ecoregion have experienced relatively minor anthropogenic changes (Riegel et al. 2006), yet the near-total agricultural conversion of the Central Valley ecoregion has severely limited the potential for fire ignition and spread in that area. By contrast, human influence has indirectly affected fire regimes in the Southeastern Deserts ecoregion, where several historically nonflammable areas now have a cover of exotic annual grasses, which constitute a new fuel condition (Brooks and Matchett 2006).

Methods

Data

Fire data sets.—We compiled wildfire databases for the State of California and the conterminous United States. The California database is maintained as part of the California Fire Plan of the Fire and Resource Assessment Program (2006). Depending on the area, the database included the mapped perimeters of wildland fires as small as 4 ha (10 acres). Prescribed fires were not added to California because they were highly clustered in a few regions and probably not exhaustive for the study period. The bulk of the source data were provided by the California Department of Forestry and Fire Protection, the U.S. Forest Service (USFS), the Bureau of Land Management (BLM), and the National Parks Service (NPS), complemented by data from the Bureau of Indian Affairs (BIA), the U.S. Department of Defense, and various local agencies. Although data are available for some regions from the early 1900s to the present, reporting is thought to have been less comprehensive before 1950. Therefore, for this database, we used data from 1950 to 2003 (Fig. 2). We also omitted smaller fires, which account for very little of the area burned in California (Strauss et al. 1989, Moritz 1997), because of inconsistent reporting (i.e., <121 ha or 300 acres). Despite these precautions, some larger fires are undoubtedly missing from the database. Furthermore, perimeter mapping is generalized such that outlines are often approximated, and unburned islands, which can be numerous, are not included. Nevertheless, this

TABLE 2. Input variables used to model the distribution of fire, including the data source and description (with units in parentheses).

Input name	Data source	Description
MaxTemp*n* [n = 1–12]	PRISM	maximum monthly temperature (°C)
MaxTempAnn	PRISM	maximum annual temperature (°C)
MaxTempWarmest†	PRISM	maximum temperature of the warmest month (°C)
MinTemp*n* [n = 1–12]	PRISM	minimum monthly temperature (°C)
MinTempAnn	PRISM	minimum annual temperature (°C)
MinTempColdest†	PRISM	minimum temperature of the coldest month (°C)
Pcp*n* [n = 1–12]	PRISM	total monthly precipitation (mm)
PcpAnn	PRISM	total annual precipitation (mm)
PcpWettest†	PRISM	total precipitation of the wettest month (mm)
PcpDriest†	PRISM	total precipitation of the driest month (mm)
PcpFrq*n* [n = 1–12]	Daymet	monthly proportion of days with precipitation (0 to 1)
PcpFrqAnn	Daymet	annual proportion of days with precipitation (0 to 1)
PcpFrqWettest†	Daymet	proportion of days with precipitation of the wettest month (0 to 1)
PcpFrqDriest†	Daymet	proportion of days with precipitation of the driest month (0 to 1)
Rad*n* [n = 1–12]	Daymet	monthly mean of daily shortwave radiation ($MJ \cdot m^{-2} \cdot d^{-1}$)
RadAnn	Daymet	annual mean of daily shortwave radiation ($MJ \cdot m^{-2} \cdot d^{-1}$)
RadHighest†	Daymet	monthly mean of daily shortwave radiation for the month with the most radiation ($MJ \cdot m^{-2} \cdot d^{-1}$)
RadLowest†	Daymet	monthly mean of daily shortwave radiation for the month with the least radiation ($MJ \cdot m^{-2} \cdot d^{-1}$)
Hum*n* [n = 1–12]	Daymet	monthly mean of daily water vapor pressure (Pa)
HumAnn	Daymet	annual mean of daily water vapor pressure (Pa)
HumHighest†	Daymet	monthly mean of daily water vapor pressure of the most humid month (Pa)
HumLowest†	Daymet	monthly mean of daily water vapor pressure of the least humid month (Pa)
Elev	USGS	digital elevation model at 1-km resolution (m)
PotVeg	USDA Forest Service	Kuchler's Potential Natural Vegetation Groups (categorical)
Nonfuel	California Gap Analysis	land cover of California, reclassified as fuel and nonfuel (categorical)

Notes: All climate variables consisted of 30-year means (1971–2000) or 18-year means (1980–1997) for PRISM and Daymet data, respectively. Individual months are identified by a numeric code, n.

† Denotes derived climate variables.

database represents the best available fire data for California, and its quality can be considered very high relative to that of similar databases worldwide.

The database for the conterminous United States included the presumed points of origin of all reported wildland fires that burned on federally managed lands from 1980 to 2003 according to source data from the following agencies: the USFS, the Department of the Interior, the BIA, the BLM, the Fish and Wildlife Service, and the NPS (Bureau of Land Management 2006). Only data for fires > 121 ha (300 acres) were used for consistency with the data available for California (Fig. 2). This database is not comprehensive, as fires on state-managed and other lands were generally not included. Fires occurring on federal lands represent most of the fires in the conterminous United States, but the completeness of information about true fire occurrence varies among states (Stephens 2005). Therefore, to increase geographic representation, we decided to include the prescribed burns, which comprise ~8% of large fires (<4% of the area burned) in the database. Both fire data sets included lightning- and human-caused fires. It is impossible to determine the exact proportion of fires ignited by cause, as this attribute was often missing, but the majority of large fires were ignited by humans in California and the United States.

Gridded climate normals.—Maps of selected climate normals for the United States were taken from two complementary sources: the PRISM Group (2004) and Daymet (Thornton et al. 1997). The PRISM and Daymet maps, presenting slightly different elevation-adjusted interpolations from weather station observations, were available as grids with cell sizes of 4 and 1 km, respectively; the climate variables that we used were climatic normals from 1971 to 2000 and from 1980 to 1997, respectively. Although the time periods spanned by these data sets did not perfectly match those of the fire data sets, they adequately represented the interannual variability of modern conditions. We selected the PRISM data for temperature and precipitation because the type of interpolation from which these data were derived is more suitable than the Daymet data for some of the mountainous areas (Daly 2006). Daymet data were used for variables that were not available through PRISM.

We used monthly and annual raster grids of each variable as "direct" climate inputs to our models (Table 2). Spatial data were manipulated in ArcGIS 9.1 (ESRI 2005) using a Lambert azimuthal equal area projection for data from California and the conterminous United States. The PRISM variables consisted of various monthly normals, as well as one annual normal, of minimum temperature (MinTemp), maximum temperature (MaxTemp), and total precipitation (Pcp). The

Daymet variables used were the proportion of days with measurable precipitation (PcpFrq), the amount of shortwave radiation (Rad), and the daily water vapor pressure (Hum), which is a measure of absolute humidity that is also proportional to the dew point temperature (Thornton and Running 1999, Thornton et al. 2000). The values of Hum vary at continental scale according to the position and influence of air masses and vary locally mainly as a function of elevation, whereby higher elevations have lower absolute humidity. The values of Rad, a measure of energy, are mainly affected by latitude and cloud cover over large areas and by aspect and elevation locally. We used precipitation frequency in addition to total precipitation because the number of days of precipitation can vary markedly among areas with similar total precipitation.

To determine the spatial patterns in the extremes of monthly climate normals, we created grids of the maximum and minimum values for each cell from the monthly normals for the following variables: total precipitation, precipitation frequency, radiation, and vapor pressure. This process yielded maps of "derived" monthly climate normals, such as the total precipitation of the wettest month (PcpWettest), and conversely, the total precipitation of the driest month (PcpDriest). For the temperature variables, we created grids of the maximum temperature of the warmest month and the minimum temperature of the coldest month; it was not relevant to produce grids of minimum values of maximum monthly temperature, and vice versa.

Elevation and land cover maps.—A raster grid (1-km resolution) from a U.S. Geological Survey (2000) digital elevation model was used as the Elev variable for all study areas. The Potential Natural Vegetation (PNV) Groups coverage (Schmidt et al. 2002) was used as a coarse-scale variable for vegetation class. This coverage consists of Kuchler's (1964) original PNV map, which was refined with elevation, hydrological information, and updated vegetation mapping. The data, which cover the entire conterminous United States, are highly generalized and, as such, are more appropriate for large-scale studies.

A nonfuel map of California was produced from the land cover types of the California Gap Analysis Project (Davis et al. 1998). "Nonfuel" was defined as any cover type where fire spread is unusual, such as those heavily modified by humans (e.g., urban and agricultural areas), natural land cover types that have sparse or no vegetation cover (e.g., most desert types and alpine tundra), and those that are considered permanent wetlands. It was impossible to obtain a true "nonfuel" map, as some pastures, wetlands, and even urban areas may burn under extreme conditions, but such cases are rare. Our models were built from wildfire data dating from 1950 to 2003 and assume no change in land cover over this period. Clearly, this is not the case, as some previously flammable areas were converted to nonfuel types during this period; however, the majority of land-use change in California, much of which has been agricultural, occurred before 1950.

We included elevation, potential natural vegetation, and nonfuel as "indirect" predictor variables in some of our models. Indirect variables do not explicitly affect the species or ecological process under study but may act as useful proxies for other information (e.g., anthropogenic factors) not included in the direct variables (Austin and Smith 1989). For example, the environmental space (and, consequently, the predicted geographic space) of wildfire should be fully captured by a combination of the climate variables, but elevation may provide additional information pertaining to the exclusion of fires from certain areas because of human land use in valleys, such as agriculture and urban uses. Although vegetation may directly affect spatial wildfire patterns, PotVeg represents the "potential" for a few generalized vegetation types at any point in the United States, regardless of its current state or non-natural land use. As such, this variable is best considered as "indirect."

Habitat distribution modeling of wildfires

Habitat distribution models can be used with observations in geographic space to describe the environmental space—also called "climate envelope" when evaluating only climate variables—required by a species (Pearson and Dawson 2003). In our case, this n-dimensional environmental variable space characterizes the environmental controls or habitat requirements for wildfire. We used two machine learning algorithms, Maxent and boosted regression trees (BRT), to describe and predict the environmental space of fire. Although conceptually similar on some levels to regression techniques such as generalized linear models and generalized additive models, machine learning does not require the specification of an a priori data model (Breiman 2001, Elith et al. 2008). This makes them particularly suitable when the nature of the process is largely unknown, presumed to be highly complex, and when an emphasis is on accurate predictions.

Maxent and BRT were used in a complementary fashion in our model building and evaluation process. Maxent was primarily used to evaluate environmental controls and make spatial predictions, whereas BRT was used to explore the interactions among variables, which is a strength of tree-based methods, and to determine whether the Maxent results were robust and consistent with those of a significantly different algorithm. From an analytical standpoint, a primary difference between algorithms is that Maxent is designed for presence-only data (Zaniewski et al. 2002, Pearce and Boyce 2006), whereas BRT is designed for presence and absence data. Although both methods output the probability of presence, presence-only models do so by discerning between the characteristics of burned areas from those of the entire landscape, as opposed to discerning between burned areas and areas where fire was assumed not to occur. Both methods are appropriate, as long as the outputs are interpreted correctly. However, presence-only models such as Maxent may be preferable for spatial predictions in areas that have experienced few large fires within a short time period, because a large fraction of the territory where fire has not been observed (fire "absences") may in fact be fire-prone.

Maxent

Maxent is designed for modeling the geographic distribution of species from the *n*-dimensional environmental variable space with presence-only data (Phillips et al. 2006). It estimates a target probability distribution (i.e., suitability) by fitting the probability distribution of maximum entropy (the one that is most uniform) to the environmental variables at each observation (i.e., the presence points). Maxent iteratively evaluates the contrasts between the values of these observations and those of a background consisting of the mean observations over the entire study area, as sampled from a large number of points. One of Maxent's (and BRT's) most important features is the capacity to fit highly complex response functions by combining several function types (linear, quadratic, product, threshold, and hinge). It can fit jagged and sharply discontinuous responses that cannot be modeled in even the most flexible regression techniques, such as generalized additive models. It also adjusts for overfitting through a process called "regularization," a mechanism that prevents the algorithm to match the data too closely. Models were computed in Maxent 3.1 (Phillips et al. 2006).

The model uses the fitted parameters of the distribution to produce a habitat suitability map containing raw probabilities that undergo a transformation to produce the so-called "logistic" output (Phillips and Dudík 2008). The mapped value of each cell of this output represents an estimate of the probability of presence, so it is possible to transform them into expected values of fire frequency, or annual area burned, per grid cell. However, we used and interpreted output probabilities as relative measures of fire suitability.

For very-long-interval and high-intensity fire regimes, the 1980–2003 period of record for the United States is probably too short to observe many fires occurring on these particular areas. However, Maxent is well suited to this situation, because the numerous wildfire observations offer ample opportunities for overlap in environmental space between these points and potentially suitable areas. Maxent may perform well with very few data points (Pearson et al. 2007), or even geographically biased data (e.g., samples collected only from areas adjacent to roads) (Kadmon et al. 2004), as long as there are sufficient data to provide a good representation of environmental space. Wildfire databases are subject to all of these shortcomings, as reliability and comprehensiveness in reporting are fairly recent phenomena.

Maxent provides a suite of model diagnostics (see *Methods: Model evaluation*), one of which is the relative contribution of each variable to the model. This measure is computed using a heuristic estimate in which the increase or decrease in model gain attributed to each variable from one iteration of the training algorithm to the next is added to the variable contribution. These values of variable importance should be interpreted carefully when the variables are correlated. Predictor variables that may in fact be informative can have a low variable contribution simply because much of the information it comprises is included in a slightly more influential variable. Alternatively, the potential usefulness of individual variables to a model can be evaluated by building a Maxent model using only this variable. Maxent also considers variable interactions through its product function type, but it is impossible to evaluate their contribution to the model.

Boosted regression trees

Boosted regression tree (BRT) models combine two simple algorithms, regression trees and boosting (Friedman 2002, Elith et al. 2008). Specifically, a large number of simple trees (100s to 1000s) are produced using recursive binary splits or nodes based on the value of a single predictor variable at each node that results in the two most homogeneous subsets of the response variable. Each tree is submitted to a set level of stochasticity by being built from a random subset of the data (termed "bagging"). The terms are fit in a stage-wise manner by building trees from the residuals of the prior collection of trees, thereby allowing the model to put more emphasis on the points that are more difficult to classify. This allows BRT to overcome the problems associated with classification and regression trees (CART), notably that of poor classification accuracy (De'ath 2007). The resulting BRT model can be viewed as an additive regression model in which every term is a tree. Recent ecological studies have showcased the strong performance of BRT compared to other modeling methods (Leathwick et al. 2006, 2008, Moisen et al. 2006, De'ath 2007).

There are three inter-dependent input settings for BRT models: the learning rate, tree complexity, and number of trees. Tree complexity consists of the number of nodes or variable interactions in each tree. The learning rate or shrinkage is a form of regularization, as it limits the amount of learning possible in each tree. Typically, its value is low and used in conjunction with a large number of trees, which enables BRT to generate highly complex response functions. Fast learning rates will require fewer trees but be subject to more noise induced by the bagging and a lack of smoothness in the response functions.

Being a tree-based method, BRT automatically takes into account interactions among variables, as every successive tree node constitutes a potential interaction. The influence of these interactions can be evaluated globally for a BRT model (Elith et al. 2008) or can be displayed graphically for an individual tree of a BRT (usually the first tree). Although the global model compounds all variable interactions, the visual display of a single tree often provide a clearer and more easily interpretable depiction of fairly complex interactions. Variable contribution in BRT is obtained by averaging over all trees in a model the number of times a variable is selected for splitting and the squared improvement resulting from these splits (Friedman 2001).

Data preparation

We generated the study sample of fire points for California and the conterminous United States using two different methods. The fire observations in California consisted of randomly selected points within the burned areas for the period 1950–2003. The number of these points was equal to the number of fires in the data set. Half of these points were used to build the models (the training points), and the remaining points were used to evaluate the accuracy of model classification (the test points). The so-called background points were the training points and an area-weighted number of points proportional to the area outside the burned areas. This resulted in 4244 training points and 41 940 background points. Because a fraction of the total fires are responsible for the majority of the area burned in California, this scheme allowed for better representation of the distribution of fire than a scheme using one data point per fire. The data points for the ecoregions were subsampled from those of California.

For the U.S. study area, it was impossible to sample observations within fire perimeters because of the point-source nature of the data. As such, all fire observations in the data set were used as data points, and these were divided in half (for training and testing). The number of background points could not be area weighted outside the burned areas as it was for the California data, because the U.S. data set was not exhaustive (accounting only for fires on federal lands), and such a scheme would have exceeded computational capabilities. Here, the background points included the training points and additional points in areas where wildfires have not been observed, equal to twice the number of training points. This fire-absence area was defined as an area no closer than 2227 m from a fire point (the radius of the mean size of fires $\geq$ 121 ha). This sampling scheme resulted in 6336 training points and 20 810 background points.

Predictive modeling of fire suitability: selection of variables

Our limited knowledge of wildfire controls at broad spatial scales required the development of models with no a priori assumptions of the relative ranking of variables and no specific tests of hypotheses. As such, this study consists of an analysis using numerous variable types that have been included in past studies of wildfire occurrence. A large number of variables is not an obstacle for machine-learning algorithms because they tend to minimize or even exclude the non-informative predictor variables from the model. However, as with other statistical methods, it was preferable to minimize the number of highly correlated variables to facilitate interpretation of the output. Similarity among continuous variables (i.e., all variables except PotVeg and Nonfuel) was evaluated in a cross-correlation matrix where correlations $\geq$ 0.9 were identified. For each correlated pair, we evaluated the relative predictive ability of the variables using univariate logistic regression that compared the data points corresponding to fire areas with those where no fires had been observed. The variable with the lowest deviance was retained. These steps were performed iteratively for each variable until no highly correlated variables remained.

Predictive modeling of fire suitability: model fitting

We selected three models for each study area. The "full" model consisted of all climate and non-climate variables that were not highly correlated. The "climate-only" models were built without indirect variables (Elev, PotVeg, Nonfuel), which have limited use for predictive modeling at large spatial scales or for extrapolation to new areas (Austin and Smith 1989, Pearson and Dawson 2003). A third model type, the "derived-climate" model, was built from only the 10 derived climate variables, which were decoupled from calendar months (Table 2). Because these derived climate variables depict the relative differences in extremes, they are more appropriate for transfer of a model to other areas. The full and climate-only model predictions are appropriate for the study area from which they were built, but not necessarily for different areas, because the ecological response to specific monthly climate values may differ strikingly among study areas.

Models using the same data were produced with Maxent and BRT. Absence points were used as such in BRT to discriminate between presence and absence, and absence points were combined to the presence points to form a "background" for presence-only modeling in Maxent. In Maxent, a random subset of points was selected for training and then testing the predictions. Because such a model prediction is subject to randomness, this process was repeated 10 times and the model outputs averaged. A regularization multiplier of "1" (the default) was determined to produce the best spatial predictions in Maxent without over-fitting the data. The BRT models were computed in R with the "gbm" package (Ridgeway 2006) using a "bernoulli" (logistic) error structure, but the input settings were identified using custom functions created by Elith et al. (2008). These custom functions identify the number of trees that produce the lowest prediction error according to a specified learning rate and tree complexity, which are selected heuristically. The most appropriate combination of learning rate and tree complexity for our models were 0.05 and 5, respectively, and the bagging fraction was set at 50% for each tree. The sum of presences and absences were given equal weights. Using these settings, the number of trees varied from 600 to 6500 among study areas. To parallel the Maxent models, 10 BRT models were built using a random sample of half of the data set.

Model performance and variable contributions were evaluated for both models, but spatial predictions were only shown for the Maxent models, as the BRT models produced similar predictions. Predictions were expressed in relative probabilities ranging from 0 to 1, and these values were mapped for the U.S. and California study areas. The outputs of the full and climate-only models were compared to assess how and where indirect variables contributed to the model outputs. The relative cell-to-cell difference in relative probability, expressed as the relative change in probability, between mapped predictions of different model types was calculated for the climate-only and the full models, as well as the derived-climate and climate-only models. The percentage contribution of the top five Maxent variables, as well as their corresponding variable contribution in BRT models, was reported for each combination of model type and study area.

To illustrate a bivariate climate envelope characterizing wildfire suitability in each study area, the relative probability calculated in Maxent climate-only models was plotted as a function of the top two climate variables, as ranked from Maxent models produced with single environmental predictors (i.e., variable interactions were not considered in the variable contribution). The relative probability and the values of the two climate variables were obtained from sampling 10 000 random points uniformly across study area in the United States and California and 2000 points in each ecoregion.

Variable interactions

To examine the nature of interactions between the main climatic variables of each study area, simple tree diagrams were produced using BRT for the derived-climate models. For this purpose, a single tree was produced with all available points (bagging fraction of 1) and the maximum learning rate. This scheme produces a "logistic regression tree" where the deviance is minimized in logit space to determine the splits in the data. For ease of interpretation, only the first three nodes were allowed. The trees were drawn and an estimate of fire probability was provided at the end of each terminal node. Note that these simplified trees would produce relatively poor spatial predictions; the fire probability values should be used as a coarse indication of fire suitability under the given minimal rule sets. Their usefulness is more in understanding variable interactions and possible threshold responses of environmental controls on fire probabilities.

Spatial transfer of models

To examine how well a model parameterized based on data from one area could be spatially transferred to another study area, Maxent models were applied as follows: (1) the derived-climate models built for California to the conterminous United States, (2) those of each ecoregion to the State of California, and (3) those of the ecoregions to one another. We quantified the spatial transfer of models with test data from the area into which each model was transferred. The test data were generated using methods described in the earlier section on model fitting.

To evaluate the capacity of models produced in large areas to transfer to smaller embedded areas, we built a model for the United States that omitted the training points for California and then transferred the model to California (henceforth called the "infill" models). The same scheme was used for each ecoregion, where a model built in California excluding the ecoregion of interest was applied to this ecoregion. Infill models can also help assess the effect of large geographic gaps in data. These models were evaluated within the ecoregion of interest using its own test points.

Sample size

The fire data sets comprise a large number of observations. To evaluate how many observations were required to build good spatially predictive models, the effect of reducing the sample size of the training points was assessed by building climate-only models in Maxent for each study area using one out of four, one out of 10, and one out of 25 of the points available for the original models. Models created from a subsample of training points can produce predictions quite different from those produced for the entire data set. We ran 10 iterations of each model, each iteration using a random selection of the data. The mean and variability of the predictions were examined for each data set. The same number of background and testing points were used for the reduced sets of sample points as for the original models (with all training points).

Model evaluation

For each Maxent and BRT model, we computed the commission error (false positive rate or 1 − specificity) and omission error (false negative rate or 1 − sensitivity) of the test points at the probability threshold that maximizes the sum of sensitivity and specificity values. If both types of error are considered equally undesirable, this point describes the threshold value where the best model performance should be observed (Liu et al. 2005). Because Maxent uses a background instead of true absences, the false positive rate cannot be truly assessed; instead, this value can be interpreted as an estimate of the fraction of the area predicted to be suitable.

A more robust and threshold-independent method for evaluating the accuracy of prediction classification is the receiver operating characteristics (ROC) curve (Fielding and Bell 1997). With this method, sensitivity (the fraction of true positives) is plotted against 1 − specificity (the fraction of false positives in BRT; the predicted suitable area in Maxent), and the area under the curve (AUC) is calculated for all possible probability

TABLE 3. Selected variables for the full model (all variables) and climate-only scenarios (all variables excluding those marked in bold), in decreasing order of importance when processed as a single environmental variable in Maxent.

Rank	United States	California	Cascades	Central coast	Northwest	Sierra Nevada	Southwest
1	**PotVeg**	**PotVeg**	PcpDriest	Hum8	Pcp8	**PotVeg**	**Nonfuel**
2	Rad7	**Nonfuel**	Pcp6	**Elev**	PcpFrq1	PcpFrq8	**Elev**
3	Pcp6	MaxTemp6	MaxTemp11	MinTemp5	**PotVeg**	Hum4	Hum9
4	**Elev**	Pcp3	PcpFrq7	MinTemp7	MaxTemp8	MaxTemp1	Rad5
5	Hum5	PcpFrq1	Pcp8	**Nonfuel**	Hum11	PcpFrq9	Hum11
6	Rad5	MinTemp7	PcpFrq8	Rad10	PcpFeq8	**Elev**	MaxTemp7
7	PcpFrq9	Hum8	PcpFrq5	Rad4	Hum8	Pcp5	MinTemp6
8	PcpFrq6	Pcp5	**PotVeg**	PcpFrq7	**Elev**	PcpFrq7	Hum6
9	PcpFrqDriest	**Elev**	**Elev**	PcpFrq9	Pcp6	Pcp6	Pcp4
10	Pcp5	Hum12	PcpFrq4	MaxTemp3	MinTemp9	Pcp7	Pcp5
11	MinTemp6	Hum7	PcpFrq9	PcpFrq6	Pcp5	**Nonfuel**	Rad6
12	PcpFrq7	Rad5	PcpFrq10	Pcp3	Pcp2	Pcp8	**PotVeg**
13	Pcp10	Rad4	RadAnn	PcpFrq11	HumLowest	Rad7	Pcp6
14	PcpFrq5	Pcp8	Pcp5	Pcp8	Rad5	Rad2	PcpFrq2
15	Pcp4	Pcp7	Pcp14	MaxTemp8	Pcp7	Rad8	PcpFrq11
16	PcpFrq10	PcpFrq8	PrcFrq1	PcpDriest	Pcp9		Pcp9
17	PcpFrq1	PcpDriest	PrcFrq11	**PotVeg**	MaxTemp2		PcpFrq12
18	Pcp11	Hum11	Rad11	Pcp9	Rad2		PcpFrq4
19	PcpFrq4	Rad8	Pcp9	PrcFrq8	**Nonfuel**		PcpFrq3
20	PcpWettest	Pcp6	**Nonfuel**	Pcp6			PcpFrq10
21	Rad9	Rad7		PcpFrqDriest			Pcp7
22	PcpFrqWettest	PcpFrq7		Pcp10			PcpFrq8
23	MaxTemp6	Pcp9		Pcp5			Rad2
24	PcpFrq11	PcpFrq9					PcpFrq5
25	Rad10						Rad8
26							PcpFrq9
27							PcpWettest
28							PcpFrq6
29							PcpFrq7

Notes: Inclusion of variables varied among study areas as a result of the selection scheme used (see *Methods*). A key to the variable names is in Table 2.

thresholds. Model AUC values typically vary between 0.5, indicating that model predictions are no better than a random classification of observations, and 1. Higher AUC values indicate a better fitting model; however, because Maxent compares the observations to background points, as opposed to absence points, the maximum possible test AUC will always be somewhat less than unity (see Phillips et al. 2006). For this reason, a comparison of AUC values among Maxent and BRT models should be interpreted with caution. Nevertheless, this is a useful index of model performance for our data. Here, we consider any model with AUC $\geq$ 0.6 to be at least more informative than a random output probability distribution.

RESULTS

Fire suitability models

We reduced the number of variables to be fitted in each full model from a possible 90 or 91 to between 15 and 29 for each of the study areas (Table 3). Bivariate plots of top climate variables for each model (Fig. 3) demonstrate response surfaces with one or more peaks reflecting more suitable conditions for fire, and surfaces were also often skewed (i.e., non-normal). Key variables included in the United States models are multimodal, with high suitability areas corresponding to known fire-prone parts of the United States. In contrast, at finer spatial scales, the California and ecoregions variables are mostly unimodal, with the exception of August maximum temperature in the Northwest ecoregion. There are also differences in the shape of the fire probability distributions among similar variables in different areas. For example, the probability distribution of the high absolute humidity of the Central Coast ecoregion (Hum8) is highly skewed or truncated, where probabilities decrease from low to high humidity values, whereas the fire distribution of a similar variable in the Southwest (Hum9) is bell-shaped (Fig. 3). For the Cascades ecoregion, the first and third variables were used because of the great similarity between the first and second variables.

At the probability threshold that maximizes specificity and sensitivity, the predicted proportion of area suitable for fire in Maxent and the associated omission rate were similar among areas and among model types (Table 4). These measures indicate that at this probability threshold, about three-quarters of the presence points will be accurately classified within the estimated one quarter of the area predicted to be most suitable. When the predicted proportion of area suitable is set at 50%, the omission error is <5% in all study areas (*data not shown*). Given the high geographic prevalence of fire in the study areas, this is a very good prediction rate. Although not comparable to the values of the Maxent models, where the absence points were not explicitly classified, the commission and omission error obtained

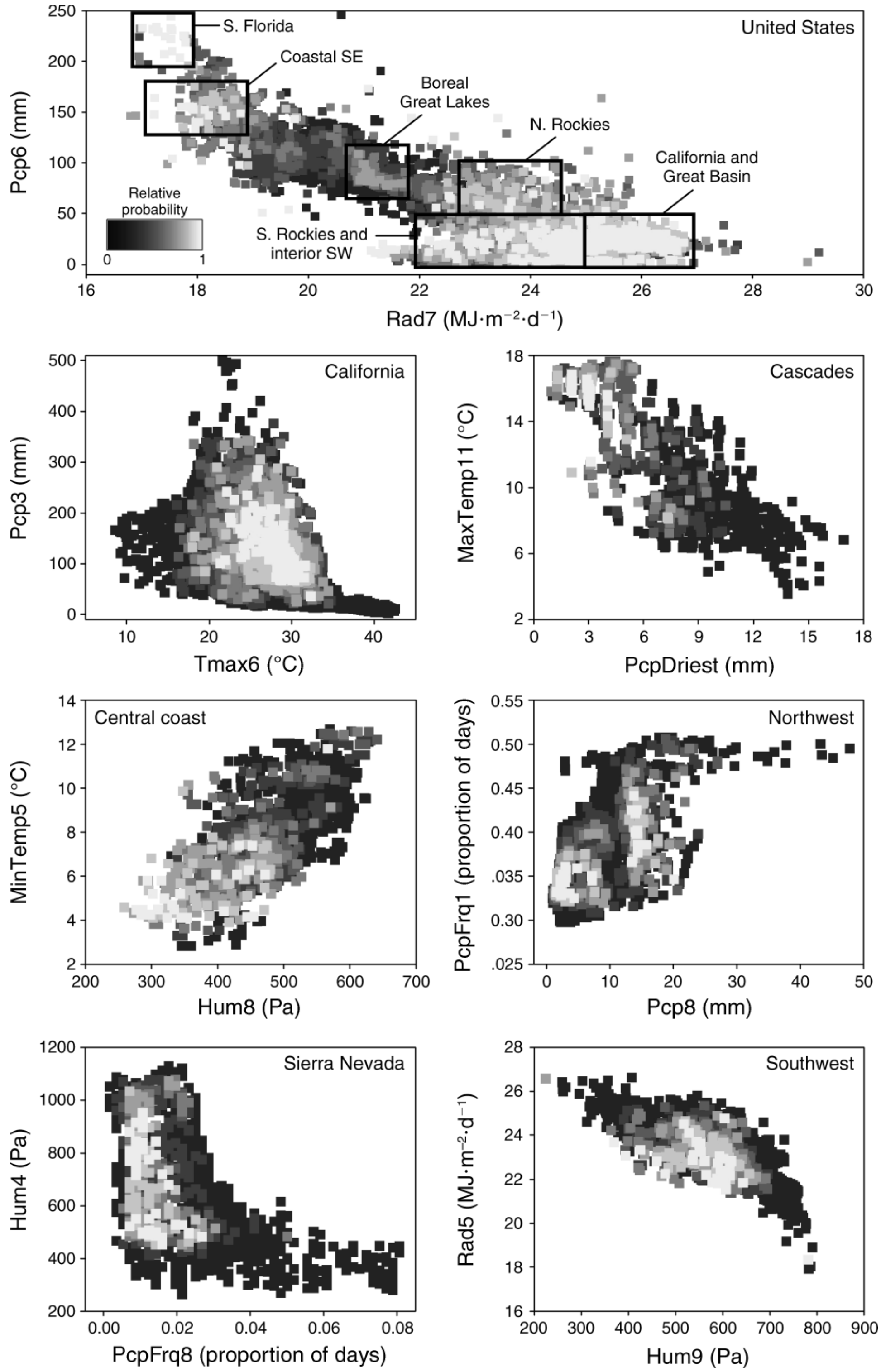

FIG. 3. The relative probability of wildfire occurrence from the climate-only model sampled uniformly over the full extent of each study area, as a function of the two top climatic variables (except for the Cascades, where the first and third variables were used). A key to the variables is in Table 2. In the United States panel, the approximate locations of geographical regions of interest are indicated.

TABLE 4. Measures of performance of the three model types for each study area produced with Maxent: models using all variables ("full"), those using all climate variables ("climate only"), and those using only the derived climate variables ("derived climate").

	Full			Climate only			Derived climate		
Location	Suitable area (%)	Omission error (%)	AUC	Suitable area (%)	Omission error (%)	AUC	Suitable area (%)	Omission error (%)	AUC
United States	25.9	29.3	0.850	26.7	29.8	0.839	28.5	20.2	0.826
California	27.9	26.6	0.871	28.6	18.6	0.856	20.1	29.6	0.842
Cascades	24.1	23.4	0.885	25.8	21.5	0.885	27.6	22.3	0.866
Central coast	26.5	21.9	0.879	25.9	23.8	0.873	26.7	27.4	0.851
Northwest	22.0	24.2	0.782	23.3	24.4	0.774	24.1	26.5	0.759
Sierra Nevada	28.2	27.9	0.792	29.0	30.0	0.781	20.8	25.9	0.785
Southwest	22.3	25.9	0.890	24.2	28.8	0.865	24.1	23.2	0.844

Notes: The "suitable area" represents the percentage of area predicted as suitable, whereas the "omission rate" is the percentage of presence points found in areas predicted to be unsuitable (false negatives). The probability threshold is minimized according to the sum of these two values. AUC is the area under the curve of the sensitivity vs. the predicted area (i.e., 1 – specificity) plot.

in the boosted regression tree (BRT) models also show strong prediction accuracy (Table 5). These measures are more variable among study areas than in Maxent and show an imbalance in the commission and omission errors in some study areas such as the United States, and the Northwest and Southwest ecoregions, suggesting a differential in BRT's classification success among presences in absences.

According to the area under the curve (AUC), model performance decreased marginally as less information was included (i.e., fewer variables) in the Maxent and BRT models (Tables 4 and 5). The mean discrepancy is somewhat larger between the climate-only and derived-climate models than between the full and climate-only models. All models are highly informative, but notably lower AUCs are observed for the Northwest and Sierra Nevada ecoregions in the three model types for both Maxent and BRT. The models for other study areas have AUC values that are similar enough to preclude any conclusive ranking in relative performance. The AUC of BRT models are always higher than those of Maxent, but this is largely due to the maximum achievable value of AUC in Maxent being <1. The training AUCs (not shown) obtained in Maxent and BRT are similar, suggesting a similar quality of fit to the data if no overfitting is assumed.

Although the three model types produce similar spatial predictions for the United States and California, there are important local and regional differences among them, as shown in the "delta" maps (Fig. 4). These comparisons suggest some over-fitting of the full model, which predicts low suitability for fire in at least two parts of the United States with a known potential for large wildfires, south-central Texas and northern Wisconsin/Michigan Upper Peninsula (the north-central and southern large dark red areas in the first "delta" map in Fig. 4). In the northeastern United States, despite the proportionally large change in relative probability from the full to the climate-only model, the probability values of the latter model remain generally low. Although the climate-only and derived-climate models may be more appropriate for identifying areas of potential wildfire suitability, much of the fine-grained quality of the predictions is lost at the scale of the United States. For example, the southeastern deserts and the Central Valley ecoregions are predicted to be more suitable for fire in the United States model than in the California models. In general, the ecoregion model predictions are similar

TABLE 5. Measures of performance of the three model types for each study area produced with boosted regression trees: models using all variables ("full"), those using all climate variables ("climate only"), and those using only the derived climate variables ("derived climate").

	Full			Climate only			Derived climate		
Location	Commission error (%)	Omission error (%)	AUC	Commission error (%)	Omission error (%)	AUC	Commission error (%)	Omission error (%)	AUC
United States	22.86	15.16	0.883	23.54	14.62	0.883	24.71	14.86	0.877
California	16.34	15.67	0.921	15.98	16.03	0.920	17.23	17.50	0.909
Cascades	15.74	16.31	0.903	16.18	17.60	0.897	14.59	19.21	0.890
Central coast	14.90	17.48	0.902	14.48	19.36	0.896	14.25	20.35	0.890
Northwest	18.37	24.92	0.851	19.33	24.94	0.847	17.65	28.37	0.842
Sierra Nevada	25.79	23.27	0.832	22.82	24.67	0.836	24.30	22.18	0.839
Southwest	19.44	13.85	0.907	19.24	16.99	0.896	20.83	17.39	0.882

Notes: The "commission error" (false positives) is the percentage of presence points misclassified as absences, whereas the "omission rate" is the percentage of presence points found in areas predicted to be unsuitable (false negatives). The probability threshold is minimized according to the sum of these two values. AUC is the area under the curve of the sensitivity vs. 1 – specificity plot.

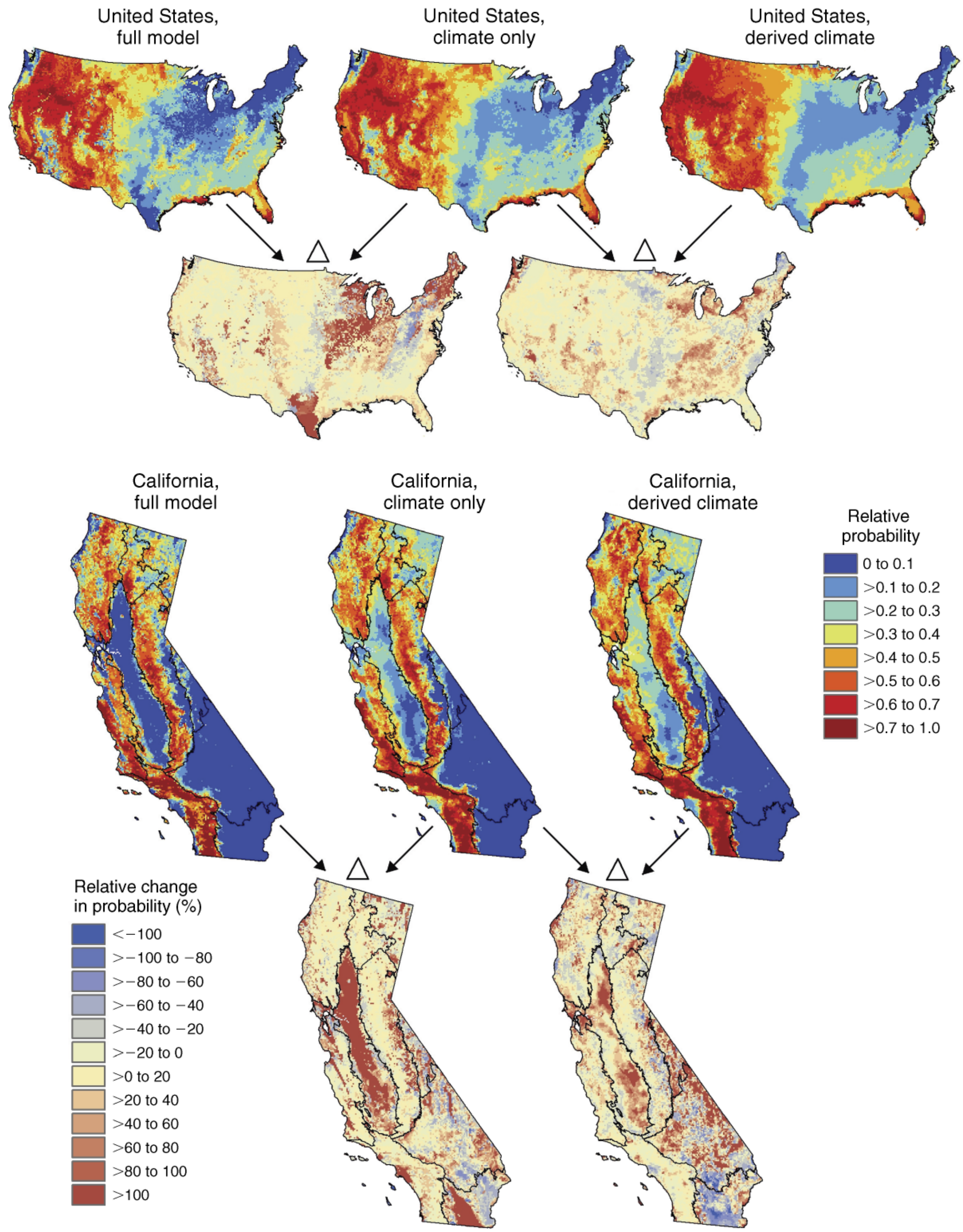

FIG. 4. Maxent outputs for the U.S. and California study areas, based on inputs from the full model, climate-only, and derived-climate scenarios (see *Methods*) and the relative difference (indicated by "delta" [Δ] symbols) between output predictions as expressed in relative change in probability (for "climate only" minus "full," and "derived climate" minus "climate only").

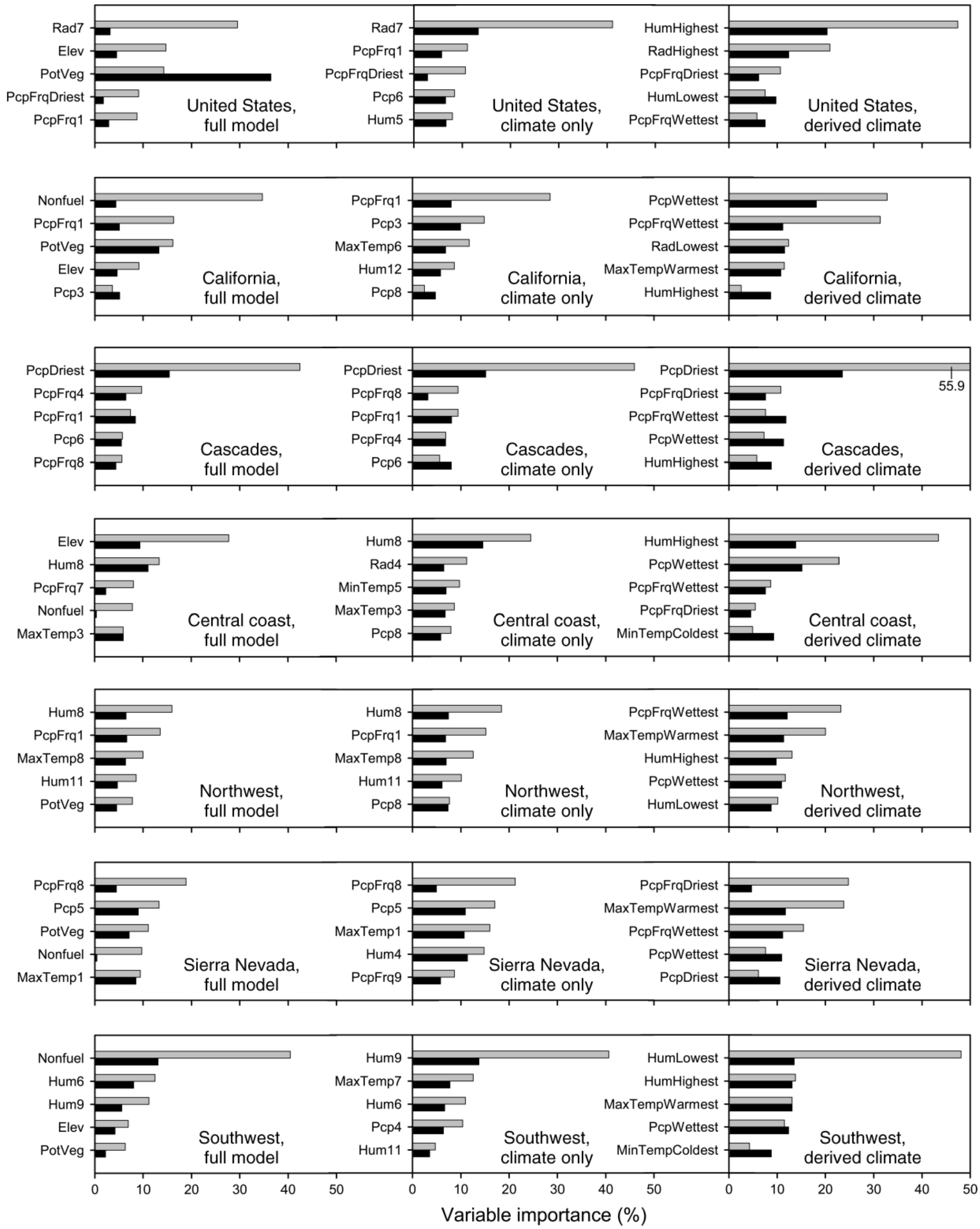

FIG. 5. The relative contribution of the top five variables of the full model, climate-only, and derived-climate scenarios produced in Maxent (gray bars) in each study area and the corresponding variable-scenario combinations values obtained in the boosted regression tree (BRT) models (black bars). Note that the technique for evaluating variable contribution differs between Maxent and BRT (see *Methods*).

to those shown for California as a whole and are therefore not presented.

Variable contribution

All variable categories emerged as top variables in at least one of the study areas (Fig. 5). There is no consistent seasonal pattern among variables of the full and climate-only models, but most of the top five Maxent monthly climate variables cover the period from June to September. The indirect variables, elevation (Elev), potential natural vegetation (PotVeg), and Nonfuel perform well, but their performance also varies among study areas. The variable contribution rankings of derived-climate models are consistent with those of the climate variable rankings of the full and climate-only models. For all derived-climate Maxent models, the most important variable is always a moisture-related variable: absolute humidity (Hum), precipitation amount (Pcp), or precipitation frequency (PcpFrq; Fig. 5). There are similarities among some ecoregions, notably the Central Coast and Southwest ecoregions, and to a lesser extent among the Cascades and Sierra Nevada ecoregions. The ranking of variables in terms of contribution is generally similar between Maxent and BRT, but the consistency varies among model types, as there is greater overall agreement within climate-only and derived-climate models than within full models.

Variable interactions

The first three nodes of the simplified BRT tree models provide relevant information regarding the dominant trends and interactions of climate–fire relationships (Fig. 6). The first node of the U.S. tree shows that areas of lower maximum absolute humidity (HumHighest $<$ 1184 Pa), corresponding to much of the western United States, are generally suitable for fire; however, within the high humidity class (HumHighest $>$ 2779 Pa) also lies an area of very high fire suitability, the Gulf of Mexico coastal region. The third node of the U.S. tree shows that in areas of moderate maximum humidity, precipitation of the driest month (PcpDriest) is the most important predictor variable. In California, the first node separates the arid and non-arid areas, as even the wettest month usually experiences fairly limited precipitation in the deserts. Within the areas of higher precipitation amounts, those experiencing more frequent precipitation events in their wettest month (PcpFrqWettest $>$ 0.25) have moderate fire suitability, whereas the areas of less frequent precipitation but high maximum temperature (MaxTempWarmest $>$ 29.5°C), chaparral-dominated areas, are highly suitable for fire.

Differences in fire environments among ecoregions are demonstrated in the ecoregion trees. The Cascades ecoregion's fire suitability appears to be highly limited by precipitation, as areas having higher precipitation in the driest month (PcpDriest $>$ 6.6 mm) have the lowest fire probability, whereas the highest fire probability is found in areas of combined low precipitation frequencies in both the driest and wettest months (PcpFrqDriest and PcpFrqWettest; mid-elevation areas in the rain shadow of the Klamath mountains). In the Central Coast ecoregion, fire is the least suitable in the areas of high maximum absolute humidity (HumHighest $>$ 823 Pa), which correspond to the low-lying coastal strip and deep inland valleys. The most suitable areas are those experiencing both relatively more rain in the wettest month (PcpWettest $>$ 130 mm) and higher minimum temperatures (MinTempColdest $>$ –1.2°C). In the Northwest ecoregion, the maximum temperature of the warmest month characterizes the main node, where cool coastal areas experience less fire (MaxTempWarmest $<$ 24.0°C). In areas experiencing maximum temperatures $>$24°C, areas of intermediate absolute humidity (i.e., mid-elevation areas) appear to be relatively more suitable for fire. In the Sierra Nevada ecoregion, the first node shows that maximum temperature of the warmest month coarsely delimits areas of cool-moist forests and shrublands/woodlands/dry forests. In each of these branches, fire is more suitable in drier areas (both splitting on PcpFrqWettest), although this discrepancy is larger in cooler areas. Fire suitability in the Southwest ecoregion is highly limited in low-lying areas of higher absolute humidity, as compared to the mountainous areas. Where fire is more suitable, its probability tends to be higher in areas of maximum precipitation frequency (PcpFrqWettest), as shown in the third node of the tree.

Model transfers to other study areas

We defined transfer of a model to another area as successful if the AUC was at least 0.60. According to this criterion, 19 of the 32 transfers and infill models are successful in predicting the environmental space of wildfire of a different study area (Table 6). Of the transfers that fail (AUC $<$ 0.60), nine produce models that perform worse than what would be expected by selecting observations at random (AUC $<$ 0.50). Models of some California ecoregions (e.g., Cascades and Southwest) are more transferable than others (i.e., Central Coast and Northwest), whereas the Sierra Nevada ecoregion model was successful in two out or four of its transfers. Only the most transferable ecoregion models are successfully projected up to the scale of California. Transfer of the California model up to the United States is almost successful according to the test AUC. The infill models produced for California, excluding the ecoregion of interest, all performed well, but this was less true for the Cascades ecoregion (Table 6). The infill model of the United States excluding California is highly successful, having an AUC higher than the non-transfer California model. This surprising result, also occurring in other transfers and infills, is largely due to a higher estimate of fire suitability where most of the presences (test points) are concentrated in those areas. For example, in the transfer from the Cascades ecoregion model to the Southwest ecoregion,

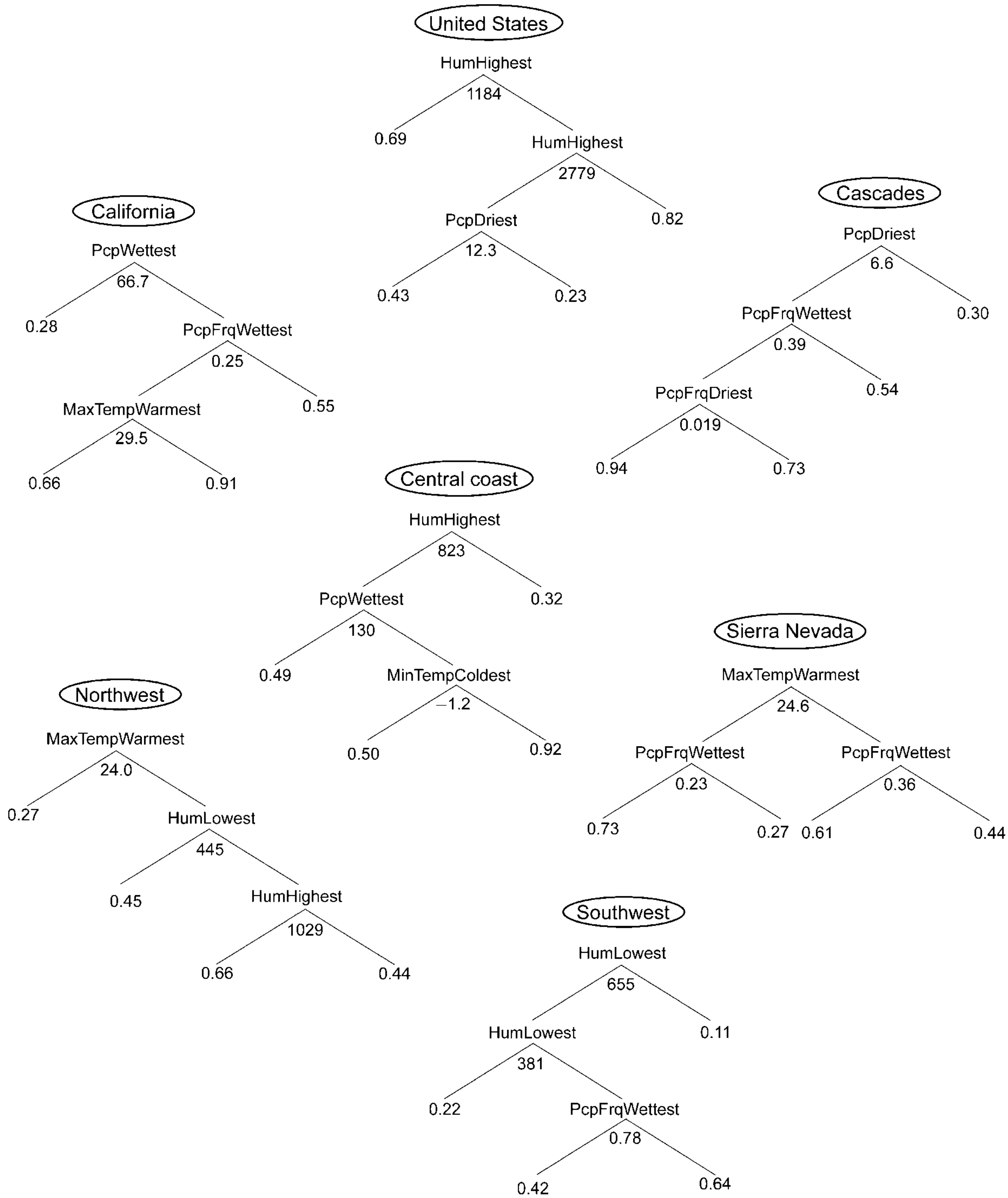

FIG. 6. Simplified versions of the first tree of the derived-climate models computed with the boosted regression tree algorithm with a Bernoulli (logistic) response function in each study area. The first three splits of each tree are shown to illustrate the interactions between key model variables. The splitting variable is shown above the node with its corresponding splitting value below the node (units are provided in Table 2). The lower values follow the left branch at each node, and the values at the end of terminal nodes represent the relative probability of fire occurrence.

TABLE 6. Measures of model performance for the projection of the Maxent derived-climate models onto other study areas.

Model type and model	Transfer area	Suitable area (%)	Test omission rate (%)	AUC	ΔAUC
Projection					
California	USA	73.19	6.01	0.576	−0.266
Cascades	California	8.80	24.17	**0.885**	+0.019
Central coast	California	9.46	76.90	0.474	−0.377
Northwest	California	10.96	73.44	0.517	−0.242
Sierra Nevada	California	21.20	49.09	**0.699**	−0.087
Southwest	California	11.18	37.06	**0.820**	−0.024
Cascades	Central coast	9.09	8.07	**0.961**	+0.095
	Northwest	14.70	40.40	**0.754**	−0.112
	Sierra Nevada	7.99	24.39	**0.899**	+0.033
	Southwest	5.72	3.15	**0.987**	+0.121
Central coast	Cascades	71.07	28.86	0.161	−0.690
	Northwest	9.35	71.73	0.494	−0.357
	Sierra Nevada	11.97	87.22	0.342	−0.509
	Southwest	11.82	79.46	0.458	−0.393
Northwest	Cascades	58.25	25.01	0.559	−0.200
	Central coast	7.41	84.68	0.398	−0.361
	Sierra Nevada	12.42	60.01	0.591	−0.168
	Southwest	55.82	29.22	0.468	−0.291
Sierra Nevada	Cascades	85.96	2.00	0.444	−0.341
	Central coast	47.08	12.30	**0.759**	−0.026
	Northwest	77.71	15.11	0.417	−0.368
	Southwest	14.82	12.91	**0.937**	+0.152
Southwest	Cascades	16.26	35.10	**0.772**	−0.072
	Central coast	4.11	23.48	**0.913**	+0.069
	Northwest	4.03	19.43	**0.920**	+0.076
	Sierra Nevada	4.96	43.26	**0.749**	−0.095
Infill					
USA	California	6.63	4.11	**0.988**	+0.162
California	Cascades	45.18	14.62	**0.690**	−0.152
	Central coast	28.87	6.50	**0.880**	+0.038
	Northwest	34.50	26.72	**0.727**	−0.115
	Sierra Nevada	35.45	15.72	**0.817**	−0.025
	Southwest	26.45	11.89	**0.881**	+0.039

Notes: The "suitable area" represents the percentage of area predicted as suitable, whereas the "omission rate" is the percentage of presence points found in areas predicted to be unsuitable (false negatives). The probability threshold is minimized according to the sum of these two values. Both of these measures are reported for the projection area. AUC is the area under the curve of the sensitivity vs. the predicted area (i.e., 1 – specificity) in the projection areas, and ΔAUC is the difference between the AUC of the projection area and the AUC of the model-building area. Test AUC values ≥0.60 are shown in bold.

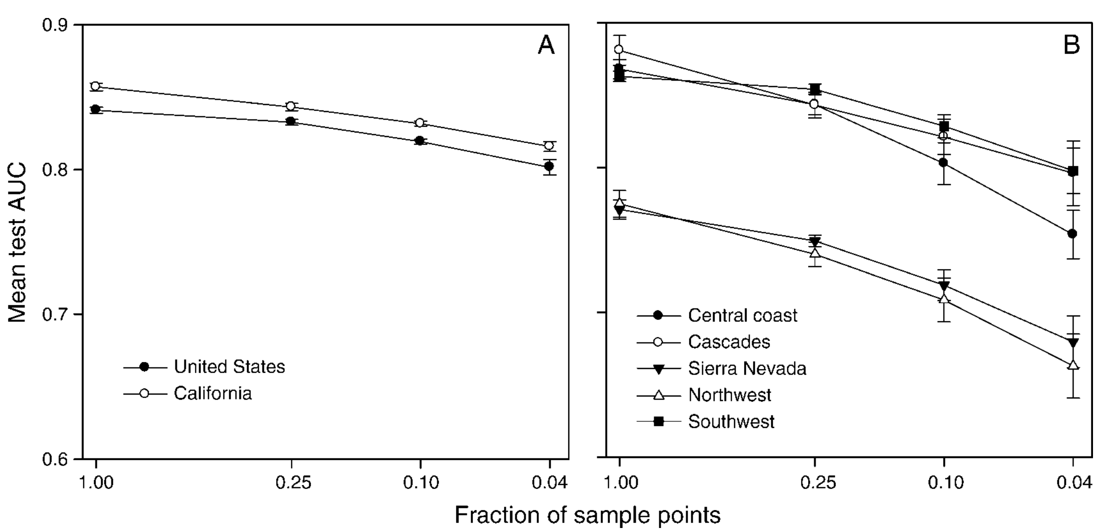

FIG. 7. The area under the curve (AUC, mean ± SD) for 10 iterations of the Maxent model as a function of decreasing fraction of sample points for (A) the United States and California and (B) the five ecoregions of California. AUC is the area under the curve of the sensitivity vs. 1 – specificity plot. Note that the *x*-axis has a log scale.

both the estimated predicted area and the omission error are very low (5.72% and 3.15%, respectively), which indicates that this model has extremely high classification accuracy at a probability thresholds that encompasses a small proportion of the total area (Table 6).

Effect of sample size on model predictions

The AUC of all models decreases as a function of sample size. Spatial model performance in Maxent relative to that of the full data set is still acceptable, or at least informative, with 1/10 of the total observations for all study areas (Fig. 7). However, the models of most ecoregions experience substantial loss of prediction accuracy when 1/25 of the available observations were used. The variability in AUC also increases as sample size decreased. This increase in variability is less pronounced in the models for the United States and California study areas, for which there are significantly more observations (289 and 170 training points for the United States and California, respectively) than in the models for ecoregions at the lowest sample size (11 to 49 training points).

Discussion

Environmental controls on wildfire

Our results support the assertion that environments unsuitable for wildfires tend to be those characterized by climatic extremes, even though this was not observed in all individual variables in all study areas. Low fire suitability areas in the United States rarely exhibit hot and dry periods, such as the northeast, the coastal northwest, and some high-elevation areas. The same is true on the other end of the climatic spectrum where climates are too hot and arid to support flammable and relatively contiguous vegetation cover, such as deserts and open shrublands. Other areas, such as parts of the Great Plains and the Midwest, which may have been suitable in the past, were probably predicted as unsuitable by our models because of the lack of fire-prone areas with analogous conditions sampled on Federal lands. Climatic constraints similar to those of the United States also apply for California, where additional relevant fine-grain patterns could be depicted at this scale. For example, it can be inferred from the relatively unsuitable coastal areas of the Northwest ecoregion that these represent a distinct wildfire climate given that they correspond to the distribution of tree species that are ill-adapted to frequent fires, such as Sitka spruce (*Picea sitchensis* [Bong.] Carr.), western hemlock (*Tsuga heterophylla* [Raf.] Sarg.), and western red cedar (*Thuja plicata* Donn; Griffin and Critchfield 1972).

Using a multiscale approach, we found that the distribution of fire is controlled by variables whose relative importance differs among study areas and spatial scales. There was limited overlap among top-ranked variables between the United States and California, and between California and its fire-prone ecoregions (Fig. 5). However, given the tight interrelationships among climate variables, the ensembles of different variables in each study area likely capture some of the same environmental space characteristics. The complexity of fire environments in areas such as the United States and California underline the importance of using several environmental predictors and their interactions to accurately describe fire's environmental space across that area. Partitioning California into ecoregions provided additional information about more localized wildfire controls, which was particularly useful in identifying the environmental conditions that limited the range of wildfire. For example, spatial fire patterns in the Cascades ecoregion are strongly linked to areas sheltered from summer precipitation events. In the neighboring Northwest ecoregion, an area of greater environmental variability, fire patterns are limited by a more complex set of environmental constraints characterized by high winter precipitation, low maximum temperatures, and high absolute humidity (i.e., dew point) patterns, as shown in the tree in Fig. 6. Our results thus agree with the claims that unique sets of environmental factors drive the fire regimes of the different mountainous areas of California (Taylor and Skinner 2003). The relationships shown in the trees are somewhat simplistic, but they provide a naïve illustration of the potential responses of fire suitability of ecoregions in a changing climate. For example, in the Sierra Nevada ecoregion, fire probabilities in the areas of cool-moist (i.e., high-elevation) forests may increase if the pulse in winter precipitation (PcpFrqWettest) decreases, as this variable is tied to summer drought severity (i.e., the length of the fire season), an important predictor of fire activity in the western U.S. forests (Westerling et al. 2006).

As opposed to the United States and California, the limited spatial extent of ecoregions may encompass only a subset of wildfire–environment responses, thus truncating this relationship before it has reached its extremes (Fig. 3). This is illustrated in the widely differing response of wildfire patterns to absolute humidity in the Central Coast and Sierra Nevada ecoregions. Clearly, absolute humidity does not have a pervasive direct impact on wildfire occurrence in California, but rather it acts as a proxy for other variables. For instance, in some parts of California this environmental predictor (and others) may effectively discriminate between low-lying areas of extensive land use (e.g., urban and agricultural) and the upland areas that still have a cover of natural flammable vegetation, as is the case in the Southwest and Central Coast ecoregions. The "natural" environmental space of some areas has experienced such drastic human-induced transformations that the relationship of fire to the true causal variable may have been weakened. For example, the Los Angeles Basin, a large coastal area where fire is now almost completely excluded, probably contributed heavily to the environmental space of fire in the Southwest ecoregion prior to human arrival, as flammable vegetation likely covered

Plate 1. (Upper) A vegetated landscape of the Central Coast ecoregion of California, USA (Las Trampas Regional Wilderness, Contra Costa County) encompassing grasslands, oak woodlands, shrublands, and exposed rock. (Lower) A vegetated landscape of the Northwest ecoregion of California (Austin Creek State Recreation Area, Sonoma County) encompassing grasslands, oak woodlands, shrublands, and mixed forests of conifers and hardwoods. As in many parts of California, the dominant vegetation of these landscapes varies according to complex environmental gradients. Photo credit: Eric Waller.

much of this area (Keeley 2002, Stephens et al. 2007). As this in no longer true, the current environmental space may be better described by variable sets that concurrently describe the natural and anthropogenic aspects of fire-proneness (Syphard et al. 2008).

Although proxy variables are correlates that may not represent the true causal environmental drivers of wildfire patterns, they may prove invaluable when the more mechanistic variables are unavailable, unknown, or not particularly useful for describing current wildfire patterns. Our results suggest that some proximal information is highly suitable for model building within a specific area. However, for the purpose of applying habitat distribution models to different areas or time periods, we must strive to identify variables that are the most portable. For example, given that our wildfire predictions are strongly linked to the presence of flammable vegetation, it might be appropriate to use synthetic variables directly tied to vegetation distribution, such as those for water balance suggested by Stephenson (1990). It is clear from our results, however, that vegetation type need not be explicitly included in the models because most of this information is contained in combinations of the climate variables. This insensitivity to land-cover variables at large spatial scales may be something of an advantage, as it may allow good models to be built without detailed land-cover data or the need to model complex vegetation–climate relationships. Alternatively, some variables directly related to fire-conducive conditions would have almost certainly provided additional useful information to our models. The most obvious of these is the mapped frequency of high-wind events that drive fire spread, information that is not yet available for many study areas. In any case, the main challenge will be to determine the variables that have the greatest generality, which will likely have the added benefit of being more scaleable (Stephenson 1998). To this end, much of our future effort will consist of identifying the best possible variables for wildfire habitat distribution models and testing their use and limitations in fire ecology.

Like species or biological communities, wildfire activity is dynamic over time. Many paleoecological studies have suggested that climate is the main driver of fire regimes at the temporal scale of centuries to millennia (Millspaugh et al. 2000, Carcaillet et al. 2001, Whitlock et al. 2003, Marlon et al. 2006). Similarly, changes in species distribution according to recent environmental change have been observed (Root et al. 2003), and these distribution changes have been used in mapping species ranges in climate change scenarios (e.g., Pearson and Dawson 2003, Thuiller et al. 2005). Successful future projections will require the inclusion of most of the future environmental space (Araújo et al. 2005, Hijmans and Graham 2006, Pearson et al. 2006), whether it comes from a modern analog to future climates or assumptions based on a new understanding of wildfire–environment relationships. These modeled changes hinge on the accurate quantification of an environment space that may be highly complex and produce rather surprising outcomes. For example, in the United States and California, many of the areas that are predicted to experience warmer and drier mean conditions may experience less, not more, wildfire, as they will no longer be able to support continuous flammable vegetation. In fact, this is precisely what may have reduced wildfire recurrence in many parts of the Sierra Nevada in the early Holocene (11 000 to 4500 years BP), a period that was significantly more arid than today (Anderson and Smith 1997). Although this area is in no imminent danger of a major vegetation shift, it is conceivable that a changing climate may cause more of its area currently suitable for wildfire to become "marginal habitat" similar to the one found in the eastern and southern parts of this range.

Habitat distribution models and spatial fire suitability predictions

The application of habitat distribution models to wildfire occurrence data appears to be a promising framework for predicting spatial wildfire suitability. Our approach was not designed to evaluate the specific sets of conditions under which fires ignite or burn in a given year. Rather, we assessed the environmental bounds within which wildfires have been observed over time. Such models can thus predict where fires are more or less likely to occur, but they are not useful for modeling when fires may happen. These two types of models provide complementary information. For example, some studies have used a correlative approach to predict area burned at very large spatial scales (subcontinental) from relationships between past fire occurrence and monthly climate normals (McKenzie et al. 2004, Flannigan et al. 2005). Although evaluated for several large regions, the outputs of these models were not predictive in a detailed spatial context. Use of the habitat distribution models could therefore complement predictions of area burned rates by predicting future geographic ranges, reflecting the potential for expansion or retraction of fire suitability. Furthermore, these models could be readily used in conjunction with those that spatially predict the annual climatic variability from oceanic climate oscillators (Gedalof et al. 2005, Kitzberger et al. 2007, Heyerdahl et al. 2008) to determine how the suitability of fire may vary annually.

Predictive habitat distribution models should be useful for numerous other theoretical or applied fire ecology questions. Because here we treated wildfire as a binary observation, similar to species presence or absence in a given location, we were not attempting to model variation inherent within natural fire regimes (e.g., fire intensity, season, frequency, size, type). However, the approach is well suited for more in-depth study of these components, and mapped inputs for fire regime types in the United States are available through the LANDFIRE Program (Schmidt et al. 2002). Wildfire habitat distribution models can aid conservation planning by predicting whether wildfire may be part of future restoration effort, taking an approach similar to that for modeling introduced species (Peterson 2003). Conversely, models could also be useful for examining potential wildfire suitability after invasion of certain species, since fire can "invade" new and unoccupied habitats as key plant species become established and alter local environmental constraints (e.g., D'Antonio and Vitousek 1992). Such change is currently happening in some of the Southeastern Deserts ecoregion, where the invasion of exotic grasses is leading to a previously rare continuous cover of fuels (Brooks and Berry 2006). This problem is particularly interesting from a theoretical standpoint, because it raises questions about whether fire regimes are "in equilibrium" with current environmental conditions, as is assumed for many species distribution modeling studies (Guisan and Thuiller 2005), and what thresholds must be crossed to alter basic wildfire–environment relationships.

The choice of algorithms for our habitat-distribution models was based on the research questions being asked and how these algorithms could overcome the limitations of our data. The Maxent and BRT algorithms were well suited for these criteria, but many other techniques could have been used. Although conceptually different, Maxent and BRT performed similarly well (e.g., Elith et al. 2006, Phillips and Dudík 2008) and produced similar rankings of variable contributions, despite different treatment of interactions. Our results also showed that both presence-only and presence–absence approaches could produce accurate classifications that were consistent with our knowledge of spatial fire patterns (Fig. 3). Presence–absence models may produce better classification when fire observations are abundant (e.g., Southwest ecoregion), whereas presence-only models are more suitable in areas of sparse observations (e.g., Northwest ecoregion), where a large fraction of the absences are "contaminated" with observations that are in reality likely to burn (i.e., false negatives).

There appears to be a trade-off between the amount of input information used in wildfire habitat distribution models and the appropriateness or realistic nature of outputs. This was chiefly due to inclusion of certain variable types, and not the number of observations (cf. Stockwell and Peterson 2002, Hernandez et al. 2006). Our results suggest that it was advantageous to use numerous data points to evaluate the environmental space of wildfire, as models produced with the most training points had a consistently better classification accuracy (AUC) without substantial overfitting (data not shown). The full models always had the best classification accuracy, albeit marginally, but their tendency to produce somewhat unrealistic predictions in some areas may reduce their general usefulness. The discrepancies between the full models and those produced solely from climate data were greatly influenced by the categorical variables of potential natural vegetation and nonfuel. These variables represent a categorical abstraction of a largely continuous phenomenon, and the way in which they are handled in habitat distribution models may sometimes be problematic. The main issue is that there is no information pertaining to similarity among classes, thus treating all classes as equally different, regardless of possible similarities. Even very high quality vegetation class data would have a tendency to "fragment" the environmental space and limit its transferability outside of each class. Although categorical variables are not always a problem in habitat distribution models, here they appear to have exacerbated the spatial bias in the U.S. wildfire data, as illustrated in the comparison of the full to the climate-only model, whereby wildfire suitability was underpredicted in some underrepresented vegetation classes that are known to be fire-prone.

It is clear that wildfire could occupy a broader geographical range than that recently observed, because our models built only from climatic variables predicted many currently fire-free areas as being relatively suitable. The most striking of these was the northern part of the Central Valley ecoregion of California. Wildfires have been absent from that area for well over a century, but our results suggest that fire was likely a common phenomenon there. There is very little lightning in this ecoregion, but it is thought that many fires could at one time have spread into the area from the adjacent mountains and that there was frequent burning by Native Americans (Wills 2006). By contrast, the southern part of the Central Valley ecoregion has a climate that is more similar to that of the deserts and may not have had a continuous cover of flammable vegetation (Davis 1999). The difference between mapped outputs produced with and without the potential natural vegetation and nonfuel variables thus provide some indication of where wildfire is currently limited by human land-use conversion or fire suppression.

The ability to transfer wildfire suitability models from one area to another has practical implications, especially for areas where information about fire ecology and fire occurrence is poor or nonexistent. The success of our infill model transfers emphasizes the importance of a spatial analog that provides necessary overlap in environmental space between the original and the projected areas. Although the transfers of wildfire habitat distribution models to new areas had mixed success, which is consistent with the results of Randin et al. (2006), there is much to be learned about the relative transferability of key fire–environment relationships used in the areas that performed well (i.e., Cascades and Southwest ecoregions). However, the failure of model transfers among some fire-prone ecoregions is not surprising, as the wildfire–environment relationships are different enough for significant parts of the environmental space of the projected area to be missing from those of the model-building area. Furthermore, transfers may suffer from the limited information related to anthropogenic influences (e.g., distance to roads, population density, land-use type), which may be prevalent in some areas (Sturtevant and Cleland 2007, Syphard et al. 2007), but this problem appears to be partly overcome at the spatial scales of the study areas by sufficient overlap in climatic space, as suggested by the success of the infill models.

Disturbance ecology and habitat suitability

Some environments are clearly more suitable for fire, an ecological disturbance process, while more marginal areas also exist. It remains to be seen whether ecological disturbances other than fire might be modeled according to their environmental requirements. Regardless, our results suggest that the habitat distribution modeling framework and habitat suitability concepts could contribute to disturbance ecology, which still lacks an inclusive general paradigm (White and Jentsch 2001). Recent consideration of fire as a "consumer" of resources and a "global herbivore" (Bond and Keeley 2005), however, indicates that new and related ideas are entering this arena of research.

The complex models of the wildfire–environment relationships used in Maxent and BRT represent a departure from the theoretical unimodal and "bell-shaped" response curves described for biota (Hutchinson 1959), a response that has been criticized as unrealistic and often assumed without basis (Austin and Smith 1989). This was particularly true for large areas such as the continental United States, where active fire regimes appear in several portions of the bivariate space for the top two climate variables. However, the complete n-dimensional environmental variable space characterizing habitat suitability is likely to be more unimodal and would resemble a distinct cloud of points with a large central nucleus and several protruding local optima. In any case, fire at this scale appears to occupy a very broad and complex environmental space, akin to a generalist organism able to inhabit many habitat types. Yet at such a broad spatial scale, we are mixing data from many different fire-regime types (e.g., frequent and low severity vs. infrequent and high severity) that may not be able to coexist. Different fire regimes may therefore be analogous to different species in the same genus, competing for similar resources and excluding one another at the same location. At finer scales, we see areas that are increasingly homogenous with respect to fire regimes, translating into less complex relationships between fire probabilities and key environmental variables.

Hierarchical environmental controls, which have been observed for the geographical distributions of species (e.g., Pearson et al. 2004), are also likely to hold for fire distributions. Over broad spatial scales one can expect climatic tolerances to regulate geographical distributions, while at finer scales the heterogeneous nature of essential resources will drive patterns (Guisan and Thuiller 2005). Ideally, the most informative environmental gradients that control habitat suitability for recurring fire at a given scale would be those capturing axes of the "fire regime triangle" shown here and in Fig. 1:

$$P(F_i) = f(\text{Vegetation}_i, \text{Climate}_i, \text{Ignitions}_i) \qquad (1)$$

where $P(F_i)$ is the long-term probability of fire being observed in location i, as a function of vegetation, climate, and ignition characteristics. For fire to play a natural role in an ecosystem, there must be sufficient biomass accumulated locally to be consumed in periodic fires. Environmental gradients capturing this aspect of the Vegetation factor in Eq. 1 are therefore analogous to resource gradients (Austin and Smith 1989) regulating productivity. Fire also has very specific environmental and rate-limiting constraints on the process of combustion, namely those involving hot, dry, and/or windy conditions (Moritz 1997, Beverly and Martell 2005). The Climate factor in Eq. 1 should therefore be represented by a direct environmental gradient (Austin and Smith 1989) in the frequency, duration, and intensity of weather conditions conducive to fire spread; it would thus account for variation in both live and dead fuel moisture conditions and also integrate wind speeds. In our study, many important climatic variables appeared to be related to vegetation productivity (e.g., summer precipitation and radiation for the United States; spring precipitation and summer temperature for California). However, it is difficult to untangle how variables act on the different factors in the equation, as precipitation variables, for example, may influence both growth of vegetation and dryness of fire-conducive weather conditions (Meyn et al. 2007). We generally lack good fire weather climatologies (Schroeder et al. 1964), and mapped environmental gradients associated with them will eventually improve the specification of the Climate factor in Eq. 1 and, subsequently, the characterization of suitable habitat for fire.

In a binary presence–absence context, for fire to exist an ignition must occur in conjunction with suitable conditions as described by Vegetation and Climate factors. Although the Ignition factor in Eq. 1 is a clearly limiting environmental constraint on fire presence, the episodic nature of ignitions might make them analogous to dispersal, which is often handled poorly in species distribution studies (Guisan and Thuiller 2005). The "infill" models consistently produced realistic predictions because of the abundance of ignitions in "infilled" regions, allowing fire to occupy suitable habitat there. Conversely, the potentially limiting nature of ignitions can help to explain discrepancies between availability of suitable fire habitat and the observed distribution of fire. For example, consider the fire-prone mediterranean-type ecosystems of the world, which contain sclerophyllous plant species adapted to summer drought. Unlike the other mediterranean-type floras, Chilean matorral contains plants that lack fire-adapted traits such as serotiny (Bond and van Wilgen 1996). The absence of summer lightning over evolutionary timescales is presumably responsible for this lack of fire adaptations, despite infrequent volcanic ignitions (Fuentes et al. 1994). Chilean matorral therefore appears to be an example of highly suitable but ignition-limited fire habitat, which was long unoccupied due to an inability of fire to disperse into this environment.

Once the minimum conditions for fire are met, different fire regime types emerge in suitable habitat as functions of covariation in Vegetation, Climate, and Ignition factors. Interactions between climate, fire regimes, and vegetation eventually result in ecological sorting on the landscape and fire–vegetation feedbacks that are to some degree self-reinforcing (D'Antonio and Vitousek 1992, Moritz et al. 2005). This feedback is what led Bond et al. (2005) to assert that it is the presence of fire, and not climate alone, that drives global vegetation patterns across large parts of the Earth.

Conclusion

Our study demonstrates that habitat distribution models can be used to characterize environmental controls on the distribution of wildfire, an abiotic ecological process. To explore fire suitability patterns at broad spatial scales, the variable selection scheme used here did not incorporate a priori knowledge of fire regimes in the study areas. Although this approach yielded good models and distinct combinations of important variables at different scales and locations, future efforts should use all available knowledge to target a few simple and causal variables (cf. Svenning and Skov 2004, Thuiller et al. 2005). Even so, the model predictions generated here were statistically robust, informative, and relatively easy to interpret. As such, this study addresses Skinner and Chang's (1996) plea for new information on large-scale wildfire patterns.

The application of concepts related to environmental gradients and habitat suitability also appears promising, and expectations of convex and unimodal fire response curves were met in many cases. Characterization of this complex "middle ground" making up the most suitable areas for wildfire indicates that the response of flammable vegetation and wildfire patterns to climate change may not be simple or linear. If key fire–vegetation–climate interactions persist, however, wildfire suitability models could be used to predict the future range of wildfire, because these models implicitly consider the potential for flammable vegetation as well as other fire-conducive environmental conditions. Although warming trends associated with climate change are generally assumed to mean more frequent and intense wildfires, much hotter and drier conditions could, in contrast, result in marginal fire habitats in many currently fire-prone locations. More research is needed to fully understand the potential for wildfire occurrence over large areas, but it is hoped that application of wildfire suitability models may eventually shed light on new aspects of wildfire–vegetation–climate dynamics.

Acknowledgments

We are grateful to John Battles, Mike Flannigan, Catherine Graham, Meg Krawchuk, Scott Stephens, Diana Stralberg, and Eric Waller for providing feedback, to Steven Phillips for his assistance with the Maxent model, and to Jane Elith for providing guidance with the boosted regression trees. This study was funded by the Canadian Forest Service and the University of California, Berkeley Center for Fire Research and Outreach.

Literature Cited

Anderson, M. K. 2006. The use of fire by Native Americans in California. Pages 417–430 *in* N. G. Sugihara, J. W. van Wagtendonk, K. E. Shaffer, J. O. Fites-Kaufman, and A. E. Thode, editors. Fire in California's ecosystems. University of California Press, Berkeley, California, USA.

Anderson, R. S., and S. J. Smith. 1997. The sedimentary record of fire in montane meadows, Sierra Nevada, California, USA: a preliminary assessment. Pages 313–327 *in* J. S. Clark, H. Cachier, J. G. Goldammer, and B. J. Stocks, editors. Sediment records of biomass burning and global change. NATO ASI Series, Vol. I 51:313–327.

Araújo, M. B., R. G. Pearson, W. Thuiller, and M. Erhard. 2005. Validation of species–climate impact models under climate change. Global Change Biology 11:1504–1513.

Austin, M. P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecological Modelling 157:101–118.

Austin, M. P., and T. M. Smith. 1989. A new model for the continuum concept. Plant Ecology 83:35–47.

Barbour, M. G., and J. Major. 1988. Terrestrial vegetation of California. John Wiley and Sons, New York, New York, USA.

Bekker, M. F., and A. H. Taylor. 2001. Gradient analysis of fire regimes in montane forests of the southern Cascade Range, Thousand Lakes Wilderness, California, USA. Plant Ecology 155:15–28.

Beverly, J. L., and D. L. Martell. 2005. Characterizing extreme fire and weather events in the Boreal Shield ecozone of Ontario. Agricultural and Forest Meteorology 133:5–16.

Bond, W. J., and J. E. Keeley. 2005. Fire as a global "herbivore": the ecology and evolution of flammable ecosystems. Trends in Ecology and Evolution 20:387–394.

Bond, W. J., and B. W. van Wilgen. 1996. Fire and plants. Chapman and Hall, London, UK.

Bond, W. J., F. I. Woodward, and G. F. Midgley. 2005. The global distribution of ecosystems in a world without fire. New Phytologist 165:525–538.

Bonn, A., and B. Schröder. 2001. Habitat models and their transfer for single- and multi-species groups: a case study of carabids in an alluvial forest. Ecography 24:483–496.

Breiman, L. 2001. Statistical modeling: the two cultures. Statistical Science 16:199–215.

Brooks, M. L., and K. H. Berry. 2006. Dominance and environmental correlates of alien annual plants in the Mojave Desert, USA. Journal of Arid Environments 67:100–124.

Brooks, M. L., and J. R. Matchett. 2006. Spatial and temporal patterns of wildfires in the Mojave Desert, 1980–2004. Journal of Arid Environments 67:148–164.

Brooks, M. L., and R. A. Minnich. 2006. Southeastern deserts bioregion. Pages 391–414 *in* N. G. Sugihara, J. W. van Wagtendonk, K. E. Shaffer, J. O. Fites-Kaufman, and A. E. Thode, editors. Fire in California's ecosystems. University of California Press, Berkeley, California, USA.

Brown, J. H. 1984. On the relationship between abundance and distribution of species. American Naturalist 124:255–279.

Bureau of Land Management. 2006. Wildland fire history. U.S. Department of the Interior, Washington, D.C., USA. ⟨ftp://ftp.nifc.gov/pub/FireHistoryData/⟩

Carcaillet, C., Y. Bergeron, P. J. H. Richard, B. Frechette, S. Gauthier, and Y. T. Prairie. 2001. Change of fire frequency in the eastern Canadian boreal forests during the Holocene: does vegetation or climate trigger the fire regime? Journal of Ecology 89:930–936.

Cardille, J. A., S. J. Ventura, and M. G. Turner. 2001. Environmental and social factors influencing wildfires in the Upper Midwest, United States. Ecological Applications 11:111–127.

Cary, G. J., R. E. Keane, R. H. Gardner, S. Lavorel, M. D. Flannigan, I. D. Davies, C. Li, J. M. Lenihan, T. S. Rupp, and F. Mouillot. 2006. Comparison of the sensitivity of landscape-fire-succession models to variation in terrain, fuel pattern, climate and weather. Landscape Ecology 21:121–137.

Collins, B. M., N. M. Kelly, J. W. van Wagtendonk, and S. L. Stephens. 2007. Spatial patterns of large natural fires in Sierra Nevada wilderness areas. Landscape Ecology 22:545–557.

Collins, B. M., and S. L. Stephens. 2007. Managing natural wildfires in Sierra Nevada wilderness areas. Frontiers in Ecology and the Environment 5:523–527.

Cyr, D., S. Gauthier, and Y. Bergeron. 2007. Scale-dependent determinants of heterogeneity in fire frequency in a coniferous boreal forest of eastern Canada. Landscape Ecology 22:1325–1339.

Daly, C. 2006. Guidelines for assessing the suitability of spatial climate data sets. International Journal of Climatology 26:707–721.

D'Antonio, C. M., and P. M. Vitousek. 1992. Biological invasions by exotic grasses, the grass/fire cycle, and global change. Annual Review of Ecology and Systematics 23:63–87.

Davis, F. W., and M. I. Borchett. 2006. Central coast bioregion. Pages 321–349 *in* N. G. Sugihara, J. W. van Wagtendonk, K. E. Shaffer, J. O. Fites-Kaufman, and A. E. Thode, editors. Fire in California's ecosystems. University of California Press, Berkeley, California, USA.

Davis, F. W., P. A. Stine, D. M. Stoms, M. I. Borchert, and A. D. Hollander. 1995. Gap analysis of the actual vegetation of California. 1. The southwestern region. Madroño 42:40–78.

Davis, F. W., D. M. Stoms, A. D. Hollander, K. A. Thomas, P. A. Stine, D. Odion, M. I. Borchert, J. H. Thorne, M. V. Gray, R. E. Walker, K. Warner, and J. Grace. 1998. The California gap analysis project final report. University of California, Santa Barbara, California, USA. ⟨http://www.biogeog.ucsb.edu/projects/gap/gap_rep.html⟩

Davis, O. K. 1999. Pollen analysis of Tulare Lake, California: Great Basin-like vegetation in central California. Review of Palaeobotany and Palynology 107:249–257.

De'ath, G. 2007. Boosted trees for ecological modeling and prediction. Ecology 88:243–251.

Elith, J., et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29: 129–151.

Elith, J., J. R. Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees. Journal of Animal Ecology 77: 802–813.

ESRI. 2005. ArcGIS 9.1. Environmental Systems Research Institute, Redlands, California, USA.

Fielding, A. H., and J. F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24:38–49.

Fielding, A. H., and P. H. Haworth. 1995. Testing the generality of bird-habitat models. Conservation Biology 9: 1466–1481.

Fire and Resource Assessment Program. 2006. Fire perimeters. California Department of Forestry and Fire Protection, Sacramento, California, USA. ⟨http://frap.cdf.ca.gov/projects/fire_data/fire_perimeters/index.asp⟩

Flannigan, M. D., K. A. Logan, B. D. Amiro, W. R. Skinner, and B. J. Stocks. 2005. Future area burned in Canada. Climatic Change 72:1–16.

Franklin, J. 1995. Predictive vegetation mapping: geographic modeling of biospatial patterns in relation to environmental gradients. Progress in Physical Geography 19:494–519.

Friedman, J. H. 2001. Greedy function approximation: a gradient boosting machine. The Annals of Statistics 29: 1189–1232.

Friedman, J. H. 2002. Stochastic gradient boosting. Computational Statistics and Data Analysis 38:367–378.

Fuentes, E. R., A. M. Segura, and H. Holmgren. 1994. Are the responses of matorral shrubs different from those in an ecosystem with a reputed fire history? Pages 16–25 *in* J. M. Moreno and W. C. Oechel, editors. The role of fire in mediterranean-type ecosystems. Springer-Verlag, New York, New York, USA.

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

Giglio, L., G. R. van der Werf, J. T. Randerson, G. J. Collatz, and P. Kasibhatla. 2006. Global estimation of burned area using MODIS active fire observations. Atmospheric Chemistry and Physics 6:957–974.

Graham, C. H., S. Ferrier, F. Huettman, C. Moritz, and A. T. Peterson. 2004. New developments in museum-based informatics and application in biodiversity analysis. Trends in Ecology and Evolution 19:497–503.

Griffin, J. R., and W. B. Critchfield. 1972. The distribution of forest trees in California. USDA, Forest Service. Research Paper PSW-82. Pacific Southwest Forest and Range Experiment Station, Berkeley, California, USA.

Guisan, A., and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters 8:993–1009.

Hardy, C. C., K. M. Schmidt, J. P. Menakis, and R. N. Sampson. 2001. Spatial data for national fire planning and fuel management. International Journal of Wildland Fire 10: 353–372.

Hernandez, P. A., C. H. Graham, L. L. Master, and D. L. Albert. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography 29:773–785.

Heyerdahl, E. K., L. B. Brubaker, and J. K. Agee. 2001. Spatial controls of historical fire regimes: a multiscale example from the Interior West, USA. Ecology 82:660–678.

Heyerdahl, E. K., D. McKenzie, L. D. Daniels, A. E. Hessl, J. S. Littell, and N. J. Mantua. 2008. Climate drivers of regionally synchronous fires in the inland Northwest (1651–1900). International Journal of Wildland Fire 17:40–49.

Hickman, J. C. 1993. The Jepson manual: higher plants of California. University of California Press, Berkeley, California, USA.

Hijmans, R. J., and C. H. Graham. 2006. The ability of climate envelope models to predict the effect of climate change on species distributions. Global Change Biology 12:1–10.

Hutchinson, G. E. 1959. Homage to Santa Rosalina or why are there so many kinds of animals? American Naturalist 93:145–159.

Kadmon, R., O. Farber, and A. Danin. 2004. Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. Ecological Applications 14:401–413.

Keeley, J. E. 2002. Native American impacts on fire regimes of the California coastal ranges. Journal of Biogeography 29: 303–320.

Keeley, J. E. 2006. South Coast bioregion. Pages 350–390 *in* N. G. Sugihara, J. W. van Wagtendonk, K. E. Shaffer, J. O. Fites-Kaufman, and A. E. Thode, editors. Fire in California's ecosystems. University of California Press, Berkeley, California, USA.

Kitzberger, T., P. M. Brown, E. K. Heyerdahl, T. W. Swetnam, and T. T. Veblen. 2007. Contingent Pacific–Atlantic Ocean influence on multicentury wildfire synchrony over western North America. Proceedings of the National Academy of Sciences (USA) 104:543–548.

Krawchuk, M. A., S. G. Cumming, M. D. Flannigan, and R. W. Wein. 2006. Biotic and abiotic regulation of lightning fire initiation in the mixedwood boreal forest. Ecology 87:458–468.

Kuchler, A. W. 1964. Potential natural vegetation of the conterminous United States. American Geographical Society, Special Publication No. 36.

Larsen, C. P. S. 1997. Spatial and temporal variations in boreal forest fire frequency in northern Alberta. Journal of Biogeography 24:663–673.

Leathwick, J. R., J. Elith, W. L. Chadderton, D. Rowe, and T. Hastie. 2008. Dispersal, disturbance, and the contrasting biogeographies of New Zealand's diadromous and nondiadromous fish species. Journal of Biogeography 35:1481–1497.

Leathwick, J. R., J. Elith, M. P. Francis, T. Hastie, and P. Taylor. 2006. Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. Marine Ecology Progress Series 321: 267–281.

Liu, C., P. M. Berry, T. P. Dawson, and R. G. Pearson. 2005. Selecting thresholds of occurrence in the prediction of species distributions. Ecography 28:385–393.

MacArthur, R. H. 1972. Geographical ecology: patterns in the distribution of species. Harper and Row, New York, New York, USA.

Marlon, J., P. J. Bartlein, and C. Whitlock. 2006. Fire–fuel–climate linkages in the northwestern U.S. during the Holocene. The Holocene 16:1059–1071.

McKenzie, D., Z. Gedalof, D. L. Peterson, and P. Mote. 2004. Climatic change, wildfire, and conservation. Conservation Biology 18:890–902.

Mermoz, M., T. Kitzberger, and T. T. Veblen. 2005. Landscape influences on occurrence and spread of wildfires in Patagonian forests and shrublands. Ecology 86:2705–2715.

Meyn, A., P. S. White, C. Buhk, and A. Jentsch. 2007. Environmental drivers of large, infrequent wildfires: the emerging conceptual model. Progress in Physical Geography 31:287–312.

Millspaugh, S. H., C. Whitlock, and P. J. Bartlein. 2000. Variations in fire frequency and climate over the last 17,000 years in central Yellowstone National Park. Geology 28:211–214.

Minnich, R. H. 2006. California climate and fire weather. Pages 13–37 *in* N. G. Sugihara, J. W. van Wagtendonk, K. E. Shaffer, J. O. Fites-Kaufman, and A. E. Thode, editors. Fire in California's ecosystems. University of California Press, Berkeley, California, USA.

Moisen, G. G., E. A. Freeman, J. A. Blackard, T. S. Frescino, N. E. Zimmermann, and T. C. Edwards, Jr. 2006. Predicting tree species presence and basal area in Utah: a comparison of stochastic gradient boosting, generalized additive models, and tree-based methods. Ecological Modelling 199:176–187.

Moritz, M. A. 1997. Analyzing extreme disturbance events: fire in Los Padres National Forest. Ecological Applications 7: 1252–1262.

Moritz, M. A. 2003. Spatiotemporal analysis of controls on shrubland fire regimes: age dependency and fire hazard. Ecology 84:351–361.

Moritz, M. A., J. E. Keeley, E. A. Johnson, and A. A. Schaffner. 2004. Testing a basic assumption of shrubland fire management: how important is fuel age? Frontiers in Ecology and the Environment 2:67–72.

Moritz, M. A., M. E. Morais, L. A. Summerell, J. M. Carlson, and J. Doyle. 2005. Wildfires, complexity, and highly optimized tolerance. Proceedings of the National Academy of Sciences (USA) 102:17912–17917.

Odion, D. C., J. R. Strittholt, H. Jiang, E. J. Frost, D. A. DellaSala, and M. A. Moritz. 2004. Patterns of fire severity and forest conditions in the western Klamath Mountains, California. Conservation Biology 18:927–936.

Parisien, M. A., and L. Sirois. 2003. Distribution and dynamics of tree species across a fire frequency gradient in the James Bay region of Québec. Canadian Journal of Forest Research 33:243–256.

Pearce, J. L., and M. S. Boyce. 2006. Modelling distribution and abundance with presence-only data. Journal of Applied Ecology 43:405–412.

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

Pearson, R. G., T. P. Dawson, and C. Liu. 2004. Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. Ecography 27:285–298.

Pearson, R. G., C. J. Raxworthy, M. Nakamura, and A. T. Peterson. 2007. Predicting species' distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. Journal of Biogeography 34:102–117.

Pearson, R. G., W. Thuiller, M. B. Araújo, E. Martinez-Meyer, L. Brotons, C. McClean, L. Miles, P. Segurado, T. P. Dawson, and D. C. Lees. 2006. Model-based uncertainty in species range predictions. Journal of Biogeography 33:1704–1711.

Peterson, A. T. 2003. Predicting the geography of species invasions via ecological niche modeling. Quarterly Review of Biology 78:419–433.

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

Phillips, S., and M. Dudík. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31:161–175.

PRISM Group. 2004. Climate normals, 1971–2000. Oregon State University, Corvallis, Oregon, USA. ⟨http://www.prismclimate.org⟩

Pyne, S. J., P. L. Andrews, and R. D. Laven. 1996. Introduction to wildfire. Second edition. John Wiley and Sons, New York, New York, USA.

Randin, C. F., T. Dirnböck, S. Dullinger, E. Niklaus, M. Z. Zimmerman, and A. Guisan. 2006. Are niche-based species distribution models transferable in space? Journal of Biogeography 33:1689–1703.

Ridgeway, G. 2006. Generalized boosted regression models. Documentation on the R package "gbm," version 1.6-3. ⟨http://cran.r-project.org/⟩

Riegel, G. M., R. F. Miller, C. N. Skinner, and S. E. Smith. 2006. Northeastern plateaus bioregion. Pages 225–263 *in* N. G. Sugihara, J. W. van Wagtendonk, K. E. Shaffer, J. O. Fites-Kaufman, and A. E. Thode, editors. Fire in California's ecosystems. University of California Press, Berkeley, California, USA.

Rollins, M. G., P. Morgan, and T. Swetnam. 2002. Landscape-scale controls over 20th century fire occurrence in two large Rocky Mountain (USA) wilderness areas. Landscape Ecology 17:539–557.

Root, T. L., J. T. Price, K. R. Hall, S. H. Schneider, C. Rosenzweig, and J. A. Pounds. 2003. Fingerprints of global warming on wild animals and plants. Nature 421:57–60.

Rothermel, R. C. 1972. A mathematical model for predicting fire spread in wildland fuels. General Technical Report INT-115, USDA, Forest Service, Intermountain Forest and Range Experiment Station, Ogden, Utah, USA.

Russell-Smith, J., C. P. Yates, P. J. Whitehead, R. Smith, R. Craig, G. E. Allan, R. Thackway, I. Frakes, S. Cridland, M. C. Meyer, and A. M. Gill. 2007. Bushfire "down under": patterns and implications of contemporary Australian burning. International Journal of Wildland Fire 16:361–377.

Schemske, D. W., B. C. Husband, M. H. Ruckelshaus, C. Goodwillie, I. M. Parker, and J. G. Bishop. 1994. Evaluating approaches to the conservation of rare and endangered plants. Ecology 75:584–606.

Schmidt, K. M., J. P. Menakis, C. C. Hardy, W. J. Hann, and D. L Bunnell. 2002. Development of coarse-scale spatial data for wildland fire and fuel management. General Technical Report RMRS-GTR-87. USDA, Forest Service, Rocky Mountain Research Station, Fort Collins, Colorado, USA.

Schoennagel, T., T. T. Veblen, and W. H. Romme. 2004. The interaction of fire, fuels and climate across Rocky Mountain forests. BioScience 54:661–676.

Schroeder, M. J., M. Glovinsky, V. Hendricks, F. Hood, M. Hull, H. Jacobson, R. Kirkpatrick, D. Krueger, L. Mallory, A. Oertel, R. Reese, L. Sergius, and C. Syverson. 1964. Synoptic weather types associated with critical fire weather. National Bureau of Standards Paper AD 449-630. U.S. Department of Commerce Institute for Applied Technology, Washington, D.C., USA.

Schulte, L. A., D. J. Mladenoff, S. N. Burrows, T. A. Sickley, and E. V. Nordheim. 2005. Spatial controls of pre–Euro-American wind and fire disturbance in northern Wisconsin (USA) forest landscapes. Ecosystems 8:73–94.

Skinner, C. N., and C. Chang. 1996. Fire regimes, past and present. Pages 1041–1069 *in* Sierra Nevada ecosystem project: final report to Congress, Volume II, Chapter 38, Assessments and scientific basis for management options. University of California Davis, Center for Wildland Resources, Davis, California, USA.

Skinner, C. N., and A. H. Taylor. 2006. Southern Cascades bioregion. Pages 195–224 *in* N. G. Sugihara, J. W. van Wagtendonk, K. E. Shaffer, J. O. Fites-Kaufman, and A. E. Thode, editors. Fire in California's ecosystems. University of California Press, Berkeley, California, USA.

Skinner, C. N., A. H. Taylor, and J. K. Agee. 2006. Klamath Mountains bioregion. Pages 170–194 *in* N. G. Sugihara, J. W. van Wagtendonk, K. E. Shaffer, J. O. Fites-Kaufman, and A. E. Thode, editors. Fire in California's ecosystems. University of California Press, Berkeley, California, USA.

Spessa, A., B. McBeth, and C. Prentice. 2005. Relationships among fire frequency, rainfall and vegetation patterns in the wet-dry tropics of northern Australia: an analysis based on NOAA-AVHRR data. Global Ecology and Biogeography 14:439–454.

Stephens, S. L. 2001. Fire history of adjacent Jeffrey pine and upper montane forests in the Eastern Sierra Nevada. International Journal of Wildland Fire 10:161–176.

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

Stephens, S. L., R. E. Martin, and N. E. Clinton. 2007. Prehistoric fire area and emissions from California's forests, woodlands, shrublands, and grasslands. Forest Ecology and Management 251:205–216.

Stephenson, N. L. 1990. Climatic control of vegetation distribution: the role of the water balance. American Naturalist 135:649–670.

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

Stockwell, D. R. B., and A. T. Peterson. 2002. Effects of sample size on accuracy of species distribution models. Ecological Modelling 148:1–13.

Strauss, D., L. Bednar, and R. Mees. 1989. Do one percent of the forest fires cause ninety-nine percent of the damage? Forest Science 35:319–328.

Stuart, J. D., and S. L. Stephens. 2006. North Coast bioregion. Pages 147–169 *in* N. G. Sugihara, J. W. van Wagtendonk, K. E. Shaffer, J. O. Fites-Kaufman, and A. E. Thode, editors. Fire in California's ecosystems. University of California Press, Berkeley, California, USA.

Sturtevant, B. R., and D. T. Cleland. 2007. Human and biophysical factors influencing modern fire disturbance in northern Wisconsin. International Journal of Wildland Fire 16:398–413.

Svenning, J.-C., and F. Skov. 2004. Limited filling of the potential range in European tree species. Ecology Letters 7: 565–573.

Syphard, A. D., V. C. Radeloff, J. E. Keeley, T. J. Hawbaker, M. K. Clayton, S. I. Stewart, and 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, and M. K. Clayton. 2008. Predicting spatial patterns of fire on a southern California landscape. International Journal of Wildland Fire 17:602–613.

Taylor, A. H., and C. N. Skinner. 2003. Spatial patterns and controls on historical fire regimes and forest structure in the Klamath Mountains. Ecological Applications 13:704–719.

Thonicke, K., I. C. Prentice, and C. Hewitt. 2005. Modelling glacial–interglacial changes in global fire regimes and pyrogenic trace gas emissions. Global Biogeochemical Cycles 19:GB 3008.

Thornton, P. E., H. Hasenauer, and M. A. White. 2000. Simultaneous estimation of daily solar radiation and humidity from observed temperature and precipitation: an application over complex terrain in Austria. Agricultural and Forest Meteorology 104:255–271.

Thornton, P. E., and S. W. Running. 1999. An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. Agriculture and Forest Meteorology 93:211–228.

Thornton, P. E., S. W. Running, and M. A. White. 1997. Generating surfaces of daily meteorology variables over large regions of complex terrain. Journal of Hydrology 190:214–251.

Thuiller, W., S. Lavorel, M. B. Araújo, M. T. Sykes, and I. Colin Prentice. 2005. Climate change threats to plant diversity in Europe. Proceeding of the National Academy of Sciences (USA) 102:8245–8250.

U.S. Geological Survey. 2000. Global 30 arc-second elevation data set. EROS Data Center Distributed Active Archive Center, Sioux Falls, South Dakota, USA. ⟨http://edcdaac.usgs.gov/gtopo30/gtopo30.html⟩

van der Werf, G. R., J. T. Randerson, G. J. Collatz, and L. Giglio. 2003. Carbon emissions from fires in tropical and subtropical ecosystems. Global Change Biology 9:547–562.

Vanreusel, W., D. Maes, and H. van Dyck. 2007. Transferability of species distribution models: a functional habitat approach for two regionally threatened butterflies. Conservation Biology 21:201–212.

Van Wagner, C. E. 1977. Conditions for the start and spread of crown fire. Canadian Journal of Forest Research 7:23–24.

van Wagtendonk, J. W., and D. Cayan. *In press*. Temporal and spatial distribution of lightning strikes in California in relationship to large-scale weather patterns. Fire Ecology.

van Wagtendonk, J. W., and J. Fites-Kaufman. 2006. Sierra Nevada bioregion. Pages 264–294 *in* N. G. Sugihara, J. W. van Wagtendonk, K. E. Shaffer, J. O. Fites-Kaufman, and A. E. Thode, editors. Fire in California's ecosystems. University of California Press, Berkeley, California, USA.

Vázquez, A., B. Pérez, F. Fernández-González, and J. M. Moreno. 2002. Recent fire regime characteristics and potential natural vegetation relationships in Spain. Journal of Vegetation Science 13:663–676.

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

White, P. S., and A. Jentsch. 2001. The search for generality on studies of disturbance and ecosystem dynamics. Progress in Botany 62:399–450.

Whitlock, C., S. H. Shafer, and J. Marlon. 2003. The role of climate and vegetation change in shaping past and future fire regimes in the northwestern U.S. and the implications for ecosystem management. Forest Ecology and Management 178:5–21.

Wills, R. 2006. Central Valley bioregion. Pages 295–320 *in* N. G. Sugihara, J. W. van Wagtendonk, K. E. Shaffer, J. O. Fites-Kaufman, and A. E. Thode, editors. Fire in California's ecosystems. University of California Press, Berkeley, California, USA.

Zaniewski, A. E., A. Lehmann, and J. M. Overton. 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. Ecological Modelling 157:261–280.