Forecasting Daily Wildfire Activity Using Poisson Regression

Casey A. Graff, Shane R. Coffield, Yang Chen, Efi Foufoula-Georgiou, James T. Randerson, and Padhraic Smyth, *Member, IEEE*

***Abstract*— Wildfires and their emissions reduce air quality in many regions of the world, contributing to thousands of premature deaths each year. Smoke forecasting systems have the potential to improve health outcomes by providing future estimates of surface aerosol concentrations (and health hazards) over a period of several days. In most operational smoke forecasting systems, fire emissions are assumed to remain constant during the duration of the weather forecast and are initialized using satellite observations. Recent work suggests that it may be possible to improve these models by predicting the temporal evolution of emissions. Here, we develop statistical models to predict fire activity one to five days into the future using Moderate Resolution Imaging Spectroradiometer (MODIS) satellite fire counts and weather data from ERA-interim reanalysis. Our predictive framework consists of two-Poisson regression models that separately represent new ignitions and the dynamics of existing fires on a coarse resolution spatial grid. We use ten years of active fire detections in Alaska to develop the model and use a cross-validation approach to evaluate model performance. Our results show that regression methods are significantly more accurate in predicting daily fire activity than persistence-based models (which suffer from an overestimation of fire counts by not accounting for fire extinction), with vapor pressure deficit being particularly effective as a single weather-based predictor in the regression approach.**

***Index Terms*— Daily fire forecasting, fire ignitions, moderate resolution imaging spectroradiometer (MODIS), smoke aerosols, vapor pressure deficit (VPD).**

Manuscript received May 2, 2019; revised October 20, 2019; accepted November 4, 2019. Date of publication February 27, 2020; date of current version June 24, 2020. This work was supported in part by the National Science Foundation under Grant 1633631 (for CG, JR, PS) as part of the UCI NRT Machine Learning and Physical Sciences (MAPS) Training Program, in part by the National Science Foundation Graduate Research Fellowship Program under Grant DGE-1839285 (for SC), in part by NASA under Award NNX15AQ06A as part of the CSULA/UCI DIRECT-STEM Project (for PS), in part by the National Science Foundation under Award CNS-1730158 (for PS), in part by the DOE Office of Science RUBISCO Science Focus Area and NASA's SMAP and CMS programs (for JR, YC), in part by the National Science Foundation under Grant DMS-1839336 (for EFG, JR, PS) as part of the TRIPODS program, and in part by NASA under Grant NNX16AO56G (for EFG) as part of the GPM Program. *(Corresponding author: Casey A. Graff.)*

Casey A. Graff and Padhraic Smyth are with the Department of Computer Science, University of California Irvine, Irvine, CA 92697 USA (e-mail: caseyagraff@gmail.com).

Shane R. Coffield, Yang Chen, and James T. Randerson are with the Department of Earth System Science, University of California Irvine, Irvine, CA 92697 USA.

Efi Foufoula-Georgiou is with the Department of Civil and Environmental Engineering, University of California Irvine, Irvine, CA 92697 USA, and also with the Department of Earth System Science, University of California Irvine, Irvine, CA 92697 USA.

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2020.2968029

I. Introduction

GIVEN that aerosol emissions from wildfires pose a significant threat to human health [1], [2], an important goal is to build national and international capacity to adaptively manage fires and their impacts on air quality. In this context, improved hourly fire forecasts on daily time scales (one to ten days) can reduce negative health outcomes and may increase the efficacy of fire suppression efforts. The integration of fire-emitted aerosols and trace gases into numerical weather forecasting systems has become possible because of the availability of reliable, high-quality active fire and fire radiative power (FRP) observations from the Moderate Resolution Imaging Spectroradiometer (MODIS) [3], Visible and Infrared Imaging spectroradiometer [4], and other satellite sensors with similar characteristics. These near-real-time satellite observations determine the initial boundary condition for emissions within the atmospheric model, which, in turn, predicts the dispersal and evolution of fire plumes and the influence of biomass burning on atmospheric composition over the duration of the weather forecast. Several operational and quasi-operational systems exist, including the European Union's Monitoring Atmospheric Composition and Climate (MACC) System [5], the U.S. Navy's Fire Locating and Monitoring of Burning Emissions (FLAMBE) Project [6], NOAA's Smoke Forecasting System [7], and NASA's GEOS-5 Forward Processing (FP) system [8].

Until recently, all of these systems have assumed that the spatial structure and intensity of fire emissions remain constant over the duration of the numerical weather forecast. Thus, the evolving impact of fires on atmospheric composition was determined by the influence of meteorology on transport, chemistry, and loss processes but not because of modification of fire behavior due to changing weather. If, for example, surface winds and vapor pressure deficits (VPDs) are predicted to abruptly increase several days into the future, the influence of fires on atmospheric composition may be underestimated because of the inability of these persistence-based algorithms to simulate an increase in emissions. Moving beyond the persistence assumption with the active fire data, Di Giuseppe *et al.* [9] explored the use of the Canadian Forest Fire Weather Index [10] as a modulator of fire activity over the duration of the forecast. By allowing emissions to covary with fire weather, the authors were able to improve estimates of aerosol optical depth in their atmospheric model compared with observations from AERONET [11].

Here, we build on work by Di Giuseppe *et al.* [9] by developing Poisson regression models to estimate the daily evolution of active fires over the duration of a weather forecast in response to weather variability and intrinsic fire dynamics. Our models rely on satellite observations and weather forecasts and operate on a 0.5° by 0.5° grid, which we chose to be comparable to the resolution of many operational smoke forecasting systems. Weather variables (temperature, precipitation, relative humidity, and wind speed) and previously observed counts are leveraged to generate fire count predictions for up to five days in the future. We focus our analysis on the spatial domain of Alaska, although conceptually the models we develop should be applicable in other regions and at the global scale. Alaska is an important region for model development because recent increases in lightning-ignited fires threaten human health and vulnerable species [12], [13].

Two broad categories of relevant methods exist for forecasting fire activity on daily time scales: 1) fire behavior and spread models and 2) statistical forecast models. Fire behavior and spread models generally focus on simulations of specific, ongoing fires to produce spatiotemporal predictions of a fire's trajectory and intensity. This category includes models such as FARSITE [14], FSPro [15], and Australis [16]. These models rely on physically based simulations that account for topography, fuel amount, structure, and moisture status, and local weather predictions [17]. Parameters in these models can be manually adjusted to better capture the past growth trajectory of the individual fire that is being targeted for decision support, thus improving future estimates of fire spread. Improvements in ground-based sensor networks are enabling near real-time fire spread models that combine simulation and statistical models, such as WIFIRE [18]. These models, while effective for informing emergency response, focus on detailed tracking of individual fire events rather than predicting emissions from multiple fires that may be simultaneously active within the regional-to-global scale domain of a smoke forecasting system. In theory, individual fire spread models could be run simultaneously to create the emissions inputs to the atmospheric model. In practice, however, computational and operational demands make this approach impractical on continental to global scales. Furthermore, in many regions, new fires are often ignited during the forecast period. These fires, which are often of short duration, can contribute significantly to the smoke burden and necessitate a different conceptual approach.

Unlike physically based fire spread models which operate on individual fires, a statistical forecasting approach can be efficiently applied to produce forecasts for multiple fires over large regions and to predict contributions from new ignitions. Specifically, this article contributes to statistical fire forecasting by demonstrating significant improvement in forecast accuracy by regression models over traditional persistence methods and by providing detailed analyses of forecast performance.

The statistical models we develop here, which separately predict existing and new fires, have the potential to increase the accuracy of operational smoke forecasting systems and can be applied globally. In particular, the use of globally available satellite datasets is critical for remote regions such as Alaska where detailed ground-based data are lacking.

The outline of the remainder of this article is as follows. Section II describes the datasets we use in this article, including satellite active fire detections and weather reanalysis products on daily timescales. In Section III, we describe the notation and general setup for the statistical models we investigate. We introduce the evaluation methods in Section IV and the predictive models in Section V. Our experimental results are divided into two major sections. We begin in Section VI by focusing on days with active fires and evaluate the predictive performance of different models specifically for active fires. We then extend this evaluation in Section VII to all of the days in the study period, allowing for the evaluation of predictive performance of different models across quiescent periods, ignition events, and active fire periods. In Section VIII, we provide a discussion of the results and present our primary conclusions.

The novel contributions of this article are as follows: 1) the development of a new model, based on mixtures of Poisson regression models, for active-fire forecasting at daily timescales, and 2) a systematic evaluation of the predictive accuracy of the proposed model that demonstrates significant improvements over current operational baselines being used in NOAA and European Union smoke forecasting systems.

II. Data

For this article, we use data for a ten-year period from 2007 to 2016 for the state of Alaska and use only the days of each year between May 14 and August 31, which we will refer to as the fire season. Specifically, we include all data bounded by 71°N by 165°W and 55°N by 138°W; excluding any areas comprising entirely of water. Although the bounding box does not cover all of Alaska, it includes more than 99% of all detections in the state for this time period. We describe below the specific datasets we used from this period.

A. Active Fire Detections

We use the daily active fire detections in the MODIS MCD14ML Fire Location Product (collection 6) [19] as our ground truth for fire detections. These data are collected by NASA's MODIS instrument on two satellites, Aqua and Terra. The spatial resolution of the fire detections is approximately 1 km^2 near nadir. The sampling frequency is one to four times per day and the effective sampling frequency (when accounting for cloud cover) can be less [5]. This is due to the fact that both Aqua and Terra are polar-orbiting, with Terra passing at roughly 10:30 A.M./P.M. and Aqua passing at roughly 1:30 A.M./P.M. local time. Fig. 1 shows all of the active fire detections for June 25, 2015, in Alaska. Fig. 2 shows the distribution of fire detections for the ten-year study period, which results from the satellite orbital paths. The MODIS data include detections from several different heat sources: vegetation fires, active volcanoes, other static land sources, and offshore gas flares. We only include detections classified as vegetation fires.

Known issues in the MODIS fire product are false-negatives that arise from obscured measurements due to cloud cover, and false-positives due to reflectance from bodies of water,

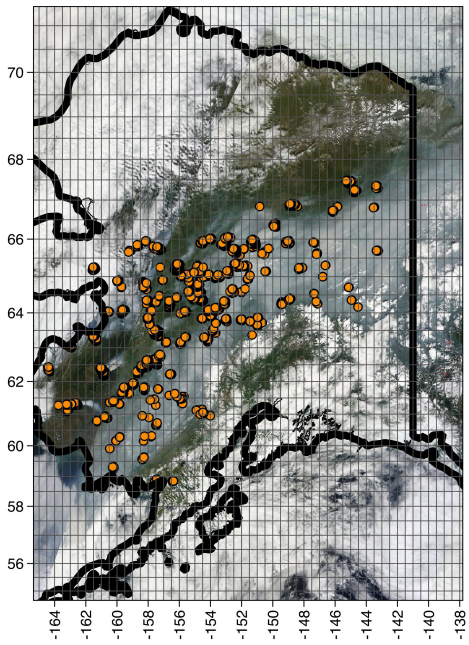

Fig. 1. All vegetation active fire detections in Alaska on June 25, 2015 from the MODIS dataset with smoke and terrain from LANDSAT with a 0.5° × 0.5° grid overlaid.

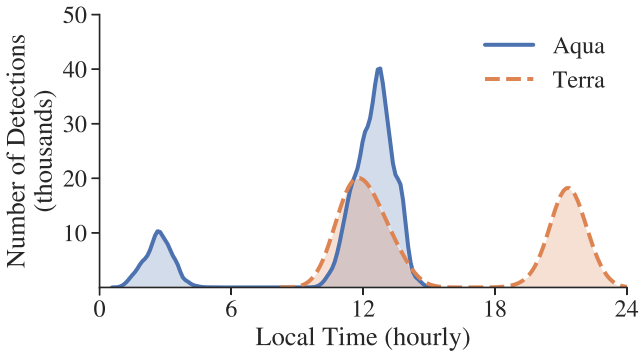

Fig. 2. Empirical density plot (kernel density estimate) of MODIS active fire detections in Alaska as a function of local solar time, for satellites Aqua and Terra from 2007 to 2016.

snow, and ice, as discussed in [3] and [20]. We chose not to threshold or filter the active fire detections, as we do not expect issues with false-positives at high latitudes. The collection 6 MODIS active fire algorithm sometimes misclassifies forest clearing as fires, but this commission error occurs mostly in tropical regions [21]. When all confidence levels are included, the mean commission error in boreal regions is often smaller than 4%.

B. Meteorology

Weather reanalysis data are obtained from the ERA-Interim (ERA) product produced by the European Centre for Medium-Range Weather Forecasts (ECMWFs) [22]. This numerical weather forecast system produces analysis data for temperature, humidity, and wind four times a day, and for precipitation twice a day. The product has a gridded spatial resolution of approximately 0.75° which we resample to 0.5°. We use the

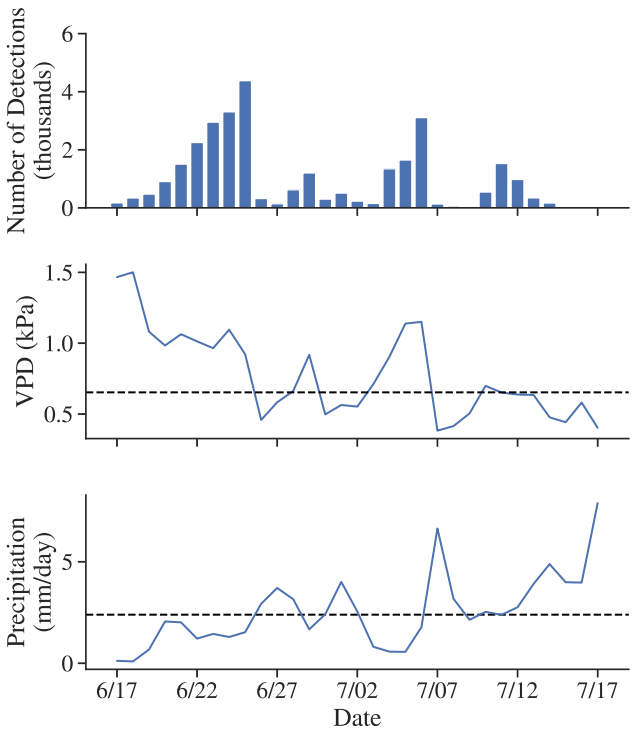

Fig. 3. Comparison of (Top) daily active fire detections, (Middle) mean VPD, and (Bottom) mean rainfall for June 17 to July 17, 2015, over the interior region of Alaska. Observed values in blue (solid line) and ten-year-mean values for interior Alaska for the same time period in black (dashed line).

time closest to mid-day for Alaska for surface temperature, humidity, and wind speed and use the 24-h accumulation for precipitation. VPD is calculated according to Tetens equation [23].

There is strong empirical evidence to support the inclusion of weather variables in our models. For example, Table I shows the correspondence between weather variables and MODIS active fire detections, in terms of anomalies from the ten-year mean. The temperature and humidity anomalies are largest for the three years with the greatest number of detections: 2009, 2013, and 2015. In all the three years, there is an increase in mean Alaskan interior temperature and a decrease in humidity during the fire season. This follows the intuition that these measurements are positive and negative drivers of fire dynamics, respectively. The relationship between fire and weather is also observed on shorter timescales, in daily comparisons between fire ignitions and weather variables within a fire season. For example, as shown in Fig. 3, in mid-June and July 2015 there are notable increases in interior VPD, with respect to the ten-year mean that coincide with several large fire events. In this period of roughly one month, over 80% of all the year's fires ignited. This positive correlation between VPD and fire events is consistent with previous research on Alaskan boreal forests [24]. In particular, we observe in Fig. 3 that all of the large increased periods of fire activity correspond to dates with the above-average VPD and that the level of fire activity tends to significantly diminish during periods of increased precipitation.

C. Land Cover

We also investigate the use of land cover information using the MCD12C1 land cover product [25]. This information is available at a 0.05° resolution and is downsampled to 0.5° by computing the fraction of each of the 17 International Geosphere-Biosphere Programme (IGBP) global vegetation classifications [26] for each 0.5° cell, producing 17 additional variables. However, we found that land cover did not significantly contribute to the predictive accuracy of the models we developed, likely due to coarse spatial resolution, and thus, we do not discuss land cover further in this article as a potential predictor variable in our models.

III. Notation and Model Structure

We spatially partition the region of interest into equal-sized grid cells of 0.5° latitude by 0.5° longitude. We use $y_l^{(t)}$ to represent the number of active fire detections in grid cell l on day t of the fire season, where $y_l^{(t)}$ takes nonnegative integer values $0, 1, 2, \ldots$. The index $t = 1, 2, \ldots, T$ is an index over the T days of each fire season, $l \in \{(i, j)\}$ is an index over the $L = H \times V$ locations, where $i = 1, 2, \ldots, H$, and $j = 1, 2, \ldots, V$ index the $H = 55$ horizontal and $V = 33$ vertical cells in the grid, respectively.[1] Similarly $\boldsymbol{x}_l^{(t)}$ defines a corresponding p-dimensional vector of explanatory covariates observed on day t in grid cell l, where the covariates can include temperature, rainfall, humidity, wind, and VPD as well as lagged versions of y itself.

The forecasting problem can be formulated as predicting the expected number of active fire detection counts $E[y_l^{(t+k)}]$, for each grid cell l, for $k = 1, \ldots, K$ days after the current day t. We use a maximum lead time of $K = 5$. When making forecasts, we assume that the forecasting model has access to: 1) measurements of y (detections) up to and including the day t that the forecast is being made and 2) weather covariate values $\boldsymbol{x}$ not only in the past and on the current day t but also for future days $t' \in \{t+1, \ldots, t+K\}$, where the $\boldsymbol{x}$ values for the future days are based on forecasts of $\boldsymbol{x}$ available on day t, e.g., from weather forecasts.

Thus, the forecast model can be written in the following general form:

$$E\left[y_l^{(t+k)}\right] = f\left(y_l^{(1)}, \ldots y_l^{(t)}, \boldsymbol{x}_l^{(1)}, \ldots, \boldsymbol{x}_l^{(t+k)}; \boldsymbol{\theta}\right), \quad 1 \le k \le K \tag{1}$$

where $\boldsymbol{\theta}$ are the parameters of the model (to be estimated from data), and f is the functional form of the model. For example, if f were linear and if there were no covariates $\boldsymbol{x}$ in the model then the model would reduce to a first-order linear autoregressive (AR) model on the observed time-series y.

In the results presented here the values of the covariates $\boldsymbol{x}$ used in our model are derived from reanalysis products, which means that these can be considered "perfect" forecasts and ignore forecast uncertainty. We would expect that if we used actual forecasts, rather than reanalysis values, the performance of the models would degrade. Nonetheless, the results using reanalysis data are useful in that they provide us with information about which variables have the most predictive power and also inform us about the relative importance of weather versus the past history of active fire detections for this prediction problem. Moreover, the use of reanalysis data allows us to isolate the fire forecast error from the weather forecast error.

[1] An additional subscript or superscript could be added to y to indicate which year each $y_l^{(t)}$ value corresponds to, but we suppress this notation for simplicity.

TABLE I

Summary Statistics of Active Fire Detections and Interior Climate for the Fire Season (May 14 to August 31)

Year	Number of Detections	Temperature Anomaly (°C)	Humidity Anomaly (%)
2007	3973	1.0	1.2
2008	887	-1.4	0.7
2009	29310	0.1	-3.0
2010	7386	0.3	2.8
2011	1635	-0.5	1.1
2012	1334	-0.6	1.3
2013	9310	0.8	-3.9
2014	1103	-0.7	-0.9
2015	34805	0.5	-1.7
2016	3710	0.5	2.4
Mean	9345	13.2	68.2
Std. Dev.	11723	0.7	2.1

The parameters $\boldsymbol{\theta}$ are assumed to be spatially and temporally constant, i.e., $\boldsymbol{\theta}$ is assumed not to vary over space or time. Nonetheless, temporal and spatial variability in the predictions enters via the y and $\boldsymbol{x}$ variables on the right-hand side of (1).

In making forecasts k days ahead using the general model formulation in (1) there are two options. The first option is the *recursive* approach, to build a single model that only predicts one day ahead (i.e., $K = 1$) and then use the forecast y for day $t+1$ as the input to recursively make predictions for day $t+2$, and so on. The second option is to build K different models, one for each value of k, $1 \le k \le K$. We use the second approach in this article since it more directly seeks to optimize predictive accuracy for $k > 1$. We find the recursive approach to be less accurate for prediction, potentially due to the propagation of errors in forecasts beyond $k = 1$.

IV. Evaluation Methods

All of the predictive models are trained and evaluated using crossvalidation. For each model, each of ten years of data is defined as a test/holdout set and the model is trained on the other nine years. The ten sets of test predictions for each model are then averaged to generate a single cross-validated score.

We use root-mean-square error (RMSE) of the model's predictions as our primary metric for evaluation. To compute the RMSE for a particular model, we compute the error over all T days and L locations in a test set and average the results from the ten cross-validation runs

$$\mathrm{RMSE}_{\mathrm{model}} = \left(\frac{1}{TL}\sum_{t=1}^{T}\sum_{l=1}^{L}\left(y_l^{(t)} - \mathbb{E}\left[y_l^{(t)}\right]\right)^2\right)^{\frac{1}{2}}.$$

We also use RMSE skill score $\mathrm{SS}_{\mathrm{model}}$ [27] using the persistence model (predicting the expected count on day t + k

as the observed count on day t) as a reference baseline

$$\mathrm{SS}_{\mathrm{model}} = \left(\frac{\mathrm{RMSE}_{\mathrm{persistence}} - \mathrm{RMSE}_{\mathrm{model}}}{\mathrm{RMSE}_{\mathrm{persistence}}}\right) * 100\%.$$

This metric shows the relative improvement of a model compared to the persistence model. A perfect forecast model would have a skill score (SS) of 100, a model equivalent to the reference would have a score of 0, and a model worse than the reference would have a negative score.

In our experimental results later in this article, we evaluate our predictive models under two different scenarios: 1) only during active fires (Fire days) and 2) on the full data set (all days). This breakdown allows us to analyze the predictive performance of the models in detail in different contexts, i.e., during active fires as well as in general. *Fire days* are defined as pairs of grid cells and days where the preceding day had at least one detection in that cell, i.e., a forecast is being made in the context of an active fire in the cell. *Non-Fire days* are defined, conversely, as grid cell/day pairs where there was no detection on the day prior to the forecast in that cell, i.e., a forecast is being made in the context of no active fire. We can further partition each category as follows, depending on the observation for the day being forecast. For Fire days the two possibilities are: 1) *sustain* (the active fire continues to have detections in that cell on the following day) or 2) *extinction* (the detections go to zero in that cell). For non-Fire days the possibilities are: 1) *quiescence* (there continues to be no detections in that cell) or 2) *ignition* (there is a new detection in that cell). Statistics for these four different combinations of events are provided in Table II as a function of the lead time k. The quiescent state dominates the statistics (consecutive days with no detections)—note that for regions other than Alaska the frequency of quiescence could be quite different. For Fire days, the likelihood of extinction increases steadily as the lead time k increases. Fig. 4 shows that, as expected, Fire days have a significant increase in the frequency of nonzero counts at all levels.

A. Baseline Persistence Model

This is a simple, frequently used baseline that uses the value on the current day t as the forecast value for day $t+k$, often referred to as a *persistence model*

$$\mathbb{E}\left[y_l^{(t+k)}\right] = y_l^{(t)}.$$

This model is limited in that it does not capture changes in fire dynamics and its prediction accuracy degrades significantly as the forecast range k is increased. This is the model deployed by most operational smoke forecasting systems [6]–[8].

B. Modulated Persistence Model [9]

This model modulates the persistence model using the change in the value of the Forest Fire Weather Index (FWI) between the current day t and the day being predicted $t+k$. In this article, we replace FWI with VPD where VPD is computed using the relative humidity and temperature from the ERA dataset. Prior work has shown that VPD and FWI often correlate similarly with Alaskan fire variability [28], so we expect the use of VPD to have a minimal impact on the model compared to using FWI.

In particular, the modulated persistence model is defined as

$$\mathbb{E}\left[y_l^{(t+k)}\right] = y_l^{(t)} * (1 + \mathcal{M}(t+k))$$
$$\mathcal{M}(t+k) = \frac{\mathrm{VPD}^{(t+k)} - \mathrm{VPD}^{(t)}}{\mathrm{VPD}^{(t)}}.$$

By incorporating this modulation factor $\mathcal{M}$, the prediction adjusts the standard persistence estimate by the percentage change in the VPD. To avoid numerical issues, if $\mathrm{VPD}^{(t)}$ is equal to zero, the model defaults to a standard persistence prediction. Furthermore, to avoid exceedingly large predictions we allow the modulation factor to be at most 2; this value was chosen by a cross-validated grid search to minimize test error.

V. Predictive Models

Below we define the different types of predictive models that we evaluate ranging from baseline models using persistence to Poisson regression models.

A. Poisson Regression

A suitable approach in general for predicting nonnegative counts is to use Poisson regression. This method assumes a conditional Poisson distribution for y, modeling the logarithm of the expected counts as a linear function of the input variables [29]. This formulation avoids the potential problem of negative predictions for counts that can occur with least-squares linear regression and a Gaussian noise assumption. We can write the Poisson regression model in exponential form as follows:

$$\mathbb{E}\left[y_l^{(t+k)}\right] = \hat{\lambda}_l^{(t+k)} = \exp\left(\theta_0 + \theta_1 h\left(y_l^{(t)}\right) + \boldsymbol{\theta}_2^T \boldsymbol{x}_l^{(t+k)}\right)$$

where $\hat{\lambda}_l^{(t+k)}$ denotes an estimate of the mean for the conditional Poisson distribution for $y_l^{(t+k)}$. The parameters θ are estimated by maximizing the log-likelihood of the observed data under a Poisson model. A transformation h is applied to the AR component of the model $y_l^{(t)}$ above because the relationship between the expected count and the previously observed count is not expected to be multiplicative (which is introduced by the use of the exponential function). To address this, we take the logarithm of $y_l^{(t)}$ as suggested in [29], with

$$h(y) = \ln(\max(y, c)), \quad 0 \le c \le 1$$

to avoid taking logarithms of zero when the previously observed count y is zero. The value of c could be estimated from the data as a parameter, but we found that setting it to a value of 0.5 (as suggested in [29]) performed well empirically.

Prior work on using Poisson regression models to predict fires includes the work of Mandallaz and Ye [30] who demonstrated that Poisson regression models provide a useful framework for predicting the total number of forest fires on daily time-scales over different large-scale regions in Central Europe, as a function of weather, physical, and socio-economic variables. In a similar vein, Preisler *et al.* [31] partitioned the state of Oregon in the USA into 1 km^2 cells and used

logistic regression to model the probability of a fire occurring in a given cell as a function of various explanatory variables. They then extended this approach to predict the total number of fires on any given day over the whole region by summing up predictions across individual cells. These approaches [30], [31] differ from this article primarily in that they are aimed at understanding different factors that can influence the number of fires recorded on any given day, rather than developing prediction schemes that can provide operational forecasts k days into the future for the purpose of smoke forecasting (which is the focus of our approach).

A common alternative to the Poisson model for forecasting count data is the use of the negative binomial model, which has a separate mean and variance, adding more flexibility to the model. However, we found that using the negative binomial model for regression did not result in improved prediction performance over the Poisson model for the datasets used in this article.

Because grid cells are measured in latitude and longitude, there is a notable decrease in the area of grid cells toward the north latitudes. To address this we formulate the model to predict fire detection intensity per unit area and then scale the prediction by the area of the cell being predicted. This approach was successfully applied to the Poisson regression of fire detections in [30]. We found that this had a negligible effect on overall model performance. This is likely attributable to the fact that the majority of detections occur in similar latitudes, near the middle of the state, and to the fact that there are other spatial differences (including climate, topography, and land cover) that vary with latitude.

B. Poisson Hurdle Model

A well-known extension to the Poisson model is the Poisson hurdle model [32], allowing the model to capture the conditional probability of additional zeros more easily. The distribution and the corresponding expected value are defined as follows, where π is the probability of a zero occurring and where f is a Poisson distribution:

$$P(y=j)=\begin{cases}\pi, & \text{if } j=0\\ \dfrac{(1-\pi)}{1-f(0)}f(j), & \text{if } j>0\end{cases}$$

$$\mathbb{E}\left[y_l^{(t+k)}\right]=\frac{\left(1-\pi_l^{(t+k)}\right)}{1-f\left(0|\hat{\lambda}_l^{(t+k)}\right)}\hat{\lambda}_l^{(t+k)}.$$

Rather than treating the probability π as a fixed parameter, a more flexible approach is to model it via regression, i.e., to estimate $\pi_l^{(t+k)}$ for each grid cell l on each day t, conditioned on observed values. Since π is a probability bounded between 0 and 1, logistic regression is a natural choice, with $\pi_l^{(t+k)}$ modeled as

$$\pi_l^{(t+k)}=\sigma\left(\alpha_0+\alpha_1 y_l^{(t)}+\boldsymbol{\alpha}_2^T\boldsymbol{x}_l^{(t+k)}\right)$$

where $\sigma(z)$ is the logistic (or sigmoid) function, defined as $\sigma(z)=1/(1+e^{-z})$. We model both π and λ as functions of the same covariates (i.e., $\boldsymbol{x}$ and past values of y), but with separate parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$ for each model.

We find in our experimental results that the Poisson hurdle model consistently performed as well as or better than the standard Poisson model (with no hurdle component), across different prediction scenarios. Given this, we only present results in the remainder of this article for the Poisson hurdle model for clarity of presentation. We also found that linear regression models tended to be competitive with Poisson hurdle regression in terms of prediction accuracy, with the Poisson hurdle being consistently slightly more accurate. However, the linear regression models suffered from the problem of occasionally producing negative predictions. Thus, given that the linear regression models do not provide any advantages in the context of predicting fire activity (relative to Poisson models), we do not discuss them further in this article.

C. Model Memory

We explored several approaches including temporal memory (additional lags) in the model, but empirical results indicated that no method produced substantial improvements in predictive accuracy. However, we found that the use of an additional lagged AR term $y^{(t-1)}$ made a qualitative improvement in model forecasts for the initial day of fire events. As such, when evaluating a model with AR terms, both $y^{(t)}$ and $y^{(t-1)}$ are used. Additional details describing the methods for incorporating memory can be found in the Appendix.

VI. Results: Forecasts for Fire Days

We begin our discussion of predictive accuracy by comparing the performance of different models on Fire days, as defined in Section IV and in Table II. As described earlier, all results reported below were obtained by evaluating the models using cross-validation across years. Code for data fetching, preprocessing, training, and prediction is publicly available.[2] This code allows for full replication of the results in this article.

A. Poisson Hurdle Variable Selection

We begin by evaluating the relative contributions of different combinations of input variables, as reflected by the predictive accuracy of the Poisson hurdle model. Table III shows the average RMSE on Fire days for different combinations of variables, including an intercept-only model for reference, models with different weather variables (humidity, temperature, VPD, rain, wind), and models with and without an AR component. The RMSE used in this table is the average RMSE across all five lead times. We see that models with an AR component tend to systematically reduce the RMSE by about 4% compared to models without an AR component and models with VPD are more accurate than models with either humidity or temperature. The addition of rain (and wind) provides negligible gains in accuracy to models with VPD. Based on these observations, in the remainder of this article we will focus on models using VPD with an AR component—this is a model that is easy to interpret and is in effect as accurate as other models with additional combinations of weather variables.

[2]https://github.com/UCIDataLab/fire_prediction

TABLE II

PERCENTAGE OF OBSERVATIONS BY CATEGORY, CONDITIONED ON FIRE/NON-FIRE AND LEAD TIME k

	Lead Time				
	1	2	3	4	5
Fire (F) Days (0.49%)					
Sustain (Active→Active)	60.33%	47.38%	38.61%	33.03%	29.67%
Extinction (Active→Inactive)	39.67%	52.62%	61.39%	66.97%	70.33%
Non-Fire (NF) Days (99.51%)					
Ignition (Inactive→Active)	0.20%	0.26%	0.30%	0.33%	0.35%
Quiescent (Inactive→Inactive)	99.80%	99.74%	99.70%	99.67%	99.65%

TABLE III

AVERAGE RMSE OF POISSON HURDLE MODELS ON FIRE DAYS

Weather Covariates	RMSE	
	Without AR	With AR
None	21.0	20.1
Temperature	19.9	19.2
Humidity	20.1	19.3
VPD	**19.8**	**19.0**
VPD, Rain	**19.8**	**19.0**

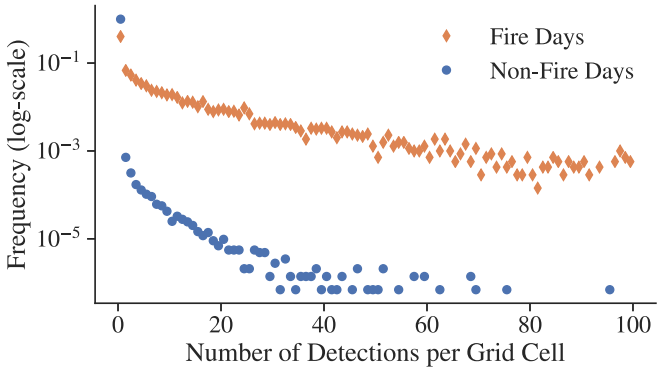

Fig. 4. Frequencies of grid cell counts from both Fire and non-Fire days; truncated at 100 on the x-axis. Fire days occur when the number of detections for a given cell on the *preceding* day was nonzero, and similarly non-Fire days occur when the number of detections for a given cell on the *preceding* day was zero. As such, Fire days can have a zero count due to extinctions and non-Fire days can have a nonzero count due to ignitions.

Fig. 5 illustrates predictive error (RMSE) as a function of lead time k, for the VPD + AR models compared to the AR-only models and the VPD-only models ($K = 5$ different models for each). These lines are the disaggregated versions (across k) of the aggregate RMSE numbers for the corresponding models in Table III. There is a notable difference in the relative errors of the different models across the lead time. The combined model (VPD + AR) clearly outperforms the other two models across all values of k. The AR-only model performs substantially better than the VPD-only model when $k = 1$ and the VPD-only models perform slightly better than AR-only for $k = 2, \ldots, 5$. This indicates that accurate weather information is more useful into

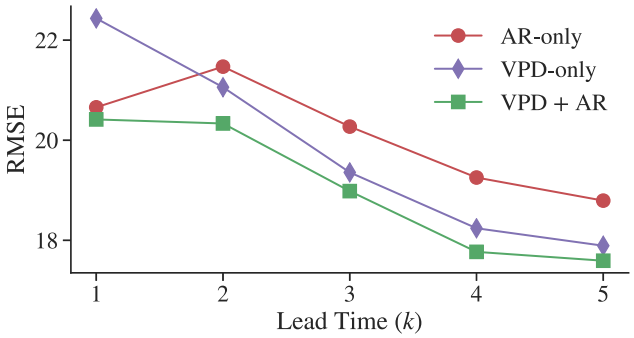

Fig. 5. RMSE for Poisson hurdle models using different weather covariates as a function of lead time on Fire days.

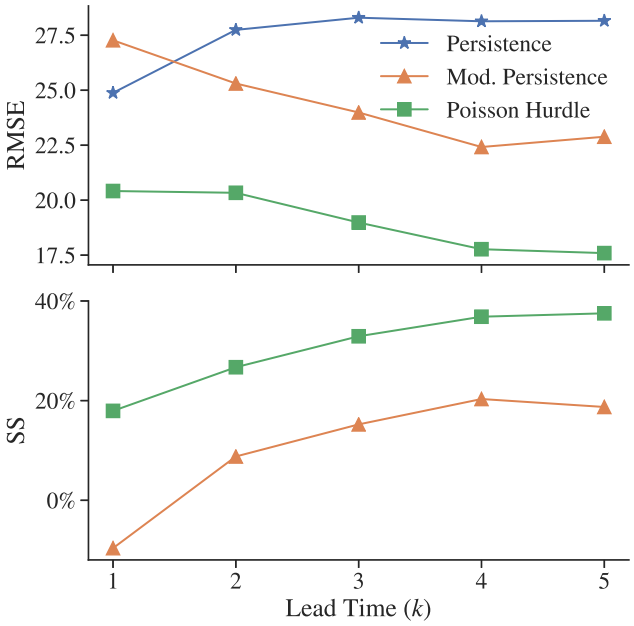

Fig. 6. Forecast performance of different models on Fire days. (Top) RMSE scores (lower scores are better). (Bottom) SSs relative to persistence (larger scores are better).

the future than information about the current fire size (bearing in mind that we use reanalysis data instead of forecasts).

The fact that the RMSE values of the different models tend to *decrease* as the lead time k increases is somewhat counterintuitive since one would expect RMSE to generally increase the further out in time the forecasts are being made. This decrease in RMSE is explained by the fact that these forecasts are being made only on Fire days. As the lead time increases, the mean number of counts per day for an active fire decreases (due to fire extinction), producing less variance in the true counts for larger lead times, making the prediction problem easier on average (and hence, lower RMSE) for days that are further out.

B. Comparing Poisson Hurdle and Persistence Models

Fig. 6 shows both the RMSE (top) and the SSs (bottom) as a function of lead time k, Poisson hurdle models with VPD + AR, for the persistence and modulated persistence baselines defined in Section V. The RMSE (top)

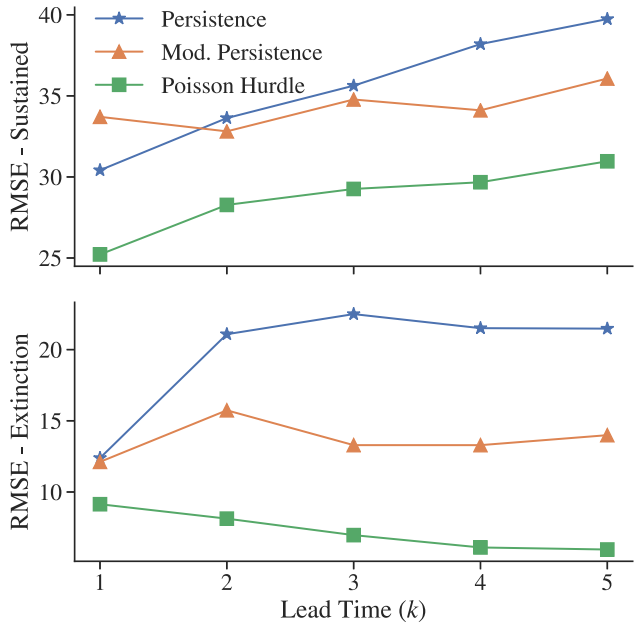

Fig. 7. Decomposition of RMSE into two components, with (Top) predictions for sustained events and (Bottom) predictions for extinctions on Fire days.

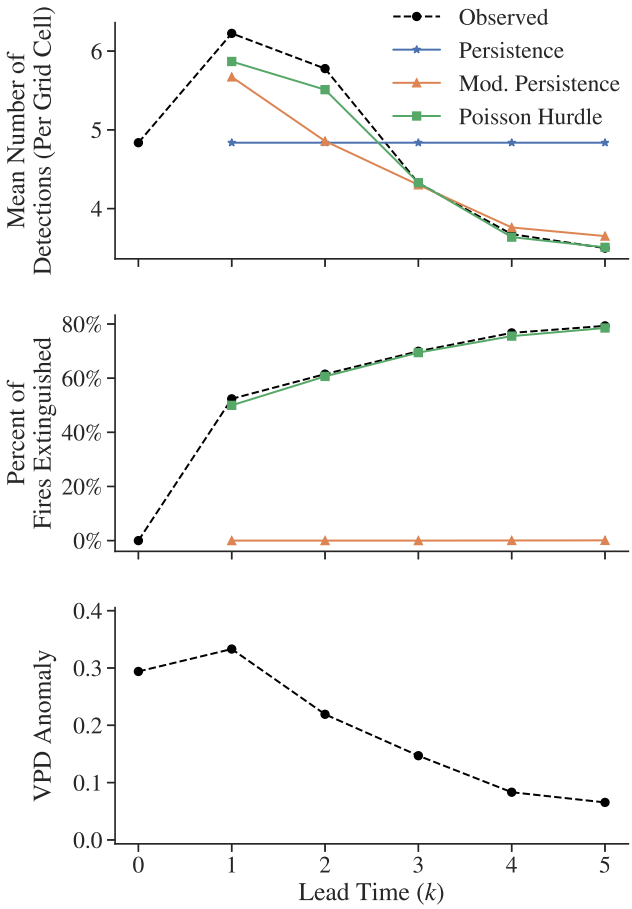

Fig. 8. Comparison of (Top) mean predicted number of detections by different models from the first day of a fire, (Middle) percent of grid cell/day pairs active on day 0 that have extinguished, and (Bottom) VPD anomaly from the ten-year fire season mean for interior Alaska (.53).

plot clearly illustrates that the persistence forecasts degrade over time. The modulated persistence and Poisson hurdle have substantially lower RMSE than persistence (except for $k = 1$ where modulated persistence is less accurate than persistence).

The bottom (SS) plot shows that both the modulated persistence and Poisson hurdle models have SSs that increase as the lead time increases, due to the fact that the persistence model's RMSE becomes increasingly worse relative to the other two models with an increasing lead time. The Poisson hurdle model has a substantially better SS than the modulated persistence model, achieving SSs above 30% at lead times of $k = 3, 4, 5$, compared to the 18%–20% SSs from the modulated persistence model for the same lead times.

Fig. 7 shows a breakdown of the RMSE scores (from the top of Fig. 6) where the RMSE has been separated into two components. The first component (top) is the RMSE of predicting sustained fires and the second component (bottom) is the RMSE contribution from predicting fire extinctions. Note that the two components represent different numbers of observations (see Table II), so the RMSE values in Fig. 6 are a weighted average of the RMSE values in the two plots in Fig. 7. The Poisson hurdle model is more accurate (much lower RMSE) at predicting both sustain events (upper plot) and extinction (lower plot) than the modulated persistence model or the persistence model. The upper plot, the RMSE for predicting sustained fire activity, demonstrates that this prediction task becomes increasingly difficult farther into the future. Much of the error here comes from the underprediction of large growing fires. The lower plot, the RMSE for predicting extinction, is quite different. As the lead time increases, a larger fraction of observations will be extinctions due to the end of a fire event (Table II), with the corresponding lower mean value and variance in the detection counts as k increases and, thus, the model can more reliably predict the smaller values.

Fig. 8 illustrates the nature of the predictions that the different models make at the start of a fire. To do this, we select all true ignition events by selecting the N grid cell/day pairs which have a nonzero count and the same grid cell on the previous day has a zero count ($y_l^{(t)} > 0$ and $y_l^{(t-1)} = 0$). We then make predictions from each of the N starting grid cell/day pairs for the next five days using the $k = 1, \ldots, 5$ models, resulting in N predicted values for each of the five following days, and compute the mean value of the N predictions for each day resulting in one mean value per model per day.

The upper plot shows the mean value per day for the same three models as before (persistence, modulated persistence, and Poisson hurdle). The observed mean for each day is also included since ideally each model should match this in expectation: the expected number of detections for an active fire tends to increase for the first two days and then decrease after that. We see that the various models track this mean behavior over time, except for the persistence model, which has a constant prediction since it does not change its prediction as a function of k.

TABLE IV
AVERAGE RMSE AND SSS ON FIRE DAYS

Model	RMSE	SS
MLP (VPD, AR)	19.3	29.57%
MLP (all weather covariates)	19.5	29.06%
MLP (with lag 2 memory)	19.6	28.59%
Hurdle Poisson	**19.0**	**30.69%**

TABLE V
AVERAGE RMSE AND SSS ON ALL DAYS

Model	RMSE	SS
Persistence	2.15	–
Modulated Persistence	1.97	8.30%
Two-Poisson	1.82	15.25%
Single-Poisson Hurdle	1.70	21.12%
Two-Poisson Hurdle	**1.68**	**22.00%**

Fig. 8 (middle) plot compares the expected fraction of counts that are zero between the Poisson hurdle model and the modulated persistence model with the observed number of extinctions for grid cell/day pairs. For the Poisson hurdle model, the conditional distribution is used instead of its predictions. The persistence model is excluded because it is incapable of predicting extinctions. The number of extinctions increases over time and this is closely tracked by the distribution from the Poisson hurdle model, whereas the modulated persistence rarely forecasts an extinction (i.e., a count of zero).

Finally, Fig. 8 (bottom) illustrates that the VPD values tend to be shifted away from their mean on the days that fires start (lead time zero) and tend to regress toward their mean over time. This helps demonstrate the effectiveness of VPD as a covariate and helps to explain why the use of VPD to modulate persistence mitigates overprediction for larger lead times.

C. Comparing Poisson Hurdle and Multilayer Perceptron (MLP)

We also perform experiments comparing the Poisson hurdle model with the MLP, a feed-forward, a fully connected neural network trained with backpropogation [33]. We use the Adam optimization algorithm [34] with a learning rate of 0.01, and explored a number of reasonably sized layer configurations. Our results (summarized in Table IV) show that the Poisson hurdle model performs comparably or superior to all evaluated MLPs. We found a two hidden-layer configuration with 16 neurons in the first hidden layer and eight neurons in the second performed best. We again found that using VPD and an AR term performed best for the MLP. We also explored adding additional combinations of covariates including all-weather covariates (temperature, humidity, wind, and rain) and including additional memory terms, but found these generally overfit and performed worse out-of-sample.

VII. RESULTS: FORECASTS FOR ALL DAYS

In this section, we evaluate the predictive accuracy of different models across all days in our dataset, i.e., not just for Fire days but across all rows in Table II. Although our primary focus in evaluating different models is in terms of their performance for forecasting on Fire days (as discussed in the previous section), for completeness we are also interested in how the models perform when also used to make predictions for ignition and quiescent events.

A. Two-Poisson Hurdle Models

We find that using two different regression models, one for Fire days and one for non-Fire days, improves predictive performance compared to using a single regression model for both. We use the same type of model for each that we used for Fire days, namely a Poisson hurdle model, with separate parameters for each. The hurdle component is helpful for the non-Fire data (as with the fire data—see Fig. 4).

The two Poisson hurdle models are trained on two separate partitions of the grid cell/day data, depending on whether they are identified as Fire or non-Fire days. At prediction time (i.e., on out-of-sample test days during crossvalidation in our experiments) each day that a forecast is being made is identified as a Fire day or a non-Fire day and the corresponding model makes the forecast.

B. Predictive Accuracy of Single- and Two-Poisson Models

Table V summarizes the reductions in RMSE obtained with various combinations of models. The use of the hurdle component or the two-Poisson model leads to significant improvements in performance over the persistence model, with SS improvements of around 15%–20%. The combined use of the hurdle component and two-Poisson model produces a small improvement relative to the single-Poisson hurdle model.

Because of the sparse nature of the detection data, most grid cells on a given day t do not contain any active fires and will not have any active fires on day $t + k$. As a consequence, when used to forecast both fire ignitions and continued activity, a single-Poisson model will tend to produce expected-valued predictions that are near zero and rely heavily on the AR term to predict activity. This produces significant overprediction during large fire events that are heavily penalized by RMSE, e.g., RMSE worse than persistence-based approaches. In contrast, the Poisson hurdle and two-Poisson models can effectively separate the prediction of the large fraction of quiescent grid cell/day pairs from the prediction of the sparse fire activity using their separate components.

We find that the non-Fire component of the model relies heavily on VPD at long forecast horizons which, when combined with the exponential from the Poisson link, leads to substantial overprediction for outlier VPD. Empirically, scaling VPD with $\log(\text{VPD} + 1)$ works well, so in all cases, the non-Fire component uses this transformation for its input.

C. Interpreting the Coefficients of Two-Poisson Models

Tables VI and VII show the estimated coefficients for each of the Fire and non-Fire components of the two-Poisson model. The covariates were standardized to have mean zero and standard deviation of 1 in order to put the coefficient values on roughly the same scale. For these tables, we do not use the

TABLE VI
COEFFICIENTS FOR THE TWO-POISSON HURDLE ($k = 1$)

Model	Intercept	VPD	$y^{(t)}$	$y^{(t-1)}$
Poisson - Sustain	2.55	0.16	0.48	0.08
Hurdle - Extinction	-0.54	-0.40	-0.85	0.00
Poisson - Ignition	1.23	0.22	–	0.02
Hurdle - Quiescent	6.83	-0.86	–	-0.15

TABLE VII
COEFFICIENTS FOR THE TWO-POISSON HURDLE ($k = 5$)

Model	Intercept	VPD	$y^{(t)}$	$y^{(t-1)}$
Poisson - Sustain	2.56	0.35	0.19	0.02
Hurdle - Extinction	1.05	-0.95	-0.36	-0.15
Poisson - Ignition	1.77	0.31	–	0.01
Hurdle - Quiescent	6.40	-0.99	–	-0.13

models scaled by the grid-cell area to ensure that the intercept terms are easily interpretable; the models with and without area scaling have almost perfect coefficient agreement except for the intercept terms.

The coefficients broadly agree with physical intuition. Specifically, the signs of the coefficients for VPD and $y^{(t)}$ are positive for the nonhurdle (count) components of both models, e.g., increases in VPD or $y^{(t)}$ on a given day correspond to higher predicted detection counts from the model for the following days. In the hurdle components of each model the coefficients for VPD and $y^{(t)}$ are negative reflecting the fact that increases in the values of VPD or $y^{(t)}$ lead to a decreased likelihood of observing an extinction or nonignition.

We also find that for a lead time of $k = 1$ (Table VI), the AR term has a relatively large coefficient value that decreases as the lead time increases to $k = 5$ (Table VII). In contrast, the coefficient on VPD becomes larger as the lead time k increases from 1 to 5, reflecting our earlier findings that weather is more important for prediction than the number of past detections as k increases. In addition, we observe that the intercept of the Poisson component of the active model increases slightly as the lead time increases, which may reflect the fact that a fire that is sustained for longer will be larger on average. Likewise, the intercept of the hurdle component of the active model increases substantially, reflecting the fact that the likelihood of extinction increases when predicting further into the future.

Coefficients for $y^{(t)}$ are omitted for the non-Fire components because all $y^{(t)}$ are zero for this component due to the definition of non-Fire days.

D. Evaluating the Non-Fire Component of the Model

Since the non-Fire model makes predictions on non-Fire days, and since the non-Fire days are dominated by the quiescent state (zero to zero, as in the last row of Table II), a baseline model that always predicts zero will do quite well in this context and be difficult to beat in terms of RMSE. As a result, the non-Fire Poisson hurdle model performed similarly to the baseline zero-only prediction ($\hat{y}^{(t+k)}_{\text{non-Fire}}$ is set to always predict zero) when predicting on non-Fire days in terms of RMSE. This zero-only prediction baseline also performs almost identically to always predicting the mean value for non-Fire days because of the large fraction of inactive days in the data (i.e., the mean is very small).

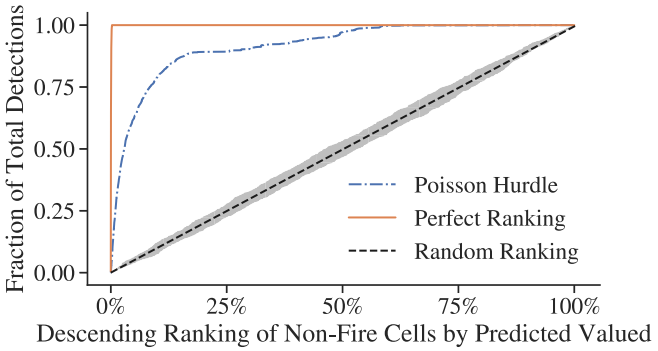

Fig. 9. Cumulative fraction of total detections (from ignitions) accounted for based on non-Fire grid cell/day pairs ranked by model prediction for lead time $k = 5$. The gray uncertainty band around the expected random ranking line is generated using the max and min values from 100 random rankings.

Despite the fact that the non-Fire Poisson hurdle model does not outperform baselines in terms of RMSE, it nonetheless does extract useful information about the likelihood of ignition events over nonignitions. Fig. 9 shows the results of comparing the rankings of the non-Fire model with the rankings of a constant predictor (which in effect generates random rankings). All observations from the non-Fire data (ignitions and quiescent days) are ranked according to the model prediction in descending order along the x-axis and each observation's contribution to the total non-Fire counts along the y-axis. An ideal model would produce a ranking in which all observations are sorted by their observed count (as shown by the "Perfect Ranking" line). We see a substantial improvement in the model's rankings compared to random. The results shown are for a lead time of $k = 5$—they are similar for all of the lead times. The ranking results indicate that the non-Fire component of the two-Poisson model is picking up a significant signal from the weather variables (specifically, VPD) to predict the likelihood of an ignition. Predicting the precise number of detections associated with ignition events (which is what RMSE is evaluating in terms of the nonzero outcomes) is a much more difficult problem where constant baselines are difficult to beat. Hence, this task may be best framed as predicting the probability of ignition instead of predicting the number of detections for an ignition.

E. Regional Performance

We can also analyze global performance of the models across the entire region by aggregating the observations and predictions across all grid locations for each day. We define a regional RMSE by computing the RMSE between the total number of true detections per day and the aggregated sum of the predictions across all grid cells per day

$$\text{Regional RMSE} = \left(\frac{1}{TL} \sum_{t=1}^{T} \left(\sum_{l=1}^{L} y_l^{(t)} - \sum_{l=1}^{L} \mathbb{E}\left[y_l^{(t)} \right] \right)^2 \right)^{\frac{1}{2}}.$$

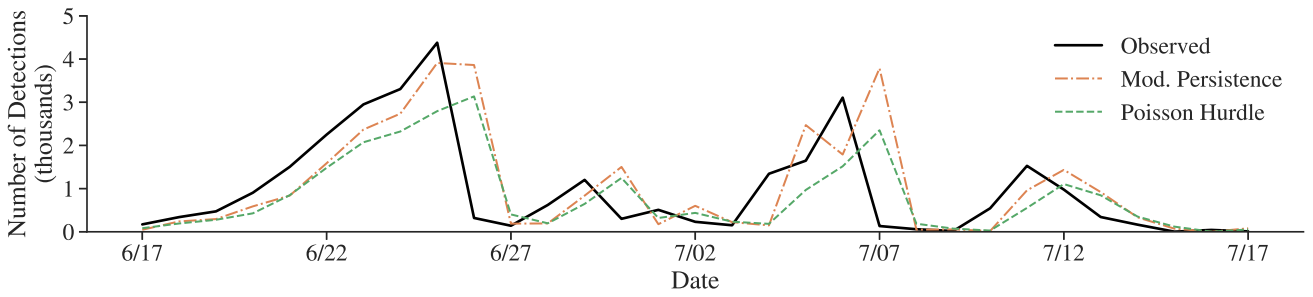

Fig. 10. Forecasts with lead time $k = 1$ from the two-Poisson Hurdle and the Modulated Persistence models, for June 17 to July 17, 2015 (a highly active period).

TABLE VIII
REGION RMSE OF DIFFERENT MODELS ($k = 1$) ON ALL DAYS

Year	Persistence	Modulated Persistence	Two-Poisson Hurdle
2007	1.26	1.10	0.99
2008	0.37	0.33	0.30
2009	9.12	10.78	7.92
2010	1.80	1.89	1.60
2011	0.79	0.82	0.76
2012	0.50	0.52	0.42
2013	2.51	2.32	2.00
2014	0.59	0.53	0.56
2015	13.53	13.60	11.50
2016	3.61	3.26	2.89
Mean	3.41	3.51	**2.89**
Std. Dev	4.19	4.47	3.58

TABLE IX
REGION RMSE OF DIFFERENT MODELS ($k = 5$) ON ALL DAYS

Year	Persistence	Modulated Persistence	Two-Poisson Hurdle
2007	1.97	1.59	1.41
2008	0.57	0.48	0.43
2009	11.88	8.37	9.95
2010	3.30	2.69	2.54
2011	1.18	1.05	0.90
2012	0.80	0.67	0.50
2013	4.15	3.43	2.92
2014	0.93	0.75	0.80
2015	20.98	18.77	16.66
2016	4.79	3.99	3.52
Mean	5.05	4.18	**3.96**
Std. Dev	6.20	5.36	5.01

TABLE X
REGION RMSE OF DIFFERENT MODELS ($k = 1$) ON ALL DAYS

Year	Persistence	Modulated Persistence	Two-Poisson Hurdle
2007	1.26	1.10	0.99
2008	0.37	0.33	0.30
2009	9.12	10.78	7.92
2010	1.80	1.89	1.60
2011	0.79	0.82	0.76
2012	0.50	0.52	0.42
2013	2.51	2.32	2.00
2014	0.59	0.53	0.56
2015	13.53	13.60	11.50
2016	3.61	3.26	2.89
Mean	3.41	3.51	**2.89**
Std. Dev	4.19	4.47	3.58

TABLE XI
REGION RMSE OF DIFFERENT MODELS ($k = 5$) ON ALL DAYS

Year	Persistence	Modulated Persistence	Two-Poisson Hurdle
2007	1.97	1.59	1.41
2008	0.57	0.48	0.43
2009	11.88	8.37	9.95
2010	3.30	2.69	2.54
2011	1.18	1.05	0.90
2012	0.80	0.67	0.50
2013	4.15	3.43	2.92
2014	0.93	0.75	0.80
2015	20.98	18.77	16.66
2016	4.79	3.99	3.52
Mean	5.05	4.18	**3.96**
Std. Dev	6.20	5.36	5.01

The resulting regional RMSE scores, shown in Tables VIII–XI, indicate that the two-Poisson hurdle models greatly outperform the persistence-based models at the region level for $k = 1$ and slightly outperforms them for $k = 5$.

Qualitatively, we also observe a systematic difference between the actual predictions of the modulated persistence and two-Poisson hurdle models. Fig. 10 shows the observed data and forecasts from each model, aggregated to the region level, for a one-month period during the 2015 fire season. Although both models tend to underpredict the growth of large fire events (e.g., June 17–25 and July 3–6), the Poisson hurdle model tends to underpredict by a larger amount. However, this more conservative approach has the advantage of decreasing the likelihood of overpredicting large growth (July 5) and also leads to less overprediction when fires rapidly extinguish (June 26 and July 7). This difference also appears with smaller fire events, but is less pronounced. The overall effect, which occurs at both the grid-level and regional RMSE, is that the increased error from underprediction is compensated for by the decreased error due to overprediction, when compared to forecasts from the modulated persistence model.

VIII. Discussion

The techniques discussed above using regression models show promise for forecasting wildfire detections and are broadly applicable to any region. Given historical data, the regression models can be trained to capture the characteristics of wildfire evolution in a particular region. For example, in regions where the wildfire statistics are quite different from those in Alaska, such as the Amazon basin or Central Africa where there are frequent small human-caused fires to clear forest and brush, models fitted to data from such regions would likely predict much higher ignition rates and shorter duration fire events.

As an example, one insight from the models fitted to the Alaska data in this article is that a key component of the regression model's improved performance over modulated persistence is the model's intercept term which encodes the fact that while individual fires (if sustained) continue to grow, on average the fires increase in size from the first to second day and then decrease at an accelerating pace. We find that an intercept-only prediction model can capture this expected behavior and that predicting this overall mean trend alone is sufficient to improve on the predictive accuracy of either persistence approach.

Upon comparing the proposed statistical approach to the baseline models (both constant persistence and modulated persistence), it is worth noting that the baseline models cannot learn from data since they have no learnable parameters. In contrast, in our proposed approach the model parameters can be reestimated and updated over time as new data arrives. In addition, although we did not leverage it in the results in this article, the statistical modeling approach can provide estimates of predictive uncertainty due to uncertainty in parameter estimates as well due to as uncertainty in the conditional predictive mixture distribution for y, in contrast to the baseline approaches which have no notion of predictive uncertainty.

Below we discuss a number of other aspects of both the modeling approach and the data employed that could potentially be improved to systematically increase the prediction accuracy of the forecasting approaches.

A. Potential Modeling Improvements

The regression-based model we propose assumes conditional independence of the y's, both temporally and spatially, given the model covariates. There are a variety of modern spatio-temporal statistical modeling approaches that allow for relaxation of this assumption and that have the potential to leverage any additional temporal and spatial dependence structure for predictive purposes (e.g., [35]–[37]). Such approaches have been applied in the wildfire context to problems such as estimating spatial fire risk maps and modeling temporal patterns of postfire vegetation recovery [38], although not to the specific problem we investigate here of predicting daily wildfire activity.

In this context, we conducted a number of explorations of the potential use of additional temporal and spatial dependence to produce more accurate count predictions. For active fire predictions from our model, we found that the linear correlation between residuals from one day to the next was -0.02 (whereas the correlation of the corresponding counts was 0.37). The fact that the lag-1 residuals from our model are temporally uncorrelated (effectively) implies that there is limited opportunity for improving predictive power via temporal residuals.

We also investigated potential spatial bias in our models by computing the mean error $E[\hat{y}_l^{(t)} - y_l^{(t)}]$ over time for each grid location l. Although the resulting spatial maps showed significant spatial dependence in these bias terms over the full ten-year study period, we also found that the specific patterns of spatial dependence are highly dependent on the period of observation. In particular, when we computed the spatial bias map for the first five and the second five years, the correlation between the biases in the two spatial maps was 0.01, i.e., the patterns of bias do not persist over time. The reason for this is that the spatial bias is strongly influenced by the presence of specific fires from year to year and there is little persistent systematic spatial pattern that is not already captured by covariates (such as weather) in the model. We validated this observation empirically by augmenting our model with a mean (intercept) random effect term $\theta_{0,l}$ per grid location l. The resulting predictions of the model out-of-sample did not improve over those of the model without the grid-specific effects.

In addition, we computed correlations between active cells at time t and their neighbors at time $t+1$, for all four count-residual pairs. We found these correlations were in the range of 0.05–0.09, indicating relatively little predictive power for single-lag, single-neighbor predictions beyond what is already predicted by the model.

The observations above should not be taken to imply that more sophisticated spatio-temporal modeling approaches might not yield better predictions. For example, in terms of temporal dependence, a model that adaptively and dynamically estimates a "local fire effect" to modulate the global parameters of our proposed model could be a promising direction, e.g., to capture systematic differences between small and large fires. However, given the relatively small amount of data available for this article (in terms of the total number of nonzero grid cell counts) we leave such extensions to possible future work.

B. Data Improvements

We conjecture that a key factor that limits the predictive accuracy of all of the models in this article is the spatial resampling of grid data. The spatial resolution of 0.5° is too coarse to capture fire spread dynamics and likely prevents the model from effectively leveraging the land cover information to improve forecasting.

To improve the spatial resolution of the model one could reduce the size of the grid cells for representing the data, but this would come with several notable drawbacks: increased sparsity of detection activity, increased crossing of cell borders by active fires, and increased computation. An alternative approach would be to spatially model fire events directly in continuous 2-D space, e.g., using clustering to group active fire detections and then make predictions for each cluster.

Another important future direction is to assess decreases in model performance when archived forecast estimates of weather are used as model drivers instead of reanalysis. If true weather forecasts were used instead of reanalysis we would expect that the degradation farther into the future would increase for the regression model. As such, some of the benefits of weather data when making predictions for larger lead times would be diminished. A comparison of forecast and reanalysis could also provide information about how the degradation of fire weather forecasts varies regionally. This information would be valuable for many fire management applications.

The fire detection data suggest that there are likely active fire detections that are unobserved. For example, there are often single-day interruptions where a cell with significant activity will become completely inactive on the following day and then resume activity following that. While there may be legitimate causes of single-day fire suppression, e.g., rain extinguishing fires, it is likely that some fraction of these interruptions is caused by false-negatives introduced by cloud cover which can obscure the signal. These false zeros artificially penalize the regression model predictions and further encourage underprediction. Future work could account for this by detecting likely false-negatives and either exclude them or attempt to interpolate the missed detections.

Another important direction for future work is to better utilize rain data. There is a strong signal for individual fire events, especially for large fire events, in that large rain events often coincide with large extinction events (e.g., compare the upper and lower panels in Fig. 3). However, for the datasets used in this article, we found that the models do not show notably improved results when incorporating rain. This may be attributable to spatial mismatches between subgrid-scale convective storms and fire perimeter locations or the high correlation between precipitation and VPD. Analysis of the data suggests that there should be predictive information about fire activity available from rain data, but that the models are not capturing it.

IX. Conclusion

We have shown that the use of regression models significantly outperforms traditional persistence-based models used in operational smoke forecasting applications at both the cell and regional levels. In particular, the use of Poisson hurdle models can effectively use past fire activity and weather data and provide significantly more accurate forecasts of fire activity up to five days into the future compared to other approaches. When predicting for all days, Fire and non-Fire, it is important to account for the significant presence of quiescent days, e.g., using a hurdle model, training separate models. We show that the non-Fire prediction task is difficult and maybe better evaluated as a ranking or ignition probability problem than with RMSE due to the skewed distribution. The use of a spatial grid is effective, but comes with several limitations that suggest a more comprehensive spatial approach is necessary to further improve model performance and utilization of available data.

Appendix
Model Memory

The models we developed can use lagged values of each of the y and $\boldsymbol{x}$ variables, up to m days of "memory." The lagged values for y available to the model for prediction are $y_l^{(t-m:t)} = \{y^{(t)}, y^{(t-1)}, \ldots, y^{(t-m)}\}$ and for $\boldsymbol{x}$ the lagged values available are $\boldsymbol{x}^{(t+k-m:t+k)} = \{\boldsymbol{x}^{(t+k)}, \ldots, \boldsymbol{x}^{(t+k-m)}\}$. Thus, the forecast models can be written in the following general form:

$$E[y_l^{(t+k)}] = f(y_l^{(t-m:t)}, \boldsymbol{x}_l^{(t+k-m:t+k)}; \boldsymbol{\theta}), \quad 1 \le k \le K. \quad (2)$$

We explored two different approaches to model memory. In the first approach, each of the previous count values $\hat{y}_l^{(t-m:t)}$ for the days $t-m$ to t, and the covariate values $\boldsymbol{x}_{i,j}^{(t+k-m:t+k)}$ for the days $t+k-m$ to $t+k$, are used as separate inputs to the model and each of these lagged values has their own coefficient. In the second approach, the $m-1$ lagged values in $\hat{y}_l^{(t-m:t-1)}$ and covariates of $\boldsymbol{x}_l^{(t+k-m:t+k-1)}$) are combined using a decaying weighted average to produce a scalar value $\bar{y}_l^{(t-m:t-1)}$ and a p-dimensional vector $\bar{\boldsymbol{x}}_l^{(t+k-m:t+k-1)}$. For the previous counts, $\hat{y}_l^{(t-m:t)}$, a scalar γ is used as the decay rate and for the vector of covariates, $\boldsymbol{x}_l^{(t+k-m:t+k)}$, each covariate has a separate decay parameter in the p-dimensional vector $\boldsymbol{\phi}$

$$\begin{aligned}
\bar{y}_l^{(t-m:t-1)} &= \frac{1}{\sum_{r=1}^{m} w_r} \sum_{r=1}^{m} w_r \hat{y}_l^{(t-r)} \\
\bar{\boldsymbol{x}}_l^{(t+k-m:t+k-1)} &= \frac{1}{\sum_{r=1}^{m} \boldsymbol{u_r}} \circ \sum_{r=1}^{m} \boldsymbol{u_r} \circ \boldsymbol{x}_l^{(t+k-r)} \\
w_r &= \gamma^{r-1}, \quad \gamma \in [0, 1] \\
\boldsymbol{u_r} &= \left\{\phi_1^{r-1}, \ldots, \phi_p^{r-1}\right\}, \quad \phi_i \in [0, 1].
\end{aligned}$$

These weighted averages are then used as inputs to the model along with $\hat{y}_{i,j}^{(t)}$ and $\boldsymbol{x}_l^{(t+k)}$, i.e., in the weighted-average version of the model there are two inputs and two coefficients in the model for each of y and $\boldsymbol{x}$. The decay parameters, γ and $\boldsymbol{\phi}$, are selected using grid search.

References

[1] F. H. Johnston *et al.*, "Estimated global mortality attributable to smoke from landscape fires," *Environ. Health Perspect.*, vol. 120, no. 5, pp. 695–701, May 2012.

[2] O. Adetona *et al.*, "Review of the health effects of wildland fire smoke on wildland firefighters and the public," *Inhalation Toxicol.*, vol. 28, no. 3, pp. 95–139, Feb. 2016.

[3] L. Giglio, J. Descloitres, C. O. Justice, and Y. J. Kaufman, "An enhanced contextual fire detection algorithm for MODIS," *Remote Sens. Environ.*, vol. 87, nos. 2–3, pp. 273–282, Oct. 2003.

[4] W. Schroeder, P. Oliva, L. Giglio, and I. A. Csiszar, "The new VIIRS 375 m active fire detection data product: Algorithm description and initial assessment," *Remote Sens. Environ.*, vol. 143, pp. 85–96, Mar. 2014.

[5] J. W. Kaiser *et al.*, "Biomass burning emissions estimated with a global fire assimilation system based on observed fire radiative power," *Biogeosciences*, vol. 9, no. 1, pp. 527–554, Jan. 2012.

[6] J. S. Reid *et al.*, "Global monitoring and forecasting of biomass-burning smoke: Description of and lessons from the fire locating and modeling of burning emissions (FLAMBE) program," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 2, no. 3, pp. 144–162, Sep. 2009.

[7] G. D. Rolph *et al.*, "Description and verification of the NOAA smoke forecasting system: The 2007 fire season," *Weather Forecasting*, vol. 24, no. 2, pp. 361–378, Apr. 2009.

[8] A. S. Darmenov and A. M. Da Silva, "The quick fire emissions dataset (QFED): Documentation of versions 2.1, 2.2 and 2.4," Ser. Global Model. Data Assimilation, NASA, Washington, DC, USA, Tech. Rep. NASA/TM-2015-104606 and GSFC-E-DAA-TN28825, 2015, vol. 38. [Online]. Available: https://ntrs.nasa.gov/search.jsp?R=20180005253
[9] F. Di Giuseppe, S. Rémy, F. Pappenberger, and F. Wetterhall, "Improving forecasts of biomass burning emissions with the fire weather index," *J. Appl. Meteor. Climatol.*, vol. 56, no. 10, pp. 2789–2799, Oct. 2017.
[10] C. E. Van Wagner and P. Forest, "Development and structure of the canadian forest fireweather index system," Canadian Forest Service, Victoria, BC, Canada, Tech. Rep., 1987.
[11] B. N. Holben *et al.*, "AERONET—A federated instrument network and data archive for aerosol characterization," *Remote Sens. Environ.*, vol. 66, no. 1, pp. 1–16, Oct. 1998.
[12] E. S. Kasischke *et al.*, "Alaska's changing fire regime—Implications for the vulnerability of its boreal forests," *Can. J. Forest Res.*, vol. 40, no. 7, pp. 1313–1324, Jun. 2010.
[13] S. Veraverbeke *et al.*, "Lightning as a major driver of recent large fire years in North American boreal forests," *Nature Climate Change*, vol. 7, no. 7, pp. 529–534, Jul. 2017.
[14] M. A. Finney, "FARSITE: Fire area simulator-model development and evaluation," USDA Forest Service, Washington, DC, USA, Res. Paper RMRS-RP-4, 1998.
[15] M. A. Finney *et al.*, "A method for ensemble wildland fire simulation," *Environ. Model. Assess.*, vol. 16, no. 2, pp. 153–167, Apr. 2011.
[16] J. K. Kelso, D. Mellor, M. E. Murphy, and G. J. Milne, "Techniques for evaluating wildfire simulators via the simulation of historical fires using the AUSTRALIS simulator," *Int. J. Wildland Fire*, vol. 24, no. 6, pp. 784–797, Jun. 2015.
[17] E. K. Noonan-Wright *et al.*, "Developing the US wildland fire decision support system," *J. Combustion*, vol. 2011, pp. 1–14, May 2011.
[18] C. Zhang, M. Rochoux, W. Tang, M. Gollner, J.-B. Filippi, and A. Trouvé, "Evaluation of a data-driven wildland fire spread forecast model with spatially-distributed parameter estimation in simulations of the FireFlux I field-scale experiment," *Fire Saf. J.*, vol. 91, pp. 758–767, Jul. 2017.
[19] L. Giglio, W. Schroeder, and C. O. Justice, "The collection 6 MODIS active fire detection algorithm and fire products," *Remote Sens. Environ.*, vol. 178, pp. 31–41, Jun. 2016.
[20] F. Di Giuseppe, S. Rémy, F. Pappenberger, and F. Wetterhall, "Using the fire weather index (FWI) to improve the estimation of fire emissions from fire radiative power (FRP) observations," *Atmos. Chem. Phys.*, vol. 18, no. 8, pp. 5359–5370, Apr. 2018.
[21] L. Giglio, L. Boschetti, D. P. Roy, M. L. Humber, and C. O. Justice, "The collection 6 MODIS burned area mapping algorithm and product," *Remote Sens. Environ.*, vol. 217, pp. 72–85, Nov. 2018.
[22] D. P. Dee *et al.*, "The ERA-interim reanalysis: Configuration and performance of the data assimilation system," *Quart. J. Roy. Meteorol. Soc.*, vol. 137, pp. 553–597, Apr. 2011.
[23] O. Tetens, "Über einige meteorologische Begriffe," *Zeitschrift für Geophysik*, vol. 6, pp. 297–309, 1930, doi: 10.23689/fidgeo-3212.
[24] F. Sedano and J. T. Randerson, "Multi-scale influence of vapor pressure deficit on fire ignition and spread in boreal forest ecosystems," *Biogeosciences*, vol. 11, no. 14, pp. 3739–3755, Jul. 2014.
[25] M. A. Friedl *et al.*, "MODIS collection 5 global land cover: Algorithm refinements and characterization of new datasets," *Remote Sens. Environ.*, vol. 114, no. 1, pp. 168–182, Jan. 2010.
[26] T. R. Loveland and A. S. Belward, "The IGBP-DIS global 1 km land cover data set, DISCover: First results," *Int. J. Remote Sens.*, vol. 18, no. 15, pp. 3289–3295, Oct. 1997.
[27] A. H. Murphy, "Skill scores based on the mean square error and their relationships to the correlation coefficient," *Monthly Weather Rev.*, vol. 116, no. 12, pp. 2417–2424, Dec. 1988.
[28] E. B. Wiggins *et al.*, "The influence of daily meteorology on boreal fire emissions and regional trace gas variability," *J. Geophys. Res. Biogeosci.*, vol. 121, no. 11, pp. 2793–2810, Nov. 2016.
[29] A. Cameron and P. Trivedi, *Regression Analysis of Count Data*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2013.
[30] D. Mandallaz and R. Ye, "Prediction of forest fires with Poisson models," *Can. J. Forest Res.*, vol. 27, no. 10, pp. 1685–1694, Oct. 1997.
[31] H. K. Preisler, D. R. Brillinger, R. E. Burgan, and J. W. Benoit, "Probability based models for estimation of wildfire risk," *Int. J. Wildland Fire*, vol. 13, no. 2, p. 133, 2004.
[32] J. Mullahy, "Specification and testing of some modified count data models," *J. Econometrics*, vol. 33, no. 3, pp. 341–365, Dec. 1986.
[33] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986.
[34] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980
[35] P. J. Diggle, J. A. Tawn, and R. Moyeed, "Model-based geostatistics," *J. Roy. Stat. Soc. C, Appl. Statist.*, vol. 47, no. 3, pp. 299–350, 1998.
[36] R. A. Davis, "Observation-driven models for Poisson counts," *Biometrika*, vol. 90, no. 4, pp. 777–790, Dec. 2003.
[37] K. Fokianos, A. Rahbek, and D. Tjøstheim, "Poisson autoregression," *J. Amer. Stat. Assoc.*, vol. 104, no. 488, pp. 1430–1439, Dec. 2009.
[38] J. Pereira and K. Turkman, *Statistical Models Vegetation Fires: Spatial Temporal Patterns*. Boca Raton, FL, USA: CRC Press, 2019, ch. 17, pp. 401–420.

Casey A. Graff received the B.S. degree in computer science from the University of California at San Diego, La Jolla, CA, USA, in 2017. He is currently pursuing the M.S. and Ph.D. degrees in computer science with the University of California Irvine, Irvine, CA.

He was a Research Assistant and Bioinformatics Programmer with the Scripps Institution of Oceanography, La Jolla, in 2016 and 2017, respectively, where he used machine learning to classify and quantify zooplankton populations from images. He is a Funded Fellow of the Machine Learning and Physical Sciences (MAPS) NSF Research Traineeship Program at UC Irvine. His research at UC Irvine focuses on applying and developing machine learning methods in the geosciences with particular interest in spatiotemporal forecasting.

Shane R. Coffield received the B.S. degree in geophysical sciences and environmental science from the University of Chicago, Chicago, IL, USA, in 2017. He is currently pursuing the M.S. and Ph.D. degrees in earth system science with the University of California Irvine (UC Irvine), Irvine, CA, USA.

In 2016 and 2017, he was an Intern with the Earth Sciences Branch, NASA Marshall Space Flight Center, Huntsville, AL, USA, where he used remote sensing to study air and water quality in the USA. Since 2017, he has been a Graduate Student Researcher with UC Irvine. His research focuses on terrestrial ecosystems, fire ecology, climate change, and machine learning. He is also interested in science outreach, education, and communication.

Mr. Coffield was a recipient of the National Science Foundation Graduate Research Fellowship in 2018 and was offered the NASA Earth and Space Sciences Fellowship in 2018. He is an Honorary Fellow of the Machine Learning and Physical Sciences (MAPS) NSF Research Traineeship Program at UC Irvine and a member of the American Geophysical Union.

Yang Chen received the B.S. and M.S. degrees in environmental engineering from Tsinghua University, Beijing, China, in 1997 and 2000, respectively, and the Ph.D. degree in atmospheric science from the University of Michigan, Ann Arbor, MI, USA, in 2006.

From 2006 to 2008, he held a post-doctoral position at the Jet Propulsion Laboratory, Pasadena, CA, USA. Since 2008, he has been a Project Scientist and an Assistant Researcher with the University of California Irvine, Irvine, CA, USA. His research interests include the relationships between wildfires and climate, as well as the impacts of fires on air quality and the terrestrial ecosystem.

Dr. Chen is a member of the American Geophysical Union.

Efi Foufoula-Georgiou received the Diploma degree in civil engineering from the National Technical University of Athens, Athens, Greece, in 1979, and the M.Sc. and Ph.D. degrees in environmental engineering from the University of Florida, Gainesville, FL, USA, in 1982 and 1985, respectively.

From 1989 to 2016, she was a Distinguished McKnight University Professor with the Department of Civil and Environmental Engineering, University of Minnesota, Minneapolis, MN, USA, and the Director of the Saint Anthony Falls Laboratory, Minneapolis, from 1999 to 2003, and the National Center for Earth-Surface Dynamics, Minneapolis, from 2008 to 2013. In 2016, she joined the University of California Irvine, Irvine, CA, USA, as a Distinguished Professor in civil and environmental engineering and the Samueli Endowed Chair in engineering. Her research interests include inverse problems in hydrology and geomorphology, river network dynamics, and data analytics for hydroclimate and extremes.

Dr. Foufoula-Georgiou is a fellow of the American Geophysical Union and the American Meteorological Society, and a member of the European Academy of Sciences and the U.S. National Academy of Engineering.

James T. Randerson was born in Fairfax, VA, USA, in 1969. He received the B.S. degree in chemistry and the Ph.D. degree in biological sciences from Stanford University, Stanford, CA, USA, in 1992 and 1998, respectively.

He conducted post-doctoral training at the Department of Earth and Planetary Sciences, University of California, Berkeley, CA, USA, and the Institute for Arctic Biology, University of Alaska, Fairbanks, AK, USA, from 1998 to 2000. He was an Assistant Professor with the Divisions of Geological and Planetary Science and Engineering and Applied Science, California Institute of Technology, Pasadena, CA, USA, from 2000 to 2003. In 2003, he moved to the Department of Earth System Science, University of California Irvine, Irvine, CA, USA, where he is currently the Ralph J. and Carol M. Cicerone Professor of Earth System Science. His research interest includes improving our understanding of contemporary and future changes in the global biosphere. His laboratory explores how global patterns of fires are changing in response to climate change and intensification of land use. Through the design and analysis of earth system model simulations, his laboratory also studies how drought and climate-carbon cycle feedbacks are likely to evolve in the future.

Dr. Randerson currently serves as a member of the Biological and Environmental Research Federal Advisory Committee for the U.S. Department of Energy Office of Science.

Padhraic Smyth (Member, IEEE) received the B.S. degree in electronic engineering from the National University of Ireland, University College Galway, Galway, Ireland, in 1984, and the M.S. and Ph.D. degrees in electrical engineering from the California Institute of Technology, Pasadena, CA, USA, in 1985 and 1988, respectively.

From 1988 to 1992, he was a member of Technical Staff and from 1992 to 1996, a Technical Group Leader with the Jet Propulsion Laboratory, California Institute of Technology. Since 1996, he has been a Professor with the Computer Science and Statistics Departments, University of California Irvine, Irvine, CA, USA. His research interests include machine learning, pattern recognition, and applied statistics.

Dr. Smyth is a fellow of the Association for Computing Machinery and the Association for the Advancement of Artificial Intelligence and a member of the American Statistical Association.