Mapping California Fires (2016-2020) and Predicting Fire Size

Westin Guthrie and Al-Ekram Elahee Hridoy
UNM wguthrie@unm.edu, UNM ahridoy@unm.edu

Dr. Liping Yang
UNM GEOG 585L Instructor lipingyang@unm.edu

1. Introduction and Motivation (Al-Ekram/Westin)

Fire in California has always been something the region has existed with symbiotically. Due to policy and human interaction with the environment California fires have changed for the worse. An increase of large fires and a larger presence of fire in wildland-urban interface areas make this project important. The goal of this work is to provide a framework for predicting fire size based on climatic variables during the month of ignition and displaying the results which can then be improved upon in the future.

2. Literature Review (Westin)

In the West, there has been a notable increase in stand-replacing fires as mean and maximum fire size and annually burned area have dramatically increased since the 1980s (Miller et al. 2009). Recent trends of rising regional temperature (Miller et al. 2009) and an autumn precipitation reduction of ~30% over the last four decades, aggregate fire weather indices have increased by +20% in California (Goss et al. 2020). As an increasing vapor pressure deficit becomes more prominent in California, fine fuels and small wood debris will experience even lower moisture content, boosting the ability of fire spread (Balch et al. 2022).

Machine learning has been used to understand fires in boreal forests of Alaska to better investigate the predictability of final fire size at the time of ignition (Coffield et al. 2019). Other studies have used machine learning in northern California accounting for the parameters of powerlines, terrain, and vegetation to better predict the risk of ignition (Malik et al. 2021). Other models have been used to better simulate forest fire spread with cellular automaton (Zheng et al. 2017) and another machine learning algorithm-based evaluation looked to understand relationships between wildfires and drought (Chen 2022). Work has been done using machine learning, but the only one study included vapor pressure deficit in their analysis and that was in Alaska.

It is important to include vapor pressure deficit in fire predictions as it is a measure of the surface drying power of the atmosphere (Seager et al. 2015). Our analysis includes vapor pressure deficit as in different ecosystems it has been found to increase the risk of lightning strike ignitions (Sedano and Randerson 2014). This relationship between lightning strike ignitions, and likely accidental ignitions as well, is important as vapor pressure deficit has increased in the west with relationships between VPD and summer forest-fire area strongly suggesting that early all the increase in summer forest-fire area during 1972–2018 was driven by increased VPD (Williams et al. 2019).

3. Data Sources and Methods (Westin)

Our goal was to develop a website that visualizes historical fire ignitions in California of large fire events (1000 acres or more) and predicts their fire size through the use machine learning techniques. To project ignitions, we used various environmental variables, including monthly maximum, minimum, and average temperature, historical fire shapes, elevation, aspect, slope, precipitation, and maximum and minimum vapor pressure to train our machine learning model.

Monthly climatic data was be downloaded from PRISM. Here we will find precipitation, mean temperature, minimum temperature, maximum temperature, minimum vapor pressure deficit, and maximum vapor pressure deficit. All these climatic aspects have impacts on fire ignitions and fire spread. The data was formatted into the form of a CSV.

We also derived a DEM from the USGS. A Western US 90m DEM will be clipped to California shapefile boundary. Once clipped the DEM was used to find aspect and slope using python using GDAL and Richdem. Elevation, slope, and aspect play an important role in fires as well so it was important to capture them.

Fire ignitions were downloaded from Karen Short (2022). Which has fire ignitions from 1992-2020. We focused on the last five years from 2016-2020 fire ignitions of a fire size of 1000 acres or larger. This resulted in 177 fires throughout California meeting those two criteria.

FIRE_YEAR	MTBS_FIRE	DISCOVERY	FIRE_SIZE	LATITUDE	LONGITUDE	FIPS_NAME	geometry	ppt_month	tmax_month	tmean_month	tmin_month	vpdmax_month	vpdmin_month	spect_degrees	elev	slope_percent
2016	CLARK	8/4/2016	2819	37.79888889	-118.9225	Mono County	POINT (-11	3.493000031	21.31399918	12.24000072	3.16899991	22.06200027	3.959000111	255.9637604	2432	521846.3438
2016	PONY	6/7/2016	2860	41.62305556	-123.5575	Siskiyou County	POINT (-12	24.9470005	28.73699951	20.5870018	12.43900013	29.13500023	5.242000103	91.53891754	1553	4241502
2016	PINE	6/30/2016	2304	34.63083333	-119.2302778	Ventura County	POINT (-11	0	27.42200089	21.46700096	15.51399994	25.9109993	5.890999794	24.77514076	1920	2174584.75
2016	JACOBSON	10/20/2016	1702	36.22777778	-118.5633333	Tulare County	POINT (-11	77.53800201	10.32600021	5.842000484	1.358999968	9.923000336	3.374000072	48.03798676	2225	6416925.5
2016	MEADOW	10/29/2016	4357	35.98666667	-118.5516667	Tulare County	POINT (-11	50.56000137	16.76399994	7.659000397	-1.444000006	1.157999992	1.157999992	109.4579773	1652	2241689.5
2016	HIDDEN	10/29/2016	2739	36.26444444	-118.6355556	Tulare County	POINT (-11	70.90899658	10.57400036	5.84800005	1.123000026	10.38199997	3.471999884	208.6011505	2455	6213182
2016	SLATE	10/4/2016	1825	36.09361111	-118.5788889	Tulare County	POINT (-11	70.17299652	10.29699993	5.920000076	1.544999957	9.925000191	3.506000042	300.2770386	2651	2535436.5

Fig 1. – example from our completed CSV with needed data to support machine learning and regressions

To use the machine learning algorithms the data must be formatted in a CSV with an x,y. We then matched each fire's climatic data to the month the fire started in. This helped the algorithm better learn the conditions that help ignitions grow into larger fires depending on slope, aspect, elevation, and our weather variables. In Fig.1 we showed the first 6 fires in our dataset with their environmental and climate variables attached. Each fire has the year, name, the discovery date, latitude and longitude, point geometry, county name, precipitation, max temperature, mean temperature, minimum temperature, max vapor pressure deficit, minimum vapor pressure deficit, aspect, slope, and elevation data. All 177 fires in the study area have the same format.

At each fire point we used python and zonal statistics to determine the climate and environmental conditions present at that specific site during the month of the fire. Initially, we created a csv that had those variables for the last 5 years, but we needed to filter it, so each fire only had the conditions present at the month of the ignition for the machine learning algorithm.

3.1 Machine Learning Methods (Westin based off Al's notebook)

To predict wildfire size in acres, Al used the python libraries numpy, pandas, and sk.learn within a notebook to execute our machine learning size predictions. Pandas were used to read in the CSV data from the before mentioned section and the explanatory climatic variables from that csv were then applied to different sk.learn properties.

Our csv data was preprocessed with *sklearn.preprocessing.StandardScaler* Al dropped non-numeric columns and then filled missing values with '0.' After this our variables were scaled with the StandardScaler which standardizes features by removing the mean and scaling to unit variance which is a important requirement for machine learning estimators where if not scaled, individual features do not appear to be normally distributed data. Al then fit the data and then transformed it, after the transformation process k-fold cross-validation was used. K-fold is used to estimate the skill of a machine learning model to estimate the skill of the model on new data. For our dataset Al then divided the csv into *K (5)* groups. The model then for each group, takes a group as a test data set and then the remaining groups as a training data set, then fits a model on the training set evaluating it on the test set, and finally it then retains the evaluation score and discards the model summarizing the skill of the model using the sample of model evaluation scores. This method was done on three different models, *sklearn.linear_model.LinearRegression*, *sklearn.tree.DecisionTreeRegressor,* and *sklearn.ensemble.RandomForestRegressor.*

Table 1 – displaying model results from ML

Model	R2 Score Mean	Std
Linear Regression	-0.1954	0.1195
Decision Tree	-1.3894	0.264
Random Forest	-0.1615	0.0899

From table 1, Al decided that Random Forest resulted as the best model and moved forward to fire size predictions using this model. For our model, Al found the best hyperparameters and the best combination of hyperparameters to the random forest algorithm enabling us to have our fire size predictions which gave us a R2 of 0.05515371905326605.

4. Discussion (Westin)

Our completed webapp can show all 2016-2020 >1000-acre ignitions within California. Through this map each point is clickable with popup information. A table below the map and graphs shows all the climatic explanatory variables at the month of fire ignition and other information which is also shown in fire point pop-ups. Pop-up information includes the discovery date, year, fire name, observed fire size, and Al's predicted fire size. This webapp can be filtered by county through a pie chart, by ignition type also through a pie chart, and finally by year. When the data is filtered the table below will show fire data by filtered query. We successfully were able to achieve a model that displays predicted fire size which can be seen in comparison to observed fire size.

This work could be increased in utility by making each point clickable and filtering the table based on clicked point. Further utility would be changing the sizes of the pie charts to have better usability, currently filtering by county is difficult due to county names being too large for pie chart slices making the user guess or mouse-over each slice to find the desired county. Within the machine learning model, being able to display the predicted fire size when compared to observed fire size will be important to improve upon in the future. Otherwise, being able to predict fire ignitions like the work originally intended to do would be the next step. After this experience the data that is needed was too difficult to obtain and create a successful model within the scope of this course.

References

Balch, J. K., J. T. Abatzoglou, M. B. Joseph, M. J. Koontz, A. L. Mahood, J. McGlinchy, M. E. Cattau, and A. P. Williams. 2022. Warming weakens the night-time barrier to global fire. *Nature* 602 (7897):442–448.

Chen, A. 2022. Evaluating the relationships between wildfires and drought using machine learning. *International Journal of Wildland Fire* 31 (3):230–239.

Coffield, S. R., C. A. Graff, Y. Chen, P. Smyth, E. Foufoula-Georgiou, J. T. Randerson, S. R. Coffield, C. A. Graff, Y. Chen, P. Smyth, E. Foufoula-Georgiou, and J. T. Randerson. 2019. Machine learning to predict final fire size at the time of ignition. *International Journal of Wildland Fire* 28 (11):861–873.

Garrison, J. D., and T. E. Huxman. 2020. A tale of two suburbias: Turning up the heat in Southern California's flammable wildland-urban interface. *Cities* 104:102725.

Goss, M., D. L. Swain, J. T. Abatzoglou, A. Sarhadi, C. A. Kolden, A. P. Williams, and N. S. Diffenbaugh. 2020. Climate change is increasing the likelihood of extreme autumn wildfire conditions across California. *Environmental Research Letters* 15 (9):094016.

Li, H., M. Kanamitsu, S.-Y. Hong, K. Yoshimura, D. R. Cayan, V. Misra, and L. Sun. 2014. Projected climate change scenario over California by a regional ocean–atmosphere coupled model system. *Climatic Change* 122 (4):609–619.
Malik, A., M. R. Rao, N. Puppala, P. Koouri, V. A. K. Thota, Q. Liu, S. Chiao, and J. Gao. 2021. Data-Driven Wildfire Risk Prediction in Northern California. *Atmosphere* 12 (1):109.

Miller, J. D., H. D. Safford, M. Crimmins, and A. E. Thode. 2009. Quantitative Evidence for Increasing Forest Fire Severity in the Sierra Nevada and Southern Cascade Mountains, California and Nevada, USA. *Ecosystems* 12 (1):16–32.

PRISM Climate Group, Oregon State University, https://prism.oregonstate.edu, data created 4 Feb 2014, accessed 16 Dec 2020.

Radeloff, V. C., D. P. Helmers, H. A. Kramer, M. H. Mockrin, P. M. Alexandre, A. Bar-Massada, V. Butsic, T. J. Hawbaker, S. Martinuzzi, A. D. Syphard, and S. I. Stewart. 2018. Rapid growth of the US wildland-urban interface raises wildfire risk. *Proceedings of the National Academy of Sciences of the United States of America* 115 (13):3314–3319.

Seager, R., A. Hooks, A. P. Williams, B. Cook, J. Nakamura, and N. Henderson. 2015. Climatology, Variability, and Trends in the U.S. Vapor Pressure Deficit, an Important Fire-Related Meteorological Quantity. *Journal of Applied Meteorology and Climatology* 54 (6):1121–1141.

Sedano, F., and J. T. Randerson. 2014. Multi-scale influence of vapor pressure deficit on fire ignition and spread in boreal forest ecosystems. *Biogeosciences* 11 (14):3739–3755.

Short, Karen C. 2022. Spatial wildfire occurrence data for the United States, 1992-2020 [FPA_FOD_20221014]. 6th Edition. Fort Collins, CO: Forest Service Research Data Archive. https://doi.org/10.2737/RDS-2013-0009.6

Steven E. Hanser, 2008, Elevation in the Western United States (90 meter DEM): Wiley.

Swain, D. L., B. Langenbrunner, J. D. Neelin, and A. Hall. 2018. Increasing precipitation volatility in twenty-first-century California. *Nature Climate Change* 8 (5):427–433.

Syphard, A. D., H. Rustigian-Romsos, M. Mann, E. Conlisk, M. A. Moritz, and D. Ackerly. 2019. The relative influence of climate and housing development on current and projected future fire patterns and structure loss across three California landscapes. *Global Environmental Change* 56:41–55.

U.S. Department of Commerce, U.S. Census Bureau, Geography Division. California State boundary in shapefile format from the US Census Bureau's 2016 MAF/TIGER database.

Williams, A. P., J. T. Abatzoglou, A. Gershunov, J. Guzman-Morales, D. A. Bishop, J. K. Balch, and D. P. Lettenmaier. 2019. Observed Impacts of Anthropogenic Climate Change on Wildfire in California. *Earth's Future* 7 (8):892–910.

Zheng, Z., W. Huang, S. Li, and Y. Zeng. 2017. Forest fire spread simulating model using cellular automaton with extreme learning machine. *Ecological Modelling* 348:33–43.