Spatial patterns and drivers for wildfire ignitions in California

Bin Chen1,2,* and Yufang Jin1

1 Department of Land, Air and Water Resources, University of California, Davis, CA 95616-8627, United States of America
2 Division of Landscape Architecture, Faculty of Architecture, The University of Hong Kong, Hong Kong SAR, People’s Republic of China
* Author to whom any correspondence should be addressed.
E-mail: binley.chen@hku.hk

Keywords: ignition risk, maximum entropy model, spatial drivers, climate change, human activities

Abstract
As a key component of wildfire activities, ignition is regulated by complex interactions among climate, fuel, topography, and humans. Considerable studies have advanced our knowledge on patterns and drivers of total areas burned and fire frequency, but much is less known about wildfire ignition. To better design effective fire prevention and management strategies, it is critical to understand contemporary ignition patterns and predict the probability of wildfire ignitions from different sources. We here modeled and analyzed human- and lightning-caused ignition probability across the whole state and sub-ecoregions of California, USA. We developed maximum entropy models to estimate wildfire ignition probability and understand the complex impacts of anthropogenic and biophysical drivers, based on a historical ignition database. The models captured well the spatial patterns of human and lightning started wildfires in California. The human-caused ignitions dominated the areas closer to populated regions and along the traffic corridors. Model diagnosis showed that precipitation, slope, human settlement, and road network shaped the statewide spatial distribution of human-started ignitions. In contrast, the lightning-caused ignitions were distributed more remotely in Sierra Nevada and North Interior, with snow water equivalent, lightning strike density, and fuel amount as primary drivers. Separate region-specific model results further revealed the difference in the relative importance of the key drivers among different sub-ecoregions. Model predictions suggested spatially heterogeneous decadal changes and an overall slight decrease in ignition probability between circa 2000 and 2010. Our findings reinforced the importance of varying humans vs biophysical controls in different fire regimes, highlighting the need for locally optimized land management to reduce ignition probability.

1. Introduction
Wildfires have become more intense and destructive across the Western States, especially in the past two decades since 2000. The increasing trend in total areas burned has been documented by numerous studies, and is mostly attributed to climatic warming (Westerling et al. 2006, Westerling and Bryant 2008, Westerling et al. 2011), drought (Swain et al. 2014, Robeson 2015), and dense fuels due to a few decades of active fire suppression (Stephens 2005). In California, devastating large wildfires became increasingly familiar scenes coincident with prolonged drought (Stephens et al. 2018), unprecedented tree mortality (Asner et al. 2016, Byer and Jin 2017, Fettig et al. 2019, Goulden and Bales 2019), and heat waves since 2012 (Williams et al. 2019, Gutierrez Aurora et al. 2021). The combination of heat waves, dry fuels, and strong winds created perfect conditions for large, fast-moving, and costly-to-fight fires (Stephens et al. 2018, Cardil et al. 2021). For example, 17 out of the top 20 largest fires in California’s history have occurred since 2000, among which, 11 occurred during the 2017–2021 period (CalFIRE 2021). A lengthening fire season has been observed, mostly due to warming (Westerling et al. 2006, Westerling 2016, Li and Banerjee 2021). The severity of burns has also been worsened due to climatic droughts, further exacerbated by
extreme heat waves in California (Huang et al 2020). These wildfire blazes destroyed tens to hundreds of thousands of structures in many fire-prone communities (Westerling and Bryant 2008, Moritz et al 2014, Jin et al 2015).

Wildfires can originate from both natural sources such as lightning and human ignitions such as those started by accident or set intentionally by arsonists. Human-started wildfires are dominant across the United States, accounting for 84% of wildfires and contributing to nearly half of the total burned areas from 1992 to 2012 (Balch et al 2017). In contrast, wildfires started by lightning were primarily located in the sparsely populated mountainous western areas. Other studies also highlighted the dominance of human ignitions in California, contributing to more than 90% of the wildfires in two-thirds of the counties over non-federal lands (Keeley and Syphard 2018, Li and Banerjee 2021). The rapid expansion of human settlements into the wilderness increased areas known as the wildland–urban-interface (WUI) (Radeloff et al 2005), placed more people living at the intersections between natural vegetation and communities, and enhanced the human accessibility to fuels along with the expansion of the road network. These development patterns increase ignition probability (Faivre et al 2014, Radeloff et al 2018, Syphard et al 2019). Wildfires started by humans expanded vastly the ‘fire niche’, in terms of both spatial extent and seasonality, across the contiguous United States (Balch et al 2017). Additionally, increases in electrical infrastructure and transmission lines with WUI expansions created further wildfire risks, especially under extreme weather condition, such as lightning storms and extreme heat-wave events (Calkin et al 2014) and downed power-lines due to wind gusts in Southern California (Syphard and Keeley 2015). The frequency of human-caused small wildfires (<500 acres) in California has increased most rapidly since 2000 (Li and Banerjee 2021).

As a key component of wildfire, ignition is regulated by complex interactions among human activities, climate, fuel, and topography (Countryman 1972, Faivre et al 2014). Under normal weather conditions without human interventions, the probability of ignitions is fuel-dependent and regulated by physical characteristics of fuels (Whelan 1995), given the natural ignition source. For example, fuel flammability of both live and dead plant materials, controlled by vegetation moisture contents through precipitation, temperature, and vapor pressure deficit, is a key factor for modulating fire ignition risks (Faivre et al 2014, Williams et al 2014, Flannigan et al 2016). Fuel amount, defined as the available combustible biomass, accounts largely for ignition propensity. For example, greater grassy fuels accumulated during two or three wet years in Mediterranean climates, e.g., in southern California, can promote fire occurrence (Jin et al 2015). Lower fuel moisture, during intensified and prolonged drought events or extreme heat conditions, contributes significantly to increased ignition probabilities by enhancing combustion initiation (Jurdao et al 2012, Chen et al 2021). Fuels including amount and continuity can be altered by management activities, such as vegetation treatments, fuel breaks, and land fragmentation due to land use, resulting in changes in ignition risk (Balch et al 2017, Radeloff et al 2018, Syphard et al 2019).

Many fire studies have advanced our understanding of the patterns, dynamics, and the associated drivers of total areas burned and frequency of large fires (Westerling and Bryant 2008, Westerling and Bryant 2008, Parisien and Moritz 2009, Westerling et al 2011, Balch et al 2017, Chen et al 2021). However, the emphasis has been on relatively large fires because the associated burned areas are well mapped. Much less is known about wildfire ignitions because of challenges in gathering reliable and complete data on ignitions of smaller fires, especially those suppressed by fire fighting. Very limited studies have addressed the controls from climate and human activities on wildfire ignitions (Faivre et al 2014), due to the following challenges (a) the stochastic nature of ignitions, especially for human started fires, (b) lack of a complete presence dataset recording the location of ignitions especially for small fires, and (c) incomplete records for attribution of fire causes, i.e. anthropogenic vs lightning. Moreover, the patterns and drivers differ substantially between human- and lightning-caused ignitions (Abatzoglou et al 2016, Keeley and Syphard 2018), but have not been systematically investigated in California at fine spatial scales.

Spatial map of ignition probability is critically important for assessing fire risk, prioritizing region-specific fuel management, and better preparing communities for fire-related emergency at local scales. Understanding what shapes the spatial patterns of contemporary wildfire ignitions is also fundamental for developing fire and land management strategies to reduce ignition risk (Faivre et al 2014, Chen et al 2021). These data and knowledge become more urgent needs, considering the continued trend of climate change in exacerbating fire risk. In this study, our goal was to improve our understanding of the spatial distribution of contemporary wildfire ignitions in California, by separating human vs. lightning caused ignition sources. We used a machine learning model to examine the complex relationships between ignitions and a suite of environmental covariates, including human settlement, climate, fuels, and topographical variables. Specifically, we aimed to address the following questions: (a) How did biophysical and anthropogenic controls shape the observed spatial patterns of fire ignitions in California? (b) How did these relationships vary across sub-ecoregions?
and (c) Is there a significant decadal change from 2000 to 2010 in human- and lightning-caused ignition probability? And if so, what contributed to the change?

2. Data and methods

2.1. Study area

California, covering an area of 423,970 km$^2$, is a topographically diverse state with elevation ranging from 0 to 4200 m (figure 1(a)). Much of California has a Mediterranean climate, characterized by hot summers and cool winters. The majority of the precipitation falls in winter, resulting in dry weather during the warmer seasons conducive to wildfires. Its climate varies widely with latitude, elevation, and proximity to the coast, e.g. the interior valley has a much hotter summer than the coastal areas and western slope of the Sierra Nevada. California is home to a high diversity of vegetation types, species, and fire regimes (Barbour et al 2019). The state can be divided into eight sub-ecoregions: North Coast, North Interior, Central Coast, Central Valley, Sierra Nevada, South Coast, South Interior, and Great Basin slope, based on California's Level III ecoregions (US Environmental Protection Agency 2012).

2.2. Ignition dataset

We used the US Forest Service Fire Program Analysis-Fire Occurrence Dataset (FPA-FOD), compiled from reporting systems of US federal, state, and local fire agencies (Short 2017). This homogenized and comprehensive dataset includes wildfire ignition records on both public and private lands from 1992 to 2015, and accounted for many small fires that are not included in many other fire datasets. Each ignition entry includes the location, discovery date, cause, and fire size (Short 2017). Specifically, the cause of fires was assigned as human, lightning, or missing data/undetermined, based on each wildfire incidence report. Anthropogenic ignitions included those started by arson/incendiarism, debris and open burning, equipment and vehicle use, firearms and explosives use, fireworks, smoking, recreation, etc. We also collected the statewide geospatial data layer of fire perimeters from California Fire and Resource Assessment Program (https://frap.fire.ca.gov/frap-projects/fire-perimeters), which recorded fires greater than 10 acres, to compare with FPA-FOD ignition data. Our analysis showed a high temporal consistency from 1992 to 2015 between these two datasets; however, the FPA-FOD recorded much more small fires, especially for human-started ones (supplementary materials, figure S1 available online at stacks.iop.org/ERL/17/055004/mmedia).

A total of 188,260 ignition records were extracted from the FPA-FOD data for California during the study period. We excluded 12,455 ignitions with missing fire cause, during the process of differentiating human- and lightning-caused ignitions. We further used the 2015 global land cover classification map at 300 m from the European Space Agency (ESA) Climate Change Initiative (ESA 2017), to generate a wildland mask by excluding water bodies, urban, and agricultural lands. This boundary mask was used to refine wildland ignitions for this study, resulting in a total of 134,115 ignitions. To analyze and model ignition patterns at a 1 km resolution, we further calculated the total number of ignitions from 1992 to 2015 (i.e. ignition frequency) over each 1 km by 1 km grid in California, for human- and lightning-caused ignitions, respectively.

2.3. Anthropogenic layers

We used spatial layers of population density, transportation road network, and nighttime lights,
## Table 1. List of anthropogenic and biophysical variables used in this study.

| Category                  | Type            | Indicator (abbreviation, unit at 1 km)                        | Time period          | Native spatial resolution | Data source                                                                 |
|---------------------------|-----------------|-----------------------------------------------------------------|----------------------|--------------------------|-----------------------------------------------------------------------------|
| Anthropogenic variables   | Population      | Population density (number km\(^{-2}\))                         | 2000–2015            | 1 km                     | Gridded population of the world version 4 (CIESIN 2018)                     |
|                           | Road density    | (Major, minor, trail roads)                                      | 2016                 | 30–500 m                 | OpenStreetMap                                                              |
|                           | Distance to     | major, minor, trail roads (m)                                    |                      |                          |                                                                             |
|                           | Human settlement| Nighttime light (mean value)                                     | 1992–2012            | 1 km                     | DMSP/OLS                                                                   |
| Biophysical variables     | Climate         | Precipitation (prcp, mm)                                         | 1992–2015            | 1 km                     | Daymet (Thornton et al 2020)                                               |
|                           | Max temperature | (tmax, °C)                                                       |                      |                          |                                                                             |
|                           | Min temperature | (tmin, °C)                                                       |                      |                          |                                                                             |
|                           | Water vapor     | pressure (vp, Pa)                                                |                      |                          |                                                                             |
|                           | Snow water      | equivalent (swe, kg m\(^{-2}\))                                 |                      |                          |                                                                             |
|                           | Incident shortwave radiation flux density (srad, W m\(^{-2}\)) |                      |                      |                          |                                                                             |
|                           | Lightning-strike density (number of strikes km\(^{-2}\))      | 1992–2012            | 1 km                     |                          | NOAA Vaisala National Lightning Detection Network (Landsat-5/7/8)           |
| Fuel amount               | Annual maximum  | normalized difference vegetation index (NDVI)                    | 1992–2015            | 30 m                     |                                                                             |
|                           | Elevation (m)    |                                                                 |                      |                          | Global multi-resolution terrain elevation data (GMTED; Danielson and Gesch 2011) |
|                           | Slope (degree)   |                                                                 |                      |                          |                                                                             |
|                           | Aspect          |                                                                 |                      |                          |                                                                             |

To quantify human settlement and accessibility (table 1). Estimates of mean population density came from the Gridded Population of the World (GPWv4) product at 1 km spatial resolution, available for 2000, 2005, 2010, and 2015 (Center for International Earth Science Information Network (CIESIN) 2018). The density of the major, minor, and trail roads within each 1 km cell was calculated from the Open Street Map (www.openstreetmap.org). We also calculated the shortest distance to major, minor, and trail roads at each 1 km cell. The annual 1 km nighttime light images during 1992–2012 were obtained from the Defense Meteorological Program (DMSP) Operational Line-Scan System as an indicator of human settlement and density of electrical infrastructure (Elvidge et al 1999, Hsu et al 2015). These variables were aggregated to a spatial resolution of 1 km at a baseline year and as long-term annual means, respectively.

### 2.4. Biophysical variables

We assembled statewide geospatial layers to evaluate the biophysical controls from topography, climate, and fuels on spatial variation of wildland ignitions (table 1). The 2010 global 250 m terrain elevation data (GMTED2010) was used to characterize slope and aspect at 1 km spatial resolution. Weather information came from the gridded Daily Surface Weather and Climatological Summaries meteorological data at 1 km (Daymet) (Thornton et al 2020), including precipitation (Prcp), minimum and maximum temperature (Tmin and Tmax), incident shortwave radiation (Srad), water vapor pressure (VP), and snow water equivalent (SWE), or the amount of water that would be released from melting snowpack. We derived long-term annual means during 1992–2015 for these meteorological variables at 1 km.

To quantify the natural ignition pressure from lightning strikes, we obtained 1 km gridded monthly lightning data (i.e. number of lightning strikes km\(^{-2}\))
from the National Oceanic and Atmospheric Administration (NOAA) Vaisala National Lightning Detection Network (www.ncdc.noaa.gov/data-access/severe-weather/lightning-products-and-services) from 1992 to 2012.

We used annual maximum normalized difference vegetation index (NDVI) from 30 m Landsat satellite imagery as an indicator of fuel amount. The Landsat surface reflectance images were filtered using the cloud mask and quality assessment information in the Landsat metadata. NDVI values were then calculated from the retained reflectance in the red and near infrared bands every 16 days. Annual maximum NDVI was calculated for each year and further aggregated to derive long-term annual means at 1 km to match the other biophysical layers.

2.5. Statistical modeling and analysis
We modeled the spatial pattern of ignition probability using the maximum entropy statistical method (MaxEnt v3.3.3k) (Phillips et al. 2004, 2006, 2021), which has been widely used for wildfire probability studies (Parisien and Moritz 2009, Moritz et al. 2012, Parisien et al. 2012, 2016). MaxEnt is a machine-learning technique originally designed to model species distribution from presence-only data using multidimensional environmental inputs (Phillips et al. 2004, 2006). It estimates a target probability distribution by iteratively searching for the probability distribution with maximum entropy (i.e. the one that is most uniform), subject to the environmental variables at each observation (i.e. presence-only point). MaxEnt allows us to model highly complex relationships while avoiding overfitting by using $l_1$-regularization (Phillips et al. 2006).

The presence-only framework of MaxEnt models was adopted here due to the following two considerations: (a) a lack of wildfire ignitions over the period 1992–2015 cannot be interpreted as a true absence in the past (i.e. before 1992); and (b) presence-only and presence-absence frameworks have been proven to provide similar model accuracies of wildfire probabilities (Parisien and Moritz 2009, Parisien et al. 2016). The ignition-presence samples were drawn from the 1 km ignition frequency maps as described in section 2.2. To derive the comparable probability maps for both human- and lightning-caused ignitions, we used the same cut-off threshold of 12 for selecting presence samples to balance the number and quality of presence samples (Chen et al. 2021), i.e. the 1 km grid cell was selected as a presence sample if a total of 12 or more wildfire ignitions occurred within the grid from 1992 to 2015. This resulted in a group of high-confidence ignition-presence samples with a total of 2388 human-caused and 165 lightning-caused ones statewide (supplementary material, figures S4 and S5).

To model the spatial variability of ignition, we used long-term mean annual climate, fuel amount, topography, and three categories of anthropogenic variables as independent variables, as shown in Table 1. Our correlation analysis showed that these six categories of variables were not highly correlated with each other, except for Tmax and Tmin (Pearson’s $r = 0.94$), elevation and Tmax/Tmin/vp (Pearson’s $r$ ranging from 0.68 to 0.79) (figure S2). Moreover, a large group of explanatory variables is not necessary an obstacle for the prediction reliability of MaxEnt models (Parisien et al. 2016, Chen et al. 2021). Therefore, we included all explanatory variables without differentiating direct and indirect causes for building MaxEnt models.

In addition to the ignition probability predicted at the statewide scale, we built a set of subecoregion specific MaxEnt models to predict the ignition probability for each one of the six sub-ecoregions (figure S3). The ignition-presence samples and corresponding anthropogenic and biophysical variables were extracted and refined using the sub-ecoregion boundaries. Due to the limited presence samples ($n < 40$) for the lightning started ignitions except for Sierra Nevada, we grouped North Coast and North Interior into one region, Central Coast, South Coast and South Interior into another region, resulting in three sub regional models (Table 2).

The MaxEnt model run was repeated ten times in a fourfold cross validation scheme, i.e. training the MaxEnt models with randomly selected 75% of the ignition-presence samples and validating the ignition probability for the remaining 25% of ignition-presence samples each time. The mean statistics on the 25% testing data from each of the ten replicates were used as quantitative measures of model performance. We used the receiver operating characteristic (ROC) curve, which was created by plotting sensitivity (i.e. the proportion of observed presences that are correctly predicted) on the $y$-axis against ‘1-specificity’ (i.e. the fractional predicted area) on $x$-axis for all possible thresholds, to quantify the model performance. Based on the ROC curves, we measured the area under the curve (AUC) value, expressed as the proportion of the total area of the square defined by the axis. The AUC can be regarded as the probability that a random ignition-presence sample is correctly predicted by the model (Phillips et al. 2006). For example, an AUC value of 0.5 indicates where prediction accuracy is no better than the scenario that samples are randomly selected, and an AUC value of 1 indicates the ideal model performance. Models with AUC values above 0.75 are typically considered robust and useful (Elith et al. 2011).

2.6. Variable contribution and marginalized response curves
We evaluated the importance of each variable in controlling the ignition spatial pattern. The relative
Table 2. Variable importance for spatial variability of human- and lightning-caused ignition probability in different sub-ecoregions of California. Results are shown for five most important variables for each sub-ecoregion, ordered by variable importance from left to right.

| Sub-ecoregion       | (Sample size) | Variable importance |
|---------------------|---------------|---------------------|
| Human-caused ignitions |               |                     |
| North Coast (367) | Nighttime light (29.9%) | Population density (25.2%) | Tmax (14.4%) | Tmin (6.2%) | Srad (4.5%) |
| North Interior (76) | Nighttime light (30.2%) | Population density (19.4%) | Major road (11.0%) | SWE (8.7%) | VP (7.5%) |
| Central Coast (253) | Nighttime light (49.3%) | NDVI (17.6%) | Tmax (8.9%) | Srad (6.4%) | VP (4.8%) |
| Sierra Nevada (1053) | Slope (21.8%) | Nighttime light (20.8%) | Population density (16.5%) | Prpc (16.0%) | Elevation (8.2%) |
| South Coast (519) | Nighttime light (26.7%) | Minor road (7.6%) | SWE (6.3%) | Prpc (5.9%) | Tmax (3.9%) |
| South Interior (121) | Nighttime light (50.7%) | Minor road (7.6%) | SWE (6.3%) | Prpc (5.9%) | Tmax (3.9%) |
| Lightning-caused ignitions |               |                     |
| North CA (56) | Lightning (40.6%) | Tmin (9.7%) | SWE (8.3%) | Trail road (7.3%) | Slope (6.0%) |
| Sierra Nevada (81) | SWE (61.1%) | Tmin (9.6%) | Lightning (5.7%) | NDVI (4.8%) | Prpc (4.4%) |
| South CA (35) | SWE (43.5%) | Prpc (20.9%) | Lightning (11.8%) | NDVI (7.3%) | Tmax (5.4%) |

3. Results

3.1. Spatial pattern of statewide ignition probability

On average, a total of 5478 fire ignitions occurred annually across the state of California, during 1992–2015. Fires started by humans were dominant, accounting for 80.6% of the total ignitions (figure 1(b)). High density of human-caused ignitions was mostly distributed nearby the populated areas at lower elevations, e.g. Orange County in Southern California, and foothills. Ignitions were also found clustered along the road network beyond the urban areas. In contrast, higher elevation mountainous areas such as Sierra Nevada and North Interior had much lower human-caused ignition probability, with only 41.5% of fire ignitions started by humans.

Wildfires ignited by lightning strikes averaged 1060 yr⁻¹ statewide, clustered more at higher elevations including Shasta-Trinity National Forest and Modoc National Forest in North Interior, Sierra Nevada, and Sequoia National Forest in Southern California (figure 1(c)). The east slopes of South Interior region had very limited fire ignitions by either human or lightning, probably due to lack of contiguous fuels in this arid region.

3.2. Controls for statewide fire ignition patterns

The MaxEnt models developed with the statewide presence samples performed well. The corresponding ROC curves had an average AUC of 0.86 and 0.96 over the ten-replicate runs for human- and lightning-caused ignition probability, respectively (figure 2). The statewide models captured the similar spatial patterns as shown by the observational records,

contribution was quantified as the increase in regularized gain by including the corresponding variable while keeping all other explanatory variables at their average sample values (Phillips et al. 2006). We further examined how each variable affected wildfire ignition probability using partial dependence plots (Phillips et al. 2006), i.e. the marginal response of ignition occurrence probability to each variable, when all other explanatory variables are kept constant at their mean values.

2.7. Decadal changes

To investigate decadal changes in ignition probability, we used the statewide MaxEnt models trained on 1992–2015 (as described in section 2.4) to estimate ignition probability at two decadal time periods, centered on the years 2000 (i.e. 1996–2005) and 2010 (i.e. 2006–2015). For example, the anthropogenic and biophysical variables from 1996 to 2005 were fed into the models for mapping circa year 2000 ignition risk.

We further designed two experimental scenarios for the circa year 2010, in order to isolate the impact from changes in anthropogenic and biophysical variables on human-caused and lightning-caused ignition probabilities, respectively. To assess the decadal change in ignition probability caused by climate and fuels, the MaxEnt models were driven by circa 2010 biophysical variables, while keeping the same anthropogenic variables as circa 2000. In this way, the predicted ignition probability change was independent of changes in human related variables. Similarly, by keeping the same circa 2000 biophysical data but replacing anthropogenic variables with circa 2010 data, the ignition probability change from the predictions can be attributed to changes in population, road network, and housing expansion.
Figure 2. The receiver operating characteristic (ROC) curves averaged over the ten replicates of the statewide MaxEnt-based ignition probability models for (a) human-caused ignitions, and (b) lightning-caused ignitions.

Figure 3. Predicted fire ignition probabilities at 1 km spatial resolution in California, based on the statewide (a), (b) and sub-ecoregion (c), (d) MaxEnt models trained by 1992–2015 ignition samples. The left panel (a) and (c) is human-caused ignition probability and the right panel (b) and (d) is lightning-caused ignition probability. The model development and prediction were done at wildland areas, excluding urban and agriculture (represented by white).

...and predicted well the difference in probability of ignitions from two types of sources (figures 3(a) and (b)).

Our analysis of the statewide models showed that mean annual precipitation was the most important driver determining the overall spatial distribution...
of human-caused ignition probability, with a relative contribution of 34.7% (figure 4(a)). The ignition probability increased rapidly with the increase in annual mean precipitation below 365 mm yr$^{-1}$ (an increase from $\sim$10% to $\sim$60%), remained fairly constant across intermediate precipitation values, and then slightly decreased beyond 1642.5 mm yr$^{-1}$ (figure 5(a)). Human-related variables, representing human settlement and accessibility to fuels, were also critical, e.g. as indicated by nighttime light contributing 23.3%, major road density contributing 3.3%, and population density contributing another 3.0% (figure 4(a)). Higher nighttime light density led to higher human-caused ignitions in general (figure 5(c)). We also found that slope had an almost equally important contribution (23.4%). The majority of human-caused ignitions occurred over areas with slope less than 40°, and the partial dependence plot showed steeper slopes enhanced ignition probability when slopes are higher than 20° (figure 5(b)). Long term mean Tmax and annual maximum NDVI contributed additional 2.9% and 2.4% (figure 4(a)).

For lightning-caused ignition probability, snow water equivalent (71.8%) was found as the most important driver controlling its spatial pattern statewide, followed by lightning strike density (9.2%) and annual maximum NDVI (4.3%) (figure 4(b)). Precipitation and elevation also contributed additional 4.1% and 2.4%, respectively (figure 4(b)). Lightning-caused ignition was rare over areas
with extremely low mean snow equivalent water (SWE), but its probability increased sharply over areas with relatively low SWE, stayed stable with SWE between 100 and 400 kg m\(^{-2}\) d\(^{-1}\), and then decreased slightly with SWE beyond 400 kg m\(^{-2}\) d\(^{-1}\), although larger variance was found among the replicates (figure 5(d)). The partial dependence on lightning strikes was stronger, with a rapid increase followed by a slower rate of change before reaching the peak over areas with two lightning strikes km\(^{-2}\) and declining gradually afterwards (figures 5(e) and (f)). Similar response to NDVI was found, and peak ignition probability occurred over areas with annual maximum NDVI of 0.6, when all other drivers were kept the same across the state.

3.3. Regional differences
The ecoregion-specific models were also found robust with sufficient accuracy, with AUCs ranging from 0.81 to 0.94 for human-caused ignitions across ecoregions and from 0.86 to 0.95 for lightning-caused ignitions (figures S6 and S7). Compared with the statewide models (figures 3(a) and (b)), predictions by sub-ecoregion specific MaxEnt models agreed better with the observations, and captured more local heterogeneity in spatial distributions of ignitions (figures 3(c) and (d)). For example, there was greater variability in predicted human-caused ignition probability along the traffic corridors in North Coast and Central Coast regions, and over large areas in Southern California in general (figure 3(c)). The ecoregion-specific models for lightning-caused ignition also captured finer spatial patterns in North California, Sierra Nevada, and Southern California than the statewide model (figure 3(d)).

Model diagnostics showed that dominant controls for the spatial pattern of human- and lightning-caused ignition probability varied across sub-ecoregions of California, as shown in table 2. As proxies of human settlement, nighttime light, followed by population density, were found as the most important drivers shaping the spatial variability of human-caused ignitions within each sub-ecoregion, contributing ~50% of relative contributions when combined (table 2). Two exceptions were Sierra Nevada, where slope contributed equally (22%) as those two human-related variables, and South Coast, where solar radiation contributed 41.8% followed by nighttime light contributing 26.7% (table 2, figure S8). The impacts of major or minor road density were found significant for North and South Interiors and South Coast, with relative contribution ranging from 7.6% to 11%. Climate variables contributed more than 20% in North Coast (mostly maximum temperature and solar radiation), and Central Coast (Tmax, solar radiation and vapor pressure), 16% in Sierra Nevada (precipitation) and 15% in North Interior (SWE and VP). Overall a positive impact of Tmax or solar radiation with high sensitivity was found (figures S8(a), (c) and (e)). Impacts of climate were much smaller in South Interior and South Coast. Fuel amount, as represented by max NDVI, was critical for Central Coast, increasing the human-started ignition probability from 20% to 75% when NDVI increases from 0.1 to beyond 0.75 (figure S8).

For lightning-caused fire ignitions, lightning strike density was the dominant factor explaining the spatial variation in Northern California, with a contribution of 40.6%, while SWE dominated the variability in Sierra Nevada (61.1%) and Southern California (43.5%) (table 2, figure S9). Tmin was found as the second important control in Northern California and Sierra Nevada, contributing about 10% (table 2, figure S9). In contrast, in southern California, mean annual precipitation ranked as the second leading controls, accounting for 20.9% and enhancing the lightning caused ignition probability, followed by lightning density (11.8%) (figure S9). Mean annual maximum NDVI, as a proxy for fuel availability, was found to contribute additional 7.3% in southern California and 4.8% in Sierra Nevada, respectively (table 2, figure S9).

3.4. Ignition probability change from 2000 to 2010
The spatial patterns of human- and lightning-caused ignition probability, as predicted by the statewide models using the circa 2000 and 2010 predictors, stayed similar in general (figure 6). 25.9% of the state (excluding agriculture and cities) experienced an increase in human-caused ignition probability, mostly clustering in North Coast, southern part of North Interior, southern Sierra, South Interior, and some more scattered patches in South Coast (figure 6(c)). This, however, was balanced out by more widespread or slight decreases in ignition probability (figure 6(c)). 5.5% of areas experienced a more significant decrease (greater than 10%) in human-caused ignition probability while only 0.7% of the region had a change greater than 10% (figures 6(c) and 7(a)). A more spatially heterogeneous pattern was found for decadal changes in lightning-caused ignition probability (figure 6(f)). Around 9.7% of wildland areas experienced a decrease of larger than 10% in probability while only 2.9% showed an increase of larger than 10% in probability (figure 7(b)). Averaged over the state, the probabilities of both human- and lightning caused ignition decreased slightly during the first decade of this century (figure 7), by roughly 2% in ignition probability.

Controlled predictions showed that the change in human-caused ignition probability during this decade was mostly due to changes in climate and fuels (figure 8(a)). However, we did find that the change in human-related variables increased human-caused ignitions over 41.2% of the areas, mainly in those
areas closer to population, such as southern California, foothills, and North Coast (figure 8(b)). This increase was partially weakened or canceled out by ignition reduction caused by biophysical variables. The lightning-caused ignition was also predominantly driven by changes in biophysical variables between circa 2000 and 2010 (figures 8(c) and (d)). The most significant increase was found in the Sierra Nevada and North Interior and northern part of North Coast (figures 6(f) and 8(c)).

4. Discussion
4.1. Controls on fire ignition probability
Our study used MaxEnt models to examine the complex relationships between ignitions and a suite of environmental drivers across scales. Across the whole state of California, the interaction among climate, topography, and human settlement and accessibility shaped the spatial patterns of human-caused ignitions. Anthropogenic factors, such as human
settlements as indicated by nighttime light, population density, and road networks, became more important at the sub-ecoregion scale, in determining localized human ignition patterns (figure S8). These reinforced the similar findings of considerable anthropogenic influences on fire regimes across space (Parisien et al 2016, Mansuy et al 2019, Chen et al 2021). For example, ignition frequency is affected significantly by distance to road and housing in southern California’s national forests (Faivre et al 2014) and by housing and human infrastructure in the northern, interior, and southern parts of California (Syphard et al 2019). Our study further indicated a negative impact from road density in North and South Interiors and South Coast, which may be due to elevated fire suppression efficiency over areas with denser road network and thus enhanced accessibility for firefighting.

We found SWE as the most important factor determining the spatial pattern of lightning-caused ignitions across the state and also in sub-ecoregions of Sierra Nevada and Southern California, although with different mechanisms (figures S9(b) and (c)). Snow melting in warm season from higher level of snowpack likely leads to higher fuel moisture, SWE, limits ignitions in the fire season of Southern California and foothills (figure S9(c), Lutz et al 2009). In contrast, higher SWE increases available water for spring vegetation growth in mountainous Sierra Nevada, which tend to be fine fuels in dry and hot summers and thus likely enhancing ignition (figure S9(b)).

4.2. Implications for management
In addition to deliberate or accidental ignitions, humans can actively alter wildfire risk through...
modification of fuel amount and continuity, e.g. by land use, fire suppression, and fuel management. Our predictions over statewide and sub-ecoregions of California provide an estimate of ignition probability at a 1 km resolution and thus serve as a valuable reference in finer details for efforts to reduce fire risks. These predicted ignition maps can help us to target areas of higher ignition probability for public education on risk awareness, policy-driven regulations, and enforcement to reduce ignition risks. Considering that the WUI expansion is expected to continue, human footprints will keep increasing (Radeloff et al 2018) and consequently intensify anthropogenic pressure along the traffic corridors. Prioritizing fuel treatments, such as reducing fine fuels and designing fuel breaks, over those high-risk areas and over ecoregions where fuels are one of the key drivers, may be more effective in mitigating the ignition risk and thus fire hazard.

The key climate drivers that we identified also helps the communities to better prepare for future fire risk. Climate warming over recent decades has advanced earlier spring snowmelt (Westerling 2016) and reduced mountain snowpacks; this trend is expected to continue and intensified over the coming decades (Gergel et al 2017, Huning and AghaKouchak 2018). The stronger dependence on Tmax in Coastal areas, for example, indicates these areas are more vulnerable to increased ignition risk under projected warming. Enhanced human ignition probability over areas with higher mean annual precipitation statewide and with higher maximum NDVI, especially in Central Coast, also suggests that fuel management such as thinning and removal, and prescribed fires can be prioritized in reducing ignition probabilities. The overall slight decadal reduction in human and lightning-caused ignition probabilities based on the MaxEnt model prediction, although counter intuitive, was consistent with the decreasing number for fires as observed in the past two decades in California (Keeley and Syphard 2018). It contradicted with the well documented increase in total area burned, which is mostly caused by larger fire size due to climate change and extreme weather events such as heatwave and wind gust (Westerling et al 2006, Westerling and Bryant 2008). This potential decoupling between ignition risk and fire size further suggested the importance of incorporating ignition probability as part of management strategies. The dominant contributions of biophysical variables to decadal ignition changes and their spatial heterogeneity also supported the complexity among climate-fuels-ignitions. Human activities increased the decadal ignition probability in some clustered areas, reinforcing the impacts of human settlements on the sub-region scale ignition patterns and thus highlighting the potential role of land use planning on localized ignition risk management.

4.3. Uncertainties

This study improved our understanding of the spatial patterns and drivers of wildfire ignition in California, but we recognized there are still a few potential limitations. First, MaxEnt models predict wildfire ignition likelihood by searching for the probability distribution with the maximum entropy, subject to the inclusive explanatory variables for the set of ignition-presence samples. Although the MaxEnt model can capture the nonlinear and complex interaction between independent variables and response variable, it cannot be regarded as a causality-based diagnosis. Since MaxEnt cannot separate the indirect cause of correlation from the direct causes, and we therefore relied on variable importance and partial dependence curves to help interpret the impact from biophysical and anthropogenic variables on human- and lightning-caused ignition likelihood.

Future ignition study should incorporate a more complete set of explanatory variables to account for all impacts of climate, fuels, and human activities (Chen et al 2021). This study targeted the ignition probability at a finer spatial resolution of 1 km, and therefore only variables from 1 km Daymet dataset were included. Impacts of fuel moisture such as those from the 4 km GridMET dataset (Abatzoglou 2013), fuel types, and fuel compositions on ignitions should be studied further. Finally, there is still some limitation in the ignition dataset (Short 2017) that we used here, although it included many smaller fires. Some fire incidences are still likely missing, i.e. those suppressed by municipal fire departments, which are often small fires close to developed areas (Carlson et al 2021).

5. Conclusions

Based on the most complete ignition database available, we developed maximum entropy models to predict the spatial distribution of long-term human- and lightning-caused ignition probability at 1 km and investigated how a set of biophysical and anthropogenic variables controlled their spatial variation in California and across its sub-ecoregions. Results showed that the integrated models with both biophysical and anthropogenic drivers predicted well the spatial patterns of both human- and lightning-caused ignitions in statewide and sub-ecoregions of California. Model diagnostics of the relative contribution and marginalized response curves showed that precipitation, slope, human settlement, and road network were the most important variables for shaping human-caused ignition probability, while snow water equivalent, lightning density, and fuel amount were the most important variables controlling the spatial patterns of lightning-caused ignition probability. The relative importance of biophysical and anthropogenic predictors differed across various sub-ecoregions of California.
This study demonstrated the capability of machine learning models in capturing spatial distribution of ignitions and diagnosing the drivers that shaped the observed patterns in human started vs natural ignitions separately. The generated fine-resolution maps of human- and lightning-caused ignition probability provide valuable inputs for ignition risk assessment. Our sub-ecoregion analysis highlighted the spatially varying determinants on ignition probability in the whole state and across sub-ecoregions of California, emphasizing the importance of region-specific management strategies to reduce future increases in wildfire risks. Our findings are expected to provide guidance on prioritizing where and which fire management strategies may be pursued in the context of ongoing and future human influences and climate change.

Data availability statement

All data used in this study are publicly available. The FOD-FPA ignition dataset is available from USDA Forest Service (www.fs.usda.gov/rds/archive/catalog/RDS-2013-0009.4). Daymet gridded estimates of daily weather and climatology variables are from Oak Radge National Laboratory (https://daymet.ornl.gov/). Monthly lightning-strike data are from the NOAA Vaisala National Lightning Detection Network (www.ncdc.noaa.gov/data-access/severe-weather/lightning-products-and-services). OpenStreetMap (www.openstreetmap.org/) provides road network data layers. Population, nighttime light imagery, Landsat and topographic data are from the Google Earth Engine data archive (https://earthengine.google.com/). The MaxEnt model software is available (https://biodiversityinformatics.amnh.org/open_source/maxent/).

The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

This work was supported by the University of California Office of the President under the UC National Laboratory Fees Research Program (Grant #LFR-18-542511) ‘The Future of California Drought, Fire, and Forest Dieback’, the California Strategic Growth Council under the Innovation Center for Advancing Ecosystem Climate Solutions, and The University of Hong Kong HKU-100 Scholars Fund. We would thank two anonymous reviewers and editors for providing valuable suggestions and comments, which are greatly helpful in improving this manuscript.

References

Abatzoglou J T 2013 Development of gridded surface meteorological data for ecological applications and modelling Int. J. Climatol. 33 121–31
Abatzoglou J T, Kolden C A, Balch J K and Bradley B A 2016 Controls on interannual variability in lightning-caused fire activity in the western US Environ. Res. Lett. 11 045005
Aasen G P, Brodlick P G, Anderson C B, Vaughn N, Knapp D E and Martin R E 2016 Progressive forest canopy water loss during the 2012–2015 California drought Proc. Natl Acad. Sci. 113 E249–55
Balch J K, Bradley B A, Abatzoglou J T, Nagy R C, Fusco E J and Mahood A L 2017 Human-started wildfires expand the fire niche across the United States Proc. Natl Acad. Sci. 114 2946
Barbour M, Keeler-Wolf T and Schoenherr A A 2019 Terrestrial Vegetation of California (Berkeley, CA: University of California Press)
Byer S and Jin Y 2017 Detecting drought-induced tree mortality in Sierra Nevada forests with time series of satellite data Remote Sens. 9 929
CalFIRE 2021 Wildfire incidents overview
Calkin D E, Cohen J D, Finney M A and Thompson M P 2014 How risk management can prevent future wildfire disasters in the wildland-urban interface Proc. Natl Acad. Sci. 111 746–51
Cardil A, Rodrigues M, Ramirez I, De-migué S, Silva C A, Mariani M and Ascoli D 2021 Coupled effects of climate teleconnections on drought, Santa Ana winds and wildfires in southern California Sci. Total Environ. 765 142788
Carlson A R, Sebasky M E, Peters M P and Radloff V C 2021 The importance of small fires for wildfire hazard in urbanised landscapes of the northeastern US Int. J. Wildland Fire 30 307–21
Center for International Earth Science Information Network (CIESIN)—Columbia University 2018 Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 11 (Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC)) (https://doi.org/10.7927/H449G7Y9)
Chen B, Jin Y, Scaduto E, Moritz M A, Goulden M L and Randerson J T 2021 Climate, fuel, and land use shaped the spatial pattern of wildfire in California’s Sierra Nevada J. Geophys. Res. Biogeosci. 126 e2020JG005786
Countryman C M 1972 The Fire Environment Concept (Berkeley, CA: Pacific Southwest Forest and Range Experiment Station)
Danielson J and Gesch D 2011 Global Multi-resolution Terrain Elevation Data 2010 (Sioux Falls, SD: US Department of the Interior and US Geological Survey)(https://doi.org/10.3133/ofr20111073)
Elith J, Phillips S J, Hastie T, Dudík M, Chee Y E and Yates C J 2011 A statistical explanation of MaxEnt for ecologists Divers. Distrib. 17 43–57
Elvidge C D, Baugh K E, Dietz J B, Bland T, Sutton P C and Kroehl H W 1999 Radiance calibration of DMSP-OLS low-light imaging data of human settlements Remote Sens. Environ. 68 77–88
ESA 2017 Land Cover CCI Product User Guide Version 2.0
Faivre N, Jin Y, Goulden M L and Randerson J T 2014 Controls on the spatial pattern of wildfire ignitions in Southern California Int. J. Wildland Fire 23 799–811
Fettig C J, Mortenson L A, Bulaon B M and Foulk P B 2019 Tree mortality following drought in the central and southern Sierra Nevada, California, US For. Ecol. Manage. 432 164–78
Flannigan M, Wotton B, Marshall G, De Groot W, Johnston J, Jurko N and Cantin A 2016 Fuel moisture sensitivity to temperature and precipitation: climate change implications Clim. Change 134 59–71
Gergel D R, Nijsen B, Abatzoglou J T, Lettenmaier D P and Stumbaugh M R 2017 Effects of climate change on
snowpack and fire potential in the western USA Clim. Change 141 287–99
Goulden M L and Bales R C 2019 California forest die-off linked to multi-year deep soil drying in 2012–2015 drought Nat. Geosci. 12 632–7
Gutierrez Aurora A, Hantson S, Langenbrunner B, Chen B, Jin Y, Goulden Michael L and Randerson J T 2021 Wildfire response to changing daily temperature extremes in California’s Sierra Nevada Sci. Adv. 7 eabf417
Hsu F-C, Baugh K E, Ghosh T, Zhizhin M and Elvidge C D 2015 DMSP-OLS radiance calibrated nighttime lights time series with intercalibration Remote Sens. 7 1855–76
Huang Y, Jin Y, Schwartz M W and Thorne J H 2020 Intensified burn severity in California’s northern coastal mountains by drier climatic condition Environ. Res. Lett. 15 104035
Huning L S and AghaKouchak A 2018 Mountain wildfire response to different levels of warming Proc. Natl Acad. Sci. 115 10932–7
Jin Y, Goulden M L, Faiivre N, Veraverbeke S, Sun F, Hall A, Hand M S, Hook S and Randerson J T 2015 Identification of two distinct fire regimes in Southern California: implications for economic impact and future change Environ. Res. Lett. 10 094005
Jordao S, Chuvieco E and Arevalillo J M 2012 Modelling fire ignition probability from satellite estimates of live fuel moisture content Fire Ecol. 8 77–97
Keeley J E and Syphard A D 2018 Historical patterns of wildfire ignition sources in California ecosystems Int. J. Wildland Fire 27 781–99
Li S and Banerjee T 2021 Spatial and temporal pattern of wildfires in California from 2000 to 2019 Sci. Rep. 11 1–17
Lutz J A, Van Wagtenendonk J W, Thode A E, Miller J D and Franklin J F 2009 Climate, lightning ignitions, and fire severity in Yosemite National Park, California, USA Int. J. Wildland Fire 18 765–74
Mansuy N, Miller C, Parisien M-A, Parks S A, Batllori E and Moritz M A 2019 Contrasting human influences and macro-environmental factors on fire activity inside and outside protected areas of North America Environ. Res. Lett. 14 064007
Moritz M A, Batllori E, Bradstock R A, Gill A M, Handmer J, Hessburg P F, Leonard J, McCaffrey S, Odion D C and Schoennagel T 2014 Learning to coexist with fire Nature 515 58–66
Moritz M A, Parisien M-A, Batllori E, Krawchuk M A, Van Dorn J, Ganz D J and Hayhoe K 2012 Climate change and disruptions to global fire activity Ecosphere 3 1–22
Parisien M-A, Miller C, Parks S A, DeLancey E R, Robineau F-N and Flannigan M D 2016 The spatially varying influence of humans on fire probability in North America Environ. Res. Lett. 11 075005
Parisien M-A and Moritz M A 2009 Environmental controls on the distribution of wildfire at multiple spatial scales Ecol. Monogr. 79 127–54
Parisien M-A, Snetsinger S, Greenberg J A, Nelson C R, Schoennagel T, Dobrowski S and Moritz M A 2012 Spatial variability in wildfire probability across the western United States Int. J. Wildland Fire 21 313–27
Phillips S J, Anderson R P and Schapire R E 2006 Maximum entropy modeling of species geographic distributions Ecol. Model. 190 231–59
Phillips S J, Dudík M and Schapire R E 2004 A maximum entropy approach to species distribution modeling Proc. Twenty-first Int. Conf. on Machine Learning (Banff, Alberta: ACM) p 83
Phillips S J, Dudík M and Schapire R E 2021 Maxent software for modelling species niches and distribution (Version 3.3.3) (available at: https://biodiversityinformatics.amnh.org/open_source/maxent/) (Accessed 1 May 2021)
Radeloff V C, Hammer R B, Stewart S I, Fried J S, Holcomb S S and McKeefry J F 2005 The wildland–urban interface in the United States Ecol. Appl. 15 799–805
Radeloff V C, Helmers D P, Kramer H A, Mockrin M H, Alexandro P M, Bar-Massada A, Butsic V, Hawbaker T J, Martiniuzzi S and Syphard A D 2018 Rapid growth of the US wildland-urban interface raises wildfire risk Proc. Natl Acad. Sci. 115 3314–9
Robeson S M 2015 Revisiting the recent California drought as an extreme value Geophys. Res. Lett. 42 6771–9
Short K C 2017 Spatial wildfire occurrence data for the United States, 1992–2015 [EPA_FOD_20170508]
Stephens S L 2005 Forest fire causes and extent on United States Forest Service lands Int. J. Wildland Fire 14 213–22
Stephens S L, Collins B M, Fettig C J, Finney M A, Hoffman C M, Knapp E E, North M P, Safford H and Wayman R B 2018 Drought, tree mortality, and wildfire in forests adapted to frequent fire BioScience 68 77–88
Swain D L, Tsiang M, Haugen M, Singh D, Charland A, Rajaratnam B and Diffenbaugh N S 2014 The extraordinary California drought of 2013/2014: character, context, and the role of climate change Bull. Am. Meteorol. Soc. 95 S3–57
Syphard A D and Keeley J E 2015 Location, timing and extent of wildfire vary by cause of ignition Int. J. Wildland Fire 24 37–47
Syphard A D, Rustigan-Romnos H, Mann M, Conliaek E, Moritz M A and Ackerly D 2019 The relative influence of climate and housing development on current and projected future fire patterns and structure loss across three California landscapes Glob. Environ. Change 56 41–55
Thornton M M, Shrestha R, Wei Y, Thornton P E, Kao S and Wilson B E 2020 Daymet: Daily Surface Weather Data on a 1 km Grid for North America, Version 4 (Oak Ridge, TN: ORNL DAAC) (https://doi.org/10.3334/ORNLDAAC/1840)
US Environmental Protection Agency 2012 Level III ecoregions of California (U.S. EPA Office of Research and Development (ORD), National Health and Environmental Effects Research Laboratory (NHEERL))
Westering A L 2016 Increasing western US forest wildfire activity: sensitivity to changes in the timing of spring Phil. Trans. R. Soc. B 371
Westering A L, Bryant B P, Reisler H K, Holmes T P, Hidalgo H G, Das T and Shrestha S R 2013 Climate change and growth scenarios for California wildfire Clim. Change 109 415–63
Westering A L and Bryant B P 2008 Climate change and wildfire in California Clim. Change 87 231–49
Westering A L, Hidalgo H G, Cayan D R and Swetnam T W 2006 Warming and earlier spring increase Western U.S. forest wildfire activity Science 313 940–3
Whelan R J 1995 The Ecology of Fire (Cambridge: Cambridge University Press)
Williams A P, Abatzoglou J T, Gershunov A, Guzman-Morales J, Bishop D A, Balch J K and Lettenmaier D P 2019 Observed impacts of anthropogenic climate change on wildfire in California Earth’s Future 7 892–910
Williams A P, Seager R, Macalady A K, Berkelhammer M, Crimmins M A, Swetnam T W, Trugman A T, Bueen N, Noone D and McDowell N G 2014 Correlations between components of the water balance and burned area reveal new insights for predicting forest fire area in the southwestern United States Int. J. Wildland Fire 24 14–26