https://doi.org/10.5194/acp-2019-885
Preprint. Discussion started: 4 November 2019

Atmospheric Chemistry and Physics Discussions

Predicting wildfire burned area in South Central US using integrated machine learning techniques

Sing-Chun Wang[1], Yuxuan Wang[1]

[1]Department of Earth and Atmospheric Sciences, University of Houston, Houston, Texas 77024, USA

Correspondence to: Yuxuan Wang (ywang246@central.uh.edu)

Abstract. Occurrences of devastating wildfires have been on the rise in the United States for the past decades. While the environmental controls, including weather, climate, and fuels, are known to play important roles in controlling wildfires, the interrelationships between fires and the environmental controls are highly complex and may not be well represented by traditional parametric regressions. Here we develop a model integrating multiple machine learning algorithms to predict gridded monthly wildfire burned area during 2002-2015 over the South Central United States and identify the relative importance of the environmental drivers on the burned area for both the winter-spring and summer fire seasons of that region. The developed model is able to alleviate the issue of unevenly-distributed burned area data and achieve a cross-validation (CV) R^2 value of 0.42 and 0.40 for the two fire seasons. For the total burned area over the study domain, the model can explain 50% and 79% of interannual total burned area for the winter-spring and summer fire season, respectively. The prediction model ranks relative humidity (RH) anomalies and preceding months' drought severity as the top two most important predictors on the gridded burned area for both fire seasons. Sensitivity experiments with the model show that the effect of climate change represented by a group of climate-anomaly variables contributes the most to the burned area for both fire seasons. Antecedent fuel amount and conditions are found to outweigh weather effects for the burned area in the winter-spring fire season, while the current-month fire weather is more important for the summer fire season likely due to the controlling effect of weather on fuel moisture in this season. This developed model allows us to predict gridded burned area and to access specific fire management strategies for different fire mechanisms in the two seasons.

1. Introduction

Wildfire is an important process maintaining the balance of terrestrial ecosystems. Wildfire occurrence is controlled by a complex interaction among fuel, weather, and climate (Bowman et al., 2009; Pausas and Keeley, 2009). In recent decades, many regions of the world have seen increasing frequency and intensity of wildfires, possibly connected to changes in regional climate (Barbero et al., 2015; Westerling et al., 2006; Westerling, 2016). More intense and more frequent wildfire activities not only heighten ecosystem vulnerability but also cause poor air quality (Jaffe et al., 2008; Pellegrini et al., 2017; Wang et al., 2018; Yue et al., 2015). Thus, it is imperative to understand how wildfires would respond to changes in environmental factors in a warming climate.

https://doi.org/10.5194/acp-2019-885
Preprint. Discussion started: 4 November 2019

Atmospheric Chemistry and Physics Discussions

Previous studies revealed the importance of several environmental factors on wildfires. Fuel availability and composition across regions can affect fire developments such as fire likelihood and spread efficiency (Nunes et al., 2005; Parks et al., 2012). Weather influences fuel moisture through changing precipitation and humidity and controls fire spread through winds. Long-term climate change can alter both fuel and weather conditions, for example by changing vegetation distributions and the frequency of fire-favorable atmospheric conditions (Heyerdahl et al., 2008; Keyser and Westerling, 2017; Morgan et al., 2008), therefore changing fire regimes. Past studies also highlighted that the complex interplay between fuel, weather, climate, and wildfires can change by spatial scale, fire size, region, and season. For instance, the relationships between fire activity and the environmental controls can exhibit complex nonlinearities across the spatial scale gradient (Peters et al., 2004). Fuel and topography mainly regulate fires at a local scale, while weather and climate control fires at a broad spatial scale (Parks et al., 2012). In terms of fire size, it was found that the major controlling factors could shift from fuel and topography to weather as fire size increases in boreal forests (Liu et al., 2013; Fang et al., 2015). For the western Mediterranean Basin where is characterized by large land heterogeneity, fuel influences can outweigh climate-weather influences on large fires (Fernandes et al., 2016). Therefore, it is challenging to examine the relative importance of the environmental drivers on wildfires due to the complex interrelationships among them.

One common method to explain the relationships between fire regimes (e.g. fire sizes or fire occurrences) and environmental factors is regression. This method is also used to evaluate the relative importance of different environmental controls (Littell et al., 2009; Slocum et al., 2010; Parisien et al., 2011; Yue et al., 2013; Liu & Wimberly, 2015; Fernandes et al., 2016). Among a wide range of regression techniques used, non-parametric machine learning algorithms have emerged as an important tool to predict wildfires because they rely on fewer pre-assumptions about the data. Bedia et al. (2013) used non-parametric multivariate adaptive regression splines (MARS) to model the monthly burned area for the phytoclimatic zones in Spain. Amatulli et al. (2013) estimated the monthly burned area in five countries in Europe using two machine learning approaches, including Random Forest (RF) and MARS. In these studies, the machine learning methods were used to estimate total burned area aggregated over a large-scale domain, e.g. on an ecoregion or a country scale. However, fewer studies have explored the utility of machine-learning methods in resolving the within-domain and grid-level relationships between fires and the environmental drivers. A particular challenge in predicting burned area of fires at the grid level across a broad region relates to the uneven distribution of burned area. The vast majority of wildlands are not burned and thus the relative portion of burned wildlands is extremely low. Among the burned wildlands, the majority of fires are small ones but the total acreage burned is usually contributed by only a few large fires. For example, Steel et al. (2015) showed that for fires in California, small fires (< 25 ha each) contributed to 87% of the total number of grids burned but only 17% of the total burned area. By contrast, large fires (> 150 ha each) accounted for only 3% of the total number of burned grids but made up 64% of the total burned area. Thus, at the grid level the majority class is non-burn wildlands, or wildlands of small fires if considering only burned wildlands alone, while the minority class is large fires. As most data-driven regression algorithms, parametric or non-parametric, would favor the majority class, large fires will be underpredicted for grid-level predictions.

https://doi.org/10.5194/acp-2019-885
Preprint. Discussion started: 4 November 2019


Atmospheric Chemistry and Physics Discussions

In this study, we develop a model integrating multiple machine learning techniques to predict wildfire burned area at the grid level over South Central United States (US) and to identify the relative importance of the environmental drivers on the burned area for that region. The integrated machine learning model aims at mitigating the problem of uneven burned area and improving the accuracy of predicting wildfire burned area at a grid-scale of 50 km x 50 km. While the 50 km-resolution is still coarse to capture individual fires, it is comparable to the resolution of historical and future climate data generated by most climate models (~10 km to 100 km), thus allowing for a practical integration with climate models for potential applications of studying climate-driven changes of wildfire behaviors. The South Central US, encompassing four states -- Texas, Oklahoma, Louisiana, and Arkansas – has experienced periodically large wildfires in recent years, such as the 2011 Texas fires. This region is projected to experience the highest risk of wildfires in 2031-2050 across the continental United States (An et al., 2015; Fann et al., 2018). We chose the vegetation-rich thus fire-prone part of the South Central US, as shown by the red box in Figure 1. The study period is from 2002 to 2015. For each year, we predict gridded wildfire burned area at the monthly scale for the typical bimodal wildfire seasons over the region: the winter-spring fire season from January to April and summer fire season from July to September (Zhang et al., 2014).

The rest of the paper is organized as follows: Section 2 describes the developed model and introduces data incorporated into the model. Section 3 presents the procedures of model validation and evaluation. In section 4, we analyze the relative importance of individual variables and the environmental controls at different spatial scales.

2. Model description and Data

2.1 Model description

One major challenge in wildfire prediction is the highly uneven distribution of burned area. For the study region (red box in Figure 1), grids without any fire occurrence or with only small fires (< 10 ha) in combination take up 70% of the total number of the grids but correspond to only 1% of the total burned area. By contrast, grids with large burned area (>100 ha) account for only 9% of the total number of grids but make up 88% of the total burned area. For uneven data like wildfire burned area, standard machine learning methods usually have a bias in favor of the majority class (i.e. non-burned or small fires), leading to low prediction accuracy of fire burned area (Krawczyk, 2016). To alleviate the bias toward large fires, we developed a model consisting of multiple steps that address the uneven data issue.

Figure 2 demonstrates the structures and processes of our model, which has four steps consisting of three machine learning algorithms. First, for each data grid, given its environmental conditions, we use the quantile regression forest (QRF) to predict a distribution of burned area at the targeted quantiles (in this step the quantiles are chosen at {0.45, 0.55, 0.65, 0.85, 0.95, 0.99}). Second, we predict if a grid is a non-burn grid or not using the logistic regression model and a variety of environmental factors. Third, for the grids that are predicted to burn, instead of predicting burned area directly, we predict the relative position of the burned area within the training set using a random forest (RF) model, which is presented by quantiles. In the last step we divided burned area data into six even-sized sub-groups according to the predicted quantiles. The six sub-

https://doi.org/10.5194/acp-2019-885
Preprint. Discussion started: 4 November 2019


Atmospheric Chemistry and Physics Discussions

groups correspond to the six quantiles predicted in the first step, given the six quantiles are the middle points between the lower and upper bounds for each sub-group, except the top three largest quantiles. For example, in the subgroup (0.39, 0.49), the corresponding quantile is 0.45. For the grids in the subgroups, they are assigned to the value at corresponding quantile as determined by the predicted distribution generated in the first step. The values represent the predicted burned areas.

Our approach alleviates the unevenness data issue for two reasons. First, the majority of zero-burn grids are first separated (i.e. step 2). Second, for the grids predicted to burn, we predict the relative position (i.e. quantiles) of the burned area based on the training set. The distribution of quantiles is less skewed compared to the burned area distribution. Thus, the unevenness of the burned area is less severe when predicting the relative position than predicting the burned area directly. We assume the probability distribution of predicted burned area for most grids is similar to those in the training set but for the grids with large burned area their probability distributions will be more right-shifted, which allows us to transfer the predicted relative position to the predicted burned area of a grid. Specifics of the machine learning algorithms and technical details of the prediction model are described below.

2.1.1 Random forest regression

Random forest (RF) is an ensemble-learning algorithm built on decision trees. Each tree is built using the best split for each node among a subset of predictors randomly selected at the node (Liaw and Wiener, 2002). The best split criterion is based on selecting the variables at the nodes with lowest Gini Index (GI), which is defined as GI $(t_x(x_i))$ = 1- $\sum_{j=1}^{m} f(t_x(x_i),j)^2$, where $f(t_x(x_i),j)$ is the proportion of samples with the value *xi* belonging to leave *j* as node *t*. Two parameters can be adjusted to optimize the RF model, including the number of trees grown (n_{tree}) and the number of predictors sampled for splitting at each node (m_{try}). The RF regression model would first draw n_{tree} bootstrap samples from the original dataset. For each sample, at each node of a tree, m_{try} predictors are randomly chosen from all predictors and then the best split from among the predictors is determined at each node according to GI. In this study, we had n_{tree} of 1200 and m_{try} of 8 for the winter-spring fire season and n_{tree} of 1500 and m_{try} of 7 for summer fire season to obtain the best prediction accuracy. The predicted value of an observation is the average of the observed values of the variable response belongs to the leaf of n_{tree} trees. Here, we used RF model to predict quantiles of burned area for the grids that are predicted to burn.

After all the predicted-burn grids obtain their predicted quantiles of burned area by RF, the whole dataset is divided into six subsets according to their predicted quantiles: {(0.39,0.49), (0.50,0.59), (0.60,0.69), (0.70,0.79), (0.80, 0.89), (>=0.90)}. The grids within each subset will be assigned with the burned area predicted from quantile regression forest (QRF; described in 2.1.2) for a chosen quantile that lies at the middle of the quantile range, which is according to the values of the quantiles at 0.45, 0.55, 0.65, 0.85, 0.95, 0.99, respectively, for the quantile subsets described above. Note that the predicted burn area quantiles belong to the subsets of (0.70, 0.79), (0.80, 0.89), and (>=0.90) are assigned to values of quantiles that are larger than their upper bounds, which is based on the assumption that grids with larger burned area will have more right-shifted burned area distribution than the distributions of the training set.

https://doi.org/10.5194/acp-2019-885
Preprint. Discussion started: 4 November 2019


Atmospheric Chemistry and Physics Discussions

The benefit of applying the RF model is that it can provide the variable importance that measures the strength of individual predictors. The variable importance is measured by the increase in MSE (%IncMSE) and the increase in node purities (IncNodePurity). The %IncMSE is calculated by comparing the mean square error with and without permuting variables for each tree, and the variables with greater values of %IncMSE are more important. As for the IncNodePurity, at each split, we can derive the changes of residual sum of square (RSS) before and after the split, and the final IncNodePurity of a variable is summed over the RSS of all the splits including the variable over all trees. Thus, a larger IncNodePurity represents higher variable importance.

2.1.2 Quantile regression forests

Quantile regression forests (QRF) are an extension of the RF. QRF develops trees in the same way as RF, but for the final leaf of each tree, instead of calculating the mean of predicted values, the QRF estimates the conditional distribution of a target variable. The conditional distribution is calculated by averaging the conditional distributions from all the trees and the predictive quantiles are derived from the final empirical distribution function. Here we chose to predict quantiles at 0.45, 0.55, 0.65, 0.75, 0.85, 0.95, and 0.99. These quantiles were selected because they can represent the full spectrum of fire sizes ranging from small to extremely large ones. The quantiles less than 0.45 are typically zero-burn while the quantile of 0.45 is the lowest quantile that can possibly record both zero-burn and very small burned area for each grid.

2.1.3 Logistic regression model

Logistic regression is used to estimate the probability of fire occurrences in a grid cell by the statistical relationships between fire occurrences and the predictor variables. Logistic regression is defined as $P_i = \frac{1}{1+e^{-\eta i}}$ and $\eta_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_\rho X_{i\rho}$, where Pi represents the probability of an occurrence of wildfire in a grid cell i; η_i is the linear combination of the predictor variables weighted by their regression coefficients (β); x_{ij} is the value of the predictor variable j of the grid i, and β_0 is the constant. The logit function can be expressed as $\log(\frac{P}{1-P}) = x_i^T \beta$, where x_i^T is the vector of the predictor variables and β is the vector of the parameters. Values of P larger than 0.4 are considered to be a fire occurrence and those with equal to or less than 0.4 are interpreted as non-occurrence of wildfires. If a grid is classified not to burn, the predicted burned area would be zero and that grid will not be processed further. On the other hand, if a grid is classified to burn, it would be analyzed by the RF model to predict the burned area quantiles.

https://doi.org/10.5194/acp-2019-885
Preprint. Discussion started: 4 November 2019


Atmospheric Chemistry and Physics Discussions

2.2 Wildfire burned area

We chose wildfire burned area as our target variable, as burned area is a widely-used parameter for quantitative assessment of fire danger and fire impact (Amatulli et al., 2013; Balshi et al., 2009; Yue et al., 2013). Wildfire information over the study period (2002-2015) was obtained from the Fire Program Analysis Fire-Occurrence Database (FPA-FOD). The FPA-FOD collects daily wildfire reports from federal, state, tribal, and local governments. The dataset includes wildfire burned area, fire location in longitude and latitude, and fire discovery date from 1992 to 2015 (Short, 2017). A known caveat of this database is that it does not include some small fires that occur on private lands and thus our model will not be able to predict those fires as such information is not in the training dataset.

The FPA-FOD wildfire data is point data of daily time step. As the prediction model deals with monthly total burned area at a 50 km spatial resolution, we aggregated the daily point burned area into 0.5°× 0.5° grids and summed the burned area in each grid by month. The resulting dataset of monthly burned area has nearly 70% of the grids with burned area less than 10 ha or non-burned. To reduce skewness and improve data symmetry, we applied the log transformation function $ln(x+1)$, where x is the gridded monthly total burned area. The log-transformed burned area was the target variable by our model.

2.3 Predictor variables

Based on previously published studies, we collected a number of predictor variables that are thought to influence wildfire burned area and grouped them into four categories of environmental controls (Table 1): weather, climate, fuel, and fixed-geospatial variables. We chose ten weather variables to represent the fire-weather conditions of month *t*, including monthly mean and maximum temperature, monthly mean zonal (U) and meridional (V) components of wind at 10 m, monthly mean of daily precipitation and monthly total accumulated precipitation, monthly mean and minimum relative humidity (RH), number of consecutive days without rainfall in a month, and Standard Precipitation and Evaporation index (SPEI). Besides current month's weather, weather conditions in the preceding months also have impacts on fire development. For example, increasing precipitation in the preceding months can promote biomass growth and can lead to widespread of larger wildfires (Fréjaville and Curt, 2017; Littell et al., 2009). To consider such lagged effects, for a given month *t*, we calculated the average of chosen weather variables from the months of *t-1* to *t-12*. We then checked if the variables have correlation coefficients (r) larger than 0.5 with fire burned area of month *t* but not are strongly correlated with the same variables of month t ($r < 0.5$). The antecedent variables that pass this criterion for the winter-spring fire season are the monthly mean of daily precipitation of months of *t-1* and the average SPEI for the months of *t-1* to *t-6*. For the summer fire season, the selected variables are the average of monthly mean temperature for months of *t-1* and *t-2*, monthly mean of daily precipitation for months *t-1* to *t-3*, and SPEI for months *t-1* to *t-3*.

Climate variables include climate anomalies of monthly mean temperature, monthly mean of daily precipitation, monthly mean relative humidity, monthly maximum temperature, monthly minimum relative humidity, and monthly total accumulated precipitation. Climate anomalies indicate the effects of changing climate on burned area, by presenting

https://doi.org/10.5194/acp-2019-885
Preprint. Discussion started: 4 November 2019

Atmospheric Chemistry and Physics Discussions

differences between monthly mean of the meteorological variables and their long-term average over 1979-2000. In addition to climate anomaly, we also included the 22-years (1979-2000) means and standard deviations of monthly mean temperature, monthly mean of daily accumulated precipitation, monthly maximum temperature, and monthly total accumulated precipitation. The long-term average of meteorology characterizes the spatial distribution of vegetation, and the standard deviation suggests climate variability that can further affect fire occurrence and size (Keyser and Westerling, 2017). Note that climate normals and standard deviations were considered as fix-geospatial variables in this study because their values do not vary with time.

To estimate the fuel effect on burned area, we included monthly mean of Leaf Area Index (LAI), sum of neighboring LAI, and soil moisture. The LAI is the ratio of the total one-sided area of green leaf area per unit ground surface area, which has been widely used to describe the structural property of a plant canopy. Additionally, LAI is correlated with important metrics of canopy fuel loads, such as canopy bulk density (Keane et al., 2005; Steele-Feldman et al., 2006). Besides local LAI values, to capture the effects of spatial autocorrelations, we consider each grid cell as the center of a 3-by-3 grid matrix and compute the summation of the LAI from the center grid's eight neighboring grids. This sum is referred to as the 'sum of neighboring LAI' and included as a predictor variable. Fuel moisture is a critical property for evaluating fire danger, but there is limited accessible data. It has been known that soil moisture is strongly correlated with fuel moisture, thus soil moisture can be approximated as an indicator of fuel moisture (Krueger et al., 2016). Here, we used the monthly surface soil moisture (0-10 cm) to represent the influence of fuel moisture. Considering the lagged effect of fuel buildup in the preceding months, we chose the averages of LAI and sum of neighboring LAI for the months of *t-5* for the winter-spring fire season, as they passed the same criterion for selecting antecedent weather variables described before. No antecedent fuel variables were included in the model of summer fire season because they did not pass the selection criteria.

Lastly, three fixed-geospatial variables were used as predictors, including land cover types, ecoregion types, and population. Land cover types and ecoregion types were chosen to capture the effects of land use and ecosystem similarity on fire burned area. Population density data in the year 2010 was used to estimate the influence of humans on wildfires (i.e., high suppression efficiency close to high population density area). All variables used in the model, their sources, spatial and temporal resolutions, and categories are listed in Table 1. Note that all variables were regridded to the same spatial resolution of 0.5°× 0.5°.

3. Model validation

3.1 Validation method

The developed model incorporates weather, climate, fuel, and fixed-geospatial variables to predict monthly burned area for each 0.5°× 0.5° grid cell over the South Central US for the two fire seasons during each year of 2002-2015. We applied a 10-fold cross-validation (CV) technique to evaluate the model performance and to avoid overfitting. The entire dataset was

https://doi.org/10.5194/acp-2019-885
Preprint. Discussion started: 4 November 2019


Atmospheric Chemistry and Physics Discussions

randomly divided into 10 equal-sized splits. For each round of CV, the model was trained with nine splits of the data and the trained model was then used to predict burned area at the remaining split. Classification of burned or non-burn grids was evaluated by the accuracy and the area under the curve (AUC) statistics. Burned area predictions were evaluated using statistical indicators such as the coefficient of determination (R^2), mean absolute error (MAE), and root mean squared error (RMSE) between the predicted and observed wildfire burned areas and the evaluation was done for the winter-spring fire season and summer fire season separately. We also quantified the prediction performance by evaluating the model ability in reproducing temporal variation of burned area for each grid and spatial patterns of burned area across all the grids of the study domain. Details of calculating the spatial and temporal correlations are described in the Supporting Information.

3.2 Validation results

The validation results are presented at both the grid-scale and the large-domain scale. The large-domain scale is defined by our study domain (red box in Figure 1) with a horizontal scale of around 700 x 700 km^2. At this large-domain scale, we will compare our model performance with prior studies which predicted total burned area of an ecoregion or a country.

Table 2 shows the model performance at the grid-scale for the winter-spring fire season and summer fire season. Evaluation statistics include the performance metrics of the classification model (accuracy, F1-score, and AUC) and the values of R^2, MAE, and RMSE between the CV-predicted and observed monthly burned area of all the grids combined over the study period. For identifying the burned grids (i.e. the second step in the model), the accuracy and the AUC for the two fire seasons attained 0.75 and 0.71. In terms of burned area prediction, the R^2 reached 0.42 and 0.40 for the winter-spring and summer fire season respectively. MAE and RMSE are 1.13 and 8.37 for winter-spring fire season; as for summer fire season, MAE and RMSE are 0.57 and 4.26. While the overall statistics demonstrate the general capability of our model predicting gridded burned area, we selected three specific years to further illustrate the model performance: 2011 with the largest mean gridded burned area and, 2008 and 2014 with the mean gridded burned area close to the 14-year-mean. Figure 3 shows the selected CV-predicted and observed monthly burned area of those years for each fire season. The R^2 is 0.42, 0.51, and 0.66 for 2011 (combing both seasons), 2014 (winter-spring season), and 2008 (summer fire season), respectively, after excluding misclassified grids. MAE of 2011, 2014, and 2008 are 5.25, 0.77, 0.43 and RMSE are 21.06, 5.87, and 1.75. The detailed statistics of model performance for each year are shown in Table S1.

To our best knowledge, our model outperforms previously published models in predicting gridded burned area and the better performance results from the advantages of the approach. First, we predicted gridded burned area at 0.5° x 0.5° grid cells and on a monthly basis, while most prior studies predicted monthly or annual total burned area on an ecoregion or a country scale. Only a few studies targeted burned area at the grid level but with a spatial resolution of 1° x 1° and a temporal resolution of one year. One study used ocean climate indices to estimate annual gridded burned area and achieved correlation coefficient (r) around 0.55 over SUS (Chen et al., 2016). That study excluded zero-burn grids from the data and thus did not confront the severe unevenness issue. Compared with their results, our model performs better despite predicting at finer

https://doi.org/10.5194/acp-2019-885
Preprint. Discussion started: 4 November 2019


Atmospheric Chemistry and Physics Discussions

temporal and spatial scale, with r of 0.60 and 0.67 for the two fire seasons, respectively. Another study (Liu and Wimberly, 2015) obtained a higher correlation R^2 of around 0.76 between climate variables and burned area over the western US using boosted regression trees, but their investigation included only extremely large fires (> 405 ha) and was at a 1° x 1° resolution and annual timestep. As discussed above, the spatial and temporal resolution of our model is finer than the prior studies and the model is capable of predicting burned area for all the grids within any large domain. Such setting makes it more practical for our prediction model to integrate with climate models. Through the integration of multiple machine learning algorithms, we are able to address the unevenness issue of burned area, which significantly improves the model ability to predict gridded burned area across the whole study domain.

We also evaluated the model performance in reproducing the spatial patterns of the burned area. We first assessed the model ability reproducing the spatial patterns of climatological burned area for the two fire seasons (Figure S1). The model reproduces the 14-year mean burned area, with a correlation coefficient between mean observed and predicted burned area of 0.82 and 0.80 for the winter-spring and summer fire season, respectively. For the whole study domain, the spatial patterns of predicted burned area (including zero burns) have correlation coefficients higher than 0.5 with the observed pattern for more than 60% of the study months. It is noteworthy that such performance is achieved without introducing any coordinate variables like longitude or latitude as predictors. This indicates the model we developed is not hardwired to geographical features of the study domain and thus can be easily adopted for other regions, a unique advantage allowing for practical incorporation into climate models. Temporally, the predicted burned area time series at 70% of the grids (combined the two fire seasons) has a correlation higher than 0.5 with the observed burned area (Figure S2). These results demonstrate the model has the certain ability in predicting both spatial and temporal variation of the burned area at the grid-scale across the study domain.

Besides the grid-scale statistics, we evaluated the model performance at the large-domain scale by adding up all the grid-level predictions to obtain the total burned area of the study domain by month. Figure 4 shows the time series of the predicted total burned area over South Central US in comparison to the observed ones for the two fire seasons. The domain-scale prediction explains 50% and 79% of the month-to-month variability of burned area for winter-spring and summer fire season, respectively. MAE of the monthly burned area across the whole domain is 251.3 km^2 for the winter-spring fire season and 100.7 km^2 for the summer fire season. As listed in Table S2, the prediction accuracy of our model in terms of R^2 is equivalent to or better than most of the published studies on the ecoregion scale or country scale.

4. Weather, climate, fuel, and fixed-geospatial controls on fire burned area

4.1 Individual variable importance at grid scale

Before discussing the environmental controls on fire burned area across our study domain, it is useful to understand the dominant factors controlling the burned area at the grid-scale. One advantage of the random forest approach is that it provides the variable importance metrics that can measure the power of predictor variables in the overall prediction. We show

https://doi.org/10.5194/acp-2019-885
Preprint. Discussion started: 4 November 2019


Atmospheric Chemistry and Physics Discussions

in Figure 5 the top 14 predictors ranked by the variable importance metrics to illustrate the intricate relationships among fires, weather, climate, and fuel. As Figure 5 shows, for both fire seasons, RH anomaly is the predictor variable with the highest rank in predicting burned area at the grid-scale. This finding broadly supports the work of other studies that highlighted the importance of RH on burned area (Riley et al., 2013; Ruthrof et al., 2016). Yet our approach particularly reveals the response of fire burned area to changes in RH anomaly, which is a climate variable as opposed to weather variable. RH anomaly indicates the changes of current RH relative to its historical climatology and it ranks higher than current RH in the variable importance metrics. While RH anomaly is identified as the top factor in both fire seasons, the two fire seasons exhibit a notable difference in that temperature anomaly and maximum temperature anomaly are included in the top 14 variables only for the summer fire season. In addition, RH anomaly and temperature anomaly have a stronger negative correlation in the summer fire season (r= -0.7) than in the spring fire season (r= -0.2). This highlights the importance of combined effects of RH and temperature on burned area during summer.

For the winter-spring fire season specifically, the long-term average of precipitation (apcp_avg and asum avg) is identified as the key climate variable (Figure 5a). Precipitation normals indicate the amount of available moisture that could affect fuel distributions and tendency of fire activities (Westerling and Bryant 2008; Keyer et al 2017). The average of SPEI for month *t-4* is the highest-ranked weather variables, exceeding the current-month SPEI in terms of variable importance, and the 3-5 months' time lag coincidentally corresponds to the interval between the two fire seasons. Our results imply that pre-fire-season drought condition is influential on fire burned area, and it is in agreement with prior studies (Scott and Burgan., 2005; Riley et al., 2013; Turco et al., 2017). Interestingly, the average of LAI and sum of neighboring LAI for months of *t-6* are the only top fuel variables being selected in the winter-spring fire season (Figure 5), indicating the importance of antecedent fuel abundance. From the ranking of variable importance, we can infer that antecedent fuel abundance together with pre-fire-season drought conditions together determines the amount of dry fuel, which likely exerts the primary controls of the burned area during the winter-spring fire season.

For the summer fire season, temperature anomalies (temp_anomaly and tmax_anomaly) are shown as the major climate factors influencing fire burned area (Figure 5b). This indicates temperature variations under a changing climate will have significant effects on the burned area in that season. Top-ranked weather variables include the average of monthly accumulated precipitation and SPEI of preceding month *t-1, t-2, and t-3.* These variables are known to affect fire burned area through changing fuel moisture. Consistently, fuel moisture as represented by soil moisture is identified as the only fuel variable among the top 14 variables in the summer fire season. These results suggest that fuel drying driven by both increasing temperature and pre-fire season drought conditions is the pivotal process determining fire burned area in the summer fire season. Similar to our findings, rising summer temperature under climate change was found to cause fast fuel dryness, which increased fire activity in the western US (Williams et al., 2013; Holden et al., 2018).

Overall, the analysis of variable importance reveals some important differences in the fire mechanisms between the two fire seasons and shows semi-quantitatively that drought conditions in the preceding months (3-5 months for the spring fire season and 1-3 months for the summer fire season) may be more important than current-month conditions on the burned area.

https://doi.org/10.5194/acp-2019-885
Preprint. Discussion started: 4 November 2019


Atmospheric Chemistry and Physics Discussions

Furthermore, we demonstrate that the effect of climate change on fire burned area is consequential, even more influential than current fire weather. This aspect has not been well documented or quantified due to a lack of long-term observations of wildfires over South Central US.

4.2 Method to decompose the relative influence of environmental controls

The variable importance metrics presented in the previous section reveal the relative importance of individual predictors. Recall these predictors were purposely selected from four broadly defined categories of environmental controls on wildfire burned area, namely climate, weather, fuel, and fixed-geospatial. As variables within the same category may work in conjunction to create conditions conducive for wildfires, in this section we examine the composite influence of predictors by category and identify which category would exert the largest contribution controlling the variability of wildfire burned area. To do so, we designed a set of sensitivity experiments using the prediction model developed to decompose the effect of different environmental controls across our study domain by perturbing variables belonging to one category at a time. The environmental control categories to be perturbed include weather, climate, and fuel; the variables of each category are listed in Table 1. The fix-geospatial factors remain unchanged in each sensitivity experiment. In addition, as the relative importance of environmental controls can vary by spatial scale (Parks et al., 2012), the results were analyzed at both the grid-scale and large-domain scale. The large-domain scale here is defined as our study domain. For instance, to examine the influence of weather, for each grid, we assigned the values of individual weather variables to their 15-year means while keeping the variation of other variables (hereafter refer to as the "weather-avg run"). The same procedure was applied to the variables in the climate and fuel category, resulting in the climate-avg run and fuel-avg run respectively. The original model with all the variables of each grid varying by time is called the full-model run. The gridded burned area predicted from each run is summed over all the grids across the study domain. The differences in resulting total burned area between the full-model run and weather-avg run represent the impact of weather control (hereafter called "weather effect"), and the same procedure was applied to derive the climate effect and fuel effect on the burned area. We also conducted the fixed run, in which for each grid, its weather, climate anomaly, and fuel variables are all fixed to their long-term mean, and the predicted burned area from this run represents the influence of geospatial variables and climate normals on the burned area (hereafter named "fix effect"). Although the calculations of deriving the effect of a given environmental category are made by assuming linearity, the machine-learning-based prediction model does not assume linearity. Thus, the summation of burned area prediction from the weather, climate, fuel, and fixed run is not necessarily equal to the burned area predicted by the full model. This difference is considered as the interaction effect among these environmental controls.

After deriving the effects of the environmental controls on the burned area, we then calculated such effects of environmental controls in the scaled absolute percentage. We first normalized the effect of an environmental control category by the number of variables in that category because the numbers of variables are different by environmental control and the

https://doi.org/10.5194/acp-2019-885
Preprint. Discussion started: 4 November 2019

Atmospheric Chemistry and Physics Discussions

category with a larger number of variables may have a larger effect on the burned area. Then, the scaled absolute percentage is defined as the normalized absolute value of the effect of one environmental control divided by the summation of the normalized absolute values of all the effects over all the categories. Thus, the scaled absolute percentage represents the average effect of a single variable in each category. For example, Equation (1) shows how we calculated the scaled absolute percentage of the weather contribution on burned area:

$$\frac{|E_w|/N_w}{|E_w|/N_w + |E_{fu}|/N_{fu} + |E_c|/Nc + |E_{fi}|/N_{fi} + |E_i|/N_t}, \quad (1)$$

, where E is the influence of the environmental controls in burned area, N indicates the number of variables in the category, N_t is the total number of variables, and the subscript *w, fu, c, fi, and, i* represent weather, fuel, climate, fixed, and interaction, respectively.

4.3 Relative importance of environmental controls at large scale

Figure S3 shows the time series of the burned area contributed by different environmental controls for the two fire seasons. According to the results, weather, fuel, climate, and fixed effect all tend to increase burned area while the interaction effect acts to reduce burned area, in particular for the large burn events (e.g. July 2011 in the summer fire season). Considering the number of variables in each environmental control category is different, it is better to compare the effects in the scaled absolute percentage that represents the average contribution of the variables in one environmental category (Section 4.2). Figure S4 presents the time series of environmental effects on the burned area in terms of the scaled absolute percentage. For both fire seasons, on average, the climate and fixed variables have larger contributions to the burned area than other types of variables, but their relative importance varies by time. To further compare the mean effect of the environmental controls between the two fire seasons, we averaged the scaled absolute percentage of each environmental controls over the whole study periods and results are shown in Table S3 and Figure 6. From Figure 6, one can clearly see the climate variable on average has a larger contribution to burned area than other variables for both fire seasons, with the mean scaled absolute contribution of 33% and 35% for the winter-spring and summer fire season, respectively. The result suggests changing climate is a significant factor that can explain wildfire burned area over our study domain. In accordance with our results, previous studies also demonstrated the significant contribution of changing climate to the total burned area of ecoregions or countries (Littell et al., 2009; Swetnam and Anderson, 2008; Yue et al., 2013). For example, increasing temperature and earlier spring snowmelt due to climate change are highly associated with increased large wildfire activity in the western US (Westerling et al., 2006). Another study showed that fire-year climate variables such as average spring temperature are predictive variables that could improve the predicting probability of high severity fires in the western US (Keyser and Westerling, 2017). Additionally, the fixed effect that comprises the geospatial variables and past climatology is ranked as the second most important control (Figure

https://doi.org/10.5194/acp-2019-885
Preprint. Discussion started: 4 November 2019


Atmospheric Chemistry and Physics Discussions

6). This is consistent with the findings of Keyser et al. (2017), which revealed the importance of long-term climate normals in controlling large fire occurrences in the western US.

Comparing effects of the environmental controls between the two fire seasons, the fuel effect is significantly more important in the winter-spring fire season, while weather and climate effects are more substantial in the summer fire season. This can probably be explained by different characteristics of the two fire seasons. Biomass growth is relatively limited in the winter-spring fire season and the wildfire fuel in this season is often provided by vegetation accumulated since the last growing season. Thus, the effect of fuel (mainly from the preceding fuel amount) is comparatively dominant in the winter-spring fire season. On the other hand, vegetation growth is relatively sufficient during the summer growing fire season and fuel abundance would not be a constraint of wildfires. Yet, fire weather that determines fuel moisture is a substantial factor in the summer fire season (Figure 6).

The above analysis focuses on the relative importance of the environmental controls at the large-domain scale. At the grid-scale, we calculated the average variable importance from RF in %IncMSE (section 2.1.1) of each category which represents the relative importance at the grid-scale and the results are presented in Table S4. Climate variables on average have the largest importance in controlling burned area for the two fire seasons, with the mean %IncMSE of 12.09 and 19.18 for the winter-spring and summer fire season, respectively. This is consistent with the results on the large-domain scale. Fuel effect outweighs weather effect on the grid level in winter-spring fire season, while weather effect is more important in summer fire season, both consistent with analyses based on the large-scale domain (Table S4). However, the fixed effect estimated at the grid-scale is not as important as at the large-scale domain (Table S4) and this is partly due to the way these variables are coded in the model. Fixed variables consist of past climatology and geospatial variables (i.e. land use, ecoregion, and population). The geospatial variables except the population were encoded as categorical variables at the grid level. For example, forest ecoregion is coded as 0 or 1 for a given grid, with 0 representing non-forest while 1 forest. For such encoding method, each categorical variable (e.g. forest v.s. non-forest) tends to have a smaller relative importance score, compared to the relative importance score of the entire variable encoded by continuous values (e.g. forest, desert, prairie, and other ecoregion types in a large domain). As RF measures the effect of a specific split on the improvement in model performance and aggregates the improvement of all the splits with a specific variable, the fragmented scores for each category are likely smaller than the scores reflecting all of the categories. Therefore, for the relative importance at the grid level measured by RF, the effect of a single geospatial variable such as a land type on the burned area is trivial. When we average the relative importance of all the fixed variables including many small scores, the resulting average importance becomes still a small value.

5. Conclusion

We presented a model integrating multiple machine learning methods to predict monthly burned area over South Central US at 0.5° x 0.5° grid cells. The developed model is able to alleviate the issue of unevenly-distributed burned area and

https://doi.org/10.5194/acp-2019-885
Preprint. Discussion started: 4 November 2019


Atmospheric Chemistry and Physics Discussions

consequently improves the model capability of predicting large burned area at a finer spatial and temporal scale. The predicted burned area showed a good agreement with the observed burned area at both the grid and large-domain scale. At the grid-scale, the model achieves a CV-R^2 value of 0.42 and 0.40 for the winter-spring and summer fire season respectively. In terms of predicting the spatial patterns of the burned area, the model has the correlation coefficient exceeding 0.5 over 60% of the study months. Across the study domain, the temporal variation of predicted burned area has a correlation coefficient larger than 0.5 with the observed burned area over around 70% of the grids. At the large-domain scale, the prediction model can explain 50% and 79% of interannual burned area for the winter-spring and summer fire season, respectively. The validation results demonstrate that the model has certain skills in predicting monthly burned area at both grid-scale and large-domain scale.

The individual variable importance was analyzed and discussed. For both fire seasons, RH anomaly followed by drought condition in the preceding months (3-5 months for the winter-spring fire season and 1-3 months for the summer fire seasons) are the two top variables in predicting burned area at the grid scale. For the winter-spring fire season specifically, the average of LAI and sum of neighboring LAI for the previous six months are the only fuel variables being identified as top important variables. The finding suggests that the antecedent fuel abundance together with the pre-fire season drought conditions regulate the abundance of dry fuel, which is the primary control of fire burned area during the winter-spring seasons. For the summer fire season, temperature anomalies, the average of monthly accumulated precipitation of the preceding three months, and fire season soil moisture are important variables in predicting burned area. Therefore, the fire burned area in the summer fire season is controlled by rising temperature and pre-fire season drought that speeds up fuel drying. The model highlights the effect of changing climate on burned area as well as the common and different critical factors controlling burned area for the two fire seasons.

Besides the relative importance of individual predictors, we also analyzed the relative importance of four environmental categories - climate, weather, fuel, and fixed-geospatial - at both the grid-scale and large-domain scale. The relative importance of these factors is generally consistent at the two scales. The climate variable on average has the largest contribution to burned area for both fire seasons, with the mean scaled absolute contribution of 33% and 35 % to the burned area at the large-domain scale for the winter-spring and summer fire season, respectively. For the winter-spring fire season, the fuel variable on average has larger importance compared to the weather variable; while for the summer fire season, the weather variable is more dominant than the fuel variable. The difference in the relative importance of the environmental controls between the large-domain scale and grid scale mainly lies in the predominance of the fixed effect. The fixed effect is ranked as the second most important controls at the large-domain scale, but it is not as important at the grid-scale.

The presented study mainly utilizes environmental factors as input variables to build the prediction model and focuses on the effects of environmental controls on burned area. The factors examined herein are not exclusive. For example, we do not examine the effects of human factors on burned area, such as anthropogenic influences that affect wildfires through ignition, suppression, or modifying fuel distribution (Syphard et al., 2007; Bowman et al., 2011; Mann et al., 2016). Future work is needed to better understand the role of human activity engaged with climate change and its implications for wildfire control.

https://doi.org/10.5194/acp-2019-885
Preprint. Discussion started: 4 November 2019

Atmospheric Chemistry and Physics Discussions

Code availability. Model code is available upon request to the first author

Data availability. All dataset used in this study are publicly accessible online at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi%3A10.7910%2FDVN%2FLRPDAA

Author contributions. SW and YW conceived the research idea. SW wrote the initial draft of the paper, performed the analyses, and model development. All authors contributed to the interpretation of the results and the preparation of the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work was funded in part with funds from an AI for Earth grant from Microsoft and from the State of Texas as part of the program of the Texas Air Research Center. The contents do not necessarily reflect the views and policies of the sponsor nor does the mention of trade names or commercial products constitute endorsement or recommendation for use.

https://doi.org/10.5194/acp-2019-885
Preprint. Discussion started: 4 November 2019

Atmospheric Chemistry and Physics Discussions

References

Amatulli, G., Camia, A. and San-Miguel-Ayanz, J.: Estimating future burned areas under changing climate in the EU-Mediterranean countries, Science of The Total Environment, 450–451, 209–222, doi:10.1016/j.scitotenv.2013.02.014, 2013.

An, H., Gan, J. and Cho, S. J.: Assessing Climate Change Impacts on Wildfire Risk in the United States, Forests, 6(9), 3197–3211, doi:10.3390/f6093197, 2015.

Balshi, M. S., McGUIRE, A. D., Duffy, P., Flannigan, M., Walsh, J. and Melillo, J.: Assessing the response of area burned to changing climate in western boreal North America using a Multivariate Adaptive Regression Splines (MARS) approach, Global Change Biology, 15(3), 578–600, doi:10.1111/j.1365-2486.2008.01679.x, 2009.

Barbero, R., Abatzoglou, J. T., Larkin, N. K., Kolden, C. A. and Stocks, B.: Climate change presents increased potential for very large fires in the contiguous United States, Int. J. Wildland Fire, 24(7), 892–899, doi:10.1071/WF15083, 2015.

Bedia, J., Herrera, S., and Gutierrez, J. M.: Assessing the predictability of fire occurrence and area burned across phytoclimatic regions in Spain, Nat. Hazards Earth Syst. Sci., 14, 53–66, doi:10.5194/nhess-14-53-2014, 2014.

Bowman, D. M. J. S., Balch, J. K., Artaxo, P., Bond, W. J., Carlson, J. M., Cochrane, M. A., D'Antonio, C. M., DeFries, R. S., Doyle, J. C., Harrison, S. P., Johnston, F. H., Keeley, J. E., Krawchuk, M. A., Kull, C. A., Marston, J. B., Moritz, M. A., Prentice, I. C., Roos, C. I., Scott, A. C., Swetnam, T. W., Werf, G. R. van der and Pyne, S. J.: Fire in the Earth System, Science, 324(5926), 481–484, doi:10.1126/science.1163886, 2009.

Camia, A., and Amatulli, G.: Weather Factors and Fire Danger in the Mediterranean, Earth Observation of Wildland Fires in Mediterranean Ecosystems, 71–82, doi:10.1007/978-3-642-01754-4_6, 2010.

Carvalho, A., Flannigan, M. D., Logan, K., Miranda, A. I. and Borrego, C.: Fire activity in Portugal and its relationship to weather and the Canadian Fire Weather Index System, Int. J. Wildland Fire, 17(3), 328–338, doi:10.1071/WF07014, 2010.

Chen, Y., Morton, D. C., Andela, N., Giglio, L. and Randerson, J. T.: How much global burned area can be forecast on seasonal time scales using sea surface temperatures?, Environ. Res. Lett., 11(4), 045001, doi:10.1088/1748-9326/11/4/045001, 2016.

Duane, A., Kelly, L., Gijohann, K., Batllori, E., McCarthy, M. and Brotons, L.: Disentangling the Influence of Past Fires on Subsequent Fires in Mediterranean Landscapes, Ecosystems, 22(6), 1338-1351, doi:10.1007/s10021-019-00340-6, 2019.

Fang, L., Yang, J., Zu, J., Li, G. and Zhang, J.: Quantifying influences and relative importance of fire weather, topography, and vegetation on fire size and fire severity in a Chinese boreal forest landscape, Forest Ecology and Management, 356, 2–12, doi:10.1016/j.foreco.2015.01.011, 2015.

Fann, N., Alman, B., Broome, R. A., Morgan, G. G., Johnston, F. H., Pouliot, G. and Rappold, A. G.: The health impacts and economic value of wildland fire episodes in the U.S.: 2008-2012, Sci. Total Environ., 610–611, 802–809, doi:10.1016/j.scitotenv.2017.08.024, 2018.

Fernandes, P. M., Monteiro-Henriques, T., Guiomar, N., Loureiro, C. and Barros, A. M. G.: Bottom-Up Variables Govern Large-Fire Size in Portugal, Ecosystems, 19(8), 1362–1375, doi:10.1007/s10021-016-0010-2, 2016.

Flannigan, M. D., Logan, K. A., Amiro, B. D., Skinner, W. R., and Stocks, B. J.: Future area burned in Canada, Climate Change, 72(1), 1–16, doi: 10.1007/s10584-005-5935-y, 2005.

https://doi.org/10.5194/acp-2019-885
Preprint. Discussion started: 4 November 2019

Atmospheric Chemistry and Physics Discussions

Fréjaville, T. and Curt, T.: Seasonal changes in the human alteration of fire regimes beyond the climate forcing, Environ. Res. Lett., 12(3), 035006, doi:10.1088/1748-9326/aa5d23, 2017.

Heyerdahl, E. K., McKenzie, D., Daniels, L. D., Hessl, A. E., Littell, J. S. and Mantua, N. J.: Climate drivers of regionally synchronous fires in the inland northwest (1651-1900), International Journal of Wildland Fire. 17: 40-49., 40–49 [online] Available from: https://www.fs.usda.gov/treesearch/pubs/29482 (Accessed 11 July 2019), 2008.

Holden, Z. A., Swanson, A., Luce, C. H., Jolly, W. M., Maneta, M., Oyler, J. W., Warren, D. A., Parsons, R. and Affleck, D.: Decreasing fire season precipitation increased recent western US forest wildfire activity, PNAS, 115(36), E8349–E8357, doi:10.1073/pnas.1802316115, 2018.

Jaffe, D., Hafner, W., Chand, D., Westerling, A. and Spracklen, D.: Interannual Variations in PM2.5 due to Wildfires in the Western United States, Environ. Sci. Technol., 42(8), 2812–2818, doi:10.1021/es702755v, 2008.

Keane, R. E., Reinhardt, E. D., Scott, J., Gray, K. and Reardon, J.: Estimating forest canopy bulk density using six indirect methods, Canadian Journal of Forest Research, 35(3), 724–739, doi:10.1139/x04-213, 2005.

Keyser, A. and Westerling, A. L.: Climate drives inter-annual variability in probability of high severity fire occurrence in the western United States, Environ. Res. Lett., 12(6), 065003, doi:10.1088/1748-9326/aa6b10, 2017.

Kirchmeier-Young, M. C., Gillett, N. P., Zwiers, F. W., Cannon, A. J. and Anslow, F. S.: Attribution of the Influence of Human-Induced Climate Change on an Extreme Fire Season, Earth's Future, 7(1), 2-10, doi:10.1029/2018EF001050, 2018.

Krawczyk, B.: Learning from imbalanced data: open challenges and future directions, Prog Artif Intell, 5(4), 221–232, doi:10.1007/s13748-016-0094-0, 2016.

Krueger, E. S., Ochsner, T. E., Carlson, J. D., Engle, D. M., Twidwell, D. and Fuhlendorf, S. D.: Concurrent and antecedent soil moisture relate positively or negatively to probability of large wildfires depending on season, Int. J. Wildland Fire, 25(6), 657–668, doi:10.1071/WF15104, 2016.

Liaw, A. and Wiener, M.: Classification and Regression by randomForest, R News, 2, 18-22, 2002.

Littell, J. S., McKenzie, D., Peterson, D. L. and Westerling, A. L.: Climate and wildfire area burned in western U.S. ecoprovinces, 1916–2003, Ecological Applications, 19(4), 1003–1021, doi:10.1890/07-1183.1, 2009.

Liu, Y., L. Goodrick, S. and A. Stanturf, J.: Future U.S. wildfire potential trends projected using a dynamically downscaled climate change scenario, Forest Ecology and Management, 294, 120–135, doi:10.1016/j.foreco.2012.06.049, 2013.

Liu, Z. and Wimberly, M. C.: Climatic and Landscape Influences on Fire Regimes from 1984 to 2010 in the Western United States, PLOS ONE, 10(10), e0140839, doi:10.1371/journal.pone.0140839, 2015.

Marcos, R., Turco, M., Bedia, J., Lalsat, M. C. and Provenzale, A.: Seasonal predictability of summer fires in a Mediterranean environment, Int. J. Wildland Fire, 24(8), 1076–1084, doi: 10.1071/WF15079, 2015.

Morgan, P., Heyerdahl, E. K. and Gibson, C. E.: Multi-season climate synchronized forest fires throughout the 20th century, Northern Rockies, USA, Ecology. 89(3): 717-728., 717–728, 2008.

Nunes, M. C. S., Vasconcelos, M. J., Pereira, J. M. C., Dasgupta, N., Alldredge, R. J. and Rego, F. C.: Land Cover Type and Fire in Portugal: Do Fires Burn Land Cover Selectively?, Landscape Ecol, 20(6), 661–673, doi:10.1007/s10980-005-0070-8, 2005.

Parisien, M.-A., Parks, S. A., Krawchuk, M. A., Flannigan, M. D., Bowman, L. M. and Moritz, M. A.: Scale-dependent controls on the area burned in the boreal forest of Canada, 1980-2005, Ecol Appl, 21(3), 789–805, doi:10.1890/10-0326.1, 2011.

Parks, S. A., Parisien, M.-A. and Miller, C.: Spatial bottom-up controls on fire likelihood vary across western North America, Ecosphere. 3(1): Article 12., doi:10.1890/ES11-00298.1, 2012.

Pausas, J. G. and Keeley, J. E.: A Burning Story: The Role of Fire in the History of Life, BioScience, 59(7), 593–601, doi:10.1525/bio.2009.59.7.10, 2009.

Pellegrini, A. F. A., Anderegg, W. R. L., Paine, C. E. T., Hoffmann, W. A., Kartzinel, T., Rabin, S. S., Sheil, D., Franco, A. C. and Pacala, S. W.: Convergence of bark investment according to fire and climate structures ecosystem vulnerability to future change, Ecology Letters, 20(3), 307–316, doi:10.1111/ele.12725, 2017.

Peters, D. P. C., Pielke, R. A., Bestelmeyer, B. T., Allen, C. D., Munson-McGee, S. and Havstad, K. M.: Cross-scale interactions, nonlinearities, and forecasting catastrophic events, PNAS, 101(42), 15130–15135, doi:10.1073/pnas.0403822101, 2004.

Riley, K. L., Abatzoglou, J. T., Grenfell, I. C., Klene, A. E. and Heinsch, F. A.: The relationship of large fire occurrence with drought and fire danger indices in the western USA, 1984–2008: the role of temporal scale, International Journal of Wildland Fire, 22(7), 894, doi:10.1071/WF12149, 2013.

Ruthrof, K. X., Fontaine, J. B., Matusick, G., Breshears, D. D., Law, D. J., Powell, S. and Hardy, G.: How drought-induced forest die-off alters microclimate and increases fuel loadings and fire potentials, Int. J. Wildland Fire, 25(8), 819–830, doi:10.1071/WF15028, 2016.

Short, K. C.: Spatial wildfire occurrence data for the United States, 1992-2015 [FPA_FOD_20170508] (4th Edition), , doi:10.2737/RDS-2013-0009.4, 2017.

Slocum, M. G., Beckage, B., Platt, W. J., Orzell, S. L. and Taylor, W.: Effect of Climate on Wildfire Size: A Cross-Scale Analysis, Ecosystems, 13(6), 828–840, doi:10.1007/s10021-010-9357-y, 2010.

Sousa, P. M., Trigo, R. M. and Pereira, M. G.: Different approaches to model future burnt area in the Iberian Peninsula, Agricultural and Forest Meteorology, 202, 11–25, doi: 10.1016/j.agrformet.2014.11.018, 2015.

Spracklen, D. V., Mickley, L. J., Logan, J. A., Hudman, R. C., Yevich, R., Flannigan, M. D. and Westerling, A. L.: Impacts of climate change from 2000 to 2050 on wildfire activity and carbonaceous aerosol concentrations in the western United States, Journal of Geophysical Research: Atmospheres, 114(D20), doi: 10.1029/2008JD010966, 2009.

Steele-Feldman, A., Reinhardt, E. and Parsons, R. A.: Fuels Management-How to Measure Success: Conference Proceedings, USDA Forest Proceedings, 283–291, 2006.

Steel, Z. L., Safford, H. D., and Viers, J. H.: The fire frequency-severity relationship and the legacy of fire suppression in California forests, Ecosphere, 6(1), 1-23, doi:10.1890/ES14-00224.1, 2015.

Swetnam, T. W. and Anderson, R. S.: Fire Climatology in the western United States: introduction to special issue, Int. J. Wildland Fire, 17(1), 1–7, doi:10.1071/WF08016, 2008.

Urbieta, I. R., Zavala, G., Bedia, J., Gutierrez J. M., Miguel-Ayanz, J. S., Camia, A., Keeley, J. E. and, Moreno, J. M.: Fire activity as a function of fire–weather seasonal severity and antecedent climate across spatial scales in southern Europe and Pacific western USA, Environ. Res. Lett., 10, 114013, doi:10.1088/1748-9326/10/11/114013, 2015.

https://doi.org/10.5194/acp-2019-885
Preprint. Discussion started: 4 November 2019

Atmospheric Chemistry and Physics Discussions

Wang, S.-C., Wang, Y., Estes, M., Lei, R., Talbot, R., Zhu, L. and Hou, P.: Transport of Central American Fire Emissions to the U.S. Gulf Coast: Climatological Pathways and Impacts on Ozone and PM2.5, Journal of Geophysical Research: Atmospheres, 123(15), 8344–8361, doi:10.1029/2018JD028684, 2018.

Westerling, A. L.: Increasing western US forest wildfire activity: sensitivity to changes in the timing of spring, Philosophical Transactions of the Royal Society B: Biological Sciences, 371(1696), 20150178, doi:10.1098/rstb.2015.0178, 2016.

Westerling, A. L., Turner, M. G., Smithwick, E. A. H., Romme, W. H. and Ryan, M. G.: Continued warming could transform Greater Yellowstone fire regimes by mid-21st century, Proceedings of the National Academy of Sciences, 108(32), 13165-13170, doi:10.1073/pnas.1110199108, 2011.

Westerling, A. L., Hidalgo, H. G., Cayan, D. R. and Swetnam, T. W.: Warming and Earlier Spring Increase Western U.S. Forest Wildfire Activity, Science, 313(5789), 940–943, doi:10.1126/science.1128834, 2006.

Williams, P. A., Allen, C. D., Macalady, A. K., Griffin, D., Woodhouse, C. A., Meko, D. M., Swetnam, T. W., Rauscher, S. A., Seager, R., Grissino-Mayer, H. D., Dean, J. S., Cook, E. R., Gangodagamage, C., Cai, M. and McDowell, N. G.: Temperature as a potent driver of regional forest drought stress and tree mortality, Nature Climate Change, 3(3), 292–297, doi:10.1038/nclimate1693, 2013.

Yue, X., Mickley, L. J., Logan, J. A. and Kaplan, J. O.: Ensemble projections of wildfire activity and carbonaceous aerosol concentrations over the western United States in the mid-21st century, Atmos Environ (1994), 77, 767–780, doi:10.1016/j.atmosenv.2013.06.003, 2013.

Yue, X., Mickley, L. J., Logan, J. A., Hudman, R. C., Martin, M. V. and Yantosca, R. M.: Impact of 2050 climate change on North American wildfire: consequences for ozone air quality, Atmospheric Chemistry and Physics, 15(17), 10033–10055, doi:https://doi.org/10.5194/acp-15-10033-2015, 2015.

Zhang, X., Kondragunta, S. and Roy, D. P.: Interannual variation in biomass burning and fire seasonality derived from geostationary satellite data across the contiguous United States from 1995 to 2011, Journal of Geophysical Research: Biogeosciences, 119(6), 1147–1162, doi:10.1002/2013JG002518, 2014.

https://doi.org/10.5194/acp-2019-885
Preprint. Discussion started: 4 November 2019

Atmospheric Chemistry and Physics Discussions

620

Table 1. Predictor variables that were used in the fire prediction models

Variables	Abbreviation	Categories	Temporal resolution	Spatial resolution	Data source
Weather variables					
Monthly mean surface temperature	temp	weather			
Monthly mean daily precipitation	apcp	weather			
Monthly total precipitation	asum	weather			
Monthly mean surface relative humidity (%)	rhum	weather			
Monthly mean U-component of wind speed	U	weather	monthly	32 km	North American Regional Reanalysis (NARR)
Monthly mean V-component of wind speed	V	weather			
Monthly maximum temperature	tmax	weather			
Monthly minimum RH	rmin	weather			
Number of consecutive days without rainfall in a month	LargeConsec	weather			
1-month SPEI	SPEI	weather	1-month	0.5°	Global SPEI database
Fuel variables					
Monthly mean Leaf Area Index (LAI)	LAI	fuel	monthly	500 m	MODerate resolution Imaging Spectroradiometer (MODIS)
Monthly mean sum of neighboring LAI	convLAI	fuel	monthly	500 m	MODerate resolution Imaging Spectroradiometer (MODIS)
Monthly mean soil moisture at 0-10 cm	soil	fuel	monthly	0.125°	North American Land Data Assimilation System (NLDAS-2)
Geospatial and population variables					
Land types	land_	fix		30 m	National Land Cover Database (NLCD)
Ecoregion types	eco_	fix			U.S. Environmental Protection Agency (EPA)
Population density	pop	fix			U.S. Census 2010
Climate variables (over 1970-2000)					
Long-term average and standard deviation of monthly temperature	temp_avg; temp_sd	fix			

https://doi.org/10.5194/acp-2019-885
Preprint. Discussion started: 4 November 2019

Atmospheric Chemistry and Physics Discussions

Long-term average and standard deviation of monthly daily precipitation	apcp_avg; apcp_sd	fix			
Long-term average and standard deviation of monthly maximum temperature	tmax_avg; tmax_sd	fix			
Long-term average and standard deviation of monthly total precipitation	asum_avg; asum_sd	fix			
Climate anomalies of monthly mean temperature	temp_anomaly	climate	monthly	32 km	North American Regional Reanalysis (NARR)
Climate anomalies of monthly mean of daily precipitation	apcp_anomaly	climate			
Climate anomalies of monthly mean RH	rhum_anomaly	climate			
Climate anomalies of monthly maximum temperature	tmax_anomaly	climate			
Climate anomalies of monthly minimum RH	rmin_anomaly	climate			
Climate anomalies of monthly total precipitation	asum_anomaly	climate			
Lagged variables					
Fire season one					
The average of antecedent monthly mean of daily precipitation for the month of t-1	apcp_mean1m	weather	monthly	32 km	North American Regional Reanalysis (NARR)
The average of antecedent SPEI for the month of t-1, t-2, t-3, t-4, t-5, and t-6	SPEI_mean1m	weather	monthly	32 km	North American Regional Reanalysis (NARR)
The average of antecedent monthly mean of LAI and sum of neighboring LAI for the month of t-5	LAI_mean5m, convLAI_mean5m	fuel	monthly	500 m	MODerate resolution Imaging Spectroradiometer (MODIS)
Fire season two					
The average of antecedent monthly mean of daily precipitation for the month of t-1, t-2, and t-3	apcp_mean1m	weather	monthly	32 km	North American Regional Reanalysis (NARR)
The average of antecedent monthly mean of temperature for the month of t-1 and t-2	temp_mean1m	weather	monthly	32 km	North American Regional Reanalysis (NARR)
The average of antecedent SPEI for the month of t-1, t-2, and t-3	SPEI_mean1m	weather	1-month	0.5°	Global SPEI database

https://doi.org/10.5194/acp-2019-885
Preprint. Discussion started: 4 November 2019

Atmospheric Chemistry and Physics Discussions

Table 2. Model performance at grid level for the two fire seasons.

Fire season	Evaluation Metrics					
	Accuracy	F1-score	AUC	R^2	RMSE (km^2)	MAE (km^2)
F1	0.74	0.64	0.71	0.42	8.37	1.13
F2	0.74	0.70	0.74	0.40	4.26	0.57

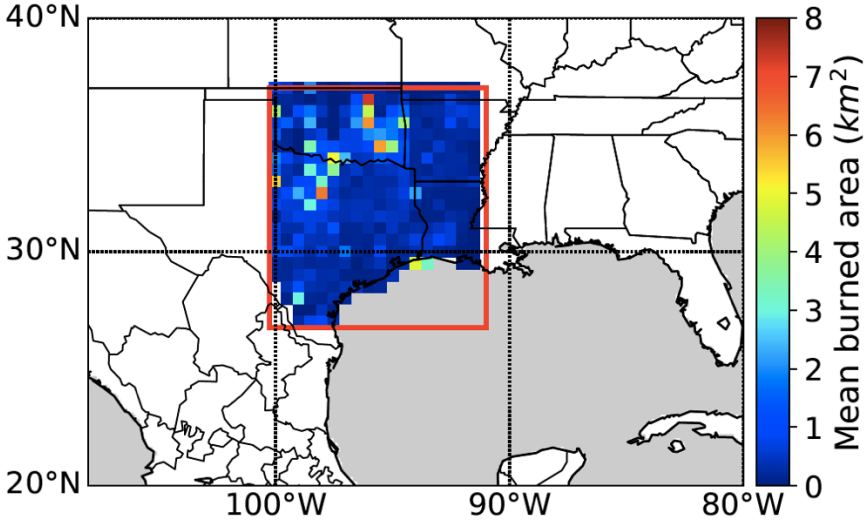
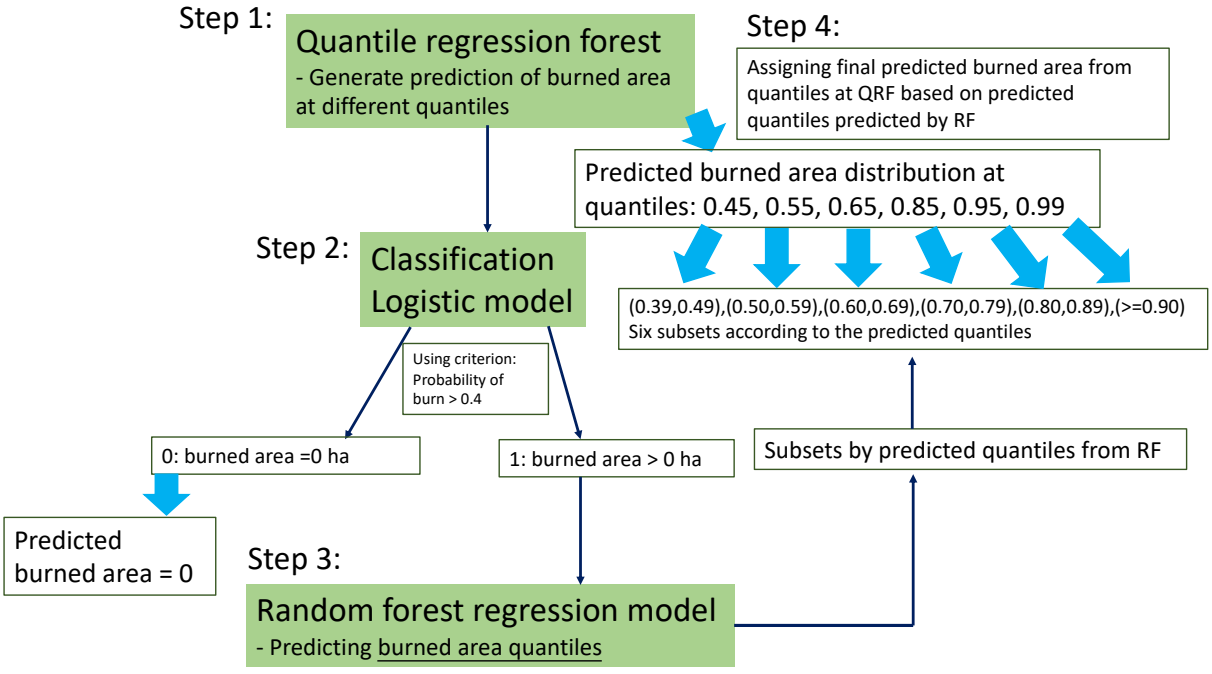


Figure 1. The colored grid boxes show the averaged burned area for the winter-spring and summer fire seasons during 2002-2015 from Fire Program Analysis Fire-Occurrence Database (FPA-FOD). The red box denotes the South Central US domain.

https://doi.org/10.5194/acp-2019-885
Preprint. Discussion started: 4 November 2019



650 **Figure 2.** Illustration of the steps in the model framework. The model framework includes a four-step model built by a classification logistical model, a random forest model predicting quantiles of burned area, and a quantile regression forests model producing conditional burned area distributions.

655

https://doi.org/10.5194/acp-2019-885
Preprint. Discussion started: 4 November 2019

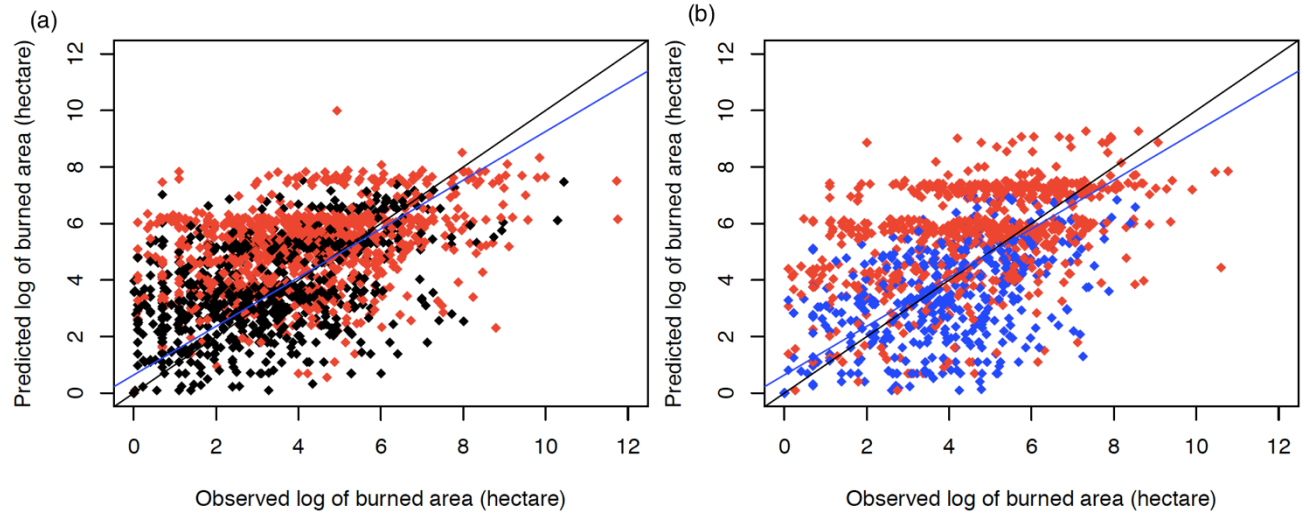


Figure 3. Comparison between log of observed and predicted burned area (hectare) for the (a) winter-spring and (b) summer fire season in selected years: 2011 (red, year of the largest burned area), 2008 (blue, year with burned area close to the 14-year mean of its season), and 2014 (black, year with burned area close to the 14-year mean of its season).

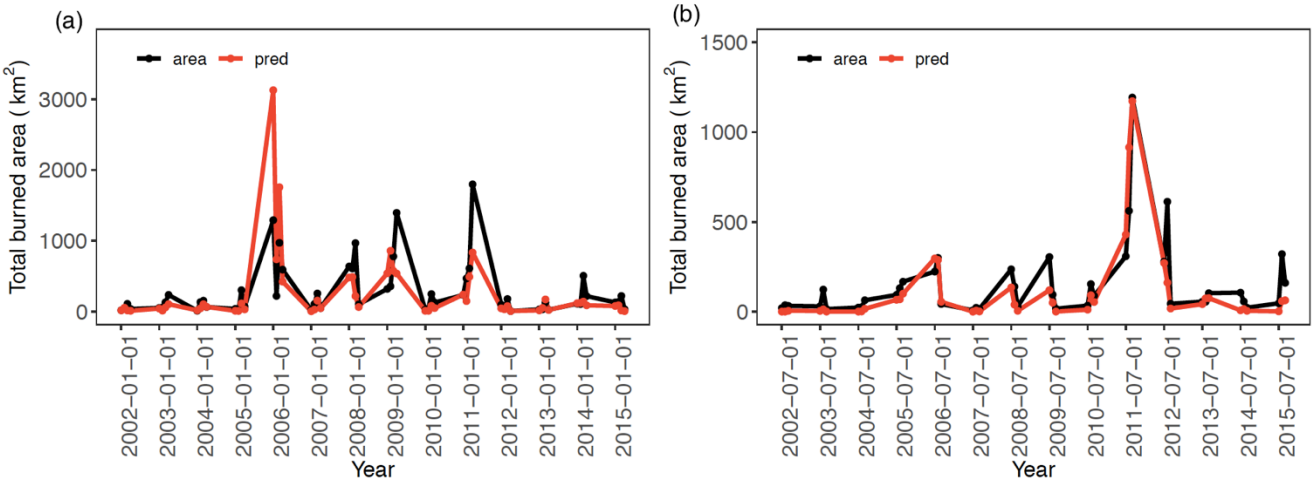
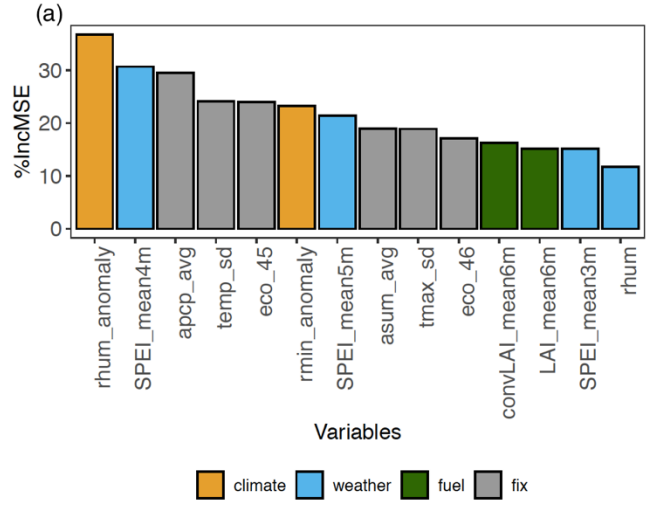
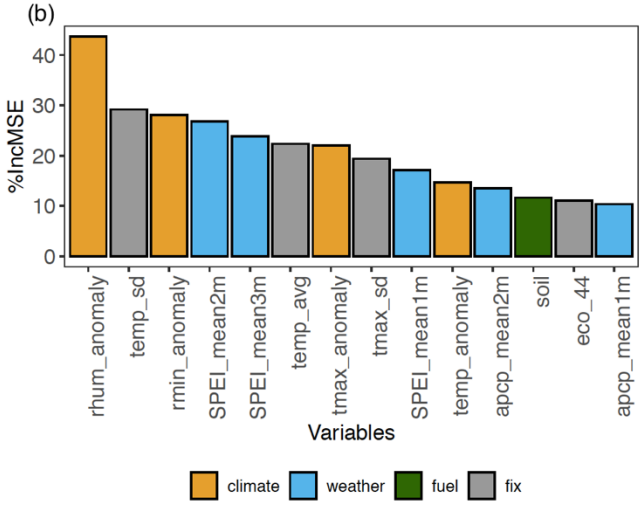


Figure 4. Timeseries of observed (black line) and predicted total burned area (red line) over South Central US for the (a) winter-spring and (b) summer fire season.

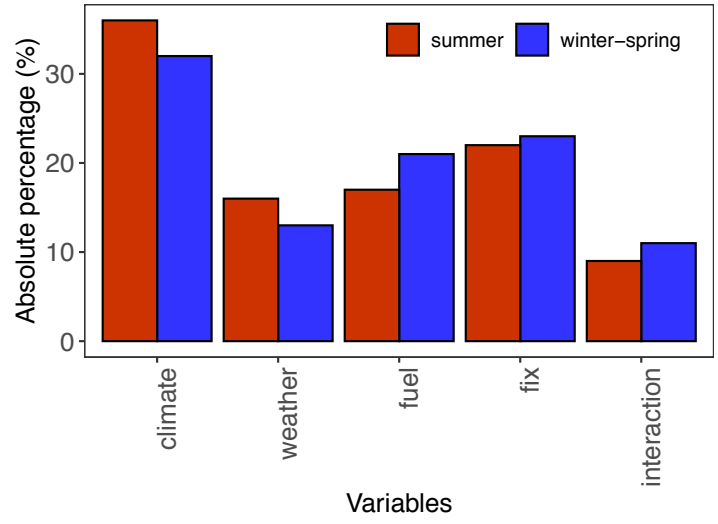
https://doi.org/10.5194/acp-2019-885
Preprint. Discussion started: 4 November 2019





670

Figure 5. Relative importance of the top 14 variables presented by increase in mean square errors (%Inc.MSE) for (a) the winter-spring fire season (b) summer fire season.



675

Figure 6. The mean scaled absolute percentage of the environmental controls for the winter-spring (blue) and summer fire season (red).