Can drought conditions be used to predict the likelyhood of utility-caused wildfires?

Anne-Laure Strong
PhD Candidate, *Civil and Evionmental Engineering*

Stanford, CA

***Abstract*—Large scale planned outages by utilities to prevent wildfires are based on a lack of understanding of the specific causes of utility-triggered wildfire. We proposed to look at the ability of neural networks to predict utility-caused wildfires based on publicly available drought maps. We used CNNs and Xception to encore the maps and CPUC publicly available utility-caused fires database. We found that the models were unable to predict accurately the likelihood of such fires or to find any link between the maps and fires, due to either lack of data or lack of correlation between drought conditions and utility-caused fires. We propose some next steps to continue to investigate the topics.**

I. Introduction

In 2019, PG&E planned several outages to prevent electric equipment from sparking wildfires At one point, one peak power outage reached 800,000 customers.[1] This outage affected more than 2 million people in California. A conservative cost estimate of the outage is at least a 2 billion dollar bill for the state.[2] In the last four years, wildfires caused by utilities have become an increasing issue in California. The three main utilities caused over 2000 wildfires between 2014 and 2017[3], and hundreds more in 2018. The causes of utility-triggered wildfires include a nonexclusive combination wind, sparks triggered by utility equipment, fuel availability, accessibility of the area, type of vegetation, land topology, temperature, relative humidity, precipitation, soil moisture, time of year… Several research projects have already focused on the use of neural networks for the prediction of wildfires[4] and wildfire risk[5], based on soil moisture[6], GIS-based tropical forest fire risk[7] or wildfires caused by lightning [8].

California has been subjected to drought conditions varying throughout the year, from historical drought to drought-free.[9]Drought conditions impact several factors like fuel availability, since the more rainfall there is the more vegetation growth is observed. Humidity, temperatures, precipitations all influence drought conditions. Finally soil moisture will be a direct result of the drought conditions as well. Therefore, we thought that drought conditions could be a good indicator of conditions leading to wildfires. Assuming that the ability of utilities infrastructure to spark a wildfire is independent of environmental conditions, we made the hypothesis that drought conditions should affect how many utilities fire were started.

We set to confirm or question this hypothesis by using neural network to predict the probability of wildfires caused by electric equipment in PG&E territory based on weekly drought conditions.

The significance of this work could improve the way that preventative outages are selected and highlights which environmental factors contribute the most to wildfires due to electric equipment.

II. Data

A. Outage data

California Investor-Owned Utilities are regulated by the California Public Utilities Commission. We accessed the data for PG&E as a CSV file formatted by the Los Angeles Times.[3] The information contained the GPS coordinate of the fire, start time, acrage and cause.

B. Drought data

We used images produced by the US Drought Monitor.[10]The Drought Monitor was created in 1999 and operated "jointly by the National Drought Mitigation Center (NDMC) at the University of Nebraska-Lincoln, the National Oceanic and Atmospheric Administration (NOAA), and the U.S. Department of Agriculture (USDA). The NDMC hosts the web site of the drought monitor and the associated data, and provides the map and data to NOAA, USDA and other agencies." [10]

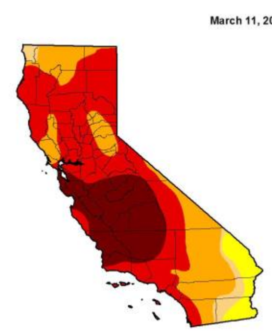


Fig. 1. Example of drought image of California as of March 11, 2014 obtained from the US Drought Monitor[11] Extreme Drought conditions were present then in the Bay Area and Central California as pictured in marroon.

USDM created a weekly map of the drought situation in the US, which is used to determine disaster situations and for insurance purposes. The drought is classified on a scale from abnormally dry (D0) to Exceptionally Dry (D4) on a scale from

Identify applicable funding agency here. If none, delete this text box.

yellow to maroon. If there is no drought then the land is shown as white. The maps are freely available online.

We created an API to collect weekly maps for California and saved them as jpeg files. Each image resolution was 912x912, as seen in Fig. 1. One image in November 2016 was corrupted, so we reused the previous week image.

III. Architecture and Hyperparameters

A. Simple Classifier

We first used a simple classifier the drought picture of the week is the input:

(1) If there is a fire during the week, then the result is positive. If there is no fire during the week, the result is negative.

(2) If there are 3/6/10 fires or more during the week then the result is positive. Otherwise, the result is negative. (3 fires brings the amount of positive vs negative weeks to approx. 50%)

The architecture of the simple classifier was as follow:

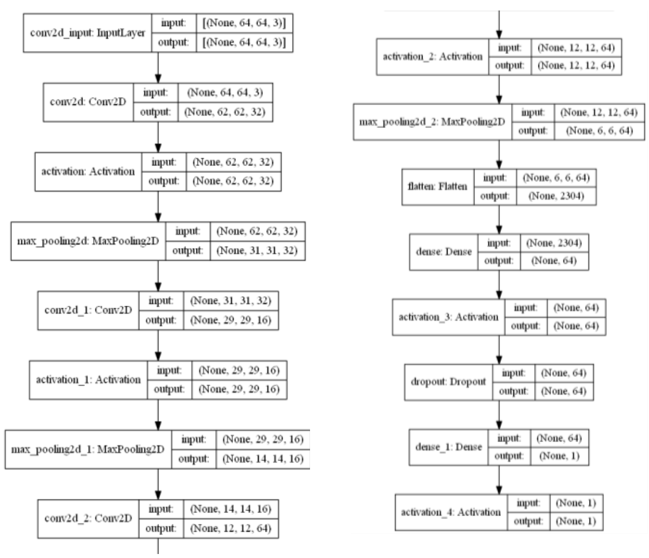


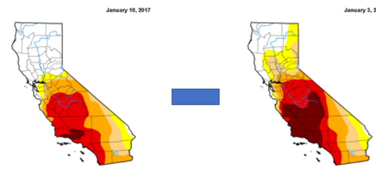
Fig. 2. Architecture of the Convolutional Neural Network for the simple classifier

B. Difference classifier

To take into account the time evolution, instead of using the image as a prediction, we took the difference between two weeks as a prediction of the wildfire risk. In order to have the images as the data, we now input the difference between two images and compare it to the likelihood of having a fire that week, as represented in equation (1) and Fig.3.

Input= Image [current week]- Image [previous week] (1)

Fig. 3. Illustration of equation (1): difference between two weekly drought images in January 2017



The network architecture used Xception as a pre-trained model. [12] Xception was picked for its efficiency and potentially better generalization to other datasets. The architecture is similar to the one presented on Fig.5 with the exception of the last dense output shape being (None, 1) since this is a simple classifier.

C. Area classifier

In this experiment, we corrected several flaws of the two precedent classifiers:

- Not leveraging as many fires since only one or three fires switched the results to positive
- Not leveraging the location of the fires

The fires were classified in 19 economic areas of PG&E using ArcGIS software as, represented on Fig. 4.

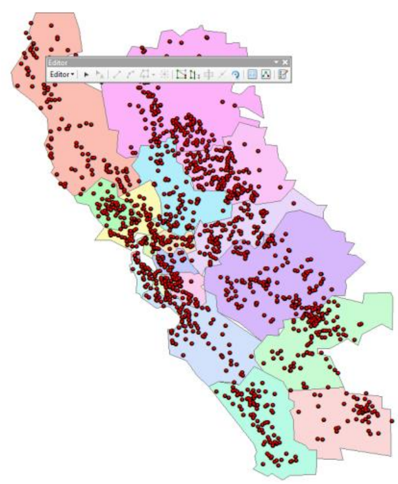


Fig. 4. Fires 2014-2018 by PG&E 19 economic areas

We represented results for each week with multi-hot encoding, as an array of shape (1,19), where 0 at position [i] represents no fire in area number i, and 1 represent a fire that week in the area i. For example, [0 1 0 1 0 0 0 0 0 ….0] would mean that there was a fire in area 1 and area 3 that week.

We kept the previous architecture with the modification of the shape of the output to an array of length 19, as represented in Fig.5.

Table I. represent the different hyperprameters. We biased the parameters toward speed initially to see if the model could learn at all. All across the models, we used the Xavier initialization since it is one of the preferred initialization method for CNNs with ReLu activation function. We used Adam optimizer to try to minimize noise in the convergence.

Since we looked at classifiers, our loss function with binary cross-entropy.

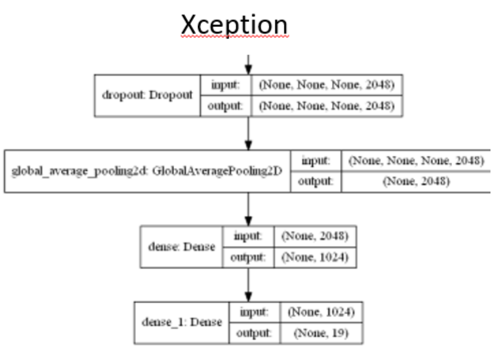


Fig. 5. Architecture of the Model for the area classifier, the architecture is identical to the difference classifier with the exception of the last dense output shape being (None,1) for the difference classifier.

TABLE I.

HYPERPARAMETERS AND FUNCTIONS

	Hyperparameters		
	Simple classifier	***Difference classifier***	***Area classifier***
Epochs	50 (not all needed)	30	30
Initialization	Xavier	Xavier	Xavier
Optimizer	RMSprop	Adam	Adam
Learning rate	0.001	0.0004	0.0004
Regularization	n/a	No dropout/ Dropout=0.5	Dropout= 0.5
Last activation function	Sigmoid	Sigmoid	Sigmoid (tanh and hard sigmoid were also tried)
Batch size	16	12	12
Image size	64	128	128
Loss function	Binary cross-entropy	Binary cross-entropy	Binary cross-entropy (we also tried hinge)

IV. RESULTS AND ANALYSIS

The metrics used to evaluate our work is primary accuracy. Our goal was to see if the model was learning any insights while increasing accuracy.

A. Simple Classifier

The simple classifier failed to learn any pattern during the training of the model. The model would oscillate around the average number of positive or negative fires, whichever is the largest, as seen in Fig. 6. The model would predict all 1s or all 0s.

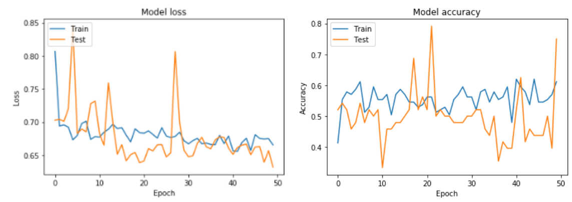


Fig. 6. Model loss and accuracy for simple classifier with threshold of 10 fires per week.

There might be a lack of data for the model to learn, but we suspect that the additional data necessary would need to be massive for the model to be able to learn from a single California map. In other words, there is too much noice and other parameters at play for the California overall conditions to play a significant role in the presence of 1 to 10 started by PG&E in California in a single week.

B. Difference Classifier

Since the computation took longer, we implemented an early exit when the validation loss would stop decreasing. Fig.7 shows the early exit after 8 epochs when the validation loss stops improving.
The model training accuracy improves without any improvement on the validation accuracy. Since the average probability in the validation set is 80%, the model guess that there will be always be fire for an accuracy that seems artificially high. However, no additional learning has occurred.

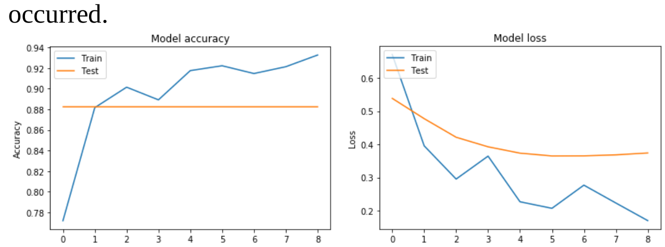


Fig. 7. Model loss and accuracy for the difference classifier

We looked at the precision and recall as shown in Fig.8. Changing the theshold to a higher number for positive results improves the precision at the expense of the recall. In our case, recall (not missing a fire prediction) is more important so that would actually be detrimental to our goal, but it did help with the precession and hence predictions.

Threshold		precision	recall	f1-score	support
Theshold 0.5	0	0.00	0.00	0.00	10
	1	0.76	1.00	0.86	32
	accuracy			0.76	42
	macro avg	0.38	0.50	0.43	42
	weighted avg	0.58	0.76	0.66	42
Threshold 0.92	0	0.38	0.30	0.33	10
	1	0.79	0.84	0.82	32
	accuracy			0.71	42
	macro avg	0.58	0.57	0.58	42
	weighted avg	0.69	0.71	0.70	42

Fig. 8. Prediction threshold and effects on precision and recall.

C. Area Classifier

We were hoping that the area classifier would force the model to learn more nuanced information. Fig. 9 shows that the model is learning and converging toward a solution since the losses are decreasing, but that the validation model accuracy is failing to improve. Analysis the output more in details, we discovered that each area had converged around the average of positive results for this area (Fig. 10). We then looked at each probability and for patterns of increase in positive results. Fig. 11 represents this analysis. Positive weeks, when fire occurs, are not more likely to have an increase in probability that negative weeks. There are as many positive weeks above the median than under the median for each area.

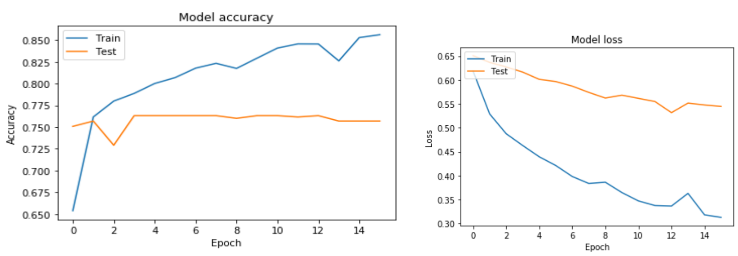


Fig. 9. Model loss and accuracy for the area classifier

Fig. 10. Output analysis in Excel, color scale from green to red represent higher fire probability. Each area probability output converged around the average probability of fire for the area

Fig. 11. Output analysis by zone in Excel. Cells in red are positive results (fire week), if the font is in black then the cell value is above the median, if the cell font is white then the cell value is below the median for the area.

D. Discussion

The main issues that we faced was the lack of large-scale data to discern any pattern. Given the amount of noise in the data, the weekly drought maps did not allow the model to discern any pattern.

However, given that we failed to see any amount of learning besides randomness does not make us confident that drought change do contribute to utility wildfire. Other elements like wind, and soil moisture more directly contribute to wildfires and do not seem to be correlated to the drought conditions like we initially thought.

Since the amount of data was relatively small, it is not unreasonable to give a human look to the data and see if there were any patterns. However, unlike our initial intuition, the researchers failed to see any patterns in the maps and data either.

Interestingly enough, the amount of wildfires caused by the utility is such that the model can reach a really high accuracy by assuming that each week there will be a fire caused by PG&E regardless of the drought conditions.

E. Limitations

The main limitations and issues of our analysis were as follow:

(1) PG&E does not cover all of California, but the drought map cover the entire state. This does create a lot of noise in our analysis.
(2) We took into account all types of fires, but some fires that were small or in urban areas may have needed to be excluded because they are unlikely to be drought-dependent.
(3) Lack of data. The maps being produced only weekly, which did reduce the importance of each single fire.
(4) One single factor. While this is was a good starting point, we believe that other factors should have been taken into account to be able to predict fires. However, the initial lightness of the model was perfect to set things up initially and for fast iterations until the model was running.

(5) Using classifier as an approximation for probabilities. Since sigmoid and the binary cross-entropy biais the model towards picking positive or negative results, it is not always a good approximation of probability.
(6) Our PG&E areas were based on economic factor rather than climate, vegetation and geological factors which may have been a better classification

Conclusions and next Steps

We initially were hoping to be able to predict an increase in wildfire caused by utilities from drought-maps in California. We used the weekly drought maps produced by US Drought Monitor and Convolutional Neural Networks to attempt to predict the likelihood of utility caused wildfires Unfortunately, the problem is more complex than drought-only. CNNs seemed to be able to handle the data well, but unable to detect any insights. Because of the limited data, it is difficult to know if the lack of results meant that we needed additional data or that the drought was not a significant enough factor. Issues affecting utilities tend to be multi-factorial, extremely noisy and strongly affected by human decisions which may change during a data collection timeline. Additional data would help to answer this question. However, at the moment, we believe that other avenues might be more fruitful than trying to gather additional data.

The next steps that we propose are as follow:

(1) Our next steps would have been to explore LSTMs as a way to capture temporal variations of the drought and see if this helped the predictions
(2) Treating the problem more as a segmentation than a classification, which was a flawed approach and would capture spatial data much better.
(3) Adding additional data, like soil moisture or wind.

Acknowledgements: There was only one team member, who did research, data collection and reporting.

References

[1] J. D. Morris, “PG&E: Massive power shut-off to hit 800,000 customers, could extend nearly a week,” SFChronicle.com, 09-Oct-2019. [Online]. Available: https://www.sfchronicle.com/california-wildfires/article/PG-E-power-shut-off-257-000-Bay-Area-residents-14500945.php. [Accessed: 07-Dec-2019].
[2] P. Stevens, “PG&E power outage could cost the California economy more than $2 billion,” CNBC, 10-Oct-2019. [Online]. Available: https://www.cnbc.com/2019/10/10/pge-power-outage-could-cost-the-california-economy-more-than-2-billion.html. [Accessed: 10-Oct-2019].
[3] “California utility equipment sparked more than 2,000 fires in over three years,” Los Angeles Times, 28-Jan-2019. [Online]. Available: https://www.latimes.com/politics/la-pol-ca-california-utilities-wildfires-regulators-20190128-story.html. [Accessed: 10-Oct-2019].
[4] Y. O. Sayad, H. Mousannif, and H. Al Moatassime, “Predictive modeling of wildfires: A new dataset and machine learning approach,” 2019.
[5] S. Lall and B. Mathibela, “The application of artificial neural networks for wildfire risk prediction,” in 2016 International Conference on Robotics and Automation for Humanitarian Applications (RAHA), 2016, pp. 1–6.
[6] J. L. Case, A. T. White, J. B. Wachter, C. J. Schultz, C. R. Hain, and K. D. White, “Applying Antecedent Land Surface Conditions and Machine Learning to Wildfire Events and Seasonal Burn Prediction,” p. 32.
[7] D. Tien Bui, H. V. Le, and N.-D. Hoang, “GIS-based spatial prediction of tropical forest fire danger using a new hybrid machine learning method,” Ecol. Inform., vol. 48, pp. 104–116, Nov. 2018.
[8] N. Read, T. J. Duff, and P. G. Taylor, “A lightning-caused wildfire ignition forecasting model for operational use,” Agric. For. Meteorol., vol. 253–254, pp. 233–246, May 2018.
[9] “California is no stranger to dry conditions, but the drought from 2011-2017 was exceptional | Drought.gov.” [Online]. Available: https://www.drought.gov/drought/california-no-stranger-dry-conditions-drought-2011-2017-was-exceptional. [Accessed: 07-Dec-2019].
[10] “What is the USDM | United States Drought Monitor.” [Online]. Available: https://droughtmonitor.unl.edu/About/WhatistheUSDM.aspx. [Accessed: 08-Dec-2019].
[11] “Map Archive | United States Drought Monitor.” [Online]. Available: https://droughtmonitor.unl.edu/Maps/MapArchive.aspx. [Accessed: 08-Dec-2019].
[12] F. Chollet, “Xception: Deep Learning with Depthwise Separable Convolutions,” ArXiv161002357 Cs, Apr. 2017.