Fire Event Prediction for Improved Regional Smoke Forecasting

Zachary Butler¹, Yang Chen², James Randerson², and Padhraic Smyth¹

Abstract—Smoke from wildfires is a significant public health concern with over 300,000 people dying annually worldwide. Given these large health impacts an important goal is to forecast fire emissions on multi-day time scales, for example, to provide higher quality forecasts for operational smoke forecasting systems. In this paper we describe initial work on statistical predictive modeling techniques that use historical satellite and weather data to predict fire activity on daily time-scales and for a regional spatial domain. Prediction results from 10 years of wildfire data in Alaska illustrate how local weather information can be used to improve the quality of multiday fire forecasts.

I. MOTIVATION AND BACKGROUND

Fire is an important and dynamic ecosystem process that responds to climate change and human modification of the land surface [1]. Fire emissions of greenhouse gases, ozone precursors, and black carbon aerosols have a warming effect on climate, whereas emissions of organic carbon aerosols and post-fire changes in species composition (and surface biophysics) may have an opposing effect. In concert, human health impacts from fire aerosols are widespread and significant [2]. Smoke impacts on health are amplified in regions downwind of large regional fire complexes [3,4]. For example, exceptionally large fire complexes in Alaska and Canada in June of 2015 generated smoke plumes that traveled thousands of miles, and significantly reduced air quality in cities across the central U.S.

To help mitigate these impacts, several federal agencies have created smoke forecasting systems, such as the European Union’s Monitoring Atmospheric Composition and Climate System (MACC) [5], the U.S. Navy’s Fire Locating and Monitoring of Burning Emissions (FLAMBE) Project [6], NOAA’s Smoke Forecasting System [7], the U.S. Forest Service’s BlueSky smoke modeling framework [8], and NASA’s GEOS-5 Forward Processing (FP) system [9]. These systems often use near real-time satellite observations of fire radiative power to estimate the spatial pattern of fire emissions. The fire emissions, in turn, are introduced into an atmospheric model that uses weather forecasts of winds and other meteorological variables to transport the smoke into downwind areas. Most of these systems assume that the spatial structure and intensity of fire emissions remain constant over the duration of the forecast. Thus, while the evolving impact of fires on atmospheric composition is determined by the influence of meteorology on aerosol transport and loss processes, increases in burned areas or modification of fire behavior due to changing weather are not considered.

In this paper we describe our work on developing models that can predict daily fire emissions over the course of a weather forecast. In our approach, we draw upon satellite data streams and online archives of weather forecasts, with a specific focus on Alaska (the methodology is, however, more broadly applicable). A primary goal of this work is to understand limits to fire prediction originating from uncertainties in weather prediction and from our ability to model fire behavior. Unlike most existing fire prediction systems that use physically-based fire behavior models to predict spread rates of individual fires and which do not track fires locally [e.g., 10, 11, 12, 13], our approach is designed to track multiple fires simultaneously and to predict new ignitions. Fire emissions forecasting at this larger spatial scale represents a novel prediction challenge and is needed for operational smoke forecasting systems operating at regional or global scales. In subsequent work we plan to couple our fire prediction algorithms to smoke forecasting systems with the goal of improving the quality of aerosol predictions from existing approaches. We expect that by allowing emissions to evolve dynamically over the duration of a forecast, we will be able to considerably improve the accuracy and value of smoke forecasting systems for public health and air quality applications.

Corresponding author: P. Smyth, smyth@ics.uci.edu.
¹Department of Computer Science, University of California, Irvine
²Department of Earth System Science, University of California, Irvine.
II. DATA

By exploiting the strong emission of mid-infrared radiation from fires, Moderate Resolution Imaging Spectroradiometer (MODIS) instruments onboard NASA’s Terra and Aqua satellites detect active fires at a ~1 km spatial resolution using a contextual algorithm [14]. Here we used the daily active fire count in the MODIS Fire Location Product (MCD14ML, from http://modis-fire.umd.edu/) as our target variable \( y \) for fire prediction. Figure 1 shows the distribution of active fire counts in Alaska for the year 2013.

![Figure 1: Left: Active fire detections in Alaska during the 2013 fire season. Right: The spatio-temporal evolution of the blue detections (which form a fire cluster – see Section III), color-coded by day of detection.](image)

The weather variables we used were from the NOAA Global Forecast System (GFS). GFS is a global numerical weather prediction system, which is run four times a day, and produces forecasts for up to 16 days in advance. Here we used gridded (0.5°) surface temperature, surface humidity, surface wind, and precipitation rate data from the GFS analysis [15] to construct the fire forecast model. These weather variables were converted to daily averages (daily accumulations for precipitation) (Figure 2). We collected both data sources for each fire season in Alaska from 2007-2016. We defined the fire season as May 14-Aug 31, as most fire detections during these years occurred within this window.

III. METHODS

The general modeling problem of interest is to predict the total number of fire detections in a fixed spatial area on day \( t+k \) given information on day \( t \), with \( k = 1, 2, 3, ... \). We decompose the overall problem into two parts: (1) spatially local predictions for clusters of fires, and (2) global prediction of new ignition events. Below we describe results on local predictions for fire clusters: in the final version of the paper we will discuss results for the complete model. We represent fires as spatio-temporal clusters of fire detections (e.g., Fig. 1). We assume two fire pixels belong to the same fire cluster if they are within 5 km of each other (or connected through a chain of points, each within 5 km of the next). This clustering yielded 1335 fire clusters over the period 2007-2016 with each cluster persisting for a mean duration of 9.7 days.

![Figure 2: Time series of active fire counts for the cluster from Figure 1 with various weather variables used as model drivers. The increase in rain and humidity results in the fire dying out.](image)

We predict the number of detections for active fires on day \( t+k \) via Poisson regression [e.g., 16]. Specifically, we model the log of the Poisson rate on day \( t+k \) (i.e., the expected number of detections on that day) for fire cluster \( i \) as

\[
\log(E[y_{t+k}]) = B_0 + B_y \log(y_t) + \sum_w B_w x_{t+k,w}^i
\]

Here \( \log y_{t+k}^i \) is the number of fire detections for cluster \( i \) on day \( t+k \). \( x_{t+k,w}^i \) represents different weather variables \( w \) for cluster \( i \) on day \( t+k \), and the \( B \)'s are global model coefficients (not dependent on \( t \) or on cluster \( i \)). The \( B \)'s are estimated by maximizing the log-likelihood, defined as the sum of the log probability of the observed data on day \( t+k \) (under a Poisson model), over clusters and days for each cluster, and where the Poisson mean is a function of the model parameters (the \( B \)'s), conditioned on the covariates \( y_t \) and \( x_{t+k,w}^i \). The weather variables \( x_{t+k,w}^i \) are estimated at the spatial centroid of fire \( i \) on day \( t+k \) (e.g., see Fig 1 (right)). (One exception is rainfall, which we found to be predictive at lag 2 - in all the results below rainfall is measured at day \( t+k-2 \)). Thus,
predictions are made locally in time and space for each fire cluster: metrics for assessing performance are then aggregated over fire clusters and over days when each fire is active. The approach above simulates having “perfect weather forecasts”, using the actual future weather data as a proxy for forecasted weather (realizing that actual forecasts will be noisier).

IV. RESULTS

Our experiments addressed three questions:
1. Can we predict fire detections more accurately than the baseline of \( y(t+k) = y(t) \)?
2. To what extent can weather covariates improve predictions beyond autoregression?
3. How does predictive performance decrease in quality as we predict further into the future?

To answer these questions we conducted two experiments. In the first experiment we fit a model to all years to predict on each day, and for each fire cluster, the number of detections on day \( t+k \) given covariates defined on day \( t \). The resulting regression coefficients (the \( B \)'s) for the log of the Poisson rate are shown in Table 1 below.

<table>
<thead>
<tr>
<th></th>
<th>Intercept</th>
<th>Counts ( x(t) )</th>
<th>Temp</th>
<th>Hum</th>
<th>Wind</th>
<th>Rain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normalized</td>
<td>1.138</td>
<td>0.935</td>
<td>0.217</td>
<td>-0.043</td>
<td>0.033</td>
<td>-0.556</td>
</tr>
<tr>
<td>Unnormalized</td>
<td>-9.473</td>
<td>0.717</td>
<td>0.035</td>
<td>-0.002</td>
<td>0.022</td>
<td>-0.163</td>
</tr>
</tbody>
</table>

Table 1: Regression coefficients for the cluster Poisson regression model. Normalized covariates were centralized to have mean of 0 and a standard deviation of 1.

The normalized coefficients above are for a model fit with standardized inputs (with a mean of zero, standard deviation of 1) and the unnormalized results are for a model without standardization. The model coefficients agree with physical intuition. The coefficients for temperature and wind are positive and those for humidity and rain are negative. For example, a unit increase in temperature (one degree Kelvin) for the unnormalized model results in a multiplicative increase in expected number of detections by a factor of \( \exp(0.035) = 1.036 \). Similarly a unit increase in rainfall (1 mm/day) corresponds to a multiplicative decrease of \( \exp(-0.163) = 0.850 \) in expected count of fire detections per cluster.

The second experiment was designed to evaluate the relative accuracy of different models via cross-validation at the yearly level. We trained models on every year but one and evaluated performance on the held-out year, then aggregating the results across all held-out years. We fit instances of each model to make predictions at day \( t+k \), \( k=1,2,3,4, \) and 5, conditioned on (a) fire detections at day \( t \), and (b) weather covariates defined on day \( t+k \). Since many existing smoke forecasting models assume that fire detections are constant over the duration of the forecast, i.e., \( y(t+k) = y(t) \), we compared against this as a baseline. The three types of models we trained are:

1. Autoregression (with lag 1): the only covariate is the number of detections on day \( t \), \( y(t) \).
2. Temp/Hum: the covariates include autoregression, \( y(t) \), and temperature and humidity on day \( t+k \).
3. All weather: This adds rain and humidity to the Temp/Hum model.

Figure 3 below shows the mean absolute error (MAE) of predicted detections compared to actual number of detections, as a function of \( k \). Models with weather covariates outperform the baseline. Model results are shown in Table 1 above, and a regression line shows the model fit for each year. Regression models built on historical data can provide systematic improvements over current forecasting practices.

![Figure 3: Cross-validated MAE from 2007-2016. The x-axis is the day we are predicting: at x=3, we are using the counts at time t and weather at time t+3 to predict counts at time t+3.](image)

V. CONCLUSIONS

We investigated the use of statistical methods for predicting fire growth over time using patterns of historical fire and weather data in Alaska. We find that the incorporation of weather variables allows for more accurate prediction compared to models solely based on temporal autoregression. Under the assumption of perfect weather forecasts the relative improvements became larger the further the model forecasted into the future, suggesting that accurate weather forecasts have the potential to significantly improve the quality of smoke forecasts. Future directions include: measuring the degradation in accuracy from actual weather forecasts relative to perfect weather information; incorporating additional local variables such as vegetation, elevation, topological features such as rivers, lakes, and roads, history of prior burned areas; adding spatial context to the models; and adding longer term memory through fire weather indices to capture ground moisture and drying effects.
ACKNOWLEDGEMENTS

The work in this paper was supported in part by the National Science Foundation under award DGE-1633631.

REFERENCES


