






The season for large fires in Southern California is projected to lengthen in a changing climate

Chunyu Dong ^{1,2}, A. Park Williams^{2,3}, John T. Abatzoglou ⁴, Kairong Lin ¹, Gregory S. Okin², Thomas W. Gillespie², Di Long⁵, Yen-Heng Lin ⁶, Alex Hall ⁶ & Glen M. MacDonald²✉

Southern California is a biodiversity hotspot and home to over 23 million people. Over recent decades the annual wildfire area in the coastal southern California region has not significantly changed. Yet how fire regime will respond to future anthropogenic climate change remains an important question. Here, we estimate wildfire probability in southern California at station scale and daily resolution using random forest algorithms and downscaled earth system model simulations. We project that large fire days will increase from 36 days/year during 1970–1999 to 58 days/year under moderate greenhouse gas emission scenario (RCP4.5) and 71 days/year by 2070–2099 under a high emission scenario (RCP8.5). The large fire season will be more intense and have an earlier onset and delayed end. Our findings suggest that despite the lack of a contemporary trend in fire regime, projected greenhouse gas emissions will substantially increase the fire danger in southern California by 2099.

¹School of Civil Engineering, Sun Yat-sen University, and Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai, China. ²Department of Geography, University of California, Los Angeles, Los Angeles, CA, USA. ³Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA. ⁴Management of Complex Systems Department, University of California, Merced, Merced, CA, USA. ⁵State Key Laboratory of Hydrosience and Engineering, Department of Hydraulic Engineering, Tsinghua University, Beijing, China. ⁶Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, Los Angeles, CA, USA. ✉email: glen@geog.ucla.edu

California has a Mediterranean climate characterised by mild, wet winters and hot, dry summers, which are conducive to wildfires. Anthropogenic warming during the past century has increased aridity and aggravated drought risk in California^{1,2}, directly contributing to increasing fuel aridity, a longer fire season, and increased wildfire activity over much of the state^{3–7}. In 2017 and 2018, California experienced consecutive exceptional fire seasons, burning a combined area of 13,255 km², and three of the seven largest fires in California's modern record occurred during this time^{5,8}. However, these years are eclipsed by 2020, as 16,907 km² have burned during this single year⁹. The annual fire suppression budget of CalFire (California Department of Forestry and Fire Protection) has also increased from less than \$30 million in the 1980s to approximately \$640 million during 2015–2019⁹.

In most of California, both large fire frequency and total area burned peak in summer, while in some years extremely large fires driven by the Santa Ana winds (SAW)¹⁰ cause the coastal southern California area (CSCA) to experience a peak in area burned in October^{11,12}. Due to the nature of the CSCA's fires, a large population, and patterns of land development at wildland-urban interfaces, the area has suffered some of the highest property losses caused by wildfires in the entire United States¹³. The annual minimum of CSCA fire activity occurs from late winter to spring (January–May) due to higher fuel moisture in response to intermittent precipitation, relatively low temperature, and low vapour pressure deficit (VPD). Recent studies suggest that warming and drying have extended the fire-season length in the western United States, including in California^{3,14–17}. However, in the CSCA, there has been no significant trend in the annual or seasonal total burned area over the past five decades, possibly due to a combination of high interannual variability in climate, reduced ignitions, improved fire suppression, and land cover change⁵.

Some researchers suggest that climate has not been and will not become a major determinant of fire activity over California's lower elevations and latitudes, such as the CSCA^{18–20}. Moreover, other researchers find that irrespective of fuel and fire management, climate change alone has driven an increase in large fires in California^{21,22}. Thus, compared with small fires, which are sensitive to human ignitions and other direct anthropogenic impacts, the recently increased large fires in California seem to be principally linked to weather and climate forcings^{14,23}. Alternatively, the interaction of climate change and continuously present human ignition sources may be responsible for the increase in large fires, i.e., climate change leads to faster drying of fuels and increased large fire risk in areas where human ignitions are prevalent^{15,23,24}. This hypothesis is supported by the fact that a recent increase in large fire frequency occurred when human-ignited fires decreased in the CSCA^{21,23}.

Moreover, climate-model projections of continued warming, increased VPD, and frequency of extreme fire-danger days raise questions as to whether an increasing trend of large fire occurrence in the CSCA will develop and persist in the future^{5,22}. However, addressing this question in the geographically small (~41,000 km²) and topographically/climatologically diverse spatial domain of CSCA (Fig. 1) requires a degree of spatial and temporal resolution typically not employed in similar predictive studies. This study applies station-based climate projection data and a machine learning-based fire modelling approach to address the following questions: What climatic conditions produce large fire days in the CSCA and how will the inter- and intra-annual variability in large fire days respond to future climate change anticipated for the mid- and late 21st centuries? To what degree is the answer to this question dependent on greenhouse gas (GHG) emission scenarios?

Results

Drivers of the large wildfire probability. To address the above questions, we statistically model the relationship between daily climate and the probability of large (> 40 hectares) wildfires at the local scale in the CSCA (see Methods). Then, we estimate the change in large fire occurrence in response to changes in climate from historical (1950–2005) and future (2006–2099) simulations from an ensemble of earth system models (ESMs) of the 5th phase of the Coupled Model Intercomparison Project (CMIP5). Potential predictors of daily large fire probability (LFP) include the meteorological variables vapour pressure deficit (VPD), wind speed (WS), and precipitation, as well as fire-danger indices from the National Fire Danger Rating System (NFDRS), including the energy release component (ERC), burning index (BI), spread component (SC), ignition component (IC) and 100-h (F100) and 1000-h (F1000) dead fuel moisture (see Methods). Meteorological records obtained from 49 weather stations (Fig. 1) in the CSCA are used in the analysis (Supplementary Table 1). Future climate simulations for a high GHG emission scenario (RCP8.5) and a moderate emission scenario (RCP4.5) were used to project future conditions. The daily climate simulations needed to calculate all of the above predictor variables for the historical, RCP8.5, and RCP4.5 scenarios are available for 14 CMIP5 ESMs (Supplementary Table 2). We downscale these models to each of the 49 CSCA weather stations.

Observations of the meteorological variables and fire-danger indices are assessed as potential predictors of daily large fire occurrence using the random forest technique²⁵. Random forest is an ensemble of decision trees, which can be understood as the sum of piecewise linear functions in contrast to global linear regression models²⁵. Random forest is a robust statistical approach in dealing with the nonlinear interactions and feedbacks between variables²⁶. Previous studies^{11,27} suggest that there are two categories of wildfires in the CSCA, i.e., the fires in the usual dry season (principally driven by hot and dry weather during April to September) and the fires in the usual shoulder and wet season (strongly affected by the Santa Ana winds during the typically wetter months of October to March). Thus, we applied random forest models separately for the dry and wet seasons (Methods; Supplementary Table 3). The relative importance of each predictor is given by its contribution to the model accuracy of LFP. The best model is selected based on a five-fold cross-validation of simulated fire presence/absence against observations during the calibration period of 1996–2010. In each cross-validation, we use independent data from three consecutive years as the out-of-bag samples and the rest of the data to train the model. Then, the selected model is applied to execute the future fire probability projections. We compute both the inter- and intra-annual time series of the multimodel ensemble means (MEMs) of meteorological variables, fire probabilities, and the number of large fire days for the historical and future periods. The 30-year mean climatologies of these variables and the variance for the late 20th century (1970–1999), mid-21st century (2040–2069), and late 21st century (2070–2099) are compared to demonstrate the seasonal changes in climate and fire regime.

The random forest model performs well at simulating the probability of large fire occurrence, with an overall accuracy of 82–84% based on cross-validation against observed data (Methods, Supplementary Fig. 1, Supplementary Table 4). The random forest models display a stable performance when the model parameters are changed (Supplementary Fig. 1, Methods). An analysis of variable importance indicates that the top four predictors of LFP for the dry (wet) season are VPD, IC, F1000, and ERC (F1000, VPD, WS, and IC) (Methods, Supplementary Fig. 2). VPD and F1000 are the most important variables driving large fires in dry and wet seasons, respectively, consistent with

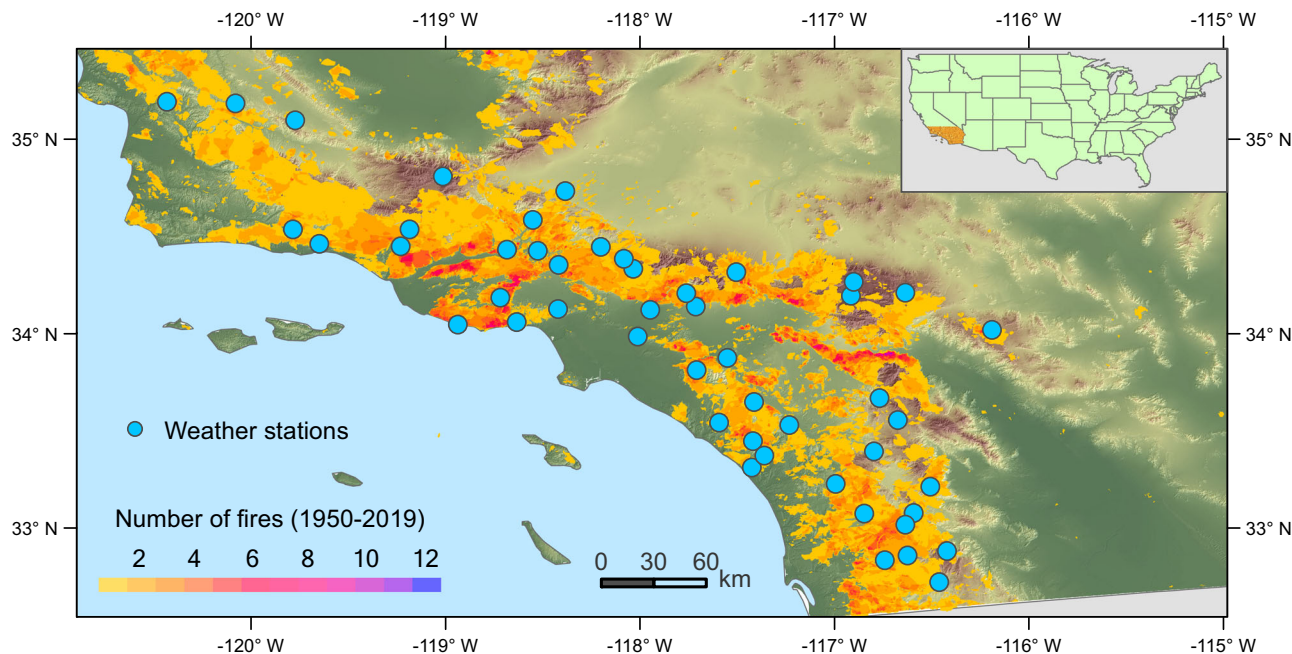


Fig. 1 Spatial domain of the study and the recorded total number of fires since 1950. Blue dots refer to the locations of the weather stations used for this study. Fire perimeter data are derived from the California Department of Forestry and Fire Protection (FRAP).

prior studies⁵. The varying ranks of the predictors' importance for the dry/wet season models likely suggest that the driving mechanisms of large fires can change with seasons, and this has been revealed by other researchers^{11,28}.

To further investigate the specific relations between LFP and the predictors, we conduct accumulated local effect (ALE) analysis for the top four key drivers of the random forest models (Fig. 2). ALE plots are powerful in describing how features influence the prediction of a machine learning model, and they are unbiased even when features are correlated²⁹. The ALE plots show that higher dry-season VPD can approximately linearly increase LFP. At the same time, there is a nonlinear relation between the dry-season F1000 and LFP. A higher F1000 only decreases dry-season fire risk above the mean by 0.0–0.8 standard deviation (s.d.). In the wet season, abnormally dry fuels (F1000 < 0.0 s.d.) can exponentially increase LFP. As a comparison, the relation between the wet-season VPD and LFP is also nonlinear, and a higher VPD only increases LFP above the mean by 0.6–0.8 s.d. These results agree with the fact that F1000 has higher importance than VPD in the wet season (Fig. 2, Supplementary Fig. 2). Ignition component (IC) ranks the second driver of the dry-season large fires, and the ALE plot suggests elevated ignition sources can always increase large fire occurrence during the warm and dry season. By contrast, the contribution of IC to LFP becomes ambiguous in wet season, as ignitions may not inevitably trigger a fire when the fuel moisture is high (Fig. 2). As a composite fuel moisture index that reflects the contribution of all live and dead fuels to potential fire intensity³⁰, ERC also displays high importance in the dry season. ERC and IC show very similar relationships with LFP, while ERC has a relatively lower effect than IC.

The altered relative importance of the predictors with seasons may reflect the local-scale fire behaviour processes. For example, the high sensitivity of LFP to negative standardised F1000 may imply the influence of Santa Ana winds, which can quickly dry the fuels and trigger large fires. This is supported by the elevated importance of wind speed (WS) in the wet season (Fig. 2). Abnormally strong winds (i.e. Santa Ana winds) can strikingly increase LFP. However, since the wet-season fuel moisture is

normally high due to frequent rainfalls, LFP is not sensitive to small declines in positive F1000 anomalies (0.0–2.0 s.d. above the mean). Similarly, it is only when the wet-season VPD increases to a very high level that it becomes a dangerous driver of large fires. In the dry season, as VPD is normally very high for most of the time, this variable can always increase LFP (Fig. 2).

Seasonal changes of the future large wildfires. The simulated seasonal variations in LFP indeed display high correspondence to the observed daily fire frequency (Supplementary Fig. 3). To improve the capability of the random forest models in capturing most of the potential large fires, we apply a resampling procedure to the training dataset, while recognising that this process inevitably induces some overestimation of LFP. Then, we employ a linear regression model to reduce the bias of the LFP estimation (Methods, Supplementary Fig. 3). The bias-correction linear regression model explains 67.4% of the variance in LFP. In general, the corrected simulations of LFP fit the observed seasonal changes in fire frequency well. Uncertainties in the LFP simulations and the bias-correction model are discussed in Methods.

Seasonal projections of climate variables suggest strong warming in spring and autumn (Supplementary Fig. 4). Precipitation is expected to increase in winter but decrease in spring and late autumn, and this seasonal shift has been revealed by another study³¹. VPD is projected to increase markedly from spring to autumn, while fuel moisture will likely decrease most in spring and autumn (Supplementary Fig. 5). IC and ERC seem to have large increases in autumn. At the same time, WS is expected to increase in summer but decrease in autumn, which is consistent with a previous study³². In addition, some previous studies suggest a future suppression of Santa Ana winds in the CSCA³³. This may imply that the contribution of the Santa Ana winds to the autumn and winter fire risk will likely be weakened in the future.

Based on the ESM ensemble simulations, we find a general increase in LFP throughout the year for both the RCP4.5 and RCP8.5 scenarios, with annual mean increases by ~39% and ~62%, respectively, by the late 21st century (Fig. 3) because the

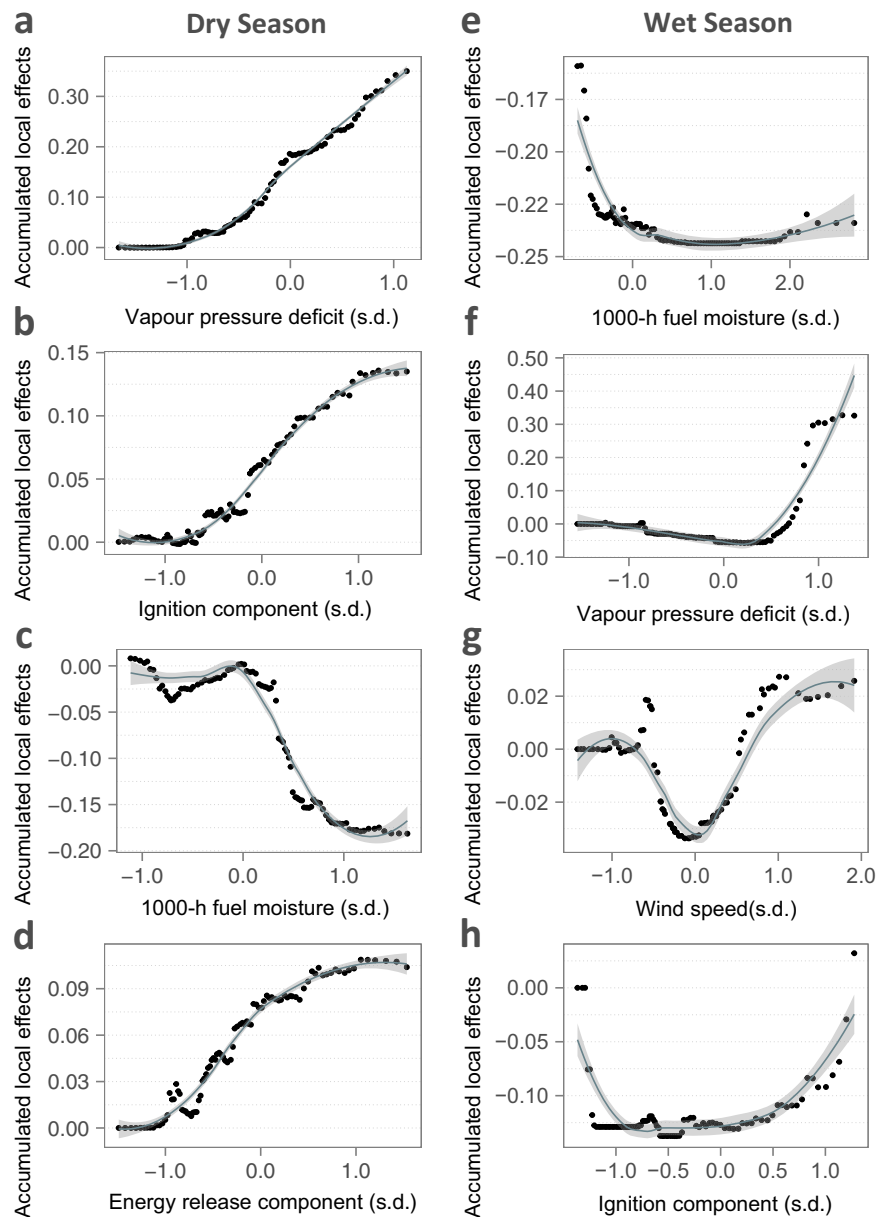


Fig. 2 Sensitivity of large fire probability (LFP) to meteorological variables and fire indices. Accumulated local effect (ALE) plots show the relationship between large fire risk and the top four key drivers in dry (a–d) and wet (e–h) seasons. The x-axes represent the independent covariates (in units of standard deviation), and the y-axes represent the size of the mean effect each covariate has on large fire probability (LFP). Variables are ranked in order of the relative importance in random forest models from high (a, e) to low (d, h). Grey lines refer to locally weighted smooth series with the 95% confidence intervals indicated.

RCP8.5 scenario leads to greater changes in the key drivers favouring large fires, e.g., higher VPD, IC, ERC, and lower F1000 (Supplementary Fig. 5).

The LFP normally peaks in summer (August) and reaches its annual minimum in spring (March–April). However, the LFP in the transition period of spring–summer (April–June) is projected to increase by 110% by the late 21st century under RCP8.5 (Fig. 3). Since the random forest model suggests that VPD plays a dominant role in driving these dry-season fires (Supplementary Fig. 2), the simulated increase in fire potential in late spring to early summer is probably mainly driven by intense warming and aridification (Supplementary Figs. 4 and 5). Apparent LFP increases in autumn–winter (November–January) are likely linked to VPD increases and fuel moisture declines (Supplementary Figs. 5).

As the LFPs in July and September are already very high, similar to that in August, a slight increase in LFP may induce more large fire days for these two months (Fig. 3). Thus, both July and September are projected to have obvious increases in large fire days under the high-emission scenario by 2070–2099 compared with the baseline period of 1970–1999. As a result, these models suggest that the large fire season will have an earlier onset and delayed end (Fig. 4).

The particularly strong relative increases in fire potential in spring and autumn are likely a response to a combination of warming, elevated aridity, and reductions in precipitation totals in autumn (Supplementary Figs. 4 and 5), in addition to reductions in daily precipitation frequency in these months³⁴. The expected slight declines in autumn and winter WS may help relieve the fire risks in this season³³ (Supplementary Fig. 5).

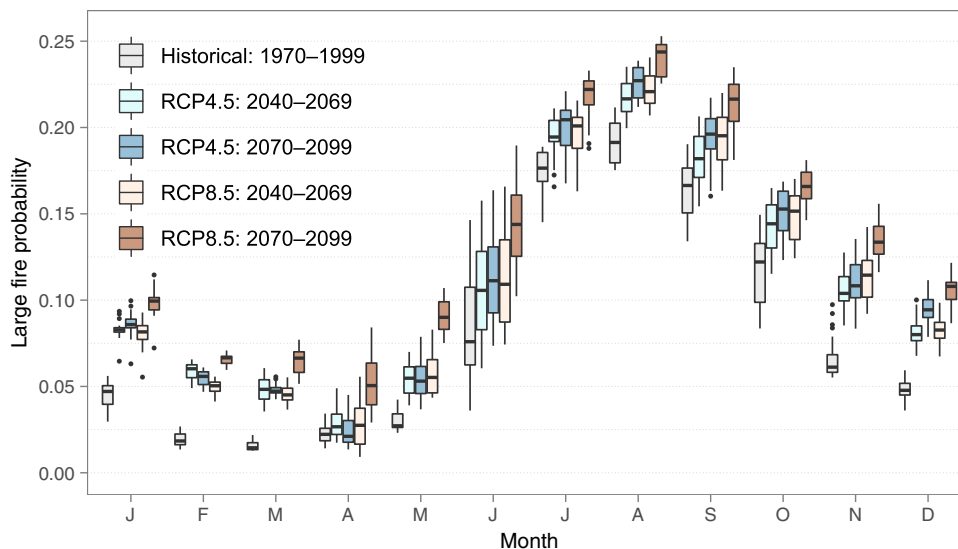


Fig. 3 Seasonal variations in the earth system model (ESM) ensemble-mean large fire probability (LFP). Monthly LFPs for the end of the 20th century (grey, 1970–1999), and the middle (light blue and light tan, 2040–2069) and end (blue and tan, 2070–2099) of the 21st century in the coastal southern California area (CSCA) are shown. LFP simulations for the moderate (RCP4.5, blue) and high (RCP8.5, tan) emission climate change scenarios are compared here. Meaning of boxplot elements: central line: median, box limits: upper and lower quartiles, upper whisker: $\min(\max(x), Q3 + 1.5 \times IQR)$, lower whisker: $\max(\min(x), Q1 - 1.5 \times IQR)$, black dots: outliers.

Inter-annual changes of the future large wildfires. Based on the CMIP5 data, we further calculate the historical and future interannual changes in the top five key drivers of large wildfires indicated by the random forest models (Fig. 4a–e). Climate projections suggest that a higher emission scenario will cause obviously elevated 21st-century warming and slightly increased precipitation in the CSCA (Supplementary Fig. 6). In 2040–2069, the two GHG emission scenarios exhibit similar degrees of warming, approximately 1.0–1.5 °C above the 1970–1999 baseline, but in 2070–2099, the RCP4.5 and RCP8.5 scenarios produce differentiable warming estimates of ~2.5 °C and ~5.5 °C above baseline, respectively. Increases in VPD, IC, and ERC and reductions in fuel moisture (F1000) are projected to follow similar trajectories, with much more substantial changes projected for the RCP8.5 scenario (Fig. 4). WS displays only small annual ensemble-mean trends for either the historical or future periods (Fig. 4). In addition, an expected increase in precipitation variability (Supplementary Fig. 6) in California may bring more extreme arid and wet years in the future³⁵ as well as prolonged periods of dry days interrupted by more extreme but less frequent storm events³⁴.

Our ESM-based simulations of LFP reveal that recent climate change has significantly ($p < 0.001$ in a t -test) increased the frequency of large fire days from ~34 days/yr in 1950–1979 to ~43 days/yr in 2000–2019 (Fig. 4f). Both scenarios are expected to increase the annual frequency of large fire days to ~55 days by 2050. By the end of the 21st century (2070–2099), climate change under a high GHG emissions scenario will likely increase the annual large fire days from ~36 days in 1970–1999 to ~71 days, while moderate GHG emissions scenario will increase it to ~58 days. This departure of the RCP8.5 climate scenario from the RCP4.5 scenario seems to begin in the mid-21st century.

Discussion

Our results indicate that the CSCA will experience striking increases in climatologically identifiable large fire days in the mid-21st century and that this trend will accelerate in the latter half of the century. Under the RCP8.5 emissions scenario, such days will nearly double in frequency by 2100, and under the more

moderate RCP4.5 scenario, they will increase by ~60% compared with the late 20th century.

In the literature, previous researchers have provided contradictory conclusions regarding future changes in wildfire risks in southern California. For example, some researchers^{5,36,37} predict a future increase in fire probability, burned area or fire-danger days in southern California, while others^{19,24,38} suggest a decrease in fire risk in this area. The opposing projections of the previous studies might be because the spatial and/or temporal resolution of such studies is generally coarse and cannot provide detailed information on fire risk changes for small regions, such as the CSCA. Here, we have developed a rather different approach from previous researchers. We applied station-based downscaling of ESM data and random forest-based local-scale fire modelling. A cluster-based resampling and buffering analysis help fully utilise the limited large fire records and capture the real relationships between meteorological stations and fire perimeters. Based on these improvements in methodology, we could simulate the local-scale changes in large fire days under different climate change scenarios.

The annual increase in large fire days reflects both an intensification of conditions during the traditional summer fire season and a lengthening of the large fire season in spring and fall. This finding is consistent with a recent large-scale study³⁹, which also estimates the Mediterranean regime mountains in California will likely have striking increases in very-large fires from spring to autumn. The elevated fire risk in the future is most likely linked to the remarkably increased VPD and decreased F1000 fuel moisture, as the two variables happen to be the top drivers of large fires for dry and wet seasons, respectively. The effects of Santa Ana winds on wildfires will probably be weakened due to the projected declines in WS in the wet season.

The long-term trends of the southern California fire weather are a likely regional feature of the large-scale circulation changes under global warming. Some researchers^{40,41} find that the strengthening and expanding Hadley Circulation due to climate warming reduces tropospheric relative humidity and increases the frequency of dry events in the subtropics. Then, the enhanced warming and drying in the southwest US exacerbates the occurrences of large wildfires⁴². Previously, it was difficult to link

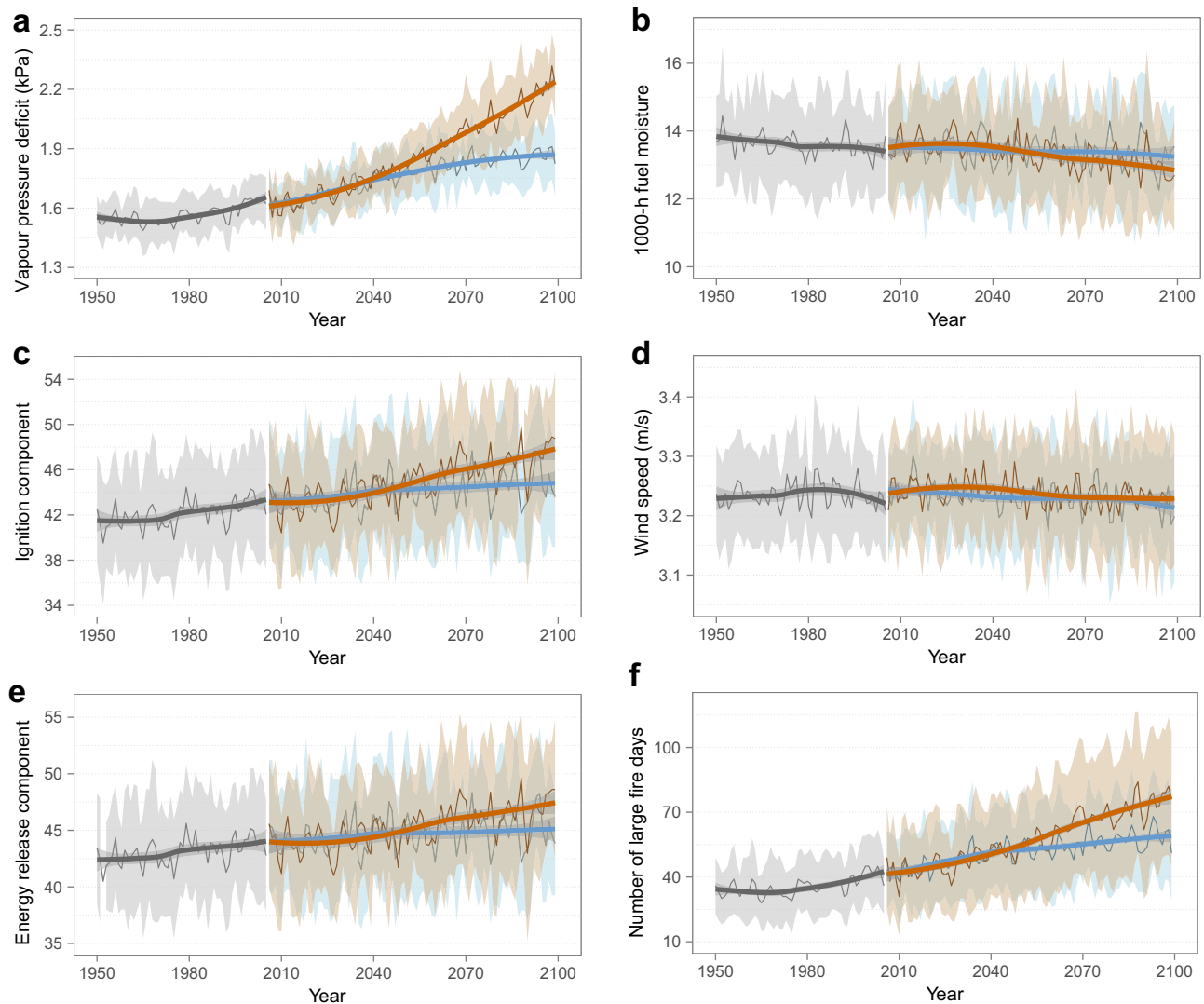


Fig. 4 Earth system model (ESM) ensemble means of the top-five key predictors and the simulated annual number of large fire days. a vapour pressure deficit (VPD), **b** 1000-h dead fuel moisture (F1000), **c** ignition component (IC), **d** wind speed (WS), **e** energy release component (ERC), **f** number of large fire days. Both the historical (grey, 1950–2005) and future (blue/tan, 2006–2099, left: moderate emission scenario, RCP4.5, right: high emission scenario, RCP8.5) variations of these variables are shown. Shaded areas represent ± 1 standard deviation. A low-pass filter was applied to remove the highest 20% frequencies to reduce noise in the time series. Bold lines (grey: historical, blue: RCP4.5, orange: RCP8.5) refer to locally weighted smooth series with the 95% confidence intervals indicated.

the general circulation model outputs and the local-scale fire risk⁴³. Here, a downscaling of the CMIP5 model outputs to station levels and a machine learning approach allow us to predict how climate change will affect the local-scale future changes in daily LFP and show what process plays a dominant role in driving the dry/wet fire risk.

Many studies indicate that fire management and human activities play an important role in altering the fire regimes^{18,44}. However, the inclusion of human factors in future fire prediction remains a major challenge, as there are large uncertainties in estimating future fire management policies and human activities. This challenge is beyond the scope of this study. As this study excluded small fires that are mainly related to human ignition, we assume that the remaining large fire records are closely linked to extreme fire weather conditions, and in speculating on future fires we also assume that fuel management will not experience radical alteration. In some circumstances, the above two assumptions may not be satisfied, which becomes a shortcoming of this study. Indeed, some studies have revealed that southern California has

displayed a shortened fire-return interval (more fires), while northern California shows opposing trends⁴⁵. The distinct fire frequency changes in the same state have implications in understanding the role of climate and fuels as drivers of wildfire risk in California.

These modelling approaches and findings should be useful in scenario development regarding the future climate change impacts on CSCA wildfires. The findings and approach may be useful for other Mediterranean climate regions and generally where fine spatial scale predictive modelling of fires is required. The CSCA region has already experienced an increase in climatic conditions that are conducive to large fires (Fig. 1), but no clear trend has been observed in annual area burned. The expected continuation of this climatic trend towards longer and more severe fire seasons and its intensification in the mid-21st century will largely enhance conditions favouring increasing magnitudes and frequency of wildfires, which may overwhelm the effect of some of the non-climatic factors acting in the recent past to moderate the annual area burned in Mediterranean-type regions.

The current wildfire management policies in these regions mainly focus on fire suppression with often limited mechanisms to address ongoing climate change and rapidly accumulated fuels due to the more frequent droughts today and in the future⁴⁶. The “novel” or “no analogue” environmental conditions caused by increased large wildfires in these Mediterranean climate ecosystems would present new challenges for natural resource and development planning and management⁴⁷.

Methods

Datasets used in this study. The fire perimeter data for the period of 1950–2019 were provided by the California Department of Forestry and Fire Protection (FRAP, <https://frap.fire.ca.gov>). The observations of Remote Automatic Weather Stations (RAWS) by the US Forest Service for 1996–2010 and the CMIP5 down-scaled weather data for the historical (1950–2005) and future (2006–2099) periods were downloaded from the website: <https://climate.northwestknowledge.net/JFSP/JFSP/pages/data.html>. Fourteen ESMs (Supplementary Table 2) were used to generate the CMIP5 dataset and then statistically down-scaled using the multi-variate adaptive constructed analogues method⁴⁸ for 49 stations in the CSCA (Fig. 1, Supplementary Table 1). Then, these observations and CMIP5 data were used to derive the daily fire indices.

National Fire Danger Rating System fire indices. The National Fire Danger Rating System (NFDRS) provides a series of fire indices that help estimate fire-danger changes for a given location³⁰. The burning index (BI) is a function of the spread component (SC), an index of the rate of fire spread, and the energy release component (ERC), an index of the amount of heat released per unit area in the flaming zone of an initiating fire³⁰. The ignition component (IC) is a rating of the probability that a firebrand will cause a fire requiring suppression action. The NFDRS 100-h (F100) and 1000-h (F1000) dead fuel moisture represent the modelled moisture content of dead fuels with different time lags. They are calculated based on the boundary conditions determined from precipitation duration, maximum and minimum temperature, and relative humidity³⁰. We calculated all the BI, IC, SC, ERC, F100, and F1000 time series using the USFS (United States Forest Service) FireFamilyPlus 5 software⁴⁹.

Fire probability modelling. We applied random forest algorithms to perform fire probability modelling. Ensemble decision-tree based approaches, such as random forest and probability estimation tree, have been shown to achieve high predictive accuracy in either classifications or regressions with large numbers of predictor variables^{25,39}. The previous studies⁵⁰ indicate that random forest has a lower risk of overfitting, as it measures the out-of-bag error for each classification or regression. However, some other researchers did find overfitting when using the random forest algorithm^{51,52}. Thus, we utilised a five-fold cross-validation in training the random forest model to avoid overfitting. In each run of the five-fold cross-validation, we selected all the data of three consecutive years within 1996–2010 as the out-of-bag samples, which helps reveal the true performance of the model in predicting LFP.

In this study, vapour pressure deficit (VPD), wind speed (WS), precipitation (Precip), ERC, BI, IC, SC, F100, and F1000 were used as predictors to estimate the probability of a large fire (>40 hectares) for each station on a given day. VPD is a useful indicator of potential burned areas in the western United States^{53,54}. VPD combines temperature and water vapour content information. Following the equations used by Seager et al.⁵⁴, we first calculated the saturation vapour pressures $e_s(T)$ for the maximum (T_{max}) and minimum (T_{min}) daily temperatures:

$$e_s(T_{max}) = e_{s0} \exp \left[17.67 \times \frac{T_{max}}{T_{max} + 243.5} \right] \quad (1)$$

$$e_s(T_{min}) = e_{s0} \exp \left[17.67 \times \frac{T_{min}}{T_{min} + 243.5} \right] \quad (2)$$

Then, we computed the daily mean e_s as follows:

$$e_s(T_a) = [e_s(T_{max}) + e_s(T_{min})] / 2 \quad (3)$$

Finally, VPD is calculated as follows:

$$VPD = e_s(T_a) (1 - RH/100) \quad (4)$$

Elevation and canopy density (representing the proportion of an area that is covered by the crown of trees) were included as predictors in the initial models. However, these predictors are ultimately excluded because they contribute minimally to the model accuracy (Supplementary Fig. 3). The RAWS observations and the FRAP fire perimeter records for 1996–2010 were used to train the random forest models. We did not use ignition coordinate data to indicate fire occurrence, as ignition coordinates cannot distinguish small/large fires and many large fires may have more than one ignition point. We assume that large fires are mainly caused by extreme fire weather and that they are sensitive to climate change, while many small fires are primarily human-caused. We applied the standardised anomalies of weather and fire index time series except for precipitation in

modelling to avoid bias induced by variability differences among stations and variables. We used percentiles of precipitation, instead of standardised anomalies, in the modelling due to its nonnormal distribution. We transferred the fire perimeter data to a binary variable (0: nonfire; 1: fire) before the modelling, and thus, it is not necessary to standardise it.

Previous studies suggest that wildland fires in southern California can be divided into two categories: autumn-winter fires typically triggered by strong offshore Santa Ana winds and summer fires principally driven by hot and dry weather with weak onshore winds¹¹. Santa Ana winds normally occur between October and March¹⁰. We assume that the above meteorological variables contribute differently to the two kinds of fires, and thus, we train and run the random forest models separately for the dry (non-Santa Ana fires, April–September) and wet (Santa Ana fires, October–March) seasons. We also tried to use both the dry and wet-season models to simulate the LFP for the months connecting the two seasons (i.e., March, April, September, and October) and averaged the results of the two models. However, this procedure decreased the model accuracy, and thus, we used random forest models to simulate LFP separately for the dry/wet seasons.

There were 579 large fires recorded in coastal southern California (CSCA) from 1996–2010 (Fig. 1). Both the meteorological stations and the historical burned areas are distributed unevenly in southern California, which has highly heterogeneous terrain, vegetation and climate. Thus, the climate data derived from one station may only be informative for fire probability estimation for a certain area. In addition, the size of this area may change with seasons and locations. Most previous studies interpolated climate data and fire records to gridded datasets^{24,37}. However, this method may induce many errors in the modelling due to the unbalanced distribution of weather stations and fire perimeters. In addition, since we only have meteorological observations at stations, the statistical downscaling of the CMIP5 data is basically station-based.

Here, we utilised a very unusual method of fire data processing. We tested buffer distances of 5, 10, 25, 50, and 100 km from each station to capture the recorded fire perimeters (Supplementary Table 3). Any fire within a specific buffer zone of a station is regarded as a fire occurrence at this station. There should be an optimal buffer distance that demonstrates the true capability of the stations in reflecting the fire weather conditions for this region. We generated 10 sets of daily fire records for the two seasons (dry and wet) and five buffer distances (5, 10, 25, 50, and 100 km). In addition to the fire data, the non-fire-day samples were used in the model to indicate meteorological conditions that have low fire risks. Model performance for different combinations of buffer distance and model parameters (maxnode, mtry, and ntree)²⁵ was compared for the calibration period of 1996–2010 (Supplementary Fig. 1). As an ensemble algorithm, random forest consists of a large number of individual decision trees. maxnode refers to the maximum number of terminal nodes trees in the forest can have; mtry determines the number of variables randomly sampled as candidates at each split; ntree means the number of trees to grow in a random forest²⁵.

Then, we used the dataset sampled with the best buffer distance (10 km) as the model input and applied the above best model parameters to train the model for the dry and wet seasons. Finally, we utilised the two models to predict the LFP for the historical and future periods.

As a large fire is an inherently rare event, the imbalanced prevalence of fire and nonfire samples can severely degrade the performance of random forest⁵⁵. In predicting these small-probability events, most existing methods tend to underestimate the minority classes to optimise the overall accuracy without considering the relative distribution of each class⁵⁶. Many researchers have suggested using cluster-based algorithms to resample imbalanced data samples and have achieved higher prediction accuracy^{56,57}. Here, we applied k -means clustering to undersample the major classes of the samples (nonfire days), and k = number of minority samples⁵⁸, which reduced the number of nonfire samples but reserved most of the information within the data.

After resampling, the results suggest that the balanced data can largely improve the model accuracy. However, to capture most of the large fires, the model tends to misclassify some nonfire days as large fire days. In other words, the predicted LFP was higher than the historical, real large fire occurrence (Supplementary Fig. 3). To overcome this problem, previous researchers²¹ suggest using a post facto calibration to correct the biased fire probability. The initially simulated LFP in our study showed a linear relation with the observed large fire occurrence (Supplementary Fig. 3). In addition, the LFP simulations displayed the highest correlations with the observed decadal mean LFP during 1950–2019 than the long-term mean during either 1996–2010 or 1950–2019. Thus, we applied a linear regression between the predicted and observed (decadal averages during 1950–2019) mean daily LFP to reduce the overestimation of the simulations (Supplementary Fig. 3b). The regression explains ~67.4% of the variance in the LFP. The bias-correction model greatly improves the simulations of LFP (Supplementary Fig. 3c).

Please note that there is still a slight seasonal departure (1–2 weeks) between the simulated and observed LFP after the bias correction. We only have limited years of fire history, which cannot represent the true fire regime of the study area. This result is reflected by the large variance in the observed LFP (~2 months in seasonal variations, Supplementary Fig. 3c). Thus, the fire observations themselves have large uncertainties, and thus, it is not reasonable to further adjust the simulated LFP to match these limited fire observations. In fact, modelling the daily scale LFP

is a very challenging task. As we increase the temporal resolution of fire modelling from annual or monthly to daily, the available records of large fires for this small area become extremely insufficient for use. Thus, a lack of data hinders the improvement of model performance.

Taking the observed annual LFP as a baseline, we identify any day with a simulated LFP that exceeds the baseline LFP threshold as a potential large fire day (LFD). Then we analysed the inter-annual changes of the number of LFD for both the moderate (RCP4.5) and high (RCP8.5) emission climate change scenarios.

We applied the widely used area under the ROC (receiver operating characteristic) curve (AUC) to evaluate the modelling performance. The AUC is recognised as a robust measure of a diagnostic test's discriminatory power, with AUCs of 1.0 and 0.5 indicating a theoretically perfect test and no discriminative value, respectively⁵⁹. Moreover, we also utilised the metrics of accuracy, false positive rate (FPrate), precision, and recall, which are derived from the confusion matrix of binary classification, to evaluate the modelling performance.

We tested the random forest parameter sets of maxnode ranging from 10–1000, mtry ranging from 2–8, and ntree ranging from 10–2000. Together with the five buffer distances, there are 31,250 cross-validation model runs for the dry and wet seasons. We selected the best parameter combinations based on the AUCs of all models. The variations in the model AUC against buffer distance and the three random forest parameters are shown in Supplementary Fig. 1. The results suggest that most model runs achieved an AUC of >0.7, indicating the good performance of random forest in predicting LFP. The parameters maxnode, mtry and ntree displayed small effects on model performance (Supplementary Fig. 1a–c). The model AUC displayed a high sensitivity to buffer distance changes (Supplementary Fig. 1d). Finally, the best parameter combination for the dry season is buffer distance = 10 km, maxnode = 500, mtry = 2, and ntree = 500; the best combination for the wet season is buffer distance = 10 km, maxnode = 100, mtry = 2, and ntree = 500 (Supplementary Table 4). The overall accuracy for the wet and dry seasons is 82% and 84%, respectively, suggesting a good performance of the models.

To quantify the contribution of each predictor to LFP, we utilised a permutation-based approach to calculate the relative importance of all predictors⁶⁰. The rationale of this metric is to measure the decrease in accuracy on out-of-bag (OOB) data when the model randomly permutes the values for that feature. A small value of decrease-in-accuracy for a feature means it is not important, and vice-versa. According to the relative importance of the predictors, VPD, IC, F1000, and ERC (F1000, VPD, WS, and IC) were the four most important variables in the dry-season (wet-season) model (Supplementary Fig. 2). The varying ranks of the predictors' importance for the dry and wet seasons may imply that the primary mechanisms driving a large fire in the two seasons have some differences.

Then we further used accumulated local effects (ALE) plots to identify the detailed relationships between LFP and the top four drivers for both wet and dry seasons. An analysis of ALE determines the effect that each predictor, isolated from all others, has on LFP. In other words, the ALE plots can isolate the change in LFP caused by a change in a single predictor⁶⁰. The ALE plots of LFP against each variable are consistent with the relative importance ranks of the predictors (Fig. 2, and Supplementary Fig. 2). For example, high VPD anomalies can always linearly increase LFP in the dry season, while the wet-season VPD mainly increases LFP when VPD is at a very high level (~0.6–0.8 s.d. above the mean). Abnormally dry fuels (lower F1000) seem to remarkably increase LFP in the wet season (Fig. 2); thus, F1000 becomes the primary fire driver in these months. WS displays a higher influence in the wet season than in the dry season (Supplementary Fig. 2). Overall, the NFDRS indices demonstrate a high capability to predict large fire risk in CSCA, and the relative contribution of these variables to wildfires shows some changes between dry and wet seasons.

Data availability

California fire perimeter data are publicly available at the GIS data portal of the California Department of Forestry and Fire Protection (FRAP, <https://rap.fire.ca.gov/mapping/gis-data/>). The Remote Automatic Weather Stations (RAWS) data and the CMIP5 downscaled weather data are publicly available through the Joint Fire Science Program (<https://climate.northwestknowledge.net/JFSP/JFSP/pages/data.html>). All the NFDRS indices were calculated using the USFS (United States Forest Service) FireFamilyPlus 5 software, which can be downloaded through the National Wildfire Coordinating Group (NWCG, <https://www.nwcg.gov/committees/fire-danger-subcommittee/nfdrs/fire-family-plus>) funded by the US government.

Code availability

The R programming codes used for the statistical analysis of this study are publicly available through the open-access repository Zenodo (<https://doi.org/10.5281/zenodo.5713530>).

Received: 16 December 2020; Accepted: 6 January 2022;

Published online: 17 February 2022

References

1. Diffenbaugh, N. S., Swain, D. L., Touma, D. & Lubchenco, J. Anthropogenic warming has increased drought risk in California. *Proc. Natl. Acad. Sci. USA* **112**, 3931–3936 (2015).
2. Williams, A. P. et al. Contribution of anthropogenic warming to California drought during 2012–2014. *Geophys. Res. Lett.* **42**, 6819–6828 (2015).
3. Westerling, A. L., Hidalgo, H. G., Cayan, D. R. & Swetnam, T. W. Warming and earlier spring increase Western U.S. forest wildfire activity. *Science* **313**, 940–943 (2006).
4. Westerling, A. L. *Wildfire simulations for California's fourth climate change assessment: Projecting changes in extreme wildfire events with a warming climate*. (California Energy Commission, 2018) CCCA4-CEC-2018-014, 1–29.
5. Williams, A. P. et al. Observed impacts of anthropogenic climate change on wildfire in California. *Earth's Futur.* **7**, 892–910 (2019).
6. Podschwit, H. & Cullen, A. Patterns and trends in simultaneous wildfire activity in the United States from 1984 to 2015. *Int. J. Wildl. Fire* **29**, 1057 (2020).
7. Yoon, J.-H. et al. Extreme fire season in California: a glimpse into the future? *Bull. Am. Meteorol. Soc.* **96**, S5–S9 (2015).
8. Balch, J. K. et al. Switching on the big burn of 2017. *Fire* **1**, 1–9 (2018).
9. California Department of Forestry and Fire Protection. *Emergency fund fire suppression expenditures*. (California Department of Forestry and Fire Protection, 2019).
10. Guzman-Morales, J., Gershunov, A., Theiss, J., Li, H. & Cayan, D. Santa Ana Winds of Southern California: their climatology, extremes, and behavior spanning six and a half decades. *Geophys. Res. Lett.* **43**, 2827–2834 (2016).
11. Jin, Y. et al. Contrasting controls on wildland fires in Southern California during periods with and without Santa Ana winds. *J. Geophys. Res. Biogeosci.* **119**, 432–450 (2014).
12. Yue, X., Mickle, L. J. & Logan, J. A. Projection of wildfire activity in southern California in the mid-twenty-first century. *Clim. Dyn.* **43**, 1973–1991 (2014).
13. Keeley, J. E., Safford, H., Fotheringham, C. J., Franklin, J. & Moritz, M. The 2007 southern California wildfires: lessons in complexity. *J. For.* **107**, 287–296 (2009).
14. Westerling, A. L. R. Increasing western US forest wildfire activity: sensitivity to changes in the timing of spring. *Philos. Trans. R. Soc. B Biol. Sci.* **371**, 20150178 <https://doi.org/10.1098/rstb.2015.0178> (2016).
15. Abatzoglou, J. T. & Williams, A. P. Impact of anthropogenic climate change on wildfire across western US forests. *Proc. Natl. Acad. Sci. USA* **113**, 11770–11775 (2016).
16. Kitzberger, T., Falk, D. A., Westerling, A. L. & Swetnam, T. W. Direct and indirect climate controls predict heterogeneous early-mid 21st century wildfire burned area across western and boreal North America. *PLoS One* **12**, e0188486 (2017).
17. Holden, Z. A. et al. Decreasing fire season precipitation increased recent western US forest wildfire activity. *Proc. Natl. Acad. Sci.* <https://doi.org/10.1073/pnas.1802316115> (2018).
18. Keeley, J. E. & Syphard, A. D. Climate change and future fire regimes: examples from California. *Geosci.* **6**, 1–14 (2016).
19. Syphard, A. D. et al. The relative influence of climate and housing development on current and projected future fire patterns and structure loss across three California landscapes. *Glob. Environ. Chang.* **56**, 41–55 (2019).
20. Keeley, J. E. & Syphard, A. D. Different historical fire-climate patterns in California. *Int. J. Wildl. Fire* **26**, 253–268 (2017).
21. Barbero, R., Abatzoglou, J. T., Steel, E. A. & Larkin, N. K. Modeling very large-fire occurrences over the continental United States from weather and climate forcing. *Environ. Res. Lett.* **9**, (2014).
22. Goss, M. et al. Climate change is increasing the likelihood of extreme autumn wildfire conditions across California. *Environ. Res. Lett.* **15**, 094016 (2020).
23. Balch, J. K. et al. Human-started wildfires expand the fire niche across the United States. *Proc. Natl. Acad. Sci. USA* **114**, 2946–2951 (2017).
24. Mann, M. L. et al. Incorporating anthropogenic influences into fire probability models: effects of human activity and climate change on fire activity in California. *PLoS One* **11**, e0153589 (2016).
25. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
26. Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B. M. & Gräler, B. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ* **2018**, e5518 (2018).
27. Kolden, C. & Abatzoglou, J. Spatial distribution of wildfires ignited under katabatic versus non-katabatic winds in Mediterranean Southern California USA. *Fire* **1**, 19 (2018).
28. Jin, Y. et al. Identification of two distinct fire regimes in southern California: implications for economic impact and future change. *Environ. Res. Lett.* **10**, 094005 (2015).
29. Apley, D. W. & Zhu, J. Visualizing the effects of predictor variables in black box supervised learning models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* <https://doi.org/10.1111/rssb.12377> (2020).

30. Bradshaw, L. S., Deeming, J. E., Burgan, R. E. & Cohen, J. D. *The 1978 National Fire-Danger Rating System: technical documentation. USDA Forest Service General Technical Report INT-169* (1984) <https://doi.org/10.2737/INT-GTR-169>.
31. Swain, D. L., Langenbrunner, B., Neelin, J. D. & Hall, A. Increasing precipitation volatility in twenty-first-century California. *Nat. Clim. Chang.* **8**, 427–433 (2018).
32. Wang, M., Ullrich, P. & Millstein, D. The future of wind energy in California: future projections with the variable-resolution CESM. *Renew. Energy* **127**, 242–257 (2018).
33. Guzman-Morales, J. & Gershunov, A. Climate change suppresses Santa Ana winds of Southern California and sharpens their seasonality. *Geophys. Res. Lett.* **46**, 2772–2780 (2019).
34. Gershunov, A. *et al.* Precipitation regime change in Western North America: the role of atmospheric rivers. *Sci. Rep.* <https://doi.org/10.1038/s41598-019-46169-w> (2019).
35. Berg, N. & Hall, A. Increased interannual precipitation extremes over California under climate change. *J. Clim.* **28**, 6324–6334 (2015).
36. Gao, P. *et al.* Robust projections of future fire probability for the conterminous United States. *Sci. Total Environ.* **789**, 147872 (2021).
37. Westerling, A. L. *et al.* Climate change and growth scenarios for California wildfire. *Clim. Change* **109**, 445–463 (2011).
38. Batllori, E., Parisien, M. A., Krawchuk, M. A. & Moritz, M. A. Climate change-induced shifts in fire for Mediterranean ecosystems. *Glob. Ecol. Biogeogr.* **22**, 1118–1129 (2013).
39. Podschwilt, H. R., Larkin, N. K., Steel, E. A., Cullen, A. & Alvarado, E. Multi-model forecasts of very-large fire occurrences during the end of the 21st century. *Climate* **6**, 1–21 (2018).
40. Lau, W. K. M. & Kim, K.-M. Robust Hadley Circulation changes and increasing global dryness due to CO₂ warming from CMIP5 model projections. *Proc. Natl. Acad. Sci. USA* **112**, 3630–3635 (2015).
41. Dai, A. Increasing drought under global warming in observations and models. *Nat. Clim. Chang.* **3**, 52–58 (2013).
42. Zhang, L., Lau, W., Tao, W. & Li, Z. Large wildfires in the Western United States exacerbated by tropospheric drying linked to a multi-decadal trend in the expansion of the Hadley circulation. *Geophys. Res. Lett.* **47**, 1–11 (2020).
43. Macias Fauria, M., Michaletz, S. T. & Johnson, E. A. Predicting climate change effects on wildfires requires linking processes across scales. *Wiley Interdiscip. Rev. Clim. Chang.* **2**, 99–112 (2011).
44. Bowman, D. M. J. S. *et al.* Vegetation fires in the Anthropocene. *Nat. Rev. Earth Environ.* **1**, 500–515 (2020).
45. Safford, H. D. & Van de Water, K. M. *Using Fire Return Interval Departure (FRID) analysis to map spatial and temporal changes in fire frequency on National Forest lands in California.* Research Paper, PSW-RP-266 1–59, Pacific Southwest Research Station (2013).
46. Moreira, F. *et al.* Wildfire management in Mediterranean-type regions: paradigm change needed. *Environ. Res. Lett.* **15**, 011001 (2020).
47. Moritz, M. A. *et al.* Learning to coexist with wildfire. *Nature* <https://doi.org/10.1038/nature13946> (2014).
48. Abatzoglou, J. T. & Brown, T. J. A comparison of statistical downscaling methods suited for wildfire applications. *Int. J. Climatol.* <https://doi.org/10.1002/joc.2312> (2012).
49. Bradshaw, L. S. & McCormick, E. *FireFamily Plus User's Guide, version 4.0.* (USDA Forest Service, 2009).
50. Breiman, L. Bagging predictors. *Mach. Learn.* **24**, 123–140 (1996).
51. Segal, M. R. *Machine learning benchmarks and random forest regression.* (UCSF Center for Bioinformatics and Molecular Biostatistics, 2004).
52. Grushka-Cockayne, Y., Jose, V. R. R. & Lichtendahl, K. C. Ensembles of overfit and overconfident forecasts. *Manage. Sci.* **63**, 1110–1130 (2017).
53. Park Williams, A. *et al.* Temperature as a potent driver of regional forest drought stress and tree mortality. *Nat. Clim. Chang.* **3**, 292–297 (2013).
54. Seager, R. *et al.* Climatology, variability, and trends in the U.S. vapor pressure deficit, an important fire-related meteorological quantity*. *J. Appl. Meteorol. Climatol.* **54**, 1121–1141 (2015).
55. Abraham, A. & Elrahman, S. M. A. A review of class imbalance problem. *J. Netw. Innov. Comput.* **1**, 332–340 (2013).
56. Rahman, M. M. & Davis, D. N. Cluster based under-sampling for unbalanced cardiovascular data. In *Proceedings of the World Congress on Engineering* 3, 3–5 (International Association of Engineers (IAENG), 2013).
57. Zhang, J. & Chen, L. Clustering-based undersampling with random over sampling examples and support vector machine for imbalanced classification of breast cancer diagnosis. *Comput. Assist. Surg.* <https://doi.org/10.1080/24699322.2019.1649074> (2019).
58. Hartigan, J. A. & Wong, M. A. Algorithm AS 136: a K-means clustering algorithm. *Appl. Stat.* **28**, 100 (1979).
59. Pepe, M. S., Longton, G. & Janes, H. Estimation and comparison of receiver operating characteristic curves. *Stata J. Promot. Commun. Stat. Stata* **9**, 1–16 (2009).
60. Molnar, C. *Interpretable machine learning publisher: a guide for making black box models explainable.* www.lulu.com (2020).

Acknowledgements

This work was financially supported by the National Natural Science Foundation of China (Grant nos. 41801254, 51822908), the National Key Research and Development Program of China (Grant no. 2021YFC3001000), the National Sciences Foundation EAR Collaborative Research Grant (No. 1702580). We also acknowledge the support from the Sustainable LA Grand Challenge, the Department of the Interior Southwest Climate Adaptation Science Center, the UCLA John Muir Memorial Endowed Chair, and the Columbia University's Center for Climate and Life, and the Zegar Family Foundation. We are grateful to the three anonymous reviewers and the editor who provided helpful comments and suggestions to this paper.

Author contributions

C.D. and G.M.M. conceived of the study and designed the analyses. A.P.W., J.T.A. K.L., and G.S.O. contributed analysis ideas. C.D. performed the experiments and analysed the data. A.P.W., J.T.A., Y.H.L., and A.H. contributed datasets. C.D. wrote the first draught of the manuscript. A.P.W., J.T.A., and G.M.M. revised the draught manuscript. K.L., G.S.O., T.W.G., D. L., Y.H.L., and A.H. contributed to the review and editing of the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s43247-022-00344-6>.

Correspondence and requests for materials should be addressed to Glen M. MacDonald.

Peer review information *Communications Earth & Environment* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Clare Davis and Heike Langenberg.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022