

# Uncertainty Aware Wildfire Management

Tina Diao,<sup>1</sup> Samridhhi Singla,<sup>2</sup> Ayan Mukhopadhyay,<sup>1</sup>  
Ahmed Eldawy,<sup>2</sup> Ross Shachter,<sup>1</sup> Mykel Kochenderfer<sup>1</sup>

<sup>1</sup>Stanford University,

<sup>2</sup>University of California, Riverside

<sup>1</sup>{tdiao, ayanmukh, shachter, mykel}@stanford.edu, <sup>2</sup>{ssing068, eldawy}@ucr.edu

## Abstract

Recent wildfires in the United States have caused many lives and billions of dollars, destroying countless structures and forests. Fighting wildfires is extremely complex. It is difficult to observe the true state of fires due to smoke and risk associated with ground surveillance. There are limited resources to be deployed over a massive area and the spread of the fire is challenging to predict. This paper proposes a decision-theoretic approach to combat wildfires. We model the resource allocation problem as a partially observable Markov decision process. We also present a data-driven model that lets us simulate how fires spread as a function of relevant covariates. A major problem in using data-driven models to combat wildfires is the lack of comprehensive data sources that relate fires with relevant covariates. We present an algorithmic approach based on large-scale raster and vector analysis that can be used to create such a dataset. Our data with over 2 million data points is the first open-source dataset that combines existing fire databases with covariates extracted from satellite imagery. Through experiments using real-world wildfire data, we demonstrate that our forecasting model can accurately model the spread of wildfires. Finally, we use simulations to demonstrate that our response strategy can significantly reduce response times compared to baseline methods.

## Introduction

In 2018, a large wildfire (named *Camp Fire*) in California resulted in the loss of 88 lives, displaced countless more, and destroyed more than 18,500 structures. The estimated cost of the destruction was a staggering \$15 billion. As we write this manuscript at the start of the wildfire season in the state, more than a million acres have already burned in California this year alone due to more than 7,000 wildfires. Wildfires have destroyed many towns and structures across the state. At one point in August 2020, the entire northern half of the state had been instructed to prepare for evacuation (State of California 2020). Crucially, the time of the year that is referred to as the “wildfire season” in the state has only just begun.

Fighting wildfires is difficult. Rapid urbanization and the effects of climate change make urban and suburban areas

particularly susceptible to wildfires. Known to be notoriously unpredictable, sudden changes in wind directions or weather can change the way fires spread (Garza 2020). Firefighters need to allocate limited resources in dynamic and uncertain environments to intervene and stop the spread of fire.

The problem of combating wildfires is an example of dynamic resource allocation. Such an approach to combat natural calamities like wildfires, floods, and earthquakes are difficult for several reasons. The dynamics of the events are unknown and complicated to model in closed-form. There are multiple covariates that affect the spread of fire like the vegetation, fuel, altitude, and wind, and the exact relationship between each covariate and fire spread is uncertain. Resource allocation is extremely challenging because of the difficulty of making predictions about the spread of fire and expected damage. Designing principled approaches to deploy resources is important to mitigate the effects of wildfires and emergencies in general (Mukhopadhyay et al. 2020).

Events like fires affect large areas, and it is difficult to correlate large-scale data from various sources to analyze and study fires. To the best of our knowledge, there exists no comprehensive data source that combines fire occurrence with geospatial features, fuel levels, and weather to allow the research community to develop approaches to combat wildfires. Finally, the problem of deploying resources to manage wildfires is full of uncertainties. As fires spread, it becomes increasingly difficult to observe the true state on the ground due to the presence of smoke. As a result, first responders only have imperfect information to allocate their resources.

**Contributions:** We make the following contributions in this paper. **1)** We present a decision-theoretic approach to dynamically allocate resources in uncertain environments to intervene against wildfires. Specifically, we model the response problem as a partially observable Markov decision process (POMDP) and present an approach to find optimal actions for suppression for a given state of the problem. Our approach accommodates the constraint that the *true* state of the fire is difficult to observe in practice. **2)** Instead of focusing on *physics-based* models, we present a data-driven approach that can simulate the spread of wildfires. We extract relevant covariates such as fuel levels, vegetation type, height of canopy, and elevation from satellite

imagery to drive the simulation. **3)** The calculation of accurate zonal statistics is a bottleneck for large scale geospatial data analysis (Singla and Eldawy 2018). We demonstrate how extremely large-scale geospatial data pertaining to wildfires can be combined with other covariates through large-scale distributed raster and vector analysis. Crucially, we release the dataset open-source for the research community. **4)** Through experiments using real-world wildfire data from California, we demonstrate that our forecasting model can accurately model the spread of wildfires and our response strategy results in significant improvement in suppression efforts compared to baseline methods that do not consider potential fire spread.

## Prior Work

The dynamics of fire spread are usually modeled using *physics-based models*. Popular fire spread models include *BehavePlus* (Andrews 1986) and *Farsite* (Finney 1998). They are based on mathematically modeled surface fire spread as a function of heat flux and fuel availability (Rothermel 1972). Such models are widely used by first responders and fire fighters to forecast the spread of fires. A relatively modern approach is to predict the rate of spread by integrating real-time information about weather from sensors (Altintas et al. 2015). Data-driven modeling has also been used to model fire spread. Supervised machine learning techniques have been applied to uncover strong associations of factors to wildfire sizes and frequency using different data sources. For example, Joseph et al. (2019) investigated weather conditions and geographic characteristics of extreme fire patterns in the contiguous United States and Ghorbanzadeh et al. (2019) examined wildfire susceptibility using geographic data in northern Iran.

Response to wildfires traditionally uses simulation-based approaches to select locations of intervention that maximize the expected utility of suppression efforts. Petrovic, Alderson, and Carlson (2012) model wildfire dynamics and examine the trade-off between multiple competing suppression efforts to compute an optimal strategy for fire responses. Stochastic simulation and multi-agent coordination has also been explored to combat wildfires (Fried, Gillies, and Spero 2006; Martin-fernández, Martínez-Falero, and Pérez-González 2002).

Griffith et al. (2017) explore how suppression efforts can be optimized by solving a mixed-integer linear program and by using Monte Carlo approaches to find optimal actions in a Markov decision process (MDP). Our approach to optimize suppression improves upon prior work (Griffith et al. 2017) to address the uncertainty in state information. We also integrate a data-driven generative model to simulate the spread of fire to aid decision-making under uncertainty.

## Problem Description

We consider a spatial area divided into a set of spatial cells  $G$ . Let  $g_i \in G$  denote the  $i$ th cell. We represent the neighbors of a cell  $g_i$  by  $N_i$ , for some definition of neighborhood (for example, neighbors of a cell can be the set of its adjacent cells). Consider that the total time in consideration

is divided into  $T$  time steps. We assume access to historical data of fire incidents  $D$ , which is a vector of tuples  $\{(t_1, \ell_1, u_1, w_1), (t_2, \ell_2, u_2, w_2), \dots, (t_n, \ell_n, u_n, w_n)\}$ , where each incident  $d_i \in D$  is identified by its time of occurrence  $t_i$ , location  $\ell_i$  (mapping to a cell in  $G$ ), intensity of fire observed  $u_i \in \mathbb{R}^+$ , and a vector of spatio-temporal features  $w_i \in \mathbb{R}^m$ . The features  $w$  capture potential determinants of fire such as weather and the type of vegetation in a cell.

Observing the true state of the world is almost always impossible when wildfires occur. Gathering real-time information by visiting the affected areas by land is naturally difficult. Therefore, information must be gathered through air surveillance obstructed by smoke. Consequently, the true dynamics of wildfires are only partially observable. We model the fire suppression problem more realistically as a partially observable Markov decision process (POMDP). A POMDP can be defined by the tuple  $\{\mathcal{S}, \mathcal{A}, \mathcal{O}, Z, T, R, \gamma\}$  (Kochenderfer 2015). We define each component of the POMDP formulation below:

1. **States:**  $\mathcal{S}$  is a finite set of states. The state at time  $t$  is denoted by  $s_t = \{X_t, F_t\}$ , where  $X_t = \{X_t^1, X_t^2, \dots, X_t^k\}$  and  $F_t = \{F_t^1, F_t^2, \dots, F_t^k\}$  denote the status of the fire and fuel level in each of the cells in  $G$ , respectively. We consider the status of the fire  $X$  as a binary variable such that

$$x_t^i = \begin{cases} 1 & \text{if } u_t^i \geq \epsilon \\ 0 & \text{otherwise} \end{cases}$$

where  $u_t^i \in \mathbb{R}^+$  denotes the measured intensity of the fire in cell  $g_i \in G$  at time step  $t$ . We consider that the fuel level  $F$  is a discrete variable, such that  $F \in \{0, 1, \dots, m\}$ . The exogenous parameters  $\epsilon$  and  $m$  can be estimated from data or through domain knowledge.

2. **Actions:**  $\mathcal{A}$  is a finite set of actions. The actions denote the different permutations of cell indices that fire suppression efforts can be applied to, up to a maximum number of cells specified as a resource constraint.
3. **State Transitions:**  $T$  defines conditional transition probabilities, with  $T(s' | s, a)$  denoting the transition probability from state  $s$  to  $s'$  when action  $a$  is taken. The transition model  $T$  includes the following three components:
  - (a) **Burning:** If cell  $g_i \in G$  is burning at time step  $t$ , we assume that the fuel level decreases by one unit in the next time step. Therefore,  $F_{t+1}^i = F_t^i - 1$  if  $x_t^i \geq \epsilon$ .
  - (b) **Action effectiveness:** At any time step  $t$ , if an action, i.e. fire suppression effort(s) is applied to a cell  $g_i \in G$  such that  $x_t^i \geq \epsilon$ , there is probability  $q$  that the effort successfully puts out the fire.
  - (c) **Fire dynamics:** We use a generative model to simulate the spread of fire. We represent spread dynamics by the probability distribution  $f(X_{t+1}^i | x_t^i, w)$ , where  $g_i \in N_j$ . Therefore, given that a specific cell  $g_j \in G$  is on fire at time step  $t$ ,  $f$  denotes the likelihood of its neighboring cells being on fire at the subsequent time step  $t + 1$ .

We assume that in each time step, fire from a cell can spread only to its neighboring cells. This assumption aids computational tractability, but is without loss of generality because the decision-maker can discretize time fine enough such that the assumption is realistic.

- Reward function:  $R : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$  is the reward function, such that  $R(s_t, a) = \sum_{g_i \in G} x_t^i U(g_i)$ , where  $U(g_i)$  denotes the utility for a cell  $g_i \in G$  to be on fire. Naturally,  $U$  varies across the cells. A cell with human occupants is presumably more *valuable* and more costly to burn than a cell composed of forested land. Cells can also carry unequal ecological utilities (Bradshaw and Lueck 2012). Without loss of generality, we consider three tiers of damage across cells, representing costs of residences (red), valuable ecological resource (yellow), and wildland (green) in decreasing values. Figure 1 shows an example grid with different types of cells.

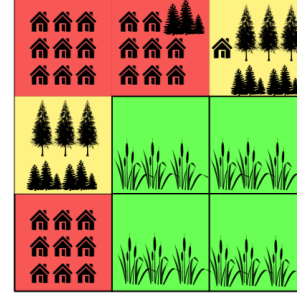


Figure 1: An example of a spatial area discretized in a grid with different costs and the colors correspond to varying costs to burn: red, yellow, and green in decreasing cost.

- Observations and Observation Transitions:  $\mathcal{O}$  represents the set of observations with  $Z(o | s, a)$  denoting the probability of receiving observation  $o$  at state  $s$  when action  $a$  is taken. We denote specific observations by  $o_t^i$ , which correspond to whether cell  $g_i \in G$  is seen to be burning or not at time step  $t$ . Its transitions  $Z(o | s, a)$  are deterministic, i.e.  $o_{t,a}^i = u_t^i$  if an action has been applied to cell  $g_i \in G$ . This is because we assume to have “eyes on location” when an action is applied to a cell. Otherwise, for cells where action for suppression is not applied (denoted by  $\bar{a}$  in the expression below), a generative representation is used based on prior work (Julian and Kochenderfer 2019) such that

$$o_{t,\bar{a}}^i = \begin{cases} 1 & \text{if } Pr_t(X_t^i) > \eta \text{ in state } s \\ 0 & \text{otherwise} \end{cases}$$

where  $\eta$  is an exogenous parameter.

- Discount factor:  $\gamma \in [0, 1]$  denotes the discount factor.

In a POMDP, the decision-maker cannot directly observe the state. Instead, they only have access to beliefs that are generated probabilistically based on the actions taken. Information about states can be inferred from the history ( $h$ ) of observations and actions. It is common to maintain a distribution over states given the history; this distribution is known as the *belief*  $B$ , such that  $B(s | h)$  denotes the probability of being in state  $s$  given history  $h$ . The goal for the decision-maker is to find a mapping from belief states to actions that maximizes the expected future discounted reward.

## Approach

### Modeling Fire Spread

In order to accurately simulate the spread of fire, we model the fire dynamics  $f$  using a data-driven model. Recall that  $f$  is a probability distribution over a cell being on fire in time step  $t + 1$ , conditional on its neighbors being on fire at the previous time step  $t$ . The goal of modeling the function  $f$  is to understand the effect of various covariates like wind, vegetation type, canopy height, altitude, etc. on fire spread. Typically, covariates in geospatial analysis are heterogeneous.

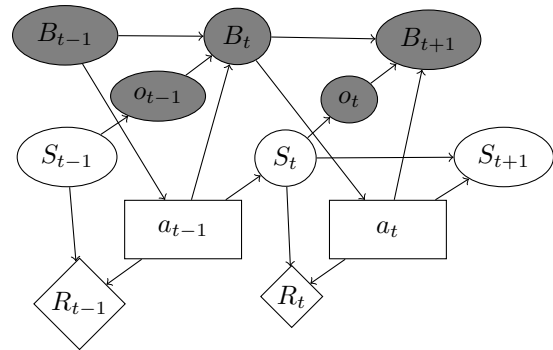


Figure 2: POMDP (factored) representation of the dynamic decision problem. Shaded ovals reflect what is known to the decision maker.

We use random forests (Liaw, Wiener et al. 2002) which involves constructing a large number of decision trees at training time and then aggregating the outputs of the trees. The aggregation technique is typically using the mode of the outputs for classification and the mean of the outputs for regression. The central idea behind using random forests is to average many noisy but (approximately) unbiased models, thereby reducing the variance of the overall forecasting model.

### Resource Allocation

The general dynamic decision framework following the POMDP formulation is shown in Figure 2. Recall that the sources of uncertainty are the current state of the fire ( $S$ ) and the dynamics of the spread (driven by the set of covariates  $w$ ). At each time step, the available set of actions ( $\mathcal{A}$ ) is a combination of the fire location(s) to suppress. The decision-maker acts based on a belief distribution of the *true* state of the world and receives an observation. Utility ( $U$ ) reflects a measure of the expected damage of having a fire in a specific cell (e.g. residence versus wildland).

Due to the large state space and action space (the action space is combinatorial), we use the sampling-based on-

line Monte Carlo tree search (MCTS) (Kochenderfer 2015). To address the computational complexity of POMDPs, approaches based on MCTS typically use a particle filter to represent beliefs in the search tree. Specifically, an approach that is of relevance to our problem is the partially observable Monte Carlo planning algorithm with observation widening (POMCPOW) (Sunberg and Kochenderfer 2017). POMCPOW differs from other online MCTS algorithms in that in the simulation step, given a state ( $s$ ), history ( $h$ ), and depth (for tree exploration), it weights the belief nodes and expands the belief updates gradually as more simulations are added. At each step, a single simulated new state is added to the particle collection, weighted to approximate the belief in every tree node, which is then used to sample the next new state.

Although POMCPOW is shown to outperform other algorithms (Sunberg and Kochenderfer 2017), an issue with directly using POMCPOW on our problem is that the observation space is large and complex. This leads to a severe sparsity of particles, i.e., the probability of sampling a relevant observation is very small. To alleviate this, we modify the routine used to update belief in POMCPOW. Specifically, we replace the weighted particle filter with the standard *particle filter without rejection* (Kochenderfer 2015). As shown in Algorithm 1, given a current belief ( $b$ ), action ( $a$ ), and observation  $o$ ,  $|b|$  samples are generated from the simulator, weighted, and subsequently resampled by its weights. Before the *resampling* step, all weights of an observation  $o$  may be 0 due to the large observation space. If all the weights of an observation are 0, the probabilities of the sampled states are normalized to be proportional to the number of states already sampled. This reweighting step makes an approximation to importance resampling that seeks to estimate properties of a target distribution by sampling from a different distribution.

---

**Algorithm 1:** UpdateBelief ( $b, a, o$ )

---

**Input** : Belief  $b$ , action  $a$ , observation  $o$

**Output:** Updated belief  $b'$

```

1  $b' \leftarrow \emptyset$ 
2 for  $i \leftarrow \{1, \dots, |b|\}$  do
3    $s_i \leftarrow$  random state in  $b$ 
4    $s'_i \sim G(s_i, a)$ 
5    $w_i \leftarrow O(o \mid s'_i, a)$ 
6 end
7 if  $\sum_{i=1}^{|b|} w_i = 0$  then
8    $w_i = \frac{1}{\text{len}(w)}$ ; // reweight states
9 end
10 for  $i \leftarrow \{1, \dots, |b|\}$  do
11   Randomly select  $k$  with probability proportional to
       $w_k$ 
12   Add  $s'_k$  to  $b'$ 
13 end
14 return normalized  $b'$ 

```

---

## Data Processing

Data needed to model the spread of wildfires comes from varied sources. The temporal and spatial resolutions of such data sources are typically different. The data sources can also be of different forms (vector or raster). The vector model uses points and line segments to identify spatial locations while the raster model uses a set of cells for the same purpose. Combining large-scale vector and raster data is known to be a difficult problem (Singla and Eldawy 2018).

We collected fire occurrence data in vector form from the Visible Infrared Imaging Radiometer Suite (VIIRS) thermal anomalies/active fire database (Schroeder et al. 2014). The spatial resolution of VIIRS data is in the form of pixels representing  $375 \times 375$  meter square cells (Schroeder et al. 2014). The latitude and longitude values correspond to the center. Evidence of fire was read from the daily fire radiative power (FRP) levels in the VIIRS dataset. The data used to build the feature space was collected in raster form from the LANDFIRE project (Ryan and Opperman 2013). The foundation of the LANDFIRE project is based on satellite imagery. The raster files had a spatial resolution of  $30 \times 30$  meter square cells. This included features like canopy base density, canopy cover, and vegetation type. We list features used and the years from which the data was collected in Table 1.

Name	Year(s)
Canopy Base Density	2012, 2014, 2016
Canopy Base Height	2012, 2014, 2016
Canopy Cover	2012, 2014, 2016
Canopy Height	2012, 2014, 2016
Existing Vegetation Cover	2012, 2014, 2016
Existing Vegetation Height	2012, 2014, 2016
Existing Vegetation Type	2012, 2014, 2016
Elevation	2016
Slope	2016

Table 1: LANDFIRE Raster Data

To reconcile the different spatial resolutions, we divide the state of California into a grid  $G$  of  $375 \times 375$  meter cells. The center of each fire pixel from the vector data can therefore overlap with exactly one cell in  $G$ . To compute the feature vector associated with each data point, we compute *zonal statistics* for the vector data using the raster data. The method of zonal statistics refers to calculating summary statistics using a raster dataset within zones defined by another dataset (typically in vector form).

Traditional systems to compute zonal statistics require the data to be converted into the same format, either raster or vector. Automated tools can then be used to either vectorize the raster dataset or rasterize the vector dataset. The first approach converts each pixel in the raster to a point and then tests the point against each polygon in the vector data to find a match. This approach has a computational complexity of  $O(n_p \log n_p \cdot c \cdot r)$ , where  $n_p$  is the number of polygons in the vector data, and  $c$  and  $r$  are the number of columns and rows in the raster data respectively. The second approach

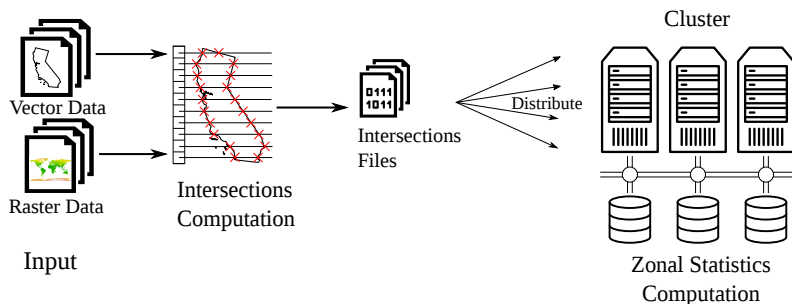


Figure 3: Architecture for calculating zonal statistics for large-scale raster and vector data

rasterizes the vector data by converting each polygon to a raster (mask) layer with the same resolution as the input raster layer. It then combines the two raster layers to compute the desired aggregate function. Most systems that use this approach keep the mask layer in memory. If the input raster layer has a very high resolution, the size of the mask layer can become too large to be kept in memory. This approach has a computational complexity of  $O(n_p \cdot c \cdot r)$ .

Neither of the two approaches mentioned above scale for the high resolution raster and vector data due to the requirement of converting between vector and raster formats. Specifically, in our problem, the vector data consist of over 3 million polygons and each of the raster data sources consists of over 1 billion entries. To deal with the large-scale geospatial data, we create a fully decentralized approach to compute large-scale zonal statistics based on Singla and Eldawy (2020). Our approach does not require data to be converted from one form to another (vector or raster). Instead, it computes an intermediate data structure, called an *intersections file* between the raster and vector data. The *intersections file* can be computed using only the vector data and the metadata of the raster data (coordinate reference, resolution, etc.). Further, our approach can leverage parallel computation by using the intersection files. Our approach, with a computational complexity of  $O(n_p \log n_p + c \cdot r)$ , is scalable and efficient for large raster and vector datasets. Furthermore, as a by-product, our approach makes it easier to find the neighborhoods of polygons by performing a spatial self-join operation using the predicate intersects on all the cells in the vector data. We show the architecture of our approach in Figure 3.

## Experiments

We used fire data from California, USA spanning 2012–2018 for the prediction modeling. We divided the state into a set of  $375 \times 375$  meter cells. Our goal was to capture how fire spreads given an initial occurrence of fire. As a result, we only considered cells and days that exhibited the possibility of fire spreading from an existing fire. Specifically, each cell in our data has a fire in its neighborhood. Our data has a total of 2,367,209 data points. Each row in our data represents a cell  $g_i$  at a specific time  $t$ , a set of spatial-temporal features  $w_{it}$ , and the status of fire  $u_i^t$  as well as the spatial-temporal features  $w_{jt+1}$  and the status of fire  $u_j^{t+1}$  of  $g_j \in N_i$ , a

neighboring cell of  $g_i$  at time step  $t + 1$ . To calculate the features for a specific cell  $g_i \in G$  at a specific time step, we calculated summary statistics (maximum, minimum, median, sum, mode, count, and mean of the feature values) using all raster cells within  $g_i$  at time step  $t$ . To the best of our knowledge, our data is the first comprehensive open-source dataset that combines fire occurrence with relevant covariates extracted from satellite imagery. The data is available at <https://wildfire-modeling.github.io>. We used data from 2012 to 2017 as our training set and data from 2018 as our test set. We set the time step for our experiments to a day, based on the minimum time fidelity of the VIIRS dataset. All experiments were run on an Intel Xeon 2.2 GHz processor with 125 GB of memory.

## Fire Spread

We label a forecast as a true positive prediction when both the predicted fire intensity and the recorded fire intensity are greater than the pre-specified threshold  $\epsilon$ . We observe that the random forest regression model is insensitive to the number of trees used (5, 50, 100, and 500). We also observed similar accuracy across training and test sets. We tested several realizations of  $\epsilon$  to examine the robustness of our forecasting approach on different sizes of fire. Our results show that while it is relatively difficult to predict spread from extremely *small* fires ( $\epsilon = 0.5$ ), our forecasting model achieves high accuracy ( $> 90\%$ ) in predicting spreads from relatively *larger* fires. We summarize the results in Table 2.

FRP Threshold ( $\epsilon$ )	Accuracy on Test Set
0.5	77%
1	81%
5	92%

Table 2: Accuracy with 5-tree Random Forest Regression

## Fire Response

Our goal is to create a pipeline for modeling response to wildfires by utilizing the data-driven model of fire spread to optimize response decisions. However, in this paper, we present experimental results based on a simple spread model. Specifically, we simulate the spread of fire using fixed environmental conditions (for example, we use a fixed

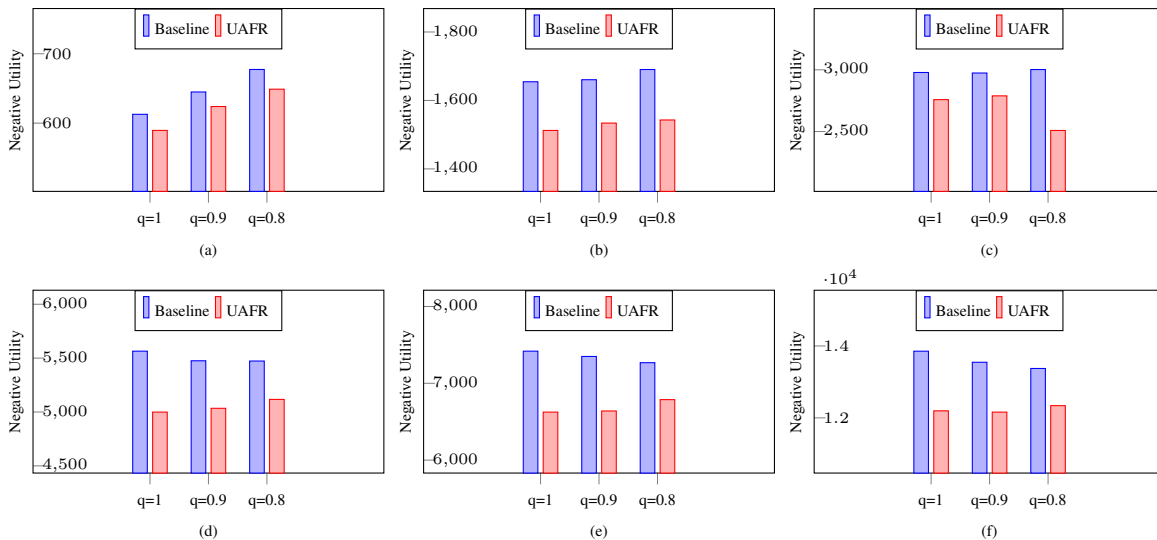


Figure 4: Effectiveness measured by negative utility on test data (lower is better) on varying grid sizes: (a)  $4 \times 4$  (b)  $6 \times 6$  (c)  $8 \times 8$  (d)  $10 \times 10$  (e)  $12 \times 12$  and (f)  $16 \times 16$ .

wind direction and rate of spread in each simulation). We calculate the probability of fire spread following a simplified version of the deterministic fire dynamics model in prior work (Griffith et al. 2017).

We conducted experiments on different sized square grids ( $4^2$ ,  $6^2$ ,  $8^2$ ,  $10^2$ ,  $12^2$  and  $16^2$ ). We varied the starting states by random initialization of fire maps and cell types. We experimented with 6 different initial states. For each initial state, we created 256 spread scenarios by varying wind and rate of spread. For each initialization, a fixed percentage (10%) of cells was considered to be on fire. The proportions of red, yellow, and green cells were set to be 20%, 30%, and 50% respectively, consistent with the distribution of land use in California (United States Department of Agriculture 2016). In all experiments, we set the fuel level in each cell to 5, i.e., each cell takes 5 days to completely burn out and deplete its fuel. We use a baseline model consistent with fire suppression strategies used in practice. Specifically, our baseline fire suppression targets the cell with the maximum utility that shows evidence of fire (through observations). Our open-source code is built using the POMDPs.jl framework (Egorov et al. 2017) and is available at [https://github.com/wildfire-modeling/response\\_model](https://github.com/wildfire-modeling/response_model).

Figure 4 shows the mean negative utility averaged across start states and spread scenarios. We also vary the action effectiveness  $q$  since in practice, suppression efforts do not always completely puts out wildfires. We see that our approach (referred to as ‘‘Uncertainty Aware Fire Response’’ or UAFR) consistently outperforms the baseline method in all scenarios with more significant improvements in larger grids and increased action effectiveness.

## Discussion

Wildfires have caused massive damage in the last decade. They are particularly challenging for first responders to

combat due to difficulties in surveillance and scarcity of resources. In this paper, we built a data-driven forecasting model by extracting relevant determinants of fire spread through satellite imagery. Then, we developed an approach to wildfire suppression that explicitly takes state uncertainty into account. To the best of our knowledge, we were the first to create a comprehensive dataset on wildfires that combines historical fire data with relevant covariates. Our dataset, with over two million data points, and our codebase are open-source for the research community to use.

There are some limitations to our current work. While our forecasting model shows high accuracy, we observe that the random forest regression model is insensitive to the number of decision trees in the ensemble. As a result, simpler methods like classification and regression trees (CART) (Breiman et al. 1984) might result in better generalization to unseen data. Second, our approach to suppression needs to be integrated with the data-driven fire spread model. Finally, while we simulated wind for our experiments, our open-source dataset does not contain information about wind. We are currently incorporating hourly wind data from the National Oceanic and Atmospheric Administration (NOAA)<sup>1</sup> with our fire data to develop a more comprehensive dataset.

## Acknowledgements

We would like to acknowledge the Department of Management Science & Engineering at Stanford University, USDA NIFA AFRI (Grant 2019-67022-29696), and the Center of Automotive Research at Stanford (CARS) for funding this research.

<sup>1</sup><https://www.climate.gov/maps-data/dataset/wind-roses-charts-and-tabular-data>

## References

- Altintas, I.; Block, J.; De Callafon, R.; Crawl, D.; Cowart, C.; Gupta, A.; Nguyen, M.; Braun, H.-W.; Schulze, J.; Gollner, M.; et al. 2015. Towards an integrated cyberinfrastructure for scalable data-driven monitoring, dynamic prediction and resilience of wildfires. *Procedia Computer Science* 51: 1633–1642.
- Andrews, P. L. 1986. *BEHAVE: fire behavior prediction and fuel modeling system: BURN subsystem, Part 1*, volume 194. US Department of Agriculture, Forest Service, Intermountain Research Station.
- Bradshaw, K.; and Lueck, D. 2012. *Wildfire policy: Law and Economics Perspectives*. Routledge.
- Breiman, L.; Friedman, J.; Stone, C. J.; and Olshen, R. A. 1984. *Classification and regression trees*. CRC Press.
- Egorov, M.; Sunberg, Z. N.; Balaban, E.; Wheeler, T. A.; Gupta, J. K.; and Kochenderfer, M. J. 2017. POMDPs.jl: A Framework for Sequential Decision Making under Uncertainty. *Journal of Machine Learning Research* 18(26): 1–5. URL <http://jmlr.org/papers/v18/16-300.html>.
- Finney, M. A. 1998. *FARSITE, Fire Area Simulator—model development and evaluation*. 4. US Department of Agriculture, Forest Service, Rocky Mountain Research Station.
- Fried, J. S.; Gilles, J. K.; and Spero, J. 2006. Analysing initial attack on wildland fires using stochastic simulation. *International Journal of Wildland Fire* 15(1): 137–146.
- Garza, A. 2020. AI Is Helping Fight Wildfires Before They Start. <https://time.com/5497251/wildfires-artificial-intelligence/>.
- Ghorbanzadeh, O.; Valizadeh Kamran, K.; Blaschke, T.; Aryal, J.; Naboureh, A.; Einali, J.; and Bian, J. 2019. Spatial prediction of wildfire susceptibility using field survey gps data and machine learning approaches. *Fire* 2(3): 43.
- Griffith, J. D.; Kochenderfer, M. J.; Moss, R. J.; Mišić, V. V.; Gupta, V.; and Bertsimas, D. 2017. Automated dynamic resource allocation for wildfire suppression. *Lincoln Laboratory Journal* 22(2).
- Joseph, M. B.; Rossi, M. W.; Mietkiewicz, N. P.; Mahood, A. L.; Cattau, M. E.; St. Denis, L. A.; Nagy, R. C.; Iglesias, V.; Abatzoglou, J. T.; and Balch, J. K. 2019. Spatiotemporal prediction of wildfire size extremes with Bayesian finite sample maxima. *Ecological Applications* 29(6).
- Julian, K. D.; and Kochenderfer, M. J. 2019. Distributed wildfire surveillance with autonomous aircraft using deep reinforcement learning. *Journal of Guidance, Control, and Dynamics* 42(8): 1768–1778.
- Kochenderfer, M. J. 2015. *Decision Making Under Uncertainty: Theory and Application*. MIT Press.
- Liaw, A.; Wiener, M.; et al. 2002. Classification and regression by randomForest. *R news* 2(3): 18–22.
- Martin-fernández, S.; Martínez-Falero, E.; and Pérez-González, J. M. 2002. Optimization of the resources management in fighting wildfires. *Environmental Management* 30(3): 352–364.
- Mukhopadhyay, A.; Pettet, G.; Vazirizade, S.; Lu, D.; Baroud, H.; Jaimes, A.; Vorobeychik, Y.; Kochenderfer, M.; and Dubey, A. 2020. A Review of Emergency Incident Prediction, Resource Allocation and Dispatch Models.
- Petrovic, N.; Alderson, D. L.; and Carlson, J. M. 2012. Dynamic resource allocation in disaster response: Tradeoffs in wildfire suppression. *PloS one* 7(4): e33285.
- Rothermel, R. C. 1972. *A mathematical model for predicting fire spread in wildland fuels*, volume 115. Intermountain Forest & Range Experiment Station, Forest Service.
- Ryan, K. C.; and Opperman, T. S. 2013. LANDFIRE—A national vegetation/fuels data base for use in fuels treatment, restoration, and suppression planning. *Forest Ecology and Management* 294: 208–216.
- Schroeder, W.; Oliva, P.; Giglio, L.; and Csiszar, I. A. 2014. The New VIIRS 375 m active fire detection data product: Algorithm description and initial assessment. *Remote Sensing of Environment* 143: 85–96.
- Singla, S.; and Eldawy, A. 2018. Distributed zonal statistics of big raster and vector data. In *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 536–539.
- Singla, S.; and Eldawy, A. 2020. Raptor Zonal Statistics : Fully Distributed Zonal Statistics of Big Raster + Vector Data. In *Proceedings of 2020 IEEE International Conference on Big Data (To Appear)*. IEEE.
- State of California. 2020. 2020 Incident Archive. <https://www.fire.ca.gov/incidents/2020/>.
- Sunberg, Z.; and Kochenderfer, M. 2017. Online algorithms for POMDPs with continuous state, action, and observation spaces. *arXiv preprint arXiv:1709.06196*.
- United States Department of Agriculture. 2016. FIA State Stats. <https://www.fs.fed.us/pnw/rma/fia-topics/state-stats/California/index.php>.