Predicting Wildfire Severity with Machine Learning

Final Report

CSE 847
Machine Learning
Professor Jiayu Zhou
Michigan State University

Andrew McDonald
mcdon499@msu.edu

April 27, 2021

1 Introduction

Climate change is a complex challenge with no straightforward solution. Yet, machine learning (ML) offers a powerful set of tools with the potential to help society respond and adapt to its effects [1]. Techniques from computer vision, natural language processing, reinforcement learning, and other subfields of ML yield promising inroads towards climate change prevention and mitigation, spurring workshops at recent ML conferences[1] and receiving coverage in mainstream media.[2] With applications to electrical grid optimization, energy management, transportation, meteorology, forestry, oceanography and beyond, the demand for ML in the fight against climate change far outpaces the supply of experts in the field.

One particularly active area of research at the intersection of ML and climate change is wildfire science. A 2020 literature review located over 300 publications applying ML to wildfire science [2] and identified six problem domains addressed in the literature: (i) fuels characterization, fire detection, and mapping; (ii) fire weather and climate change; (iii) fire occurrence, susceptibility, and risk; (iv) fire behavior prediction; (v) fire effects; and (vi) fire management. In spite of the expansive body of existing research, the authors note that each domain still offers a number of open problems, and that relatively newer technologies from deep learning [3] and interpretability [4] have yet to be applied.

In this project, we focus on the fourth problem domain of fire behavior prediction, specifically aiming to predict the ultimate severity of a wildfire as measured by burned area, perimeter, and fire duration. We first address the problem with standard linear models, then extend our approach to consider nonlinear, ensemble, and neural network-based methods to account for the inherently complex and chaotic nature of wildfire spread. All code associated with this project, from data preprocessing to hyperparameter selection, cross-validation and model evaluation, is available online and in D2L.[3]

2 Problem Description

Given the month and location of a wildfire ignition, along with local geophysical, meteorological and climatological information associated with the ignition, we aim to predict three target variables characterizing the ultimate severity of the wildfire:

- **Size.** The total area (km^2) burned by a wildfire is perhaps the most natural measure of severity. Large wildfires pose a significant risk to people, property and ecosystems, release more carbon into the atmosphere, are more likely to result in economic damage, and require more resources to contain [5, 6]. Thus, a model capable of predicting the ultimate size of a wildfire at ignition-time would be of great value to firefighters, government officials and the public at large, empowering leaders to make informed decisions when planning fire response and mitigation efforts, evacuation procedures and recovery plans.

- **Perimeter.** The perimeter (km) of the land burned by a wildfire is another proxy of wildfire severity, adding contextual information not necessarily captured by size. Though correlated with size, the perimeter-to-area ratio of a wildfire allows one roughly estimate its shape and eccentricity, informing firefighting, risk mitigation, and evacuation strategies. Long-perimeter (elliptic) wildfires may be more difficult to contain than short-perimeter (circular) wildfires of the same size, and may threaten more linear infrastructure such as roads or electrical transmission lines. Hence, predicting the ultimate perimeter of a wildfire at ignition-time is valuable for the reasons outlined above.

- **Duration.** The duration (days) of a wildfire's lifetime relates to its severity in a nuanced way—on one hand, slow-burning wildfires may pose less immediate danger to communities than fast-moving wildfires—yet on the other hand, long-lasting wildfires may cause a larger adverse impact on local air quality [7] and economics [8]. Thus, predicting the ultimate duration of a wildfire at ignition-time is valuable for the same reasons previously discussed.

[1] e.g., the Tackling Climate Change with Machine Learning and AI for Earth Sciences workshops at NeurIPS 2020.
[2] e.g., National Geographic, Forbes and The Verge, to name a few.
[3] https://github.com/andrewmcdonald27/CSE847Public

Traditional wildfire growth models are generally physics-based [9], simulating the progression of complex, nonlinear convective processes with differential equations and finite-element methods. A significant amount of research has gone into the development of today's state-of-the-art physical models [10], with the FARSITE simulator [11] in particular being the most popular among wildfire scientists. While powerful, such physical models are not perfect: they become computationally expensive at fine resolutions, require an expert to fine-tune parameters, are sensitive to initial conditions, and pose a steep learning curve to new users. As such, we see opportunity for a data-driven approach in which the relationship between total size, perimeter, and duration of a wildfire may be learned from historical examples, supplemented by meteorological and climatic information.

Formally, let $\mathcal{R} \subset \mathbb{R}^2$ be an area of land in which one or more wildfires occur, and suppose wildfires are uniquely indexed by the natural numbers $i = 1, 2, \ldots$. Let $\text{burn}_i : \mathcal{R} \to \{0, 1\}$ be an indicator function which takes the value 1 at points $\mathbf{r} \in \mathcal{R}$ that were burned by fire i, and 0 elsewhere, i.e.,

$$\text{burn}_i(\mathbf{r}) = \begin{cases} 1 & \mathbf{r} \text{ was burned by wildfire } i \\ 0 & \text{else.} \end{cases}$$

The total area A_i burned by fire i is given by the double integral

$$A_i = \iint_{\mathcal{R}} \text{burn}_i(\mathbf{r}) dA,$$

while the perimeter P_i burned by fire i is given by the line integral

$$P_i = \int_{\partial\{\text{burn}_i(\mathbf{r})=1\}} 1 ds,$$

where $\partial\{\text{burn}_i(\mathbf{r}) = 1\}$ is the boundary of the region burned by fire i. Further, we may define the map $\text{alive}_i : [0, T] \to \{0, 1\}$ to be an indicator function defined over the time window $[0, T]$ which takes the value 1 at all $t \in [0, T]$ for which the fire is actively burning, and 0 at other times. Then the total duration D_i of the i-th wildfire is given by

$$D_i = \int_0^T \text{alive}_i(t) dt.$$

In practice, we assume that a wildfire has only a single lifetime, hence the duration of the i-th wildfire reduces to $D_i = t_{\text{end}_i} - t_{\text{start}_i}$. Moreover, the integrals for A_i and P_i are computed over a discrete grid due to the rasterized storage of geospatial data.

We assume a vector $\mathbf{x} \in \mathbb{R}^d$ is associated to each wildfire such that $\mathbf{x}_i$ contains information about the ignition of fire i. This vector may include, for example, (i) the month of ignition, (ii) the location of ignition, and (iii) local geophysical (e.g., terrain type, land cover), meteorological (e.g., instantaneous temperature, wind, precipitation) and climatological (e.g. 30-year-average temperature, wind, precipitation) information relevant to wildfire dynamics. It may even include information from human experts, or information from physics-based forecasts.

Our goal is to construct three models, $f_A,\ f_P,\ f_D : \mathbb{R}^d \to \mathbb{R}$ where each model maps ignition information to a prediction of total area burned (f_A), perimeter of total burn (f_P) and duration of burn (f_D) with a good degree of accuracy. That is, given $i = 1, \ldots, N$ wildfires, we hope to construct models $f_A^*,\ f_P^*,\ f_D^*$ for which

$$f_A^* = \min_{f_A} \sum_{i=1}^N \mathcal{L}_A(A_i, f_A(\mathbf{x}_i)), \quad f_P^* = \min_{f_P} \sum_{i=1}^N \mathcal{L}_P(P_i, f_P(\mathbf{x}_i)), \quad f_D^* = \min_{f_D} \sum_{i=1}^N \mathcal{L}_D(D_i, f_D(\mathbf{x}_i)),$$

where $\mathcal{L}_A,\ \mathcal{L}_P,\ \mathcal{L}_D$ are suitably-chosen regression loss functions scoring predictions against ground truths. For example, the functions

$$\mathcal{L}_{\text{MSE}}(y, \hat{y}) = (y - \hat{y})^2, \quad \mathcal{L}_{\text{MAE}}(y, \hat{y}) = |y - \hat{y}|, \quad \mathcal{L}_{\text{Pearson}} = -\rho(y, \hat{y}) = \frac{-\text{cov}[y, \hat{y}]}{\sqrt{\text{var}[y]}\sqrt{\text{var}[\hat{y}]}}.$$

would all make for suitable losses, rewarding different model objectives.

In this project, we consider mean-squared losses

$$\mathcal{L}_A = (A_i - f_A(\mathbf{x}_i))^2, \quad \mathcal{L}_P = (P_i - f_P(\mathbf{x}_i))^2, \quad \mathcal{L}_D = (D_i - f_D(\mathbf{x}_i))^2$$

for two key reasons. First, we seek a loss function that penalizes the extreme cases more heavily—a large-sized, long-perimeter, long-lasting wildfire poses a much graver threat than a small, brief fire—so the fact that mean-squared loss penalizes outliers disproportionately is an advantage in this setting. Second, and more practically, mean-squared losses are easier to optimize against, and are inherently baked into the assumptions of many popular machine learning models (the most common being ordinary least-squares linear regression).

3 Data

3.1 Wildfire Data

We leverage the Global Fire Atlas [12] as the central dataset in this project. Offering high-resolution GIS data from over 13.3 million individual wildfires observed between 2003-2016 via NASA MODIS[4] imagery from the Terra and Aqua satellites, the Global Fire Atlas provides a rich characterization of wildfire dynamics on Earth in a form conducive to machine learning. In this project, we limit our consideration to the 11,524 unique wildfires observed in the United States in 2016, but emphasize that the vast volume of data in the Global Fire Atlas makes it a ripe venue for future research.

Data in the Global Fire Atlas is separated by year, and comprises a combination of vectorized and rasterized GIS data. Vectorized data characterizing the point of ignition and polygonal boundaries of burned area for each wildfire in a single year are stored in a shapefile (.shp) format, and rasterized data characterizing the direction, speed, and day-by-day extent of wildfire spread are stored in GeoTIFF (.tif) format.[5] Fires are uniquely identified with an ID number such that information from the vectorized (.shp) and rasterized (.tif) file formats may be linked and used in tandem.

The vectorized (.shp) data allow for a more straightforward application of standard machine learning techniques, as it is packaged in rectangular form with the attributes presented in Table 1. On the other hand, the rasterized (.tif) data provide a useful means for visualization of fire dynamics, packaging 500m-resolution pixels in four separate files measuring (i) the day of year in which a pixel burned, (ii) the primary direction of wildfire spread within a pixel, (iii) the average speed of wildfire spread within a pixel, and (iv) the day of year in which a pixel was on the fire line (wildfire perimeter). A sample of the rasterized and vectorized data associated to the 2016 Soberanes Fire[6] south of Monterey, CA is presented in Figure 1 and Table 2, respectively. Additional information about the dataset can be found in the user guide[7] associated with the publication [12].

3.2 Climate Data

In addition to the predictors included in the Global Fire Atlas dataset presented in Table 1, we seek climate-related predictors to include in the feature vector $\mathbf{x}_i$ associated to each wildfire. Though one might argue latitude, longitude, and landcover are, to some degree, proxies of climate from the model's perspective, it is reasonable to assume that incorporating a higher-dimensional representation of climate including 30-year averages of temperature, humidity, precipitation, etc., will ultimately benefit the model's performance. Further, incorporating present month-scale anomalies from 30-year averages (e.g., the current precipitation deficit with respect to the climatic average) will improve model performance.

[4] https://modis.gsfc.nasa.gov/

[5] Shapefile data is referred to as being vectorized in the GIS community because it is associated to a specific point (vector) in Earth's coordinate space, rather than a pixel on a gridded representation of Earth.

[6] https://en.wikipedia.org/wiki/Soberanes_Fire

[7] https://glihtdata.gsfc.nasa.gov/files/fire_atlas/Fire_Atlas_user_guide.pdf

Attribute	Type	Description
fire_ID	int	Unique identifier of each observed wildfire
latitude	float	Latitude of ignition location
longitude	float	Longitude of ignition location
size	float	Total area burned (km^2)
perimeter	float	Final fire perimeter length (km)
start_date	date	Date of fire ignition
start_DOY	int	Day of year of fire ignition
end_date	date	Date of fire extinction
end_DOY	int	Day of year of fire extinction
duration	int	Number of days between ignition and extinction
expansion	float	Average fire growth rate (km^2/day)
fire_line	float	Average fire perimeter length (km)
speed	float	Average rate of fire perimeter movement (km/day)
direction	int	Primary direction of fire spread (encoded 0-8)
direction_s	string	Primary direction of fire spread (string)
landcover	int	Primary landcover burned by fire (encoded 0-16)
landcover_s	string	Primary landcover burned by fire (string)
tile_ID	string	Unique MODIS tile identifier in which fire occurred
geometry	shape	Point/Polygon characterizing ignition location/spread extent

Table 1: An overview of the attributes available in rectangular shapefile form from [12]. Our ultimate goal is to predict wildfire size A_i, perimeter P_i, and duration D_i given information that would be available at the time of ignition of a wildfire, including latitude, longitude, start_DOY, landcover_s and early estimates of speed, expansion and direction.

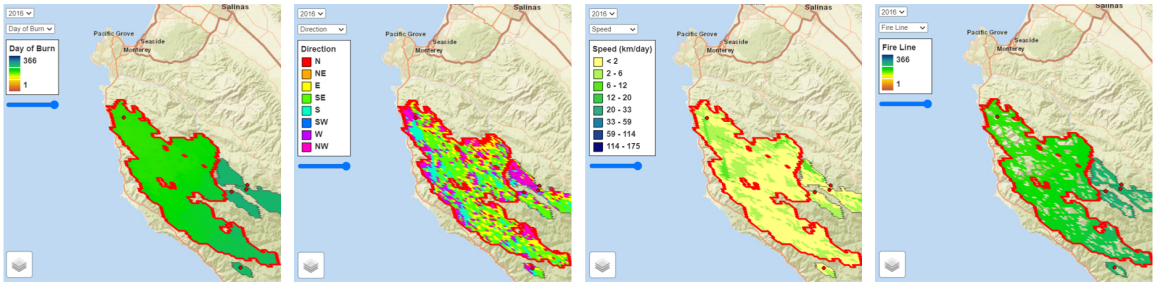

Figure 1: Sample GeoTIFF data from [12] associated to the 2016 Soberanes Fire south of Monterey, CA. Visualizing the 500m-pixel-level (i) day of burn, (ii) primary direction of spread, (iii) average speed of spread and (iv) day of fire line builds intuition behind wildfire dynamics, but does not immediately lend itself to machine learning methods. Significant feature transformation and feature engineering, or perhaps spatio-convolutional representation learning would be necessary for this data to be useful in a machine learning model. Rectangular data better suited to standard machine learning techniques are presented in Table 2.

fire_ID	latitude	longitude	size	perimeter	start_date	start_DOY	end_date	end_DOY	duration
107	36.4604	-121.894	447.39	196.31	7/21/2016	203	9/17/2016	261	59

expansion	fire_line	speed	direction	direction_s	landcover	landcover_s	tile_ID	geometry
7.58	12.96	1.89	4	southeast	8	Woody savannas	h08v05	PT(-121.89 36.46)

Table 2: Sample shapefile data from [12] associated to the 2016 Soberanes Fire south of Monterey, CA. These attributes represent the average values over the spatial and temporal extent of the wildfire, as visualized in Figure 1.

Such predictors are available from the NOAA National Climate Data Center (NCDC)[8] programmatically through the Climate Data Online (CDO) API.[9] Specifically, thirty-year (1981-2010) monthly climate normals are accessible via the API from the `NORMAL_MLY` dataset,[10] while present monthly data is accessible via the API from the Global Summary of the Month, or `GSOM`, dataset.[11] Several hundred features are available in each dataset, though many contain incomplete data coverage over space and time—after dropping variables with sparse spatiotemporal coverage in preprocessing, a total of 77 features from `NORMAL_MLY` and 27 features from `GSOM` remained viable for use as predictors of wildfire severity. A representative subset of these 107 climate-related variables is included in Table 3 to give a sense of the breadth of information contained within.

To associate this wealth of 30-year monthly normal and present monthly anomaly data to each wildfire in the Global Fire Atlas dataset, a significant amount of feature engineering is necessary. First, we locate all weather stations within the NOAA network containing (nearly) full records of the features we are interested in obtaining. Unfortunately, the weather stations which report `NORMAL_MLY` and `GSOM` datasets are exclusive from one another, so we must locate stations separately for each dataset. Figure 2 shows the geographic distribution of stations reporting to the `NORMAL_MLY` and `GSOM` datasets in North America, operational in 2016. Because we consider only wildfires in the United States in 2016 in this project, these stations suffice to provide necessary climate data.

Dataset	Attribute	Type	Description
GSOM	TAVG	float	Average temperature in Celsius over a month
GSOM	DX70	float	# days with $t_{\max} > 70$F/21.1C over a month
GSOM	DX90	float	# days with $t_{\max} > 90$F/32.2C over a month
GSOM	PRCP	float	Total precip in millimeters over a month
GSOM	DP01	float	# days with at least 0.1in/2.54mm precip over a month
GSOM	DP10	float	# days with at least 1.0in/25.4mm precip over a month
NORMAL_MLY	MLY-TAVG-NORMAL	float	Long-term average of monthly average temperature
NORMAL_MLY	MLY-TMAX-AVGNDS-GRTH070	float	Long-term avg # d/mo where $t_{\max} > 70$F
NORMAL_MLY	MLY-TMAX-AVGNDS-GRTH090	float	Long-term avg # d/mo where $t_{\max} > 90$F
NORMAL_MLY	MLY-PRCP-50PCTL	float	50th percentile of monthly precip totals
NORMAL_MLY	MLY-PRCP-AVGNDS-GE010HI	float	Long-term avg # d/mo with precip > 0.1in/2.54mm
NORMAL_MLY	MLY-PRCP-AVGNDS-GE100HI	float	Long-term avg # d/mo with precip > 1.0in/25.4mm

Table 3: A representative subset of the climate-related features pulled from the NOAA NCDC API. Given our goal of predicting wildfire size A_i, perimeter P_i, and duration D_i, it stands to reason that incorporating present monthly weather (`GSOM`) and 30-year normals (`NORMAL_MLY`) will improve model performance.

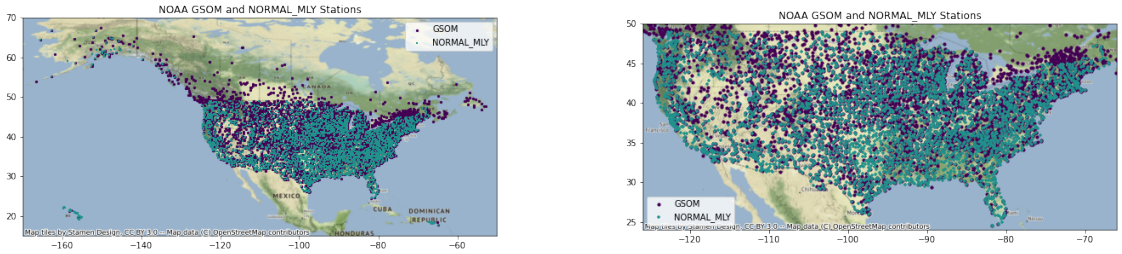

Figure 2: All operational NOAA weather stations reporting to the `NORMAL_MLY` and `GSOM` datasets in North America, operational in 2016. Data from approximately 1000 different stations in each dataset was used to create a 107-dimensional climate embedding associated to each wildfire.

[8] https://www.ncdc.noaa.gov/
[9] https://www.ncdc.noaa.gov/cdo-web/webservices/v2
[10] https://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/climate-normals
[11] https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ncdc:C00946

We construct a KD-tree using the latitude/longitude coordinates of each weather station, and query this KD-tree once for each wildfire in our dataset to determine the nearest `NORMAL_MLY` and `GSOM` station (which are usually different), and associate these stations to the wildfire. We then construct a list of all unique `NORMAL_MLY` and `GSOM` stations from which we must pull data ($\sim$1000 stations in each dataset), and query the NOAA CDO API once for each station to obtain the feature values for all 107 climate-related variables over the 12 months of 2016. Finally, for each wildfire in our dataset, we locate these feature values at the nearest station for the month in which the wildfire ignited, in effect obtaining a 107-dimensional embedding of the wildfire in climate-space. While somewhat complex, if this wildfire prediction model were to be deployed in practice, a pipeline could be configured to carry out this preprocessing from end-to-end in a fully-automated manner.

3.3 Preprocessing

To encourage reproducibility of our work and foster future research in machine learning-driven wildfire prediction, we summarize the preprocessing steps used on the Global Fire Atlas and NOAA NCDC data below:

- Global Fire Atlas data in the `2016.zip` file was downloaded from the NASA Goddard Space Flight Center host.[12] Next, the `Global_fire_atlas_V1_ignitions_2016.shp` file contained within was filtered to consider only wildfires originating in the USA using the `geopandas.overlay` function, leaving 11,524 unique fires. Shapefile geometry associated to each fire was dropped, leaving latitude and longitude, and the remaining features summarized in Table 1 were exported to a `.csv` for easier manipulation.

- NOAA NCDC data was joined to each fire using the process described previously, matching fires with features by the nearest station's data for the month in which the ignition occurred. Distance from the wildfire's point of ignition to the nearest NOAA station is computed using the `geopy.distance.geodesic` function and included as a feature, allowing models to determine the relevance of included climate information (climate data from stations close to the point of ignition is more relevant to prediction than data from distant stations).

- Fires for which incomplete NOAA NCDC data was available were dropped from the dataset, leaving 9,925 fires. Target columns (size, perimeter and duration) are saved into a separate `y.csv` file, then dropped. It is worth noting that although speed (average rate of fire perimeter movement, km/day), expansion (average rate of fire growth, km^2/day), fire line (average perimeter length over the course of the fire, km) and direction (predominant one-of-eight cardinal direction of spread) are not fully determined until the extinction of a wildfire, we include them as predictors in our analysis, as initial estimates of each would be available shortly after wildfire ignition.

- Redundant temporal information associated to each fire is dropped, leaving only the month of ignition in 2016 as a temporal feature. This is then featurized using a sinusoidal embedding to encourage similarity between December and January, mapping the month m to $\cos(2\pi m/12)$ and $\sin(2\pi m/12)$. Features from the NCDC `GSOM` dataset measuring the day of month in which an extreme event occurred (e.g., maximum precipitation of the month, high temperature of the month, etc.) are converted from datetimes to integers.

- Landcover and direction of spread from the Global Fire Atlas dataset in string form is one-hot-encoded to vectors of length 13 and 7, respectively, using the `pandas.get_dummies` function. All numeric (i.e., non-one-hot-encoded) features are standardized by subtracting the feature mean and dividing by the feature standard deviation, using the `sklearn.StandardScaler` class. The resulting 140 predictors are saved to an `X.csv` file, concluding the preprocessing pipeline.

[12] https://glihtdata.gsfc.nasa.gov/files/fire_atlas/

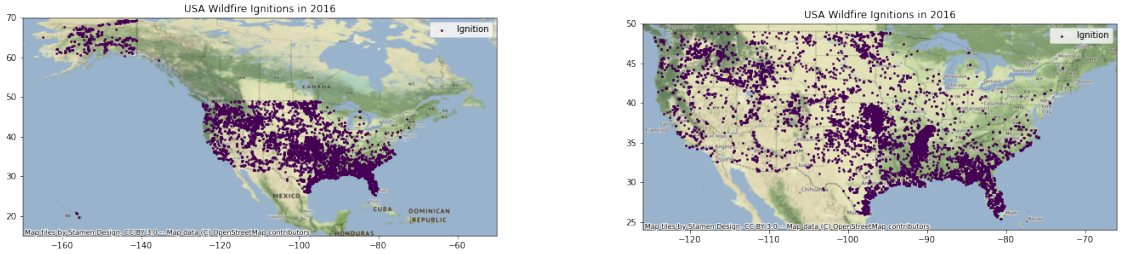

Figure 3: Geographic distribution of the 9,925 wildfires from the Global Fire Atlas occurring in the United States in 2016 which remain after preprocessing. Note that points represent locations of ignition, and do not convey significance with respect to final size, perimeter, or duration.

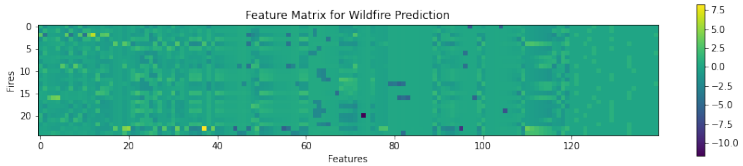

Figure 4: A random sample of 25 rows from the 9,925 $\times$ 140 dimensional $\mathbf{X}$ matrix following preprocessing. Columns correspond to features, and rows correspond to instances of fires. The diversity of feature values in the matrix is encouraging, and suggests a sufficient amount of variation in our features is available for models to leverage. Columns 0-4 correspond to fire metadata from the Global Fire Atlas dataset; Columns 5-12 correspond to weather station metadata; Columns 13-39 correspond to present monthly anomalies from the NOAA `GSOM` dataset; Columns 40-116 correspond to 30-year climate normals from the NOAA `NORMAL_MLY` dataset; Columns 117-119 correspond to temporal information of the ignition; Columns 120-126 correspond to one-hot-encoded fire spread direction information, and Columns 127-139 correspond to one-hot-encoded landcover information.

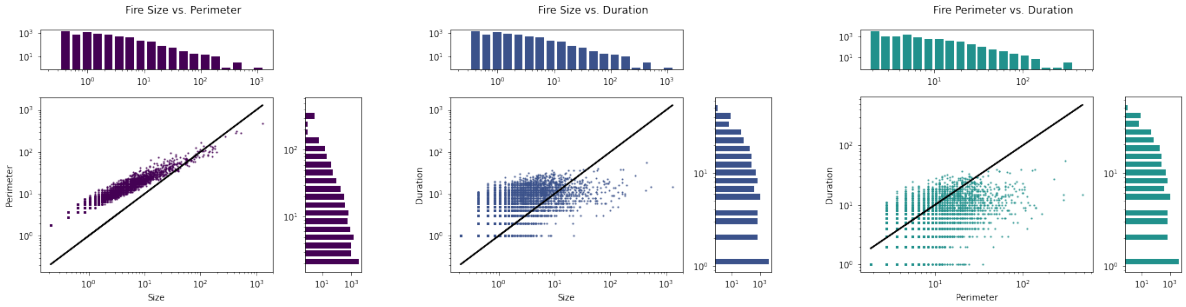

Figure 5: Pairwise log-scaled joint distributions of the three target variables considered in this project: wildfire size A_i, wildfire perimeter P_i, and wildfire duration D_i. While correlated, it is clear that a model suitable for predicting one target may not suffice to predict other targets. Thus, we construct a model for each target separately. We encourage future research which frames the prediction of the joint distribution as a multitask learning problem.

The geographic distribution of the 9,925 wildfires in the final dataset is visualized in Figure 3. Moreover, a sample of rows from the complete data matrix $\mathbf{X}$ is provided in Figure 4, and conforms to our expectations—there do not seem to be any anomalies following preprocessing and standardization. Pairwise log-scaled joint

distributions of the three target variables (area, perimeter and duration) are presented in Figure 5. Though the three targets are correlated, the figures justify our motivation to construct a model for each target individually—being able to predict one variable does not necessarily guarantee a model's ability to predict others. Ultimately, the problem may be best framed from the perspective of multi-task learning, but for now, we treat each target variable individually.

4 Models

In this section, we review the mechanics behind a number of popular machine learning models. In the following section, we compare the performance of these models on the separate tasks of predicting wildfire size A_i, wildfire perimeter P_i, and wildfire duration D_i. For the sake of generality in this section, let $\mathbf{x}_i \in \mathbb{R}^d$ denote a feature vector, $\mathbf{X} \in \mathbb{R}^{N \times d}$ denote a design matrix of N feature row vectors, $y_i \in \mathbb{R}$ denote a scalar target value, and $\mathbf{y} \in \mathbb{R}^N$ denote a target vector with entries y_i. Further, let $\hat{y}_i \in \mathbb{R}$ and $\hat{\mathbf{y}} \in \mathbb{R}^N$ denote a scalar prediction and vector of predictions, respectively.

4.1 Linear Models

Linear Regression. Perhaps the simplest and most popular regression model used in machine learning, least-squares linear regression [13] aims to find a weight vector $\mathbf{w} \in \mathbb{R}^d$ for which the linear combination $\mathbf{x}_i^\top \mathbf{w}$ is close to y_i. In matrix-vector form, least-squares linear regression solves the optimization problem

$$\mathbf{w}_* = \min_{\mathbf{w}} \frac{1}{N} ||\mathbf{X}\mathbf{w} - \mathbf{y}||_2^2$$

and predicts the target vector $\hat{\mathbf{y}} = \mathbf{X}\mathbf{w}_*$. In our experiments, we utilize the `sklearn.linear_model.LinearRegression` implementation from Scikit-Learn [14].

Ridge Regression. Designed to address the tendency of least-squares linear regression to overfit a training set, ridge regression [13] aims to find a weight vector $\mathbf{w} \in \mathbb{R}^d$ which solves the optimization problem

$$\mathbf{w}_* = \min_{\mathbf{w}} \frac{1}{N} ||\mathbf{X}\mathbf{w} - \mathbf{y}||_2^2 + \frac{\alpha}{2} ||\mathbf{w}||_2^2 .$$

The additional $\frac{\alpha}{2} ||\mathbf{w}||_2^2$ regularization term penalizes large weights in $\mathbf{w}$, and often improves a model's ability to generalize in practice. We use the `sklearn.linear_model.Ridge` implementation from Scikit-Learn [14].

Lasso Regression. Lasso regression [13] adds a 1-norm penalty term to the loss of least-squares regression, aiming to regularize $\mathbf{w}$ and prevent overfitting while simultaneously encouraging sparsity in the components of $\mathbf{w}$. Specifically, lasso regression aims to find a weight vector $\mathbf{w} \in \mathbb{R}^d$ which solves the optimization problem

$$\mathbf{w}_* = \min_{\mathbf{w}} \frac{1}{N} ||\mathbf{X}\mathbf{w} - \mathbf{y}||_2^2 + \alpha ||\mathbf{w}||_1 .$$

We use the `sklearn.linear_model.Lasso` implementation in our experiments.

ElasticNet. First proposed in [15], the ElasticNet combines ridge and lasso regularization penalties to encourage sparsity in a more stable manner. Whereas ridge regression fails to achieve sparsity and lasso regression is susceptible to instability and saturation in high-dimensional data, ElasticNet finds a happy medium, converging to a weight vector solving the optimization problem

$$\mathbf{w}_* = \min_{\mathbf{w}} \frac{1}{N} ||\mathbf{X}\mathbf{w} - \mathbf{y}||_2^2 + \frac{\alpha}{2} ||\mathbf{w}||_2^2 + \beta ||\mathbf{w}||_1 .$$

In the `sklearn.linear_model.ElasticNet` implementation, a slightly different parameterization is provided: the user specifies an overall regularization weight α and an `l1_ratio` parameter which determines β.

4.2 Nonlinear Models

Kernel Ridge Regression. Kernel ridge regression [13] can be viewed as a generalization of feature engineering for linear regression: a feature vector $\mathbf{x}_i$ is mapped into a high-dimensional space, where it has coordinates $\phi(\mathbf{x}_i)$, and a linear model is fit in this high-dimensional space. It is not necessary to actually compute this high-dimensional projection in order to make predictions, however, so long as we may compute the inner product $\phi(\mathbf{x}_i)^\top \phi(\mathbf{x}_j)$ in the high-dimensional space. Hence, we may formulate kernel ridge regression using a kernel function $k(\mathbf{x}_i, \mathbf{x}_j)$ which measures the similarity of vectors and gives a closed form for their inner product in the high-dimensional space. Letting $\mathbf{K} = [k_{ij}]$ be the $N \times N$ matrix with entry $k_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ measuring the similarity between training vectors i and j, and $\mathbf{k}$ be the $N \times 1$ vector measuring the similarity between a test point $\mathbf{x}_*$ and all N training points, kernel ridge regression predicts

$$\hat{y} = \mathbf{k}^\top [\mathbf{K} + \lambda I]^{-1} \mathbf{y}.$$

In our experiments, we leverage the `sklearn.kernel_ridge.KernelRidge` implementation with linear and Gaussian kernels, respectively given by

$$k_{\text{Linear}}(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^\top \mathbf{x}_j, \quad k_{\text{Gaussian}}(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \, ||\mathbf{x}_i - \mathbf{x}_j||_2^2\right).$$

Random Forest Regression. Random forest regression [16] incorporates stochasticity in an ensembled framework by constructing bootstrap samples of the training dataset, dropping a random subset of features from each bootstrap, and fitting a decision tree regressor on each bootstrap. Splits at each step in the decision tree construction process are made to minimize the residual sum of squares between left and right children L and R:

$$RSS = \sum_{i \in L} (y_i - \overline{y}_L)^2 + \sum_{i \in R} (y_i - \overline{y}_R)^2.$$

Of interest are the hyperparameters governing the forest construction process. In the `sklearn.ensemble.RandomForestRegressor` implementation, we tune `max_features` $\in [0, 1]$ and `max_samples` $\in [0, 1]$, which control the proportion of features kept in each tree and the proportion of observations used to construct bootstraps, respectively. We leave the default setting of `n_estimators`$= 100$ alone, finding it to perform well in practice.

Gradient Boosting Regression Trees. Gradient-boosting regression trees [16] fit a sequence of models which sequentially correct the errors of earlier models. At each step $t = 1, \ldots, T$, a base model b parameterized by γ is fit to a weighted set of observations $w_i \mathbf{x}_i$, and a cumulative model f_t is updated via the rule

$$f_t(\mathbf{x}) = f_{t-1}(\mathbf{x}) + \eta \beta_t b(\mathbf{x}, \gamma_t)$$

where β_t and γ_t are parameters chosen to minimize the error of the prediction by the model f_t, and η is a learning rate parameter. We leverage the `sklearn.ensemble.RandomForestRegressor` implementation, and tune the learning rate η, number of rounds of boosting `n_estimators`, the proportion `max_features` $\in [0, 1]$ controlling features kept when fitting base models, and the proportion `subsample` $\in [0, 1]$ controlling observations used when fitting base models.

Multilayer Perceptron. Multilayer perceptrons [3] are simple feedforward, fully-connected neural networks consisting of layers of neurons which linearly combine inputs from the previous layer and apply a nonlinear activation function, then pass a scalar forward to the next layer of neurons. Specifically, if k denotes the number of neurons per layer, h_i^ℓ denotes the output of the i-th neuron at layer ℓ, the output of h_i^ℓ is given by

$$h_i^\ell = \sigma\left(\sum_{i=1}^{k} w_i h_i^{\ell-1} + b\right),$$

where σ is a suitably-chosen nonlinear activation function (e.g., relu, sigmoid, tanh, etc.) and w_i, b are parameters trained by backpropagation. Inputs to the first layer h_i^0 are simply components of an input vector $\mathbf{x}$, and in the case of univariate regression, the last layer consists of a single neuron whose output is taken to be $\hat{y}$. Notably, multilayer perceptrons of arbitrary size are able to approximate any function [17]—but in practice, they can be difficult to design and optimize.

In our experiments, we leverage the `sklearn.neural_network.MLPRegressor` implementation, and cross-validate against the number of layers and depth of network. We hope to extend our work to leverage PyTorch in the future, and believe more sophisticated deep learning architectures will outperform the approaches studied in this project.

5 Results

All models described in the previous section were evaluated on a fixed 70-30% train-test split of the 9,925-fire dataset, trained with mean-squared error loss for reasons described in Section 2. Where relevant, hyperparameters were optimized by applying 3-fold cross-validation on the training set and selecting the value for which average mean-squared error across validation sets was lowest. Models were evaluated with respect to each target variable—wildfire size A_i, wildfire perimeter P_i, and wildfire duration D_i—completely independently, though we encourage future research into the joint prediction of all three targets as a multitask learning problem.

To compare models, we report their mean-squared error $\mathcal{L}_{\text{MSE}}$ on the test set—and to gain additional insight into the relative strengths and weaknesses of each model, we report the mean-absolute error $\mathcal{L}_{\text{MAE}}$ and the Pearson correlation loss[13] between predicted values $\hat{y}$ and true values y in the test set, where

$$\mathcal{L}_{\text{MSE}}(y, \hat{y}) = (y - \hat{y})^2, \quad \mathcal{L}_{\text{MAE}}(y, \hat{y}) = |y - \hat{y}|, \quad \mathcal{L}_{\text{Pearson}} = -\rho(y, \hat{y}) = \frac{-\text{cov}[y, \hat{y}]}{\sqrt{\text{var}[y]}\sqrt{\text{var}[\hat{y}]}}.$$

The performance of each model on the task of predicting size, perimeter and duration with respect to these losses is summarized in Tables 4, 5 and 6, respectively. Hyperparameters chosen via cross-validation are also included, where relevant.

In general, we see that the spatiotemporal, meterological and climatic predictors in $\mathbf{X}$ provide information relevant to the task at hand, and are able to construct models for A_i, P_i and D_i which would be sufficiently accurate for real-world use by firefighters, policymakers and the public at large. It remains to be seen how the accuracy of our models compares to that of state-of-the-art physics-based models, and while such a comparison is beyond the scope of this project, we encourage future research in this direction. That said, we emphasize that the data-driven nature of our model offers inherent advantages over a physics-based model, particularly in terms of computational complexity, accessibility and scalability.

High-capacity nonlinear models tended to outperform simple linear models and were not prone to overfitting, suggesting this dataset is ripe for the application of deep learning. Although the multilayer perceptron did not perform as well as the random forest and gradient boosting tree models, we suspect that further improvements to network architecture and tuning (e.g., dropout, skip-connections, spatiotemporal convolutions, pre-training, longer training time) would ultimately propel a deep learning approach to achieve superior performance.

6 Conclusions and Future Work

As machine learning researchers and practitioners, it is our responsibility to unite with the scientific community at large in the fight against climate change. Hotter, drier summers paired with more frequent severe

[13] We consider the negative Pearson correlation for consistency with other losses, i.e., lower is better.

	$\mathcal{L}_{\text{MSE}}$	$\mathcal{L}_{\text{MAE}}$	$\mathcal{L}_{\text{Pearson}}$	Hyperparameters
LinearRegression	35.6151	1.8845	-0.7727	NA
Ridge	35.4597	1.8251	-0.7738	$\alpha = 2.0236$
Lasso	35.4698	1.8133	-0.7736	$\alpha = 1e\text{-}3$
ElasticNet	34.5095	1.6922	-0.7796	$\alpha = 1e\text{-}2$, `l1_ratio` $= 0.8$
KernelRidgeLinear	35.5137	1.8305	-0.7734	NA
KernelRidgeRBF	27.0661	2.0285	-0.8087	$\gamma = 1e\text{-}4$
RandomForestRegressor	21.7992	1.0731	-0.8414	`max_features` $= 0.4$, `max_samples` $= 0.4$
GradientBoostingRegressor	**20.0036**	**0.974**	$\mathbf{-0.8514}$	$\eta = 1e\text{-}1$, `n_estimators` $= 250$, `max_features` $= 0.8$, `subsample` $= 0.8$
MLPRegressor	24.3134	1.4547	-0.8186	$\eta = 3e\text{-}4$, `activation` $=$ `"relu"`, `solver` $=$ `"adam"`, `hidden_layer_width` $= 256$, `n_hidden_layers` $= 4$, `batch_size` $= 256$, `max_iter` $= 200$

Table 4: Performance of various models f_A^* trained to predict wildfire size A_i given $\mathbf{x}_i$, a vector of spatiotemporal, meterological and climatic predictors associated to fire i. Data involving 9,925 fires were divided using a 70-30% train-test split. Hyperparameters were selected with 3-fold CV on the training set. Models were trained on the training set, then evaluated on the test set. Lower losses are better.

	$\mathcal{L}_{\text{MSE}}$	$\mathcal{L}_{\text{MAE}}$	$\mathcal{L}_{\text{Pearson}}$	Hyperparameters
LinearRegression	26.2715	2.3908	-0.8551	NA
Ridge	25.7818	2.2213	-0.8575	$\alpha = 356$
Lasso	26.2002	2.0643	-0.855	$\alpha = 1e\text{-}1$
ElasticNet	25.8696	2.0798	-0.8569	$\alpha = 1e\text{-}1$, `l1_ratio` $= 0.6$
KernelRidgeLinear	26.1789	2.3635	-0.8556	NA
KernelRidgeRBF	29.6917	2.6639	-0.845	$\gamma = 1e\text{-}4$
RandomForestRegressor	**19.2019**	**1.6615**	$\mathbf{-0.8964}$	`max_features` $= 0.4$, `max_samples` $= 0.6$
GradientBoostingRegressor	22.2639	1.8045	-0.8828	$\eta = 1e\text{-}1$, `n_estimators` $= 250$, `max_features` $= 0.6$, `subsample` $= 0.6$
MLPRegressor	20.8388	2.0485	-0.8875	$\eta = 3e\text{-}4$, `activation` $=$ `"relu"`, `solver` $=$ `"adam"`, `hidden_layer_width` $= 256$, `n_hidden_layers` $= 8$, `batch_size` $= 256$, `max_iter` $= 200$

Table 5: Performance of various models f_P^* trained to predict wildfire perimeter P_i given $\mathbf{x}_i$, a vector of spatiotemporal, meterological and climatic predictors associated to fire i. Data involving 9,925 fires were divided using a 70-30% train-test split. Hyperparameters were selected with 3-fold CV on the training set. Models were trained on the training set, then evaluated on the test set. Lower losses are better.

	$\mathcal{L}_{\text{MSE}}$	$\mathcal{L}_{\text{MAE}}$	$\mathcal{L}_{\text{Pearson}}$	Hyperparameters
LinearRegression	12.5785	2.2078	-0.6956	NA
Ridge	12.4835	2.1738	-0.6992	$\alpha = 54$
Lasso	12.4699	2.1494	-0.6994	$\alpha = 1e\text{-}2$
ElasticNet	12.455	2.152	-0.7	$\alpha = 1e\text{-}2$, `l1_ratio` $= 0.8$
KernelRidgeLinear	12.4898	2.1944	-0.6984	NA
KernelRidgeRBF	11.9549	2.1366	-0.7186	$\gamma = 3.2e\text{-}3$
RandomForestRegressor	**7.6284**	**1.3594**	$\mathbf{-0.8295}$	`max_features` $= 0.4$, `max_samples` $= 0.8$
GradientBoostingRegressor	7.7322	1.4623	-0.8271	$\eta = 1e\text{-}1$, `n_estimators` $= 250$, `max_features` $= 0.6$, `subsample` $= 0.8$
MLPRegressor	11.2239	1.9289	-0.741	$\eta = 3e\text{-}4$, `activation` $=$ `"relu"`, `solver` $=$ `"adam"`, `hidden_layer_width` $= 64$, `n_hidden_layers` $= 2$, `batch_size` $= 256$, `max_iter` $= 200$

Table 6: Performance of various models f_D^* trained to predict wildfire duration D_i given $\mathbf{x}_i$, a vector of spatiotemporal, meterological and climatic predictors associated to fire i. Data involving 9,925 fires were divided using a 70-30% train-test split. Hyperparameters were selected with 3-fold CV on the training set. Models were trained on the training set, then evaluated on the test set. Lower losses are better.

storms and continued urban sprawl at the metropolitan-wildland boundary have increased the frequency and severity of wildfires in the United States and beyond [18, 19], bringing an increased risk to the communities, ecosystems and economies that make up our world.

Informed by this context, the ability to forecast a wildfire's severity at the time of ignition—as characterized by its ultimate size, perimeter and duration—delivers clear value to firefighters, policymakers, and the public at large, providing an incredible opportunity for machine learning to save lives, money, and stress. Whereas traditional physics- and rule-based models may fail, rely on precise hyperparameter settings, require expert insight, or operate at a prohibitive computational cost, data-driven techniques offer a promising alternative. Accessible, scalable, automatic, and computationally cheap in comparison, we encourage further research into their application across wildfire science in general.

In this project, we successfully applied machine learning to the Global Fire Atlas dataset [12] paired with 30-year monthly normal and present monthly climatic data from the NOAA NCDC to predict wildfire size, perimeter and duration at the time of ignition. We see vast opportunity for improvement upon our work by incorporating additional climatic and meteorological information from NOAA, extending the region and period of study beyond the United States in 2016, approaching the problem in a multitask learning framework, and by leveraging more sophisticated techniques from deep learning. Given that the distribution of wildfire sizes is long-tailed and that large wildfires pose a more significant risk to the public than small wildfires, we also encourage research into predictive techniques that perform well in extreme cases—it is easy to predict the average case, but far more difficult to predict the most destructive fires. Machine learning techniques incorporating statistical extreme value theory are of particular interest.

We encourage the reader to build upon our open-source code,[14] and to do their part in the fight against climate change.

References

[1] D. Rolnick, P. L. Donti, L. H. Kaack, K. Kochanski, A. Lacoste, K. Sankaran, A. S. Ross, N. Milojevic-Dupont, N. Jaques, A. Waldman-Brown *et al.*, "Tackling climate change with machine learning," *arXiv*, 2019, arXiv preprint: https://arxiv.org/abs/1906.05433.

[2] P. Jain, S. C. Coogan, S. G. Subramanian, M. Crowley, S. Taylor, and M. D. Flannigan, "A review of machine learning applications in wildfire science and management," *Environmental Reviews*, vol. 28, no. 4, pp. 478–505, 2020.

[3] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*. MIT Press Cambridge, 2016.

[4] A. McGovern, R. Lagerquist, D. J. Gagne, G. E. Jergensen, K. L. Elmore, C. R. Homeyer, and T. Smith, "Making the black box more transparent: Understanding the physical implications of machine learning," *Bulletin of the American Meteorological Society*, vol. 100, no. 11, pp. 2175 – 2199, 2019.

[5] D. Thomas, D. Butry, S. Gilbert, D. Webb, and J. Fung, "The costs and losses of wildfires," *NIST Special Publication*, vol. 1215, no. 11, 2017.

[6] T. J. Venn and D. E. Calkin, "Accommodating non-market values in evaluation of wildfire management in the united states: challenges and opportunities," *International Journal of Wildland Fire*, vol. 20, no. 3, pp. 327–339, 2011.

[7] D. A. Jaffe, S. M. O'Neill, N. K. Larkin, A. L. Holder, D. L. Peterson, J. E. Halofsky, and A. G. Rappold, "Wildfire and prescribed burning impacts on air quality in the united states," *Journal of the Air & Waste Management Association*, vol. 70, no. 6, pp. 583–615, 2020.

[8] E. J. Davis, C. Moseley, M. Nielsen-Pincus, and P. J. Jakes, "The community economic impacts of large wildfires: A case study from Trinity County, California," *Society & Natural Resources*, vol. 27, no. 9, pp. 983–993, 2014.

[9] A. L. Sullivan, "Wildland surface fire spread modelling, 1990–2007. 1: Physical and quasi-physical models," *International Journal of Wildland Fire*, vol. 18, no. 4, pp. 349–368, 2009.

[10] G. D. Papadopoulos and F.-N. Pavlidou, "A comparative review on wildfire simulators," *IEEE systems Journal*, vol. 5, no. 2, pp. 233–243, 2011.

[11] M. A. Finney, *FARSITE, Fire Area Simulator–model development and evaluation*. US Department of Agriculture, Forest Service, Rocky Mountain Research Station, 1998, no. 4.

[14] https://github.com/andrewmcdonald27/CSE847Public

[12] N. Andela, D. C. Morton, L. Giglio, R. Paugam, Y. Chen, S. Hantson, G. R. Werf, and J. T. Randerson, "The Global Fire Atlas of individual fire size, duration, speed and direction," *Earth System Science Data*, vol. 11, no. 2, pp. 529–552, 2019, dataset available: https://www.globalfiredata.org/fireatlas.html.

[13] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[14] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[15] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the royal statistical society: series B (statistical methodology)*, vol. 67, no. 2, pp. 301–320, 2005.

[16] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer, 2009.

[17] K. Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural networks*, vol. 4, no. 2, pp. 251–257, 1991.

[18] S. W. Running, "Is global warming causing more, larger wildfires?" *Science*, 2006.

[19] J. D. Miller, H. Safford, M. Crimmins, and A. E. Thode, "Quantitative evidence for increasing forest fire severity in the sierra nevada and southern cascade mountains, california and nevada, usa," *Ecosystems*, vol. 12, no. 1, pp. 16–32, 2009.