Daily Local-Level Estimates of Ambient Wildfire Smoke PM$_{2.5}$ for the Contiguous US

Marissa L. Childs,* Jessica Li, Jeffrey Wen, Sam Heft-Neal, Anne Driscoll, Sherrie Wang, Carlos F. Gould, Minghao Qiu, Jennifer Burney, and Marshall Burke
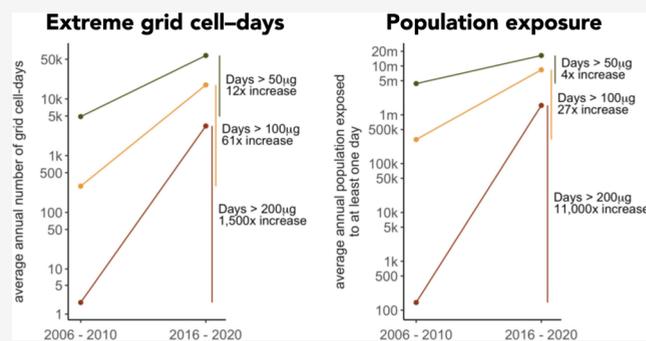
Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Smoke from wildfires is a growing health risk across the US. Understanding the spatial and temporal patterns of such exposure and its population health impacts requires separating smoke-driven pollutants from non-smoke pollutants and a long time series to quantify patterns and measure health impacts. We develop a parsimonious and accurate machine learning model of daily wildfire-driven PM$_{2.5}$ concentrations using a combination of ground, satellite, and reanalysis data sources that are easy to update. We apply our model across the contiguous US from 2006 to 2020, generating daily estimates of smoke PM$_{2.5}$ over a 10 km-by-10 km grid and use these data to characterize levels and trends in smoke PM$_{2.5}$. Smoke contributions to daily PM$_{2.5}$ concentrations have increased by up to 5 $\mu g/m^3$ in the Western US over the last decade, reversing decades of policy-driven improvements in overall air quality, with concentrations growing fastest for higher income populations and predominantly Hispanic populations. The number of people in locations with at least 1 day of smoke PM$_{2.5}$ above 100 $\mu g/m^3$ per year has increased 27-fold over the last decade, including nearly 25 million people in 2020 alone. Our data set can bolster efforts to comprehensively understand the drivers and societal impacts of trends and extremes in wildfire smoke.

KEYWORDS: *particulate matter, wildfires, smoke, aerosols, machine learning*

INTRODUCTION

Exposure to ambient fine particulate matter (PM$_{2.5}$) has long been recognized as a leading environmental determinant of health outcomes and is estimated to cause millions of premature deaths globally and 48,000 deaths in the US every year.[1] Efforts to reduce this health burden require understanding the main sources of PM$_{2.5}$ and how pollutant concentrations from these sources are distributed across human populations. While many countries have made remarkable strides in reducing key anthropogenic emission sources and thus PM$_{2.5}$ concentrations, albeit sometimes from very high levels,[2,3] other PM$_{2.5}$ sources appear to be growing rapidly. Of key concern in many parts of the world is the growing contribution of wildfire smoke to pollution concentrations.[4−9] Increases in wildfire smoke are being driven in a substantial part by a warming climate, which increases the flammability of fuels and in turn the occurrence of larger, more extreme wildfires which emit more smoke.[10−12] Lofted high into the atmosphere, resulting pollutants can travel long distances from their source fire, potentially affecting ground pollution levels and health outcomes thousands of kilometers away.[13] Because smoke from wildfires can also generate abrupt extreme pollutant concentrations and because emitted pollutants might differ in their toxicity relative to analogous non-wildfire species,[14] accurate measurement of the location and level of smoke-attributable surface pollutant concentrations is key for understanding the societal impacts of growing wildfire risk.

However, accurate attribution and measurement remain empirically challenging. Ground monitors measure concentrations of total pollutants and cannot be easily used to attribute pollutants to their sources without additional information. Atmospheric chemical transport models (CTMs) offer one common approach to linking wildfire activity to concentrations of key pollutants and downstream human impacts,[15−20] but their use faces a number of difficulties.[21] First, uncertainties in wildfire emission inventories have been shown to lead to substantial differences in CTM-estimated smoke PM$_{2.5}$ concentrations, with up to 20-fold differences in estimated regional wildfire-attributed PM$_{2.5}$ concentrations[22] when different leading inventories are fed
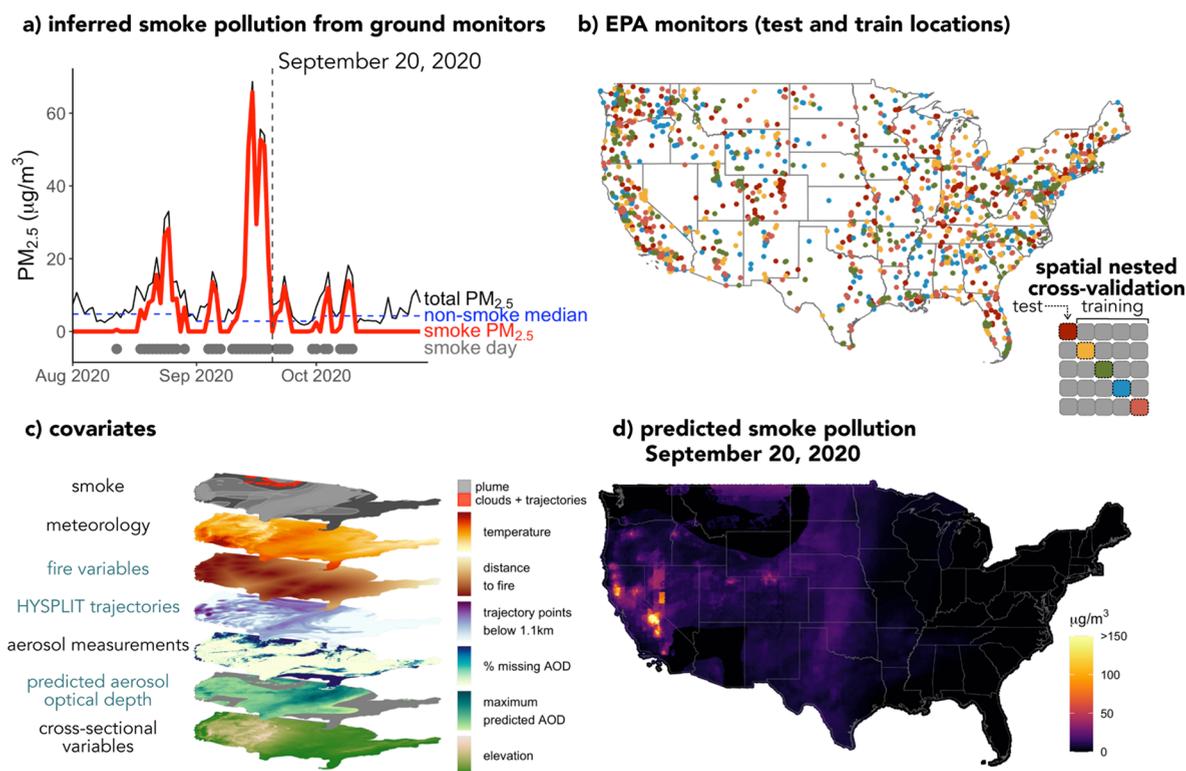
Figure 1. Model predictions are derived from inferred smoke pollution at ground monitors and remotely sensed covariates. (a) Total fine particulate matter (PM$_{2.5}$, black line) is observed at ground monitors. Plumes and trajectories from fires are used to identify days with smoke (gray points). Non-smoke median PM$_{2.5}$ (blue dashed line) is the station- and month-specific median of PM$_{2.5}$ on non-smoke days, and smoke PM$_{2.5}$ (red line) is total PM$_{2.5}$ above the non-smoke median on days with smoke overhead. (b) Smoke PM$_{2.5}$ is inferred at EPA monitor stations (points) throughout the US, which are then grouped into five disjoint folds for model training and validation to perform spatial nested cross-validation (inset). (c) Smoke days (top layer) are defined based on the presence of smoke plumes (light gray areas) or the presence of 50 HYSPLIT trajectory points below 1.1 km and AOD missingness over 75% (red areas), as a proxy for cloud cover that would prevent the identification of smoke plumes. Other covariates included in the model are used to predict smoke PM$_{2.5}$ on smoke days and include both time-varying and cross-sectional variables. Covariates labeled in black are directly measured by other products, while those in blue are derived. (d) Resulting predictions are a 10 km resolution gridded product of daily smoke PM$_{2.5}$ over the contiguous U.S. from 2006 to 2020. All time-varying covariates and predictions shown are from September 20, 2020. State outlines are reproduced with permission from the US Census Bureau TIGER/Line Shapefiles.[51]

into the same transport model, although recent efforts have shown that satellite-based emission calibration can improve the simulation accuracy.[23] Second, the complex chemistry regimes during wildfire (in particular, production of organic aerosols) and hyperlocalized meteorology are often not well-captured by CTMs, leading to concerns over their ability to correctly model the transport of wildfire smoke.[24−26] Third, parameterized heights of plume injections and the vertical allocations of emissions differ considerably across different CTMs and often compare poorly with observed injection heights, leading to discrepancies in surface PM$_{2.5}$ concentrations when compared with in situ observations.[27,28]

As a complement to CTM-based approaches, recent efforts have sought to use statistical approaches to directly estimate smoke PM$_{2.5}$ concentrations without explicitly parameterizing the underlying physical processes. The rapid expansion in the frequency and availability of satellite imagery has led to numerous satellite-based estimates of wildfire smoke movement,[29,30] but commonly used satellite products (e.g., aerosol optical depth or satellite-derived plume boundary estimates) do not provide precise information on where in the atmospheric column the smoke occurs, making it difficult to use this information alone to infer surface concentrations. Instead, non-CTM-based approaches typically include the use of ground monitor data to infer wildfire-attributable pollutant

concentrations. The main empirical challenges of non-CTM-based approaches for estimating wildfire-attributable air pollution are the accurate partitioning of smoke- and non-smoke-sourced pollutants and the estimation of pollution concentrations in locations where ground monitors are not available. Studies have taken multiple approaches to solving the partitioning problem, including building statistical models of all the main sources of pollution and estimating the model with and without wildfire variables,[4] estimating smoke PM$_{2.5}$ at a given monitor location as anomalous deviations from estimated seasonal non-smoke background PM$_{2.5}$ concentrations,[5,31] using particle trajectory models to estimate wildfire-influenced and non-influenced surface pollution,[32] and simply attributing extreme positive PM$_{2.5}$ anomalies during summer and fall wildfire seasons in the US West to wildfires.[6] Alternate approaches have also been used to estimate smoke pollutant concentrations for locations without monitors, including fitting statistical models that relate monitor observations to input variables and then using the model to estimate concentrations in non-monitor locations[4] or interpolating monitor observations across space and inferring the smoke contribution using one of the abovementioned techniques.[5,6,31] The primary goal of most of these analyses has been to shed light on the longer-term trends in smoke-

derived $PM_{2.5}$ and how smoke contributions to total $PM_{2.5}$ have changed over space and time.

Here, we build on these non-CTM-based approaches to estimate daily, local-level wildfire-smoke-attributable surface $PM_{2.5}$ across the contiguous US. We combine information from satellite-based smoke plume identification and simulations of air trajectories from fire locations to identify when smoke is in the air. Then, similar to other studies, we attribute anomalous deviations from background levels of $PM_{2.5}$ to smoke when smoke is plausibly in the air. To estimate smoke $PM_{2.5}$ at locations without monitors, we take advantage of the additional available information from satellite and reanalysis products and build a machine learning model to predict daily smoke $PM_{2.5}$ over the contiguous US. This combines the strengths of non-CTM-based approaches, which carefully attribute observed ground $PM_{2.5}$ concentrations to smoke, and satellite data which provide additional information in locations without monitors. In contrast to many past machine learning-based efforts predicting total $PM_{2.5}$, our goal is to isolate smoke-driven changes in $PM_{2.5}$ from other changes in $PM_{2.5}$. Our efforts are thus a complement to recent machine learning-based efforts at predicting variation in total $PM_{2.5}$, for example, Di et al.[33] and Reid et al.[34]

These smoke $PM_{2.5}$ predictions are designed with downstream applications and inference tasks in mind. We sought predictions that accurately captured local short-run variation in smoke $PM_{2.5}$ levels and extremes over sufficiently long time series and across a large spatial domain. Simultaneously, we sought a tractable model with inputs that would be easy to update in the future. These dual goals led us to produce daily predictions of smoke $PM_{2.5}$ at a 10 km resolution using a model with high predictive performance at both low and extreme levels of smoke $PM_{2.5}$, which we then used to understand the levels, trends, and changing daily extremes in smoke $PM_{2.5}$ concentrations across the contiguous US. We also connected these estimates to data on income, race, and ethnicity to understand differential patterns of exposure and how these patterns are evolving over time.

■ METHODS

Our approach has three main steps. First, we identify days when smoke was overhead ("smoke days") from satellite imagery-based plume classification and when clouds may be obscuring plumes, from simulated air trajectories originating at fires (Figure 1c, smoke). Second, we construct a ground-based measure of smoke $PM_{2.5}$ by calculating $PM_{2.5}$ anomalies at Environmental Protection Agency (EPA) monitoring stations (Figure 1a,b), measured as deviations from recent location- and month-specific median $PM_{2.5}$ on non-smoke days and then attribute positive anomalies to smoke if our approach in the first step indicated that there was smoke overhead on that day. Third, because pollution monitoring stations are infrequent in space and typically have many missing observations, we train a model that predicts the station-based ground smoke $PM_{2.5}$ estimated in the second step, based on a set of spatially and temporally consistent inputs (Figure 1c), validating the model on daily time series from held-out EPA stations (i.e., stations not used in model training) and on daily time series from an independent set of private monitors that were not used in model development. Finally, we use the model to produce daily 10 km gridded estimates of smoke pollution over the contiguous US (Figure 1d). Unless noted otherwise, analyses were performed using the R programming language.[35]

Defining Smoke Days. To understand when smoke from fires may be affecting ground pollution levels, we construct a binary classification of smoke days for each cell of a 10 km grid covering the contiguous US based on two sources of information: hand-annotated smoke plumes from satellite imagery and modeled air particle trajectories from fire locations to aid in identifying smoke when clouds may obscure plume identification. First, we use data on smoke plumes from the National Oceanic and Atmospheric Administration (NOAA) Hazard Mapping System (HMS),[36] which are analyst-identified plume boundaries based on visible bands of satellite imagery.[30,37,38] A grid cell is classified as a smoke day if it has any intersection with a smoke plume on a given day (Figure S2a,b). The first full year for which the HMS plume data are available is 2006, which limits the start date of our study period.

Second, given the potential for smoke plume boundaries to miss areas affected by smoke, especially when clouds hinder identification, we build on recent work[29,32] and simulate air packet trajectories from smoke-producing fire points detected by HMS using the Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model.[29,38,39] Using these runs, we classify grid cell days as smoke days if they are missing more than 75% of 1 km aerosol optical depth observations (AOD, 0.47 μm; based on a moderate resolution imaging spectroradiometer (MODIS) multi-angle implementation of atmospheric correction (MAIAC) daily observations) and have a count of at least 50 HYSPLIT trajectory points in the lowest height quintile (0−1.1 km above ground level) within a 50 km buffer of the location of interest. These thresholds are selected to yield a set of smoke days with anomalous $PM_{2.5}$ values similar to the original set of days (Figure S2e,f; see the Supporting Information for more details).

Calculating Ground-Based Measures of Smoke $PM_{2.5}$. We then combine classification of smoke days with data on daily average $PM_{2.5}$ concentrations from 2019 EPA monitoring stations throughout the US to define daily time series of smoke $PM_{2.5}$ at each station.[40] We first define $PM_{2.5}$ anomalies as deviations from recent month- and location-specific median values on non-smoke days (Figure 1a)

$$\widetilde{PM}_{idmy} = PM_{idmy} - \overline{PM}_{imy}^{NS} \tag{1}$$

where PM_{idmy} is the $PM_{2.5}$ at station i on day d in month m and year y and $\overline{PM}_{IMY}^{NS}$ is the 3 year location- and month-specific median $PM_{2.5}$ on non-smoke days. This median is calculated as

$$\overline{PM}_{IMY}^{NS} = \text{median}(\{PM_{idmy}|i = I, m = M, Y - 1 \leq y$$
$$\leq Y + 1, \text{smoke}_{idmy} = 0\}) \tag{2}$$

with smoke_{idmy} being a binary variable indicating smoke day classification. We use medians rather than means to prevent days with extreme $PM_{2.5}$ that are not smoke days from affecting the background $PM_{2.5}$ estimates, as is occasionally the case in our data. Furthermore, using 3 year medians, we allow the measure of background non-smoke $PM_{2.5}$ to change over time in each location to capture trends in non-smoke $PM_{2.5}$ over time, including potential declines in anthropogenic emissions. We then define ground-based smoke $PM_{2.5}$ as anomalies above the median on days in which smoke was overhead

$$\text{smokePM}_{idmy} = \max(\widetilde{PM}_{idmy} \times \text{smoke}_{idmy}, 0) \tag{3}$$

Our approach to identifying smoke $PM_{2.5}$ at monitoring stations is similar to other recent studies.[5,31] Similar to previous non-CTM efforts, these estimates rely on the HMS plume boundaries for identification of when smoke is affecting $PM_{2.5}$. We find that having a smoke plume overhead is associated with an average of 4.5 $\mu g/m^3$ increase in $PM_{2.5}$ after controlling for monitor-specific averages and average differences in $PM_{2.5}$ between states, months, and years (Table S1). We also find that in monitor time series, plumes align temporally with spikes in $PM_{2.5}$ (Figure S5). Furthermore, to counter omissions when clouds may obscure smoke plumes,[29] we included locations with potential cloud cover and air trajectories from known fires (see the "Defining Smoke Days" section mentioned above). Nevertheless, smoke days may be a conservative estimate of the locations with air quality impacted by smoke due to undetected plumes under cloud cover, during nighttime when satellite-based plume segmentations are unavailable, or where smoke is dilute and difficult to identify in satellite imagery.[5,29] See more detailed discussion in the Supporting Information ("Plume accuracy").

To further confirm that our method of constructing smoke $PM_{2.5}$ from ground station anomalies is indeed picking up $PM_{2.5}$ from smoke and not from local time-varying sources of $PM_{2.5}$ unrelated to smoke, we apply our method to harmonized speciated data from Interagency Monitoring of Protected Visual Environments (IMPROVE)[41] and Chemical Speciation Network (CSN)[42] monitors.[43−46] If our approach is identifying smoke-sourced $PM_{2.5}$ and not $PM_{2.5}$ from other sources, then species most likely to be present in smoke $PM_{2.5}$—which includes organic carbon throughout the US and perhaps additionally sulfates from agricultural fires in the southeastern US[47,48]—will increase on smoke days but other non-fire-associated species will not increase. This is indeed what we find: the share of anomalous $PM_{2.5}$ made up by organic carbon increases substantially, particularly on days that our method would predict are very high smoke days and particularly in settings where background sources of other $PM_{2.5}$ are low (such as in the western US) (Figure S6). In areas where background $PM_{2.5}$ levels are higher and where agricultural burning is more common, such as in much of the southeastern US and in the southern Central Valley of CA, both organic carbon and SO_4 increase in importance on smoke days—although, as expected,[21] their contribution to total $PM_{2.5}$ on that day is still smaller than in Western areas where background $PM_{2.5}$ is low. In both settings, non-wildfire sources of $PM_{2.5}$ (such as dust and elemental carbon) do not increase on smoke days. These results provide confirmatory evidence that our method of assigning anomalous $PM_{2.5}$ to smoke on days with smoke plumes overhead is indeed picking up wildfire-sourced $PM_{2.5}$ and not some other correlated $PM_{2.5}$ source.

Predicting Smoke $PM_{2.5}$. Because of the limited spatial and intermittent temporal coverage of ground-based $PM_{2.5}$ monitors in the US, estimates of $smokePM_{idmy}$ from eq 3 with time series for more than 5 years are only available at roughly 1400 ground locations. In addition, the average person in the US lives more than 20 km from the nearest pollution monitor.[49] To capture local short-run variation in smoke-PM_{idmy} over space and time, we therefore incorporate additional sources of information. One option is to use existing high-resolution gridded daily estimates of $PM_{2.5}$, from which anomalized estimates of $\widetilde{PM}_{idmy}$ could be constructed to plug into eq 3. Numerous machine-learning-based efforts have succeeded in generating such estimates, but using them faces a number of challenges. First, many estimates are not updated after their initial publication and are available only through earlier years (e.g., Di et al.[33]) and/or for a geographic region such as the Western U.S (e.g., Reid et al.[34]), making them difficult to use to estimate up-to-date countrywide smoke pollution. Second, these $PM_{2.5}$ predictions are from a model trained to predict total $PM_{2.5}$, not smoke $PM_{2.5}$, and as such could potentially not be optimized for our particular task of interest. Finally, estimates could be hard to update, as highest performing model runs depend on a large number of inputs or outputs from other models that themselves are infrequently updated, making updating computationally burdensome.[33,34]

A second option is to train a model to predict $smokePM_{idmy}$ directly, using inputs from satellite, ground-measured, and reanalysis data sets that are straightforward to update. Because in this approach the model can be optimized to predict the outcome of interest—including focusing on inputs that predict smoke $PM_{2.5}$ (e.g., distance to fire) rather than total $PM_{2.5}$ (e.g., road density)—and can be carefully validated with downstream inference tasks in mind, we take this second option as our main approach.

Model Inputs. Building on earlier statistical-model-based smoke pollution prediction efforts,[4,5] we extract or compute a set of model inputs that include meteorology, derived fire variables from HMS fire points, counts of HYSPLIT trajectory points, direct aerosol measurements, AOD predictions, and cross-sectional land use and elevation (Figure 1c and Table S3). Model inputs were merged to a consistent 10 km grid for smoke $PM_{2.5}$ training and prediction. The grid is constructed to cover the contiguous US, based on US Census Bureau TIGER/Line county borders.[50,51]

Detailed information on model inputs is provided in the Supporting Information. Briefly, meteorological inputs to the model include daily mean, minimum, and maximum of the planetary boundary layer, mean wind speed in the eastward and northward directions, mean air temperature and dewpoint temperature at 2 meters, total precipitation, and mean sea level and surface pressure, all drawn from European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis 5th Generation (ERA5) land or global products.[52,53] Second, to capture the potential for characteristics of proximate fires to affect smoke $PM_{2.5}$, we used fire points from NOAA HMS[36] identified by trained analysts and used distance to these points as a model input. Third, counts of HYSPLIT trajectory points within a 50 km buffer by quintiles of height above the ground level were used as the input. Fourth, we used two complementary approaches for measuring aerosols: a reanalysis-based approach with consistent data but coarse spatial resolution, based on Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2) aerosol optical thickness (AOT) data,[54] and a satellite-based approach with higher resolution but many missing observations from MODIS MAIAC AOD.[55−57] Observations in the MAIAC product were first predicted using a gradient boosted tree model[58] with AOT anomalies on smoke days, meteorology, fire variables, elevation, and land cover as inputs (Table S2), and the resulting predicted AOD measures were used as the input to the smoke prediction model. Finally, we used time-invariant data on land use and elevation from USGS National Elevation Dataset[59] and National Land Cover Dataset[60] as additional inputs.

a) predicted and observed smoke PM$_{2.5}$

b) feature importance

c) out-of-sample R^2 by monitor

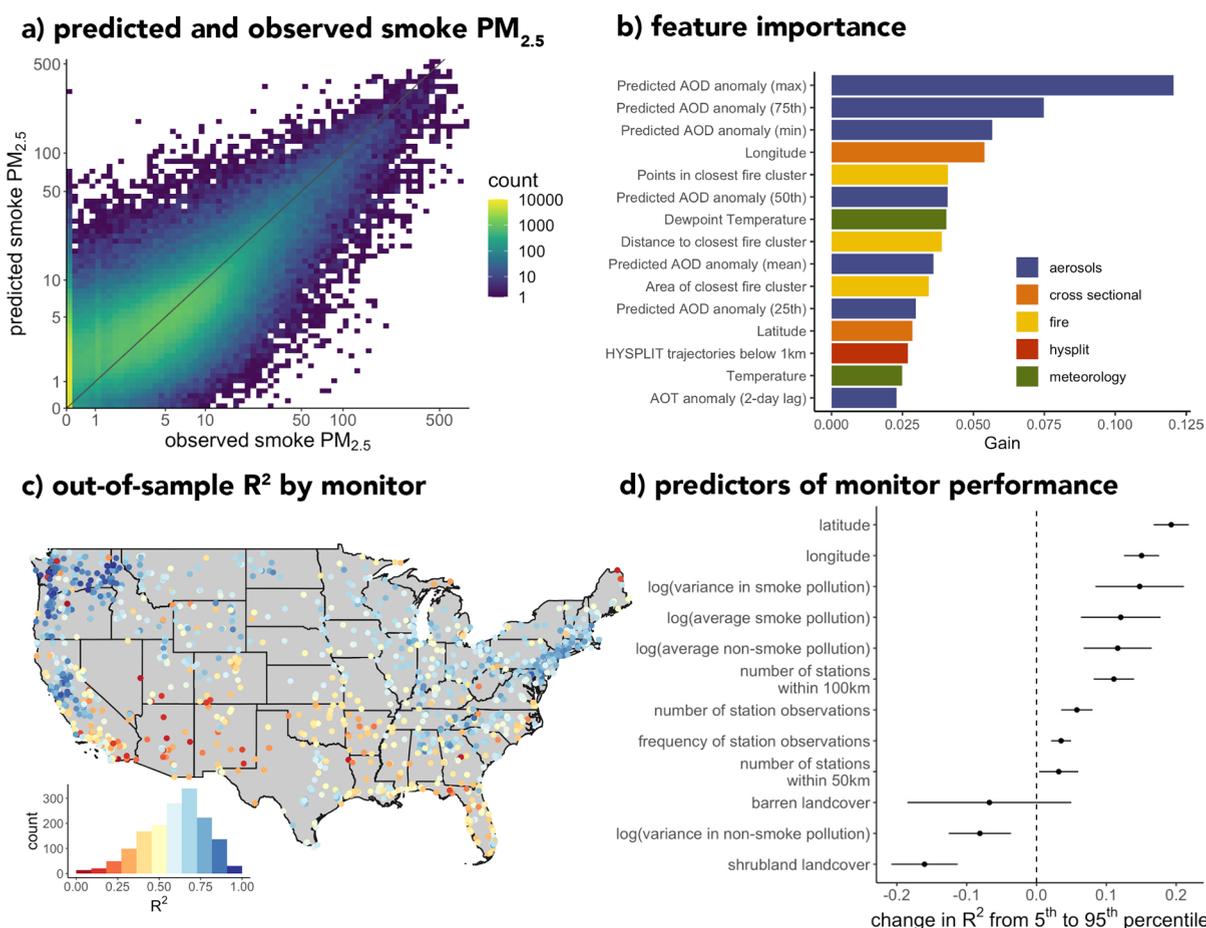## d) predictors of monitor performance

Figure 2. Model performs well out-of-sample and across the range of observed smoke PM$_{2.5}$ with predictable spatial variation in performance. (a) Color indicates count of monitor smoke days within bins of observed (horizontal axis) and predicted (vertical axis) smoke PM$_{2.5}$. Axes are pseudo-log-transformed, and color scale is log-transformed. The black line indicates the 1−1 line or where predictions perfectly match observations. (b) Feature importance for the top 15 features (vertical axis) from the final model, as measured by gain (horizontal axis) and colored by a broad category of the feature type, includes aerosols, cross-sectional information on elevation and land cover, fire variables, HYSPLIT, and meteorology. (c) For each monitor with at least 50 observations (points on map), R^2 is calculated over all smoke days with available smoke PM$_{2.5}$ using predictions from the model in which that station was out-of-sample. The inset histogram shows the distribution of R^2 values over monitors. (d) To explain the variation in model performance across monitors shown in (c), we fit a cross-sectional linear model relating monitor-specific performance (R^2) shown in (c) to monitor/location characteristics. Predictive power of each characteristic is calculated as the estimated change in monitor R^2 when each characteristic is increased from the 5th to 95th percentile of its distribution. Points show central estimates, and line segments show 95% confidence intervals. State outlines are reproduced with permission from the US Census Bureau TIGER/Line Shapefiles.[51]

Model Tuning and Validation. We aimed to predict smoke pollution for a 10 km grid over the conterminous US, so our measure of model performance needed to account for the fact that we would be predicting for out-of-sample locations and that those locations would have a median distance of about 50 km (population-weighted average of about 18 km) from EPA stations (Figure S9). Given concerns of information leakage between training and testing sets for proximate EPA stations, we used grid cells from the coarsest input (MERRA-2, 0.5° latitude × 0.625° longitude, ~50 km) to define spatial folds, that is, the disjoint sets of training and testing stations used, respectively, for model training and model evaluation. Splitting train and test data sets by station rather than the more conventional method of random splitting by observation (in which a given station can contribute data to both train and test) is a more demanding prediction task[34,61]—and using coarser spatial blocks further increases the difficulty of the prediction task[62]—but is a more realistic test of how well the

model would do predicting time series of smoke PM$_{2.5}$ in a new location with no training data.

For model architecture, we fit gradient boosted trees,[58] which balance nicely our twin goals of model parsimony and high predictive power in related applications (e.g., Chen et al.,[63] Zamani Joharestani et al.,[64] and Xu et al.[65]). We performed fivefold nested spatial cross-validation, with hyperparameter tuning in the inner loop performed with fourfold cross-validation (Figure 1b; see the Supporting Information). All model tuning and training was performed on observations with smoke overhead. Based on preliminary tests, we used regression with the root mean squared error (RMSE) as the objective function, which performed better than pseudo-Huber loss and Tweedie regression—loss functions that, in principle, could respectively help with outliers in the data that the squared error is sensitive to and help with skewed non-negative distribution of outcomes. To measure variable importance, we use gain or the contribution to improvements in model performance from splits on each feature.
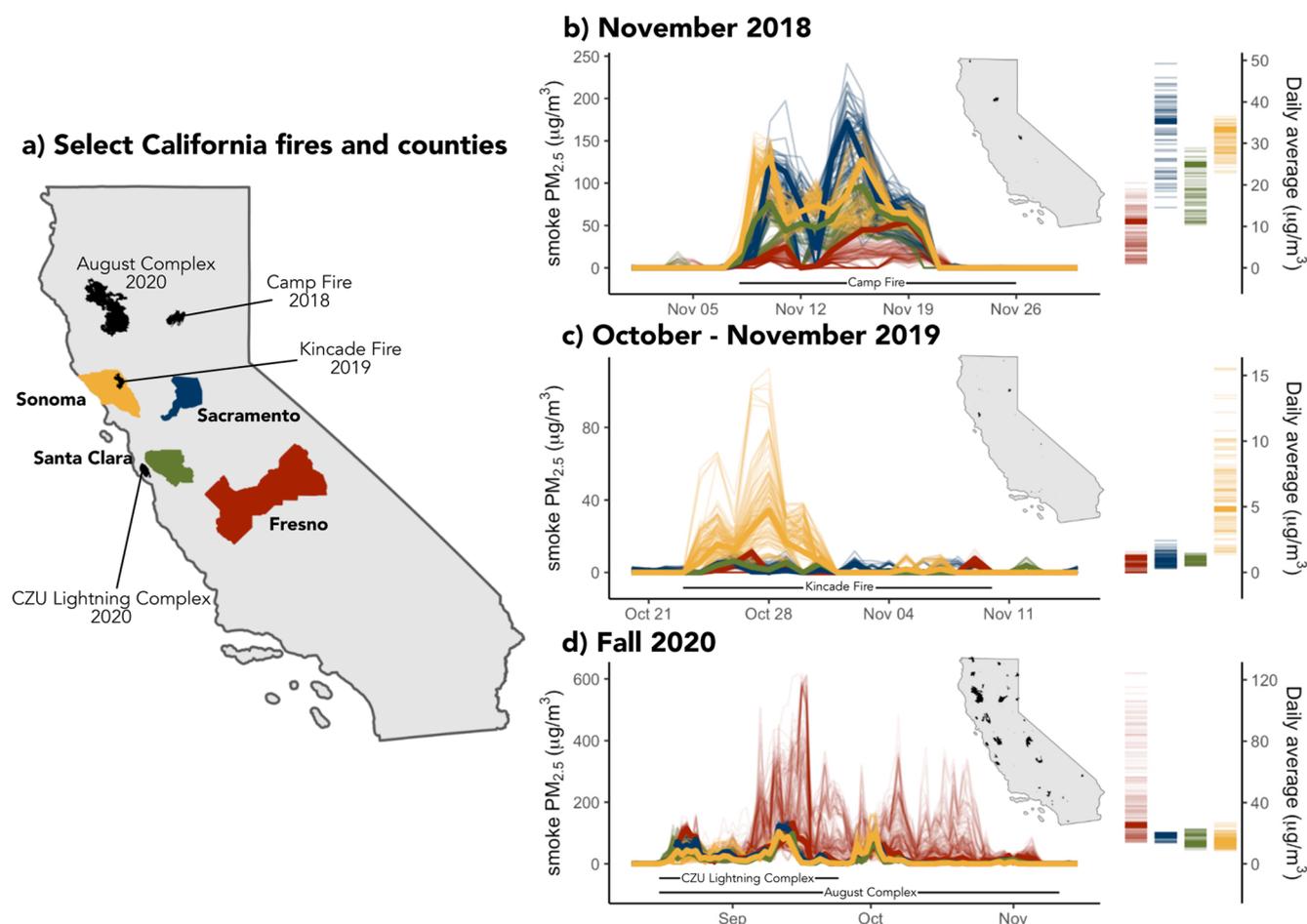
Figure 3. Predictions show remarkable spatial and temporal variation in smoke pollution from recent California fires. (a) Black shapes are select fires from November 2018, October to November 2019, and Fall 2020, and colored areas are four select California counties. (b−d) Daily smoke pollution varies between and within counties, and increases in smoke pollution temporally match with fire events. Thin lines show smoke pollution in grid cells within the county, while thick lines show population-weighted average for the county. Time ranges of select fires are shown along the horizontal axis. Inset maps show all fires burning during the time range of the panel. To the right, dashes show average smoke pollution over the time range in the grid cell (thin dashes) and county (thick line). Fire perimeters and time ranges are reproduced with permission from the CAL FIRE Fire Perimeter data set.[72] State and county outlines are reproduced with permission from the US Census Bureau TIGER/Line Shapefiles.[51]

Assessing Model Performance. We measured model performance by comparing observed smoke $PM_{2.5}$ with model predictions after truncating model predictions at zero. We calculate three metrics for evaluation: the overall R^2, the "within" R^2 after netting out time-invariant differences in smoke $PM_{2.5}$ across stations and changes in smoke over time common to all stations, and the RMSE on the held-out test set for each of the five folds. The "within" R^2, which is calculated by regressing observed smoke $PM_{2.5}$ on predicted smoke $PM_{2.5}$ while including sets of station and year fixed effects (i.e., separate intercepts for each station and each year), is meant to measure performance in the context of a common downstream inference task in which within-location variation over time, rather than variation between locations, is used to isolate the impact of pollutants or other environmental exposures from other factors that might also shape human outcomes. To understand performance across days with any smoke, and across days with dense smoke, we calculate these metrics on the full sample of smoke days and the set of observations with smoke pollution of at least 50 $\mu g/m^3$.

RESULTS

Model Performance. Spatial out-of-sample model performance, measured over all observations using predictions from models trained on the other four folds, was an R^2 of 0.67, that is, our model explained two-thirds of the variation in smoke $PM_{2.5}$ at stations not used in training. Our "within" performance ($R^2 = 0.65$) approached our overall R^2, indicating that our model is capably predicting local, temporal variation in smoke and not simply differences in average smoke across locations or between years. The model performed well over the entire range of observed smoke $PM_{2.5}$ (Figure 2a and Table S4), including accurate prediction without saturation at very high daily $PM_{2.5}$ levels, which has historically been a challenge for both numerical models and statistical models that often focus on $PM_{2.5}$ observations below a certain threshold[33,61,66] (although, see Reid et al.,[34] for an exception). We also compared model predictions to an alternative data set of ambient $PM_{2.5}$ at private, non-reference-grade PurpleAir monitors[67] and found similar performance ($R^2 = 0.70$) (Figure S18); this additional data set was not used in training, so these results again demonstrate that our model performs well on completely new data and locations. Although the presence of
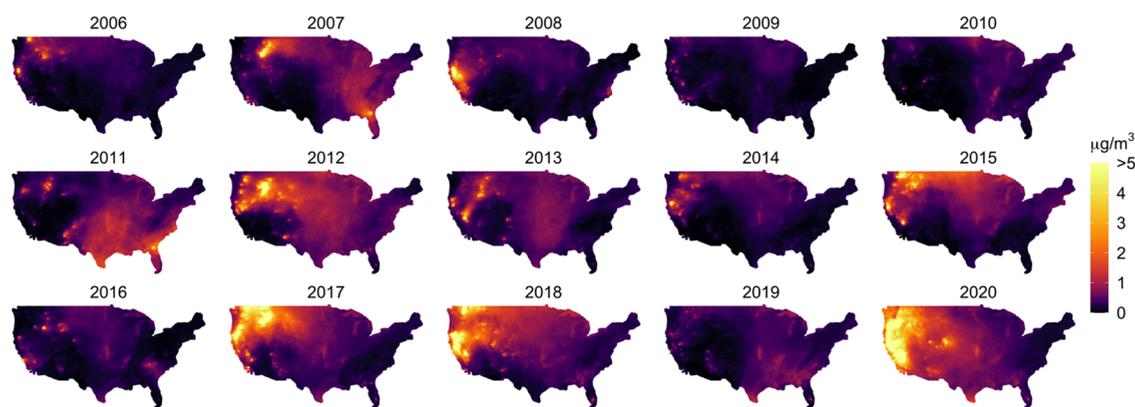
Figure 4. Daily average $PM_{2.5}$ from smoke by year. Each panel shows average daily smoke $PM_{2.5}$ in each location and each year, computed as the average over all days in each grid cell in each year.

satellite-based smoke plumes alone can be an imperfect indicator of surface-level pollution,[68] the performance of the model in predicting held-out surface $PM_{2.5}$ data suggests that the included features allow us to link changes in satellite measurements of aerosols and meteorological conditions to changes in surface $PM_{2.5}$. However, there remain a subset of smoke days where our model struggles to predict variation in surface $PM_{2.5}$, including days when a plume is overhead but observed smoke $PM_{2.5}$ values are near zero—likely cases when smoke is overhead but not mixing to the surface in a way that is predictable with our model features (Figures 2 and S18). However, among days with observed smoke $PM_{2.5}$ <1 $\mu g/m^3$, 94% of out-of-sample predictions are less than 5 $\mu g/m^3$ and less than 1% of predictions exceed 10 $\mu g/m^3$.

Predicted AOD contributed most to model performance, followed by longitude, size of the nearest fire cluster, and dewpoint temperature (Figure 2b). To test the added predictive value from the two most computationally intensive feature sets, we also fit models without AOD predictions and without HYSPLIT features. Including AOD predictions had a higher positive impact on model performance (increase of 0.057 in R^2) than HYSPLIT features (increase of 0.014 in R^2, Figure S10).

We calculated monitor-specific performance when each monitor was not used in model training. We found that the median station performance was comparable to overall model performance (R^2 = 0.63), with 90% of monitor-specific R^2 values falling between 0.21 and 0.88. Model performance varied spatially, with often very high performance at EPA monitors throughout California, the Pacific Northwest, the upper Midwest, and the Northeast and lower performance at monitors in the Southwest and the South (Figure 2c). To better understand this spatial variation in model performance, we then fit simple linear models between station-level R^2 and cross-sectional covariates, including new covariates from an alternative land cover classification (Figure S11).[69−71] We found that after latitude and longitude, shrubland land cover and average and variance in smoke pollution were most predictive of model performance at individual stations, as measured by the absolute predicted change in R^2 when a covariate is changed from its 5th to 95th percentile (Figure 2d). Model performance was higher in locations with higher average and variance in smoke $PM_{2.5}$ and lower in locations with higher percentage shrubland land cover. Higher variance in non-smoke pollution was also associated with worse model performance. Shrubland areas may have less variation in smoke

$PM_{2.5}$, more variation in non-smoke $PM_{2.5}$, leading to less precise estimates of smoke $PM_{2.5}$, or a different relationship between satellite AOD and ground $PM_{2.5}$ due to land cover, any of which could result in lower performance in smoke $PM_{2.5}$ predictions. These results help account for why performance was substantially higher in the Pacific Northwest (high average and variance in smoke $PM_{2.5}$, forest landcover) versus the desert Southwest (low average and variance in smoke $PM_{2.5}$, shrubland landcover); other efforts to predict total $PM_{2.5}$ have also performed substantially worse in the desert Southwest.[5,33,34] By examining individual station time series, we also found that our lowest performing stations were often characterized by a small number of extreme outliers in the ground measurements (Figure S12).

Patterns and Trends in Predicted Smoke $PM_{2.5}$. Predicted daily smoke $PM_{2.5}$ shows remarkable temporal and spatial variation around fire events, consistent with observed smoke behavior around individual fires (Figure S13). For example, Figure 3 shows model predictions for time periods surrounding three specific fire events[72] in California over multiple years. Predicted smoke $PM_{2.5}$ showed increases corresponding to the onset of known fires, with variation in smoke $PM_{2.5}$ within fire events over time and across locations within counties (Figure 3b−d). These heterogeneities aggregate to marked differences in total smoke $PM_{2.5}$ concentrations between nearby locations and highlight the differences lost when using binary smoke measures or when interpolating between monitors.

Daily gridded estimates can also be aggregated to larger temporal or spatial scales to characterize broader patterns and trends in smoke concentrations. We measured the contribution of smoke to annual $PM_{2.5}$ by aggregating daily predictions to the annual level across the contiguous US and found that some locations experienced over 2000 $\mu g/m^3$ of daily accumulated $PM_{2.5}$ from smoke in the span of a year, equivalent to average annual $PM_{2.5}$ levels 5 $\mu g/m^3$ higher (Figure 4). This level is equivalent to roughly half of the overall annual average $PM_{2.5}$ concentrations from all sources across much of the US.[73] Annual hotspots of high smoke $PM_{2.5}$ correspond to locations of large fires (Figures S15 and S16), although impacts often extend beyond the burned areas. The locations most affected varied by year but were primarily concentrated in the West, with larger areas affected in recent years, particularly 2017, 2018, and 2020. The Southwest and Northeast experienced the least smoke pollution over the study period.
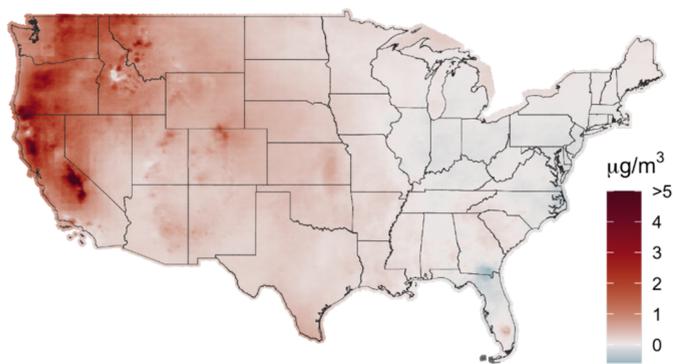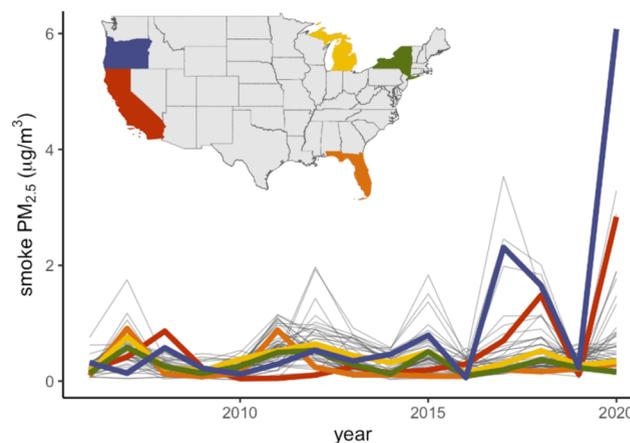
a) Decadal change in smoke PM$_{2.5}$

b) Annual smoke PM$_{2.5}$ by state

Figure 5. Over the last decade, smoke PM$_{2.5}$ has increased in much of the US, particularly in Western states, but some areas in the South and East have seen modest declines. (a) Decadal change in smoke PM$_{2.5}$ is the difference in daily average smoke PM$_{2.5}$ during 2006−2010 and 2016−2020. (b) Population-weighted average smoke PM$_{2.5}$ shows marked increases in Western states and little change in Eastern states. Each line indicates a state, with the inset map showing highlighted states whose colors match the lines corresponding to the states in the main panel. State outlines are reproduced with permission from the US Census Bureau TIGER/Line Shapefiles.[51]
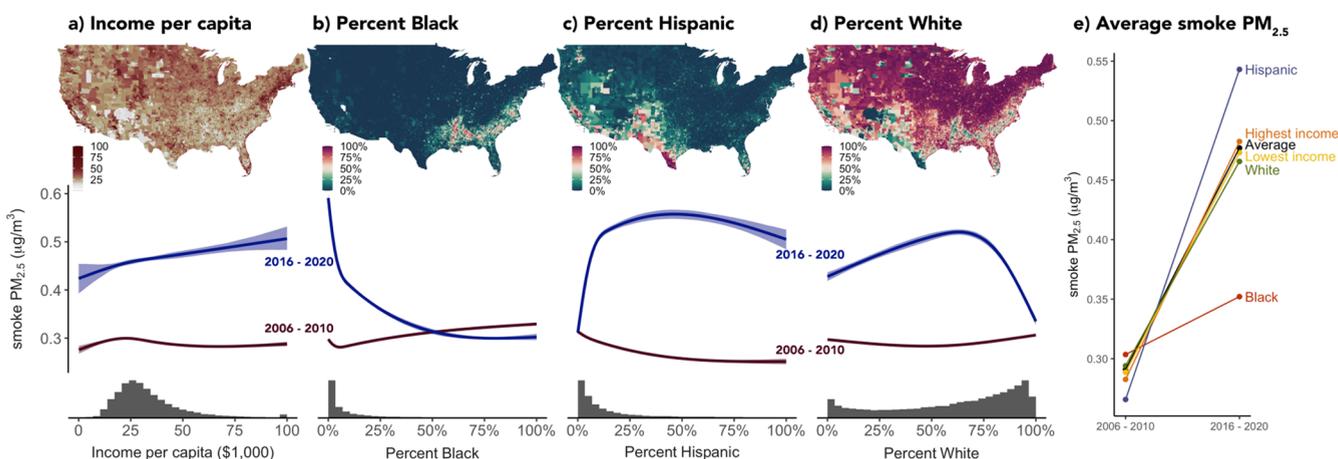

Figure 6. One decade ago, smoke PM$_{2.5}$ levels did not differ meaningfully by income, race, or ethnicity, but recent exposure patterns now differ by race, ethnicity, and, to a smaller extent, income. (a) All income levels experienced similar smoke PM$_{2.5}$ concentrations from 2006 to 2010 and increases in smoke PM$_{2.5}$ to recent years (2016−2020), with the greatest increases occurring at the highest income levels. (b) Smoke PM$_{2.5}$ concentrations have fallen in heavily non-Hispanic Black tracts, consistent with the modest declines seen in the Southeastern region. (c) Concentrations have increased most in tracts with Hispanic populations, corresponding to higher percent Hispanic in Western and Southwestern regions. (d) Smoke PM$_{2.5}$ increases were the smallest in tracts that were heavily non-Hispanic white, concordant with the more modest smoke PM$_{2.5}$ increases in the Midwest and Northeast. For all panels, lines are spline fits between income/race/ethnicity and average smoke concentrations in 2006−2010 (purple) and 2016−2020 (blue) in US Census tracts, with shaded areas showing bootstrapped 95% confidence intervals, while histograms show the distribution across tracts. Tract-level smoke PM$_{2.5}$ is an area-weighted average of smoke PM$_{2.5}$ in intersecting grid cells. Income per capita is top-coded to $100,000. (e) We also calculate the average annual smoke PM$_{2.5}$ concentrations for the average Hispanic individual (blue), White individual (green), Black individual (red), and US resident (black) and individuals living in census tracts in the highest (orange) and lowest (yellow) income quintiles. Census tract outlines are reproduced with permission from the US Census Bureau TIGER/Line Shapefiles.[51]

To better understand longer-run changes in smoke PM$_{2.5}$ concentrations, we calculated the decadal change in smoke pollution as the difference in average annual smoke PM$_{2.5}$ from 2006−2010 to 2016−2020, with 5 year windows used to avoid undue influence from extreme years, such as 2020. Locations west of the Mississippi River saw increases in smoke pollution in the last decade, with the greatest increases in Washington, Oregon, and California (Figure 5a). Some of these Western regions saw decadal increases in an annual smoke PM$_{2.5}$ of 5 μg/m³ or greater, an amount comparable in absolute magnitude to the reduction in PM$_{2.5}$ brought about by the

Clean Air Act in the US.[74,75] Some locations in the Southeast and Northeast saw modest (<1 μg/m³) declines in smoke PM$_{2.5}$, consistent with a decline in burned area observed in existing fire products in many states in the Southeast, East, and Midwest (i.e., Monitoring Trends in Burn Severity,[76] Figure S14) and a decline in days with heavy smoke plumes in the air in the Southeast.[31] The notable declines on the Georgia-Florida border and in central Idaho may be due to the 2007 fires in those areas, which were associated with elevated annual smoke PM$_{2.5}$ (Figures 4 and S16). We also quantify population-weighted concentrations by state for each year
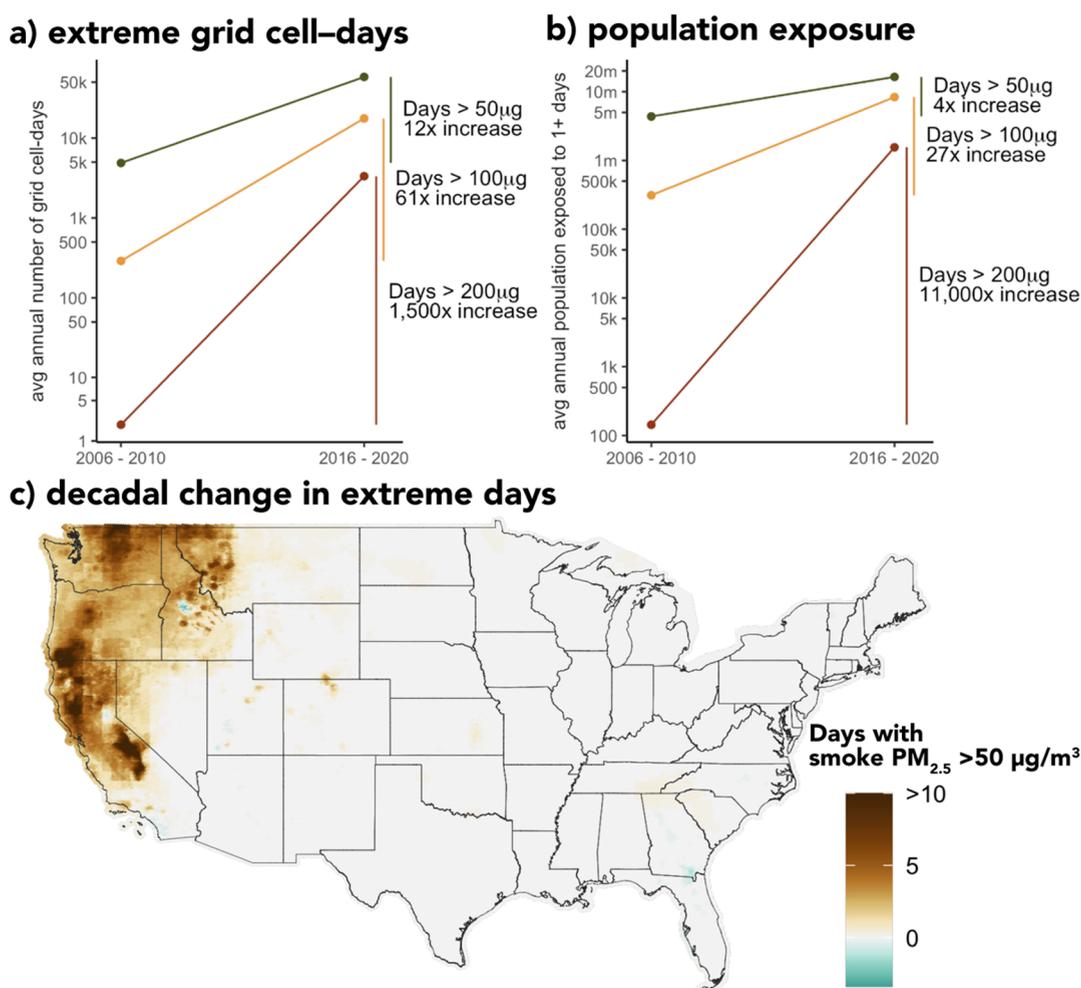
a) extreme grid cell–days

b) population exposure

c) decadal change in extreme days

Figure 7. Changing occurrence of smoke $PM_{2.5}$ extremes. (a) Decadal change in the number of grid cell days with smoke $PM_{2.5}$ above 50 (green), 100 (orange), and 200 $\mu g/m^3$ (red), computed as the difference in average annual counts during 2006−2010 and during 2016−2020. Annotations show the change in multiples of the baseline rate. (b) Decadal change in the population in the US living in places with at least 1 day per year of smoke above 50, 100, or 200 $\mu g/m^3$. For instance, over 16 million people on average in the US lived in places with at least 1 day above 50 $\mu g/m^3$ per year in the last 5 years, a fourfold increase from a decade ago. (c) Map of decadal change in extreme days with smoke $PM_{2.5}$ over 50 $\mu g/m^3$. State outlines are reproduced with permission from the US Census Bureau TIGER/Line Shapefiles.[51]

and find that concentrations were worst for the average Oregon resident in 2020, with annual smoke $PM_{2.5}$ levels over 50 times higher than a decade previous (Figure 5b). On the other hand, residents in Midwestern, Northeastern, and Southern states had similar levels of smoke $PM_{2.5}$ over the last 15 years.

Differences in Smoke Pollution by Race, Ethnicity, and Income. Although $PM_{2.5}$ from smoke is growing throughout most of the country, the geographic patterns in smoke $PM_{2.5}$ may result in differential exposure for different populations. Understanding the magnitudes and sources of disparities in environmental exposures across income and racial/ethnic groups is a key question in environmental justice. We merge our gridded data with census tract level data on income, race, and ethnicity from the US Census Bureau American Community Survey (ACS)[77] to quantify the levels and trends in ambient concentrations. From 2006 to 2010, average annual smoke $PM_{2.5}$ did not differ meaningfully by income or by the percent of tract population represented by non-Hispanic black, non-Hispanic white, or Hispanic individuals (Figure 6). Over the last decade, smoke $PM_{2.5}$ concentrations have grown fastest for predominantly Hispanic

and non-Hispanic white populations, while predominantly non-Hispanic black census tracts have seen a slight decline in smoke $PM_{2.5}$ (Figure 6). Across all income levels, smoke $PM_{2.5}$ has increased, but that increase has been slightly larger in higher income census tracts. These changes reflect the broad geographic patterns in smoke $PM_{2.5}$ trends (Figure 5), as well as the geographic concentrations of different racial and ethnic groups in the country. We also compare average annual smoke $PM_{2.5}$ levels to gross domestic product (GDP) per capita by county[78] and to population density per grid cell.[79] We find that smoke $PM_{2.5}$ concentrations have increased more in counties with higher GDP per capita and have been the smallest in intermediate population density areas, larger in high population density areas, and the largest in very low density areas (Figure S17).

Changing Occurrence of Daily Extremes. Finally, the primary goal of our analysis was to ensure that our model had low bias in predicting not only moderate (and more common) daily smoke $PM_{2.5}$ concentrations but also extreme levels of daily smoke $PM_{2.5}$. Quantifying exposures to extremes is critical for understanding health and related impacts, both because impacts might be non-linear in exposure and because

compensatory behavior (e.g., staying indoors) differs substantially by the level of exposure.[31] Whether resulting concentration−response functions are concave, convex, or linear is not well established in the literature[4] but has important implications for the magnitude of societal impacts and the optimal policy response.

We estimate a substantial increase in exposure to extremes in daily smoke $PM_{2.5}$ over the last decade. Using thresholds of 50, 100, or 200 $\mu g/m^3$ of daily smoke $PM_{2.5}$ concentrations to denote extremes, we calculate a remarkable 12−1500× decadal increase in the annual average count of grid cell days with concentrations above these thresholds (Figure 7a). Because many of these grid cells might be in relatively unpopulated areas of the Western US, we also calculate the number of people residing in locations where at least a day per year exceeded these thresholds, using population estimates from WorldPop.[79] We calculate that each year, on average, 16.4 million people in the US lived in places with at least 1 day of smoke $PM_{2.5}$ above 50 $\mu g/m^3$ during the 2016−2020 period, a fourfold increase from the 2006 to 2010 average (Figure 7b). While less than half a million people lived in areas experiencing a day per year of smoke $PM_{2.5}$ concentrations over 100 $\mu g/m^3$ a decade ago, this number increased to over 8 million in recent years. In proportional terms, increases were the largest for the most extreme threshold, with almost nobody experiencing a day per year above 200 $\mu g/m^3$ a decade ago, but now, over 1.5 million people are residing in locations routinely exposed to these conditions. The locations experiencing the greatest increase in days with extreme smoke $PM_{2.5}$ are concentrated in the Western US, particularly Washington, Oregon, California, Idaho, Montana, and Nevada (Figure 7c).

■ DISCUSSION

Here, we produce 10 km resolution estimates of smoke $PM_{2.5}$ over the contiguous US from 2006 to 2020 on a daily basis. These predictions perform well over the range of observed smoke $PM_{2.5}$ and vary predictably in performance by region (Figure 2). These predictions capture small-scale variation in smoke $PM_{2.5}$ within counties and within fire events (Figure 3) and larger scale interannual variation in smoke pollution (Figure 4). We find steep increases in average smoke pollution and days with smoke $PM_{2.5}$ extremes over the last decade, particularly in the Western US (Figures 5 and 7), resulting in a changing profile of smoke $PM_{2.5}$ concentrations (Figure 6). Consistent with recent work,[4−6] these predictions suggest that the increases in $PM_{2.5}$ from wildfire smoke are rapidly undoing much of the improvements in air quality from the last few decades, at least throughout many areas of the Western US.

In comparison to many existing efforts to estimate smoke $PM_{2.5}$ using CTMs with and without emissions from fires or through sensitivity analyses embedded in CTMs,[80] our approach uses statistical, machine-learning models fit to inferred smoke $PM_{2.5}$ at EPA monitors. Although CTM-based smoke $PM_{2.5}$ estimates trade-off between the computational cost, spatial coverage, and spatial resolution, our statistical approach can achieve high resolution (10 km) estimates over the entire contiguous US, can be validated against long time series of monitor observations, and can easily scale to large areas due to the fast prediction speeds after models have been trained. On the other hand, CTM-based estimates can characterize subdaily variation in smoke $PM_{2.5}$, while our estimates are currently limited to daily resolution. Finally, CTMs can provide speciated pollutant estimates of

$PM_{2.5}$, albeit with substantial uncertainty in concentration estimates,[22] while our focus is on total smoke $PM_{2.5}$.

Our work is related to a recent interpolation-based approach used to estimate wildfire-attributable $PM_{2.5}$ across the US.[5] While we take a similar approach to measuring smoke $PM_{2.5}$ in ground data, we diverge in using a range of satellite- and model-based inputs, rather than interpolation, to estimate smoke $PM_{2.5}$ in locations without stations. Because interpolation by design smooths the input data, such an approach could understate extreme concentrations, both in locations without monitors and on the large number of smoke days when specific monitors do not report concentrations. Finally, we validate our approach against 15 year time series of held-out station observations and against independent private sensor data, which stands in contrast to the single-season evaluation in earlier work.[5] Our work complements recent machine learning-based efforts to use satellites and other inputs to predict total $PM_{2.5}$ concentrations over time.[33,34]

As with most available data sets on specific pollutants, variation in estimated smoke $PM_{2.5}$ in our data is likely correlated with variation in other pollutants including NO_2 and ozone,[47,81] although the emission, formation, and evolution of those other pollutants are an evolving area of research.[82,83] Downstream applications that wish to use our data to isolate the impact of $PM_{2.5}$, specifically on outcomes, must then take care to account for co-varying pollutants that could also affect outcomes. Such applications should also consider accounting for other covarying meteorological variables such as temperature and precipitation, although how best to account for the role of these variables is not necessarily clear, as these variables are plausibly both a driver of wildfire activity (and thus not a confound) and an independent source of impact (and thus possibly a confound).

Future advancements to smoke $PM_{2.5}$ estimates could include improvements to both the monitor-based training data and to our machine learning-based smoke prediction pipeline. Precision of smoke $PM_{2.5}$ estimates at monitors—our training data—is a function of the precision of the non-smoke counterfactual, that is, the estimated background rate of $PM_{2.5}$ absent smoke. We currently use median $PM_{2.5}$ from days without smoke, which accounts for station-specific differences, seasonal variation within the station, and trends over time in $PM_{2.5}$ levels but does not account for more granular variation in other time-varying non-smoke sources of $PM_{2.5}$. This may result in imprecision in our training smoke $PM_{2.5}$ estimates, as anomalous $PM_{2.5}$ may be a noisy estimate of the concentrations from smoke, especially in locations with high variation in background non-smoke $PM_{2.5}$ which was associated with poorer model performance (Figure 2). Future efforts could more explicitly model background non-smoke $PM_{2.5}$, perhaps using machine learning and other information on time variation in other $PM_{2.5}$ sources. Future model development might also benefit from the increasing availability of monitor data as large numbers of private sensors, such as PurpleAir monitors, come online (Figure S18). In this study, we use these latter data as a source of independent evaluation, but such data could also be useful in model training, although their use should consider that these monitors are neither reference grade nor regularly calibrated. Several procedures have been developed for improving $PM_{2.5}$ estimates derived from PurpleAir monitors.[84−88] While the most appropriate calibration likely depends on the setting, we utilized the Barkjohn correction[84] because we found in previous work[31] that it

performed best at predicting out-of-sample PM$_{2.5}$ at EPA reference grade monitors. As higher spatial and temporal densities of training data were predictive of increased model performance (Figure 2d), future inclusion of these data sources should improve the accuracy of estimates. Other features and data sources such as smoke plume height, spatial lags of meteorology, and indicators of atmospheric mixing such as air temperature at different vertical heights have been found in other settings to improve total PM$_{2.5}$ estimates, predict variation in the relationship between PM$_{2.5}$ and AOD, or have the potential to improve the model's ability to identify when smoke mixes to the surface, something the current model occasionally struggles with, as evidenced by the range of predicted values on days with very low observed smoke PM$_{2.5}$ values.[89−92] Future advances could also include alternative machine learning models, such as convolutional neural networks, that take advantage of the spatial information instead of features at a single point and have been found to provide good performance on total PM$_{2.5}$.[93] While our estimates rely on plume boundaries drawn by NOAA analysts over the contiguous US, automation of plume identification—a task for which early computer vision work has shown promise[94−96]—could allow for generalization of this approach to other geographic regions, an effort of increasing importance as wildfires grow in many parts of the world.[11,97−99] Finally, uncertainty quantification from machine learning models is an active area of research, and future improvements to these estimates could include more granular quantification of uncertainty.

A growing literature estimates the impact of smoke on a range of outcomes including health outcomes such as mortality, hospitalizations, and birth outcomes,[100,101] economic outcomes including labor productivity and economic output,[102] and cognitive outcomes such as test scores and mental health.[94] Many of these studies have been limited to short time spans, small geographic ranges, or binary measures of smoke concentrations, in part due to lack of broad-scale validated data on pollutant concentrations attributable to wildfire smoke.[103] Our new estimates should enable improvements in the breadth and precision of these studies and help refine our understanding of how growing smoke PM$_{2.5}$ concentrations affect a range of societal outcomes and are well suited for studying the impact of more common, low-level smoke concentrations and increasingly frequent periods of extreme concentrations. Our model's strong ability to predict within-location variation in smoke PM$_{2.5}$ over time and to generate long time series of concentration estimates are particularly beneficial for understanding societal impacts, given the frequent reliance of modern causal inference approaches on temporal rather than spatial variation in estimating the impact of environmental exposures on societal outcomes. Our estimates can thus help establish smoke-specific concentration−response functions across a range of societal responses of interest and, by more accurately measuring extreme exposures, help better characterize potential non-linearities in these functions. Finally, smoke PM$_{2.5}$ can also be linked to antecedent climate conditions to better understand the past local and regional influence of climate on smoke PM$_{2.5}$ concentrations and to inform how smoke pollution might evolve under future climate.

■ DATA AVAILABILITY

Data and code to replicate all results in the main text and supplementary materials are available at https://github.com/echolab-stanford/daily-10km-smokePM.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.2c02934.

Additional model details including model tuning and model inputs, assessment of smoke plume accuracy and smoke PM$_{2.5}$ estimates using speciated data, additional model performance comparisons, and trends in and annual burned areas (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Marissa L. Childs − *Emmett Interdisciplinary Program in Environment and Resources, Stanford University, Stanford, California 94305, United States;* ⓞ orcid.org/0000-0002-8597-2161; Email: marissac@stanford.edu

Authors

Jessica Li − *Center on Food Security and the Environment, Stanford University, Stanford, California 94305, United States*

Jeffrey Wen − *Department of Earth System Science, Stanford University, Stanford, California 94305, United States;* ⓞ orcid.org/0000-0001-9391-6939

Sam Heft-Neal − *Center on Food Security and the Environment, Stanford University, Stanford, California 94305, United States*

Anne Driscoll − *Center on Food Security and the Environment, Stanford University, Stanford, California 94305, United States*

Sherrie Wang − *Goldman School of Public Policy, UC Berkeley, Berkeley, California 94720, United States*

Carlos F. Gould − *Department of Earth System Science, Stanford University, Stanford, California 94305, United States*

Minghao Qiu − *Department of Earth System Science, Stanford University, Stanford, California 94305, United States;* ⓞ orcid.org/0000-0002-5224-570X

Jennifer Burney − *Global Policy School, UC San Diego, San Diego, California 92093, United States;* ⓞ orcid.org/0000-0003-3532-2934

Marshall Burke − *Center on Food Security and the Environment and Department of Earth System Science, Stanford University, Stanford, California 94305, United States; National Bureau of Economic Research, Cambridge, Massachusetts 02138, United States*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.est.2c02934

Author Contributions

All authors contributed to the conception and design of the study. M.L.C., J.L., J.W., S.H.-N., and A.D. contributed to data extraction and processing. M.L.C. conducted AOD prediction, and J.L. conducted HYSPLIT simulations. M.L.C. and J.L. performed model tuning and evaluation. M.L.C. led the analyses of model performance, smoke PM$_{2.5}$ patterns, and differential exposures, and M.B. led the analysis of daily

extremes. All authors contributed to analyzing results and writing the paper.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) GBD 2019 Risk Factors Collaborators. Global burden of 87 risk factors in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* **2020**, *396*, 1223−1249.

(2) Zhang, Q.; Zheng, Y.; Tong, D.; Shao, M.; Wang, S.; Zhang, Y.; Xu, X.; Wang, J.; He, H.; Liu, W.; Ding, Y.; Lei, Y.; Li, J.; Wang, Z.; Zhang, X.; Wang, Y.; Cheng, J.; Liu, Y.; Shi, Q.; Yan, Y.; Geng, G.; Hong, C.; Li, L.; Liu, F.; Zheng, B.; Cao, J.; Ding, A.; Gao, J.; Fu, Q.; Huo, J.; Liu, B.; Liu, Z.; Yang, F.; He, K.; Hao, J. Drivers of improved PM 2.5 air quality in China from 2013 to 2017. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 24463−24469.

(3) Aldy, J. E.; Auffhammer, M.; Cropper, M.; Fraas, A.; Morgenstern, R. Looking Back at 50 Years of the Clean Air Act. *J. Econ. Lit.* **2022**, *60*, 179−232.

(4) Burke, M.; Driscoll, A.; Heft-Neal, S.; Xue, J.; Burney, J.; Wara, M. The changing risk and burden of wildfire in the united states. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118*, No. e2011048118.

(5) O'Dell, K.; Ford, B.; Fischer, E. V.; Pierce, J. R. Contribution of wildland-fire smoke to us pm2. 5 and its influence on recent trends. *Environ. Sci. Technol.* **2019**, *53*, 1797−1804.

(6) Jaffe, C. D.; Jaffe, D. A. Us particulate matter air quality improves except in wildfire-prone areas. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, 7901−7906.

(7) Meng, J.; Martin, R. V.; Li, C.; van Donkelaar, A.; Tzompa-Sosa, Z. A.; Yue, Y.; Xu, J.-W.; Weagle, C. L.; Burnett, R. T. Source contributions to ambient fine particulate matter for canada. *Environ. Sci. Technol.* **2019**, *53*, 10269−10278.

(8) Liu, J. C.; Mickley, L. J.; Sulprizio, M. P.; Dominici, F.; Xu, Y.; Ebisu, K.; Anderson, G. B.; Khan, R. F. A.; Bravo, M. A.; Bell, M. L. Particulate air pollution from wildfires in the western us under climate change. *Climatic Change* **2016**, *138*, 655−666.

(9) Chen, Y.; Cao, C.; Cao, Y.; Bashir, B.; Xu, M.; Xie, B.; Wang, K. Observed evidence of the growing contributions to aerosol pollution of wildfires with diverse spatiotemporal distinctions in china. *J. Clean. Prod.* **2021**, *298*, 126860.

(10) Xie, Y.; Lin, M.; Decharme, D.; Delire, C.; Horowitz, L. W.; Lawrence, D. M.; Li, F.; Séférian, S. Tripling of western us particulate pollution from wildfires in a warming climate. *Proc. Natl. Acad. Sci. U.S.A.* **2022**, *119*, No. e2111372119.

(11) Williams, J. T.; Williams, A. P. Impact of anthropogenic climate change on wildfire across western us forests. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 11770−11775.

(12) Hurteau, M. D.; Westerling, A. L.; Wiedinmyer, C.; Bryant, B. P. Projected effects of climate and development on california wildfire emissions through 2100. *Environ. Sci. Technol.* **2014**, *48*, 2298−2304.

(13) O'Dell, K.; Bilsback, K.; Ford, B.; Martenies, S. E.; Magzamen, S.; Fischer, E. V.; Pierce, J. R. Estimated mortality and morbidity attributable to smoke plumes in the united states: Not just a western us problem. *GeoHealth* **2021**, *5*, No. e2021GH000457.

(14) Aguilera, R.; Corringham, T.; Gershunov, A.; Benmarhnia, T. Wildfire smoke impacts respiratory health more than fine particles from other sources: observational evidence from southern california. *Nat. Commun.* **2021**, *12*, 1493.

(15) Liu, J. C.; Mickley, L. J.; Sulprizio, M. P.; Dominici, F.; Yue, Y.; Ebisu, K.; Anderson, G. B.; Khan, R. F. A.; Bravo, M. A.; Bell, M. L. Particulate air pollution from wildfires in the western US under climate change. *Climatic Change* **2016**, *138*, 655−666.

(16) Rappold, A. G.; Reyes, J.; Pouliot, P.; Cascio, W. E.; Diaz-Sanchez, D. Community vulnerability to health impacts of wildland fire smoke exposure. *Environ. Sci. Technol.* **2017**, *51*, 6674−6682.

(17) Munoz-Alpizar, R.; Pavlovic, R.; Moran, M. D.; Chen, J.; Gravel, S.; Henderson, S. B.; Ménard, S.; Racine, J.; Duhamel, A.; Gilbert, S.; Beaulieu, P.-A.; Landry, H.; Davignon, D.; Cousineau, S.; Bouchet, V. Multi-Year (2013−2016) PM2.5 wildfire pollution exposure over north america as determined from operational air quality forecasts. *Atmosphere* **2017**, *8*, 179.

(18) Koman, P. D.; Billmire, M.; Baker, K. R.; de Majo, R.; Anderson, F. J.; Hoshiko, S.; Thelen, B. J.; French, N. H. F. Mapping modeled exposure of wildland fire smoke for human health studies in california. *Atmosphere* **2019**, *10*, 308.

(19) Larsen, A.; Yang, S.; Reich, B. J.; Rappold, A. G. A spatial causal analysis of wildland fire-contributed pm2. 5 using numerical model output. March 12, **2020**, Statistics, Methodology, arXiv:2003.06037 accessed on April 1, 2022.

(20) Neumann, J. E.; Meredith, A.; Anenberg, S.; Kinney, P. L.; Marcus, S.; Martinich, J.; Lukens, J.; Xu, J.-W.; Henry, R. Estimating PM2.5-related premature mortality and morbidity associated with future wildfire emissions in the western US. *Environ. Res. Lett.* **2021**, *16*, 035019.

(21) Jaffe, D. A.; O'Neill, S. M.; Larkin, N. K.; Holder, A. L.; Peterson, D. L.; Halofsky, J. E.; Rappold, A. G. Wildfire and prescribed burning impacts on air quality in the united states. *J. Air Waste Manage. Assoc.* **2020**, *70*, 583−615.

(22) Koplitz, S. N.; Nolte, C. G.; Pouliot, G. A.; Vukovich, J. M.; Beidler, J. Influence of uncertainties in burned area estimates on modeled wildland fire PM2.5 and ozone pollution in the contiguous U.S. *Atmos. Environ.* **2018**, *191*, 328−339.

(23) Li, Y.; Tong, D.; Ma, S.; Zhang, X.; Kondragunta, S.; Li, F.; Saylor, R. Dominance of wildfires impact on air quality exceedances during the 2020 record-breaking wildfire season in the united states. *Geophys. Res. Lett.* **2021**, *48*, No. e2021GL094908.

(24) Herron-Thorpe, F. L.; Mount, G. H.; Emmons, L. K.; Lamb, B. K.; Jaffe, D. A.; Wigder, N. L.; Chung, S. H.; Zhang, R.; Woelfle, M. D.; Vaughan, J. K. Air quality simulations of wildfires in the pacific northwest evaluated with surface and satellite observations during the summers of 2007 and 2008. *Atmos. Chem. Phys.* **2014**, *14*, 12533−12551.

(25) Gunsch, M. J.; May, N. W.; Wen, M.; Bottenus, C. L. H.; Gardner, D. J.; VanReken, T. M.; Bertman, S. B.; Hopke, P. K.; Ault, A. P.; Pratt, K. A. Ubiquitous influence of wildfire emissions and secondary organic aerosol on summertime atmospheric aerosol in the forested great lakes region. *Atmos. Chem. Phys.* **2018**, *18*, 3701−3715.

(26) Carter, T. S.; Heald, L. H.; Jimenez, J. L.; Campuzano-Jost, P.; Kondo, Y.; Moteki, N.; Schwarz, J. P.; Wiedinmyer, C.; Darmenov, A. S.; da Silva, A. M.; Kaiser, J. W. How emissions uncertainty influences

the distribution and radiative impacts of smoke from fires in north america. *Atmos. Chem. Phys.* **2020**, *20*, 2073.

(27) Paugam, R.; Wooster, M.; Freitas, S.; Val Martin, M. V. A review of approaches to estimate wildfire plume injection height within large-scale atmospheric chemical transport models. *Atmos. Chem. Phys.* **2016**, *16*, 907−925.

(28) Ye, X.; Arab, P.; Ahmadov, R.; James, E.; Grell, G. A.; Pierce, P.; Kumar, A.; Makar, M.; Chen, J.; Davignon, D.; Carmichael, G. R.; Ferrada, G.; McQueen, J.; Huang, J.; Kumar, R.; Emmons, L.; Herron-Thorpe, F. L.; Parrington, M.; Engelen, R.; Peuch, V.-H.; da Silva, A.; Soja, A.; Gargulinski, E.; Wiggins, E.; Hair, J. W.; Fenn, M.; Shingler, S.; Kondragunta, S.; Lyapustin, A.; Wang, Y.; Holben, H.; Giles, D. M.; Saide, P. E. Evaluation and intercomparison of wildfire smoke forecasts from multiple modeling systems for the 2019 williams flats fire. *Atmos. Chem. Phys.* **2021**, *21*, 14427−14469.

(29) Brey, S. J.; Ruminski, M.; Atwood, S. A.; Fischer, E. V. Connecting smoke plumes to sources using hazard mapping system (hms) smoke and fire location data over north america. *Atmos. Chem. Phys.* **2018**, *18*, 1745−1761.

(30) Ruminski, M.; Kondragunta, S.; Roland, D.; Zeng, J. Recent changes to the hazard mapping system. *Proceedings of the 15th International Emission Inventory Conference*, 2006; Vol. 15, p 18.

(31) Burke, M.; Heft-Neal, S.; Li, J.; Driscoll, A.; Baylis, P.; Stigler, M.; Weill, J. A.; Burney, J. A.; Wen, J.; Childs, M. L.; Gould, C. F. Exposures and behavioural responses to wildfire smoke. *Nat. Human Behav.* **2022**, 1.

(32) Schneider, S. R.; Lee, K.; Santos, G.; Abbatt, J. P. D. Air Quality Data Approach for Defining Wildfire Influence: Impacts on PM2.5, NO2, CO, and O3 in Western Canadian Cities. *Environ. Sci. Technol.* **2021**, *55*, 13709−13717.

(33) Di, Q.; Amini, H.; Shi, L.; Kloog, I.; Silvern, R.; Kelly, J.; Sabath, M.; Choirat, C.; Koutrakis, P.; Lyapustin, A.; Wang, Y.; Mickley, L. J.; Schwartz, J. An ensemble-based model of pm2.5 concentration across the contiguous united states with high spatiotemporal resolution. *Environ. Int.* **2019**, *130*, 104909.

(34) Reid, C. E.; Considine, E. M.; Maestas, M. M.; Li, G. Daily PM2.5 concentration estimates by county, ZIP code, and census tract in 11 western states 2008-2018. *Sci. Data* **2021**, *8*, 112.

(35) R Core Team *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020. https://www.R-project.org/ (accessed on June 22, 2020).

(36) National Oceanic and Atmospheric Administration Hazard mapping system fire and smoke product. https://www.ospo.noaa.gov/Products/land/hms.html#about (accessed on Nov 22, 2021).

(37) Schroeder, W.; Ruminski, M.; Csiszar, I.; Giglio, L.; Prins, E.; Schmidt, C.; Morisette, J. Validation analyses of an operational fire monitoring product: The hazard mapping system. *Int. J. Rem. Sens.* **2008**, *29*, 6059−6066.

(38) Rolph, G. D.; Draxler, R. R.; Stein, A. F.; Taylor, A.; Ruminski, M. G.; Kondragunta, S.; Zeng, J.; Huang, H.-C.; Manikin, G.; McQueen, J. T.; Davidson, P. M. Description and verification of the noaa smoke forecasting system: the 2007 fire season. *Weather Forecast.* **2009**, *24*, 361−378.

(39) Stein, A. F.; Draxler, R. R.; Rolph, G. D.; Stunder, B. J. B.; Cohen, M. D.; Ngan, F. NOAA's HYSPLIT Atmospheric Transport and Dispersion Modeling System. *Bull. Am. Meteor. Soc.* **2015**, *96*, 2059−2077.

(40) US Environmental Protection Agency Outdoor air quality data. https://www.epa.gov/outdoor-air-quality-data/download-daily-data (accessed on June 8, 2021).

(41) US Environmental Protection Agency. Interagency monitoring of protected visual environments (improve). http://vista.cira.colostate.edu/Improve/ (accessed on Dec 10, 2020).

(42) US Environmental Protection Agency. Chemical speciation network (csn). https://www.epa.gov/amtic/chemical-speciation-network-csn (accessed on March 11, 2021).

(43) Malm, W. C.; Sisler, J. F.; Huffman, D.; Eldred, R. A.; Cahill, T. A. Spatial and seasonal trends in particle concentration and optical extinction in the united states. *J. Geophys. Res. Atmos.* **1994**, *99*, 1347−1370.

(44) Gorham, K. A.; Raffuse, S. M.; Hyslop, N. P.; White, W. H. Comparison of recent speciated pm2.5 data from collocated csn and improve measurements. *Atmos. Environ.* **2021**, *244*, 117977.

(45) Hand, J. L.; Schichtel, B. A.; Malm, W. C.; Frank, N. H. Spatial and temporal trends in pm2. 5 organic and elemental carbon across the united states. *Adv. Meteorol.* **2013**, *2013*, 1.

(46) Hand, J. L.; Prenni, A. J.; Schichtel, B. A.; Malm, W.; Chow, J. C. Trends in remote PM2.5 residual mass across the United States: Implications for aerosol mass reconstruction in the IMPROVE network. *Atmos. Environ.* **2019**, *203*, 141−152.

(47) Liu, X.; Zhang, Y.; Huey, L. G.; Yokelson, R. J.; Wang, Y.; Jimenez, J. L.; Campuzano-Jost, P.; Beyersdorf, A. J.; Blake, D. R.; Choi, Y.; St Clair, J. M.; Crounse, J. D.; Day, D. A.; Diskin, G. S.; Fried, A.; Hall, S. R.; Hanisco, T. F.; King, L. E.; Meinardi, S.; Mikoviny, T.; Palm, B. B.; Peischl, J.; Perring, A. E.; Pollack, I. B.; Ryerson, T. B.; Sachse, G.; Schwarz, J. P.; Simpson, I. J.; Tanner, D. J.; Thornhill, K. L.; Ullmann, K.; Weber, R. J.; Wennberg, W.; Wisthaler, A.; Wolfe, G. M.; Ziemba, L. D. Agricultural fires in the southeastern U.S. during SEAC 4 RS: Emissions of trace gases and particles and evolution of ozone, reactive nitrogen, and organic aerosol. *J. Geophys. Res. Atmos.* **2016**, *121*, 7383−7414.

(48) Rickly, P.; Guo, H.; Campuzano-Jost, P.; Jimenez, J. L.; Wolfe, G. M.; Bennett, R.; Bourgeois, I.; Crounse, J. D.; Dibb, J. E.; DiGangi, J. P.; Diskin, G. S.; Dollner, M.; Gargulinski, E. M.; Hall, S. R.; Halliday, T. F.; Hanisco, R. A.; Hannun, J.; Liao, R.; Moore, A.; Nault, J. B. N.; Nault, C. E.; Nowak, T.; Robinson, K. J.; Ryerson, M.; Sanchez, A. J.; Schöberl, M.; Soja, S.; St Clair, K. L.; Thornhill, U.; Ullmann, P. O.; Wennberg, B.; Weinzierl, E. B.; Wiggins, E. L.; Winstead, A. W.; Rollins, A. W. Emission factors and evolution of SO$_2$ measured from biomass burning in wild and agricultural fires. *Atmos. Chem. Phys. Discuss.* **2022**, 1.

(49) Martin, R. V.; Brauer, M.; van Donkelaar, A.; Shaddick, S.; Narain, U.; Dey, S. No one knows which city has the highest concentration of fine particulate matter. *Atmos. Environ.: X* **2019**, *3*, 100040.

(50) Walker, K. tigris: Load Census TIGER/Line Shapefiles, 2021. R package version 1.4. https://CRAN.R-project.org/package=tigris (accessed on May 16, 2021).

(51) US Census Bureau 2019 TIGER/Line Shapefiles, 2019. Accessed using the R package tigris. Shapefiles are also available from https://www.census.gov/cgi-bin/geo/shapefiles/index.php.

(52) Hersbach, H.; Bell, B.; Berrisford, P.; Biavati, G.; Horányi, A.; Sabater, J. M.; Nicolas, J.; Peubey, C.; Radu, R.; Rozum, I.; Schepers, D.; Simmons, A.; Soci, C.; Dee, D.; Thépaut, J.-N. Era5 hourly data on pressure levels from 1979 to present. *Copernicus Climate Change Service (c3s) Climate Data Store (cds)*, 10, 2018 (accessed on February 20, 2022).

(53) Sabater, J. M. *Era5-land Hourly Data from 1981 to Present, Copernicus Climate Change Service (C3s) Climate Data Store (Cds) [data Set]*, 2019.

(54) Global Modeling and Assimilation Office (GMAO) *MERRA-2 tavg1_2d_aer_Nx: 2d,1-Hourly, Time-averaged, Single-Level, Assimilation, Aerosol Diagnostics V5.12.4*; Goddard Earth Sciences Data and Information Services Center (GES DISC): Greenbelt, MD, USA, 2015.

(55) Li, L.; Franklin, M.; Girguis, M.; Lurmann, F.; Wu, J.; Pavlovic, N.; Breton, C.; Gilliland, G.; Habre, R. Spatiotemporal imputation of maiac aod using deep learning with downscaling. *Remote Sens. Environ.* **2020**, *237*, 111584.

(56) Lyapustin, A.; Wang, Y.; Korkin, S.; Huang, D. Modis collection 6 maiac algorithm. *Atmos. Meas. Tech.* **2018**, *11*, 5741−5765.

(57) Bi, J.; Belle, H. B.; Wang, Y.; Lyapustin, A. I.; Wildani, A.; Liu, Y. Impacts of snow and cloud covers on satellite-derived PM2.5 levels. *Remote Sens. Environ.* **2019**, *221*, 665−674.

(58) Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference*

on Knowledge Discovery and Data Mining, KDD '16; ACM, 2016; pp 785−794.

(59) US Geological Survey *National Elevation Dataset 1/3 Arc-Second (Ned 1/3)*; Courtesy of The US Geological Survey, 2009.

(60) Dewitz, J. *National Land Cover Database (Nlcd) 2016 Products*; US Geological Survey Data Release, 2019.

(61) Hu, X.; Belle, H. B.; Meng, M.; Wildani, A.; Waller, L. A.; Strickland, M. J.; Liu, Y. Estimating PM2.5 Concentrations in the Conterminous United States Using the Random Forest Approach. *Environ. Sci. Technol.* 2017, 51, 6936−6944.

(62) Just, A. C.; Arfer, K. B.; Rush, J.; Dorman, M.; Shtein, A.; Lyapustin, A.; Kloog, I. Advancing methodologies for applying machine learning and evaluating spatiotemporal models of fine particulate matter (PM2.5) using satellite data over large regions. *Atmos. Environ.* 2020, 239, 117649.

(63) Chen, Z.-Y.; Jin, J.-Q.; Zhang, R.; Zhang, T.-H.; Chen, J.-J.; Yang, J.; Ou, C.-Q.; Guo, Y. Comparison of Different Missing-Imputation Methods for MAIAC (Multiangle Implementation of Atmospheric Correction) AOD in Estimating Daily PM2.5 Levels. *Remote Sens.* 2020, 12, 3008.

(64) Zamani Joharestani, M.; Cao, C.; Ni, X.; Bashir, B.; Talebiesfandarani, S. PM2.5 Prediction Based on Random Forest, XGBoost, and Deep Learning Using Multisource Remote Sensing Data. *Atmosphere* 2019, 10, 373.

(65) Xu, Y.; Ho, C.; Wong, M. S.; Deng, C.; Shi, S.; Chan, T.-C.; Knudby, K. Evaluation of machine learning techniques with multiple remote sensing datasets in estimating monthly concentrations of ground-level PM2.5. *Environ. Pollut.* 2018, 242, 1417−1426.

(66) Geng, G.; Murray, N. L.; Tong, J. S.; Fu, X.; Hu, P.; Lee, M.; Meng, H. H.; Chang, Y.; Liu, Y. Satellite-Based Daily PM 2.5 Estimates During Fire Seasons in Colorado. *J. Geophys. Res. Atmos.* 2018, 123, 8159−8171.

(67) Purpleair sensor data. https://thingspeak.com/, Via JSON in accordance with PurpleAir terms and conditions (accessed on Feb 11, 2021).

(68) Buysse, C. E.; Kaulfus, A.; Nair, U.; Jaffe, D. A. Relationships between particulate matter, ozone, and nitrogen oxides during urban smoke events in the western us. *Environ. Sci. Technol.* 2019, 53, 12519−12528.

(69) Friedl, M.; Sulla-Menashe, D. *MCD12Q1-MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006 [Data Set] (NASA Eosdis Land Processes Daac, 2019)*, 2020.

(70) Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, I.; Thau, D.; Moore, R. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* 2017, 202, 18−27.

(71) Aybar, C.; Wu, Q.; Bautista, L.; Yali, Y.; Barja, B. rgee: An r package for interacting with google earth engine. *J. Open Source Software* 2020, 5, 2272.

(72) CAL FIRE, California Department of Forestry and Fire Protection Fire perimeters. https://frap.fire.ca.gov/frap-projects/fire-perimeters/ (accessed on Feb 7, 2022).

(73) US Environmental Protection Agency. Particulate matter (pm2.5) trends, 2022. https://www.epa.gov/air-trends/particulate-matter-pm25-trends (accessed on March 24, 2022).

(74) Chay, K.; Greenstone, M. *Air Quality, Infant Mortality, and the Clean Air Act of 1970*, NBER Working Paper, 2003.

(75) Heft-Neal, S.; Burney, J.; Bendavid, E.; Burke, M. Robust relationship between air quality and infant mortality in africa. *Nature* 2018, 559, 254−258.

(76) Monitoring Trends in Burn Severity (MTBS) Project (USDA Forest Service/U.S. Geological Survey). MTBS data access: Burned areas boundaries dataset, 1984−2020 https://www.mtbs.gov/direct-download (accessed on July 22, 2022).

(77) US Census Bureau American community survey 5-year estimates. Accessed via R package tidycensus.

(78) Regional Economic Accounts *Cagdp9: Real gdp in Chained Dollars by County and msa*; Bureau of Economic Analysis, US Department of Commerce, 2022. https://apps.bea.gov/regional/downloadzip.cfm (accessed on July 6, 2022).

(79) Stevens, F. R.; Gaughan, A. E.; Linard, C.; Tatem, A. J. Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PLoS One* 2015, 10, No. e0107042.

(80) Huang, R.; Qin, M.; Hu, Y.; Russell, A. G.; Odman, M. T. Apportioning prescribed fire impacts on PM2.5 among individual fires through dispersion modeling. *Atmos. Environ.* 2020, 223, 117260.

(81) Jaffe, C. D.; Jaffe, D. A. Investigation of high ozone events due to wildfire smoke in an urban area. *Atmos. Environ.* 2018, 194, 146−157.

(82) Calahorrano, J. F. J.; Lindaas, J.; O'Dell, K.; Palm, B. B.; Peng, Q.; Frank, F.; Pollack, I. B.; Garofalo, L. A.; Farmer, D. K.; Pierce, J. R.; Collett, J. L., Jr.; Weinheimer, A.; Campos, T.; Hornbrook, R. S.; Hall, S. R.; Kirk, U.; Pothier, M. A.; Eric, C. A.; Wade, P.; Hu, L.; Hills, A. J.; Montzka, D.; Tyndall, G.; Thornton, J. A.; Fischer, E. V. Daytime oxidized reactive nitrogen partitioning in western us wildfire smoke plumes. *J. Geophys. Res. Atmos.* 2021, 126, No. e2020JD033484.

(83) Xu, L.; Crounse, J. D.; Vasquez, K. T.; Allen, H.; Wennberg, P. O.; Bourgeois, I.; Brown, S. S.; Campuzano-Jost, P.; Coggon, M. M.; Crawford, J. H.; DiGangi, J. P.; Diskin, G. S.; Fried, A.; Gargulinski, E. M.; Gilman, J. B.; Gkatzelis, G. I.; Guo, H.; Hair, J. W.; Hall, S. R.; Halliday, H. A.; Hanisco, T. F.; Hannun, R. A.; Holmes, C. D.; Huey, L. G.; Jimenez, J. L.; Lamplugh, A.; Lee, Y. R.; Liao, J.; Lindaas, J.; Neuman, J. A.; Nowak, J. B.; Peischl, J.; Peterson, D. A.; Piel, F.; Richter, D.; Rickly, P.; Robinson, M. A.; Rollins, A. W.; Ryerson, T. B.; Sekimoto, K.; Selimovic, V.; Shingler, S.; Soja, A. J.; St Clair, J. M.; Tanner, D. J.; Ullmann, U.; Veres, P. R.; Walega, W.; Warneke, C.; Washenfelder, R. A.; Weibring, P.; Wisthaler, A.; Wolfe, W. M.; Womack, C. C.; Yokelson, R. J. Ozone chemistry in western U.S. wildfire plumes. *Sci. Adv.* 2021, 7, No. eabl3648.

(84) Barkjohn, K. K.; Gantt, B.; Clements, A. L. Development and application of a United States-wide correction for PM2.5 data collected with the PurpleAir sensor. *Atmos. Meas. Tech.* 2021, 14, 4617−4637.

(85) Wallace, L.; Bi, J.; Ott, W. R.; Sarnat, J.; Liu, Y. Calibration of low-cost PurpleAir outdoor monitors using an improved method of calculating PM. *Atmos. Environ.* 2021, 256, 118432.

(86) Bi, J.; Wallace, L. A.; Sarnat, J. A.; Liu, Y. Characterizing outdoor infiltration and indoor contribution of PM2.5 with citizen-based low-cost monitoring data. *Environ. Pollut.* 2021, 276, 116763.

(87) Krebs, B.; Burney, J.; Zivin, J. G.; Neidell, M. Using crowd-sourced data to assess the temporal and spatial relationship between indoor and outdoor particulate matter. *Environ. Sci. Technol.* 2021, 55, 6107−6115.

(88) Singer, W. W.; Singer, B. C. Wildfire Smoke Adjustment Factors for Low-Cost and Professional PM2.5 Monitors with Optical Sensors. *Sensors* 2020, 20, 3683.

(89) Lyapustin, A.; Wang, Y.; Korkin, S.; Kahn, R.; Winker, D. Maiac thermal technique for smoke injection height from modis. *Geosci. Rem. Sens. Lett. IEEE* 2019, 17, 730−734.

(90) Cheeseman, M.; Ford, B.; Volckens, J.; Lyapustin, A.; Pierce, J. R. The relationship between maiac smoke plume heights and surface pm. *Geophys. Res. Lett.* 2020, 47, No. e2020GL088949.

(91) Qiu, M.; Zigler, C.; Selin, E. S. Statistical and machine learning methods for evaluating trends in air quality under changing meteorological conditions. *Atmos. Chem. Phys.* 2022, 22, 10551−10566.

(92) Tai, A. P. K.; Mickley, L. J.; Jacob, D. J.; Leibensperger, L.; Zhang, J. A.; Fisher, H. O. T.; Pye, H. O. T. Meteorological modes of variability for fine particulate matter (PM2.5) air quality in the United States: implications for PM2.5 sensitivity to climate change. *Atmos. Chem. Phys.* 2012, 12, 3131−3145.

(93) Park, Y.; Kwon, B.; Heo, J.; Hu, X.; Liu, Y.; Moon, T. Estimating PM2.5 concentration of the conterminous United States via interpretable convolutional neural networks. *Environ. Pollut.* 2020, 256, 113395.

(94) Wen, J.; Burke, M. Wildfire smoke plume segmentation using geostationary satellite imagery. September 3, 2021. arXiv (Computer Science, Computer Vision and Pattern Recognition),

arXiv:2109.01637. https://doi.org/10.48550/arXiv.2109.01637 (accessed April 14, 2022).

(95) Ramasubramanian, M.; Kaulfus, A.; Maskey, M.; Ramachandran, R.; Gurung, I.; Freitag, B.; Christopher, S. Pixel level smoke detection model with deep neural network. *Image and Signal Processing for Remote Sensing XXV*; International Society for Optics and Photonics, 2019; Vol. *11155*, p 1115515.

(96) Larsen, A.; Hanigan, I.; Reich, B. J.; Qin, Y.; Cope, M.; Morgan, G.; Rappold, A. G. A deep learning approach to identify smoke plumes in satellite imagery in near-real time for health risk communication. *J. Expo. Sci. Environ. Epidemiol.* **2021**, *31*, 170−176.

(97) Gillett, N. P.; Weaver, A. J.; Zwiers, F. W.; Flannigan, M. D. Detecting the effect of climate change on canadian forest fires. *Geophys. Res. Lett.* **2004**, *31*, L18211.

(98) Balch, J. K.; Abatzoglou, J. T.; Joseph, M. B.; Koontz, M. J.; Mahood, A. L.; McGlinchy, M. G.; Cattau, M. E.; Williams, A. P. Warming weakens the night-time barrier to global fire. *Nature* **2022**, *602*, 442−448.

(99) Jolly, W.; Cochrane, M. A.; Freeborn, P. H.; Holden, Z. A.; Brown, T. J.; Williamson, G. J.; Bowman, D. M. J. S. Climate-induced variations in global wildfire danger from 1979 to 2013. *Nat. Commun.* **2015**, *6*, 7537.

(100) Reid, C. E.; Brauer, M.; Johnston, F. H.; Jerrett, M.; Balmes, J. R.; Elliott, C. T. Critical review of health impacts of wildfire smoke exposure. *Environ. Health Perspect.* **2016**, *124*, 1334−1343.

(101) Heft-Neal, S.; Driscoll, A.; Yang, W.; Shaw, G.; Burke, M. Associations between wildfire smoke exposure during pregnancy and risk of preterm birth in california. *Environ. Res.* **2022**, *203*, 111872.

(102) Borgschulte, M.; Molitor, D.; Zou, E. *Air Pollution and the Labor Market: Evidence from Wildfire Smoke*, NBER Working Paper Series, Rev Econ Stat, 2018.

(103) Liu, C.; Pereira, G.; Uhl, S. A.; Bravo, M. A.; Bell, M. L. A systematic review of the physical health impacts from non-occupational exposure to wildfire smoke. *Environ. Res.* **2015**, *136*, 120−132.

Supplemental Information

Daily local-level estimates of ambient wildfire smoke PM$_{2.5}$ for the contiguous US

Marissa L. Childs[1,*], Jessica Li[2], Jeffrey Wen[3], Sam Heft-Neal[2], Anne Driscoll[2], Sherrie Wang[4], Carlos F. Gould[3], Minghao Qiu[3], Jennifer Anne Burney[5], and Marshall Burke[2,3,6]

[1]Emmett Interdisciplinary Program in Environment and Resources, Stanford University, Stanford CA 94305, USA

[2]Center on Food Security and the Environment, Stanford University, Stanford CA 94305, USA

[3]Dept. of Earth System Science, Stanford University, Stanford CA 94305, USA

[4]Goldman School of Public Policy, UC Berkeley, Berkeley CA 94720, USA

[5]Global Policy School, UC San Diego, San Diego CA 92093, USA

[6]National Bureau of Economic Research, Cambridge MA 02138, USA

[*]Communicating author: marissac@stanford.edu

31 pages

18 figures

4 tables

Defining smoke days and calculating ground-based PM$_{2.5}$ As described in the main text, our smoke day designation relies heavily on HMS smoke plume data. As the HMS product in turn relies on imagery from the GOES-West and GOES-East satellites[1], changes to these satellites—including the launches of GOES-16 and GOES-17—may have corresponding changes in plume detection, which we find some evidence of in declines in the 5th and 10th percentiles of plume size (Fig. S3). For days when HMS plume information is unavailable (24 dates during 2006 - 2020), we assign grid cell-days as smoke days based on the temporally-nearest adjacent days with available plume classification, defaulting to non-smoke days, e.g., if a day with missing plume information is preceded by smoke but followed by no smoke, it is classified as a non-smoke day (more detail on this is given below).

HYSPLIT simulations   The HYSPLIT model is a system for computing particle transport trajectories using gridded meteorology data developed by the NOAA Air Resources Laboratory[2]. We use the HYSPLIT model to simulate the forward trajectories of smoke particles emitted at smoke-producing fire points ("HYSPLIT points"). Fire points are hotspots that are detected automatically and confirmed manually or detected manually by trained analysts, while HYSPLIT points are a subset of fire points confirmed to be smoke-producing by an analyst using visible satellite imagery[3,4]. Analysts assign co-located or nearly co-located HYSPLIT points in clusters to reflect the amount of smoke coming from a single fire. Each HYSPLIT point is assigned a number of hours over which it produced smoke ("duration"), and HYSPLIT point start time may vary within a cluster.

HYSPLIT points, fire points, and smoke plumes from January 1, 2006 to December 31, 2020 are downloaded from NOAA HMS[1]. NOAA HMS data are available from the HMS Fire and Smoke Data portal (`https://www.ospo.noaa.gov/Products/land/hms.html#data`) or its backend (`https://satepsanone.nesdis.noaa.gov/pub/FIRE/web/HMS/`) and the HMS Archive (`https://satepsanone.nesdis.noaa.gov/pub/volcano/FIRE/HMS_ARCHIVE/`) or its backup (`https://satepsanone.nesdis.noaa.gov/pub/FIRE/HMS/hms_backup/`).
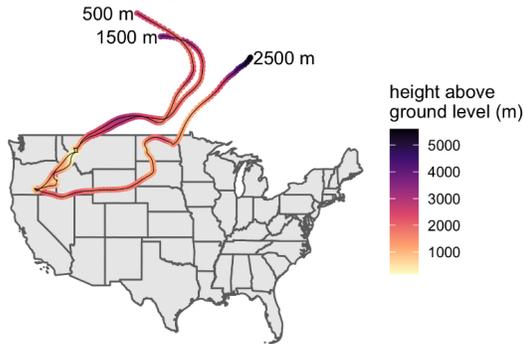
To generate HYSPLIT trajectories, we use the R package *splitr*, executed in parallel on a high-performance computing cluster, with Global Data Assimilation System (GDAS1) meteorology data[5]. Following Brey et al.[4], we generate trajectories covering the first 6 days of particle transport. We initialize trajectories at 500, 1500, and 2500 meters above ground level, given uncertainty about smoke injection heights (Fig. S1a). Trajectories are initialized at HYSPLIT points periodically throughout the observed duration. Specifically, for each injection height:

- if a HYSPLIT point has duration in $[0, 6]$ hours, then 1 trajectory is initialized at midway through the duration;

- if a HYSPLIT point has duration in $(6, 12]$ hours, then 2 trajectories are initialized at $\frac{1}{3}$ and $\frac{2}{3}$ through the duration;

- if a HYSPLIT point has duration in $(12, 18]$ hours, then 3 trajectories are initialized at $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{3}{4}$ through the duration;

- if a HYSPLIT point has duration in $(18, 24]$ hours, then 4 trajectories are initialized at $\frac{1}{5}$, $\frac{2}{5}$, $\frac{3}{5}$, and $\frac{4}{5}$ through the duration;

- if a HYSPLIT point has duration in $(24, 30]$ hours, then 5 trajectories are initialized at $\frac{1}{6}$, $\frac{1}{3}$, $\frac{1}{2}$, $\frac{2}{3}$, and $\frac{5}{6}$ through the duration;

- if a HYSPLIT point has duration above 30 hours (the maximum observed is 33 hours), then 6 trajectories are initialized at $\frac{1}{7}$, $\frac{2}{7}$, $\frac{3}{7}$, $\frac{4}{7}$, $\frac{5}{7}$, and $\frac{6}{7}$ through the duration.
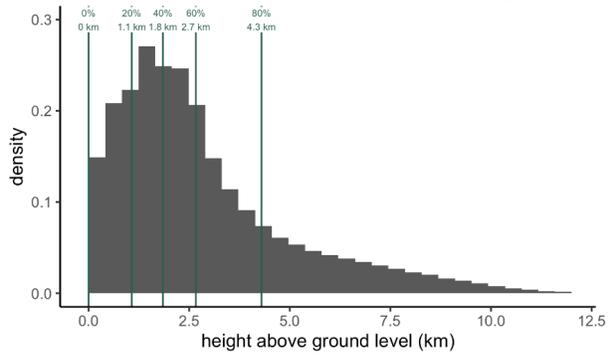
As durations may exceed the maximum duration observed in Brey et al. of 24 hours, this differs slightly from Brey et al. We round trajectory start times to the nearest hour, to conform with the capabilities of the *splitr* package. To avoid unnecessary computation, we generate the sample of unique trajectories and from this produce a representatively weighted sample by duplicating trajectories according to the number of duplicate initialization point-hours. We discard trajectory points that occur after a trajectory hits the ground.

To convert HYSPLIT trajectories into features for the smoke $PM_{2.5}$ model, for each 10km grid cell-day, we obtain counts of trajectory points within a 50km circular buffer centered at the cell centroid and each trajectory point height quintile (Fig. S1c). We use a 50km buffer to account for the fact that discrete trajectory points reflect neither the continuous travel nor dispersion of smoke particles and because 50km is approximately the 80th percentile value of path distance between consecutive trajectory points in our data. To identify smoke days obscured by cloud cover, we use the count of trajectory points within the lowest height quintile (0 to approximately 1.1km above ground level).
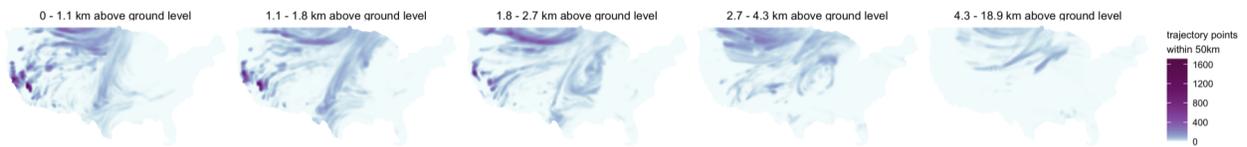
Figure S1: a) For each HYSPLIT point and starting time, we initialize three simulations with injections heights at 500m, 1500m, and 2500m. Colored points show individual trajectory points and are colored by their height above ground level. Example point is taken from September 20, 2020. b) Over all HYSPLIT trajectory points, we calculate the 0th, 20th, 40th, 60th, 80th, and 100th percentiles and use these to construct five height bins. Distribution of trajectory point heights is shown for a random sample of 100,000 points. c) For each height bin, we accumulate the number of trajectory points within 50km of each grid cell centroid to use as features in the smoke $PM_{2.5}$ model. Gridded trajectory counts are shown for September 20, 2020, and may include HYSPLIT trajectories initialized as early as September 15, 2020. State outlines are reproduced with permission from the US Census Bureau TIGER/Line Shapefiles[6]
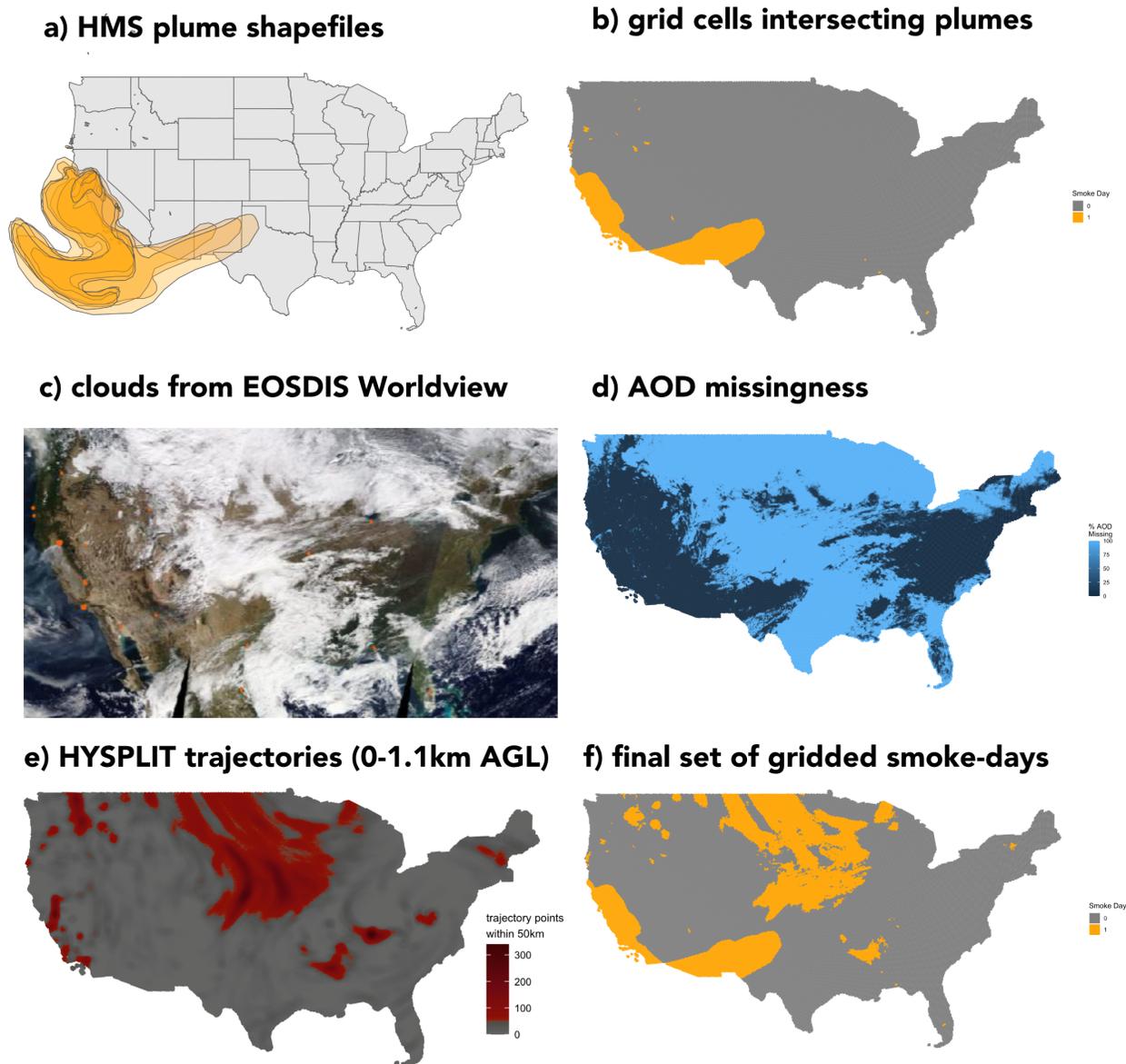
a) HMS plume shapefiles

b) grid cells intersecting plumes

c) clouds from EOSDIS Worldview

d) AOD missingness

e) HYSPLIT trajectories (0-1.1km AGL)

f) final set of gridded smoke-days

Figure S2: a) Shapefiles of HMS smoke plumes[1] are shown in orange over the contiguous US on November 11, 2018. b) From plumes, all grid cells intersecting a plume are identified as smoke days (orange). c) Clouds may obscure identification of smoke plumes, as seen from EOSDIS Worldview[7] on the same date. d) Missingness of MAIAC AOD corresponds to this cloud cover, making it possible to use missingness to indicate when clouds interfere with plume identification. e) HYSPLIT trajectory point counts within a 50km buffer in the lowest height quintile (0 - 1.1 km above ground level (AGL)) on the same date show high density of trajectory points in the Midwest, an area covered by clouds. f) Grid cells with at least 50 HYSPLIT trajectory points and more than 75% missingness of MAIAC AOD are also classified as smoke days, leading to the final set of smoke days on November 11, 2018. State outlines are reproduced with permission from the US Census Bureau TIGER/Line Shapefiles[6].
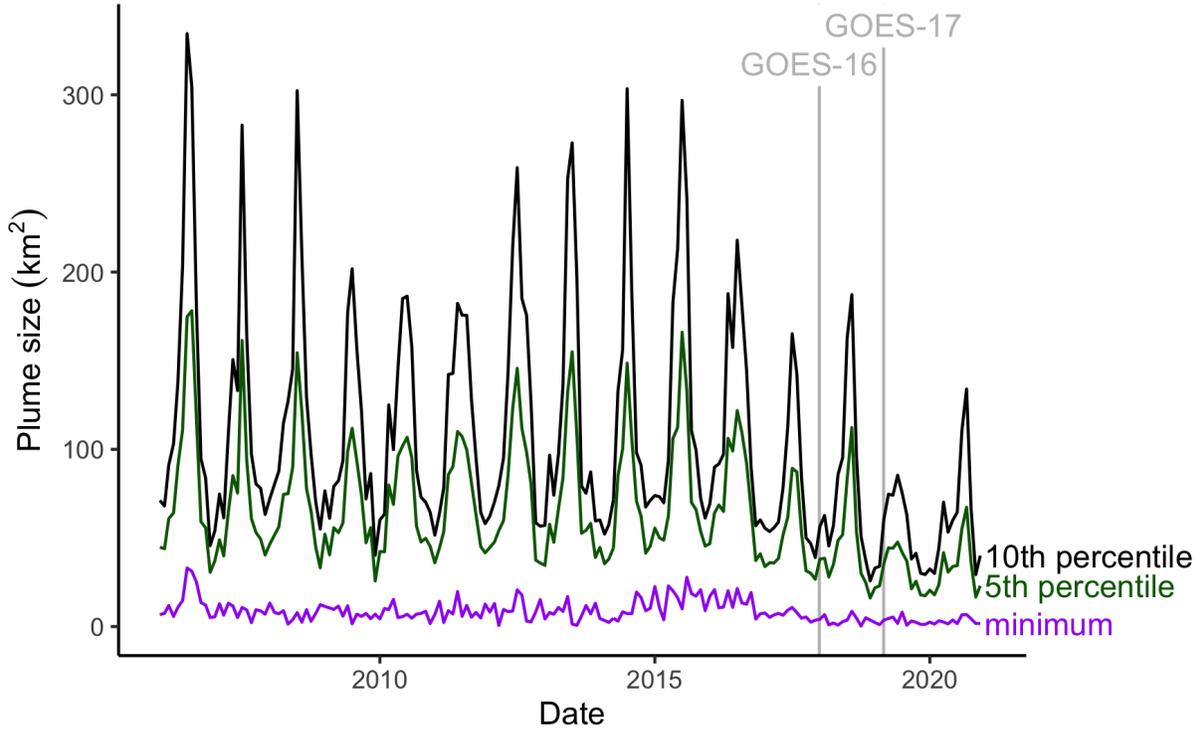
Figure S3: Quantiles of smoke plume size by month of sample. Grey lines indicate start dates of GOES-16 and GOES-17 satellites in the HMS products[1].

Treatment of missing HMS data There are 3 sources of HMS data – smoke plume shapefiles, HYSPLIT points, and fire points. Both smoke plumes and simulated trajectories from HYSPLIT points are used to define smoke days, while simulated trajectories and features from clustered fire points are used as features in the smoke $PM_{2.5}$ model. Below we explain our handling of missing observations for each of these inputs and their respective uses.

HMS does not provide smoke plume shapefiles for 24 dates, likely due to cloud obscuration of smoke plume boundaries. We exclude these dates from calculation of non-smoke day $PM_{2.5}$ medians and model training and validation. At inference time, we use the preceding day and following day, or the 3 days in a 2-day window centered at a given day d that are not missing smoke plume data if either the preceding or following day's smoke plume data are also not available ("temporal nearest neighbors (NN)"), to assign smoke days, defaulting assignment to non-smoke day when any of the 2-3 non-missing days is not a smoke day.

Smoke plumes may sometimes be recorded as invalid geometries. We subset to valid geometries,

and we repair geometries for 9 dates on which all geometries are invalid by connecting the last point in each plume polygon back to the first point in the respective plume polygon.

HMS does not provide HYSPLIT points for 46 dates and does not provide durations or start times before April 19, 2006. Shapefiles are corrupt for 2 dates, and 50 dates are found to be oddly empty, as they do not have any HYSPLIT points recorded despite having at least one smoke plume observed that day (notably including all days in October 2018). We treat the trajectory point counts on these dates, as well as the 6 affected days following these dates, as missing in attribute (MIA), allowing the boosted regression trees to define a default path in trees for when values are missing. When defining smoke days, we do not use dates missing trajectory point counts to attribute smoke over any grid cells.

HMS does not provide fire points for 26 dates. For the 7 dates that are missing fire data but not smoke data, we assume the data to be missing due to lack of fire detection. For these 7 dates, as well as the 11 dates for which all fire clusters were smaller than the threshold for minimum fire cluster size and the 1 date for which data were available but no points were recorded, we treat distance to fire as MIA and set the area and number of fire points to 0. For the 19 dates that are missing both fire and smoke data, we assume the data to be missing due to technical issues. We exclude these 19 dates from model training and validation. We obtain feature values for predicting smoke $PM_{2.5}$ on these dates by averaging the values of distance to fire, area, and number of fire points, respectively, from the temporal nearest neighbors.
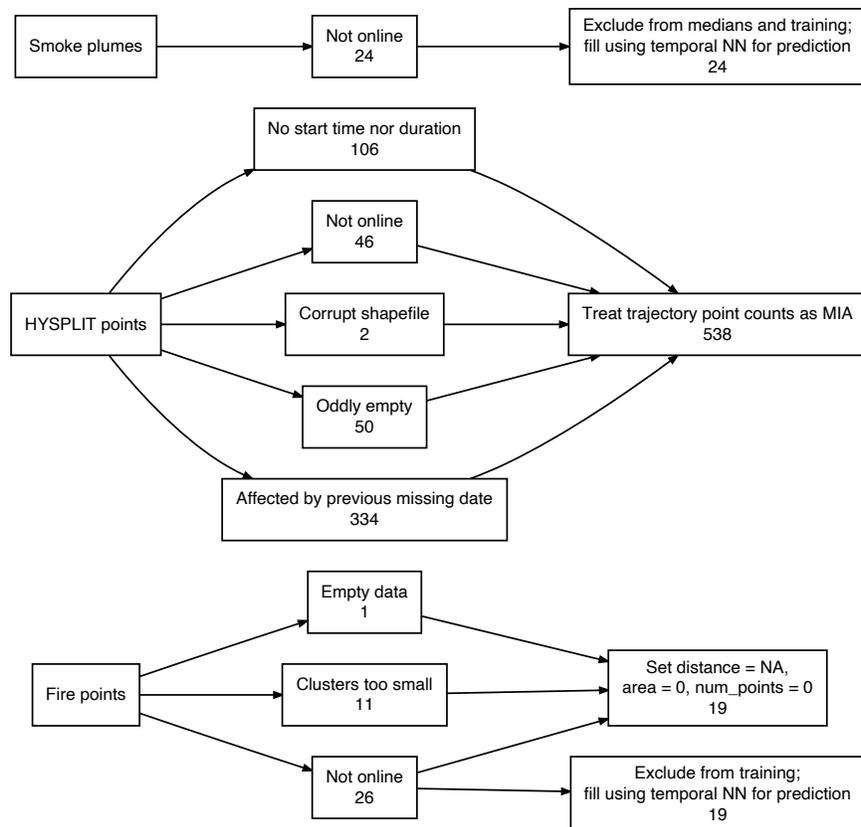
Smoke plumes → Not online 24 → Exclude from medians and training; fill using temporal NN for prediction 24

No start time nor duration 106

Not online 46

HYSPLIT points → Corrupt shapefile 2 → Treat trajectory point counts as MIA 538

Oddly empty 50

Affected by previous missing date 334

Empty data 1

Fire points → Clusters too small 11 → Set distance = NA, area = 0, num_points = 0 19

Not online 26 → Exclude from training; fill using temporal NN for prediction 19

Figure S4: HMS data inputs, causes of missingness, and handling of missing data. Numbers indicate the number of days to which the cause of missingness or procedure for handling missing data applied.

Plume accuracy   As in earlier studies[8,9], we rely substantially on HMS smoke plume data to understand whether wildfire smoke is in the atmosphere over a given location and time period. Thus any errors in the plume locations could have implication for the accuracy of our estimates. As noted in the main text, temporal variation in plumes is highly predictive of ground $PM_{2.5}$ concentrations at a given locations, and to counter omissions when clouds may obscure smoke plumes, we included locations with potential cloud cover and air trajectories from known fires as smoke days (Fig. S2). Nevertheless, our derived smoke days measure may still be a conservative estimate of the locations with air quality impacted by smoke due to undetected plumes under cloud cover, during nighttime when satellite-based plume segmentations are unavailable, or where smoke is dilute and difficult to identify in satellite imagery[4,8].

Consistent with this concern, we find that $PM_{2.5}$ monitors near but not underneath HMS-determined smoke plumes on a given day have elevated $PM_{2.5}$ relative to average after controlling for monitor-month and state-year-month fixed effects (Table S1, plumes), and these effects persist after in-

cluding smoke-days defined from HYSPLIT trajectory points and AOD missingness (Table S1, smoke days). Further controlling for meteorological conditions explains half to three-quarters of these effects outside of smoke days, suggesting that weather conditions conducive to high $PM_{2.5}$ may co-occur with or near smoke (Table S1, plumes w/ met. and smoke days w/ met.). Defining smoke days using the HMS plumes may therefore result in errors of omission, in which case our estimates of smoke $PM_{2.5}$ will be conservative, both because a plume-based approach sets smoke $PM_{2.5}$ to zero when a plume is absent, as well as because the estimated non-smoke median $PM_{2.5}$ would be too high. There alternatively could be errors of inclusion where locations unaffected by smoke are labeled as being under a plume, in which case ground monitor data—or, in the case of model predictions, other model inputs (e.g. AOD or reanalysis data)—should be helpful in identifying low $PM_{2.5}$ anomalies.

Even if plume boundaries are correct or overgenerous, an additional threat to the validity of these inferred smoke $PM_{2.5}$ estimates is if other factors correlated with the presence of smoke drive up $PM_{2.5}$. In this case, we might misattribute the increase in $PM_{2.5}$ to smoke when other factors— e.g., variation in meteorology –are the source of this increase. This is a subtle issue however, as meteorological factors are also certainly a "cause" of the emission and fate of wildfire pollutants – e.g when rainfall puts out fires or hot temperatures expand them – and thus a key part of the overall "effect" of wildfire on surface $PM_{2.5}$; in this setting, controlling for meteorology would be incorrect, since meteorology is what is driving the variation in wildfire activity and $PM_{2.5}$ concentrations.

In the case that plume boundaries are correct, we approximate that this effect is at most 1.14 $\mu g/m^3$, which we estimate by calculating the difference between $PM_{2.5}$ concentrations at monitors nearby (but not under) plumes on a given day when controlling for meteorology versus not; this difference is plausibly an upper bound—and perhaps a substantial overstatement, given the reasons just noted—of the elevation in surface $PM_{2.5}$ due to conditions correlated with smoke but not smoke itself. We note that 1.14 $\mu g/m^3$ is low relative to the very large spikes in $PM_{2.5}$ we often observe when plumes are overhead, suggesting this source of bias is likely small.
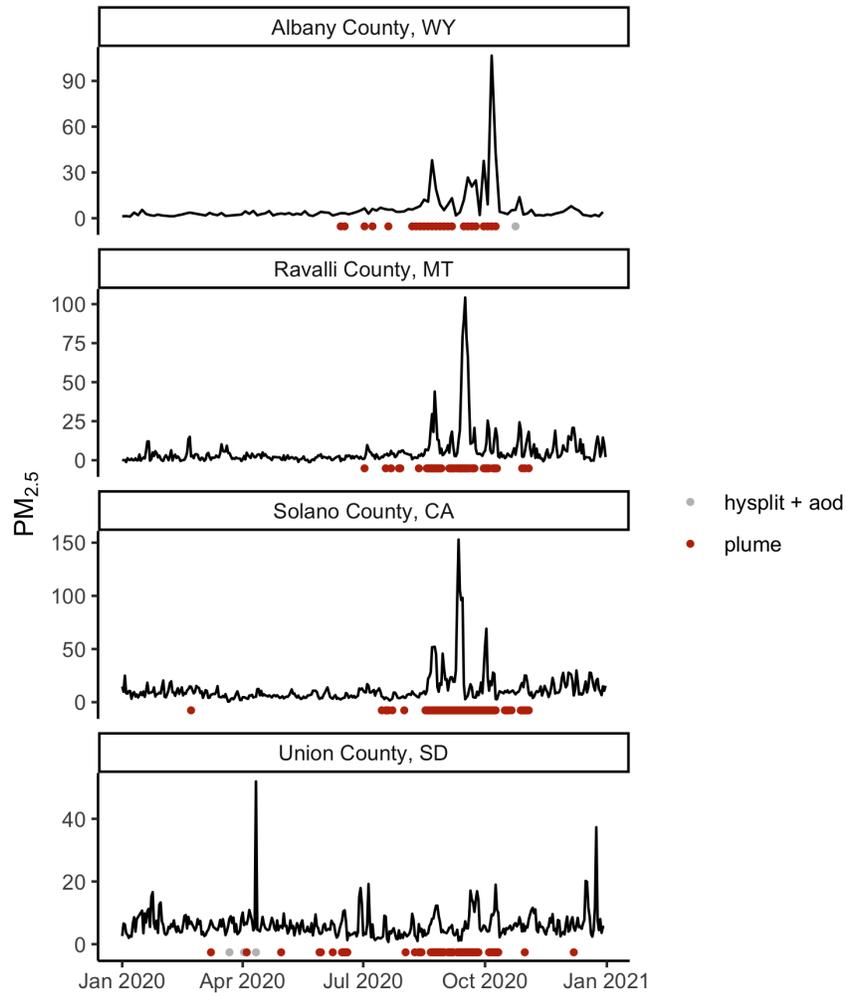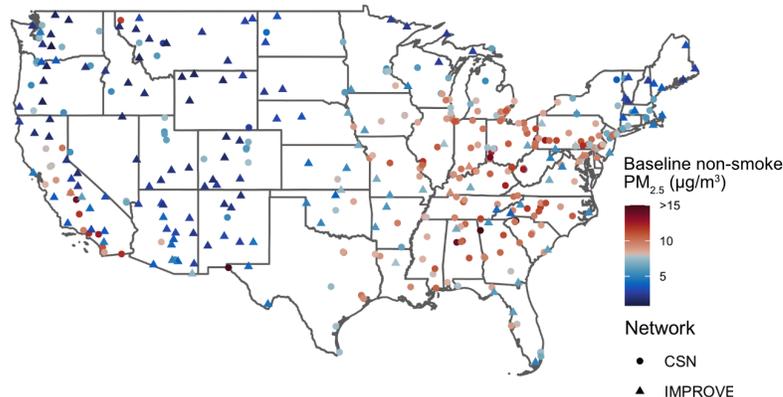
Figure S5: Time series of PM$_{2.5}$ at EPA stations (black lines) in 2020, labeled by county in which the monitor is located. Red point indicate dates with a plume overhead, and grey points indicate days additionally classified as smoke-days based on AOD missingness and HYSPLIT trajectories. Spikes in PM$_{2.5}$ commonly co-occur on days with plumes overhead, with most exceptions around holidays (e.g. early July 2020 and late December 2020 in Union County, SD).

Table S1: Estimated effect on monitor $PM_{2.5}$ of being underneath or near a plume or smoke day. The third and fourth columns (labeled w/ met.) control flexibly for meteorological variables including temperature, dewpoint temperature, planetary boundary layer minimum, maximum and mean height. Plume variables are based solely on the presence of HMS plumes, while smoke variables also include days identified as smoke-days based on AOD missingness and HYSPLIT trajectories. All regressions include monitor-month and state-month-year fixed effects.

	plumes	smoke days	plumes w/ met.	smoke days w/ met.
plume dist: overhead	4.504***		3.099***	
	(0.0510)		(0.0410)	
plume dist: 0-100	1.272***		0.3164***	
	(0.0203)		(0.0226)	
plume dist: 100-250	1.012***		0.2041***	
	(0.0150)		(0.0139)	
plume dist: 250-500	0.7020***		0.1632***	
	(0.0127)		(0.0104)	
plume dist: 500-750	0.3203***		0.1121***	
	(0.0117)		(0.0094)	
plume dist: 750-1000	0.0608***		0.0619***	
	(0.0105)		(0.0091)	
smoke dist: overhead		4.934***		3.399***
		(0.0482)		(0.0385)
smoke dist: 0-100		1.947***		0.8033***
		(0.0202)		(0.0211)
smoke dist: 100-250		1.449***		0.5029***
		(0.0160)		(0.0141)
smoke dist: 250-500		0.9806***		0.3568***
		(0.0140)		(0.0113)
smoke dist: 500-750		0.4757***		0.2037***
		(0.0127)		(0.0103)
smoke dist: 750-1000		0.1570***		0.1093***
		(0.0113)		(0.0100)
Fixed-effects				
id×month	Yes	Yes	Yes	Yes
state×year×month	Yes	Yes	Yes	Yes
Observations	4,442,067	4,442,067	4,442,067	4,442,067
Within R^2	0.021599	0.02626	0.125643	0.127239

One-way (id×month) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

a) IMPROVE and CSN monitors
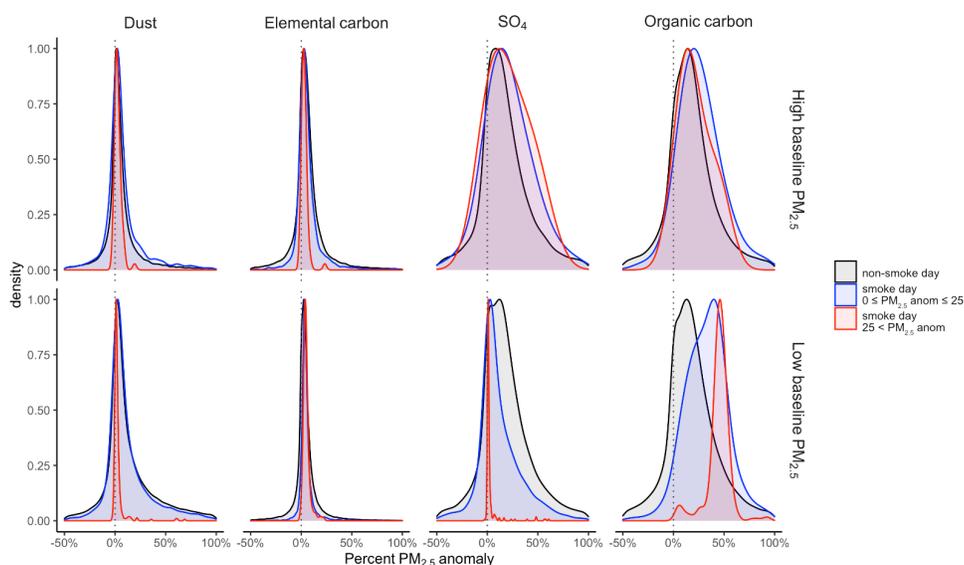

b) speciated comparison

Figure S6: Speciated data from IMPROVE[10] and CSN[11] monitors. a) Locations of monitors in to conterminous US with at least one smoke day, colored by median $PM_{2.5}$ on non-smoke days, with red stations indicating locations classified as high baseline and blue stations classified as low baseline. Shapes indicate whether the monitor is in the IMPROVE or CSN network. b) We apply our approach (Eq. 1 - 2) to calculate smoke anomalies for total $PM_{2.5}$, dust, elemental carbon, SO_4, and organic carbon. We then calculate the percent of the total $PM_{2.5}$ anomaly that is accounted for by each species anomaly. Distributions show the percent of $PM_{2.5}$ anomaly (horizontal axis) for non-smoke days (black), low smoke days with $0 \le PM_{2.5}$ anomalies ≤ 25 (blue), and high smoke days with $PM_{2.5}$ anomalies > 25 (red), for high and low baseline $PM_{2.5}$ locations (top and bottom rows, respectively). High and low baseline $PM_{2.5}$ locations correspond to the classification in a). Distributions are truncated to between -50% and 100% for visibility. IMPROVE and CSN data are harmonized based on existing recommendations[12–14]. State outlines are reproduced with permission from the US Census Bureau TIGER/Line Shapefiles[6].

ERA5 meteorology extraction   Meteorology variables are retrieved from ERA5 and ERA5-Land datasets (see Table S3 for the source of each meteorological variable used). ERA5 is a gridded global climate reanalysis product available at 0.25-degree resolution, and ERA5-Land is developed from replaying the ERA5 land component and available at 0.1-degree resolution[15,16]. Measurements are provided natively at hourly time scale and aggregated to UTC-06:00 daily time scale (except in the case of total precipitation, for which we use UTC+00:00 due to the cumulative way in which this variable is recorded) using the "Daily statistics calculated from ERA5 data" application (`https://cds.climate.copernicus.eu/cdsapp#!/software/app-c3s-daily-era5-statistics?tab=overview`) via the Copernicus Climate Data Store (`https://cds.climate.copernicus.eu/#!/home`) API.

We obtain ERA5 and ERA5-Land variables over our 10 km grid by matching each 10 km grid centroid to its overlapping ERA5 grid cell and its overlapping ERA5-Land grid cell. In the case of coastal 10 km grid cells located beyond the spatial coverage of ERA5-Land, we instead match to the nearest ERA5-Land grid cell within 1 degree.

Fire clusters   We created fire clusters from fire points to identify large fires likely to produce large quantities of smoke. We generate fire clusters from fire points (see "HYSPLIT simulations above for more detail on fire points) by creating a 5.8 km-by-5.8 km square buffer around each fire point on the collapsed cross-section of a given day d plus the three previous days and merging intersecting buffers. We exclude fire clusters that are not at least three times the minimum fire cluster size. We calculate distance to fire as the distance from the centroid of each grid cell on day d to the centroid of the nearest fire cluster on day d or 0 if the nearest fire cluster intersects the grid cell. Grid cells are also assigned the area and the number of fire points composing the matching fire cluster.

MERRA-2 AOT extraction   The AOT data is retrieved from the MERRA-2 data portal (`https://disc.gsfc.nasa.gov/datasets/M2T1NXAER_5.12.4/summary`) and is subset to the "total aerosol extinction aot [550 nm]" variable for the continental US from 2006-2020. The reanalysis data is natively provided hourly starting at UTC-00:30 at approximately 50km resolution (0.5° latitude x 0.625° longitude)[17]. We subset and aggregate the data using the Goddard Earth Sciences Data and Information Services Center (GES DISC) subsetter tool to obtain the gridded daily average AOT.

Predicted AOD anomalies Satellite-based MAIAC AOD (0.47 μm) has a much higher resolution (1 km) but has a high rate of missing observations (e.g., 41% missing over California[18]) due to a variety of factors including cloudiness and surface reflectivity[19,20]. Recent work on total PM$_{2.5}$ predictions has also found higher AOD levels but lower PM$_{2.5}$ on days with missing AOD observations[21], so we calculate the percent of MAIAC AOD observations that are missing in each 10km grid cell as an additional model input. We use Google Earth Engine to access MAIAC AOD data and perform calculations[22,23].

Consistent with previous work on total PM$_{2.5}$ predictions[21,24], preliminary tests on a subset of our data suggested that training an initial model on 1km MODIS MAIAC AOD and then using predicted AOD as an input to predict smoke PM$_{2.5}$ improved performance. For the first stage model, we used grid cells consistent with 1km MAIAC tiles that cover our 10km grid over the contiguous US. While training at locations consistent with the ground monitors should result in the best performance for the model in-sample, we instead optimize for out-of-sample performance in the highest population areas by sampling 5,000 grid cells based on WorldPop mean population estimates from 2013 (the midpoint of our study period), with WorldPop data accessed from Google Earth Engine[22,23,25]. We then selected a population-weighted sub-sample of 20% of those grid cells, ensuring that all selected grid cells were at least 5 km away from each other (Fig. S7). For model labels, we used AOD anomalies from non-smoke medians (as with AOT anomalies and PM$_{2.5}$ anomalies, Eq. 1) after extracting MAIAC AOD (0.47 μm) observations using the Google Earth Engine platform[19,22,23]. Model inputs included AOT anomalies on smoke-days, meteorology, fire variables, elevation, and land cover (Table S2). We use values from the 10km grid cell covering the centroid of the MAIAC 1km grid cell for time-varying features and smoke-day assignment and values extracted over the 1 km grid cells for cross-sectional features (elevation and land cover). Models were trained on the set of smoke days with non-missing AOD anomalies at 1,000 training locations (260,857 observations). We fit gradient boosted trees[26], which have performed well in similar tasks previously (e.g., Chen et al.[21]), using Bayesian optimization and 4-fold spatial cross-validation for hyperparameter tuning.

For hyperparameter tuning, we use Bayesian optimization with 24 randomly chosen initial points and 24 subsequent iterations, and searched over the following xgboost parameter ranges: learning rate from 0.0001 to 0.2, minimum split loss from 0 to 100, maximum tree depth from 2 to 50, fraction of columns to sample for each tree from 0.5 to 1, fraction of observations to subsample on each boosting iteration from 0.25 to 1, and minimum child weight from 1 to 50. We first performed hyperparameter tuning using internal 4-fold cross-validation to identify the optimal number of trees for each parameter set, we train a model on all observations and predict AOD anomalies for all 1km smoke-day grid cells. The predictions explain 79.3% of out-of-sample variation

in AOD anomalies, and MERRA-2 AOT anomalies are the most important features (Fig. S8). We then predicted for all 1km grid cells over the contiguous US, and aggregated from 1km to 10km grid cells by taking all 1km grid cells whose centroids fell within a 10km grid cell and calculating the mean, minimum, maximum, and 25th, 50th, and 75th percentiles of predicted AOD anomalies.

Table S2: AOD anomaly model inputs. NED = National Elevation Database, NLCD = National Land Cover Database, PBL = planetary boundary layer.

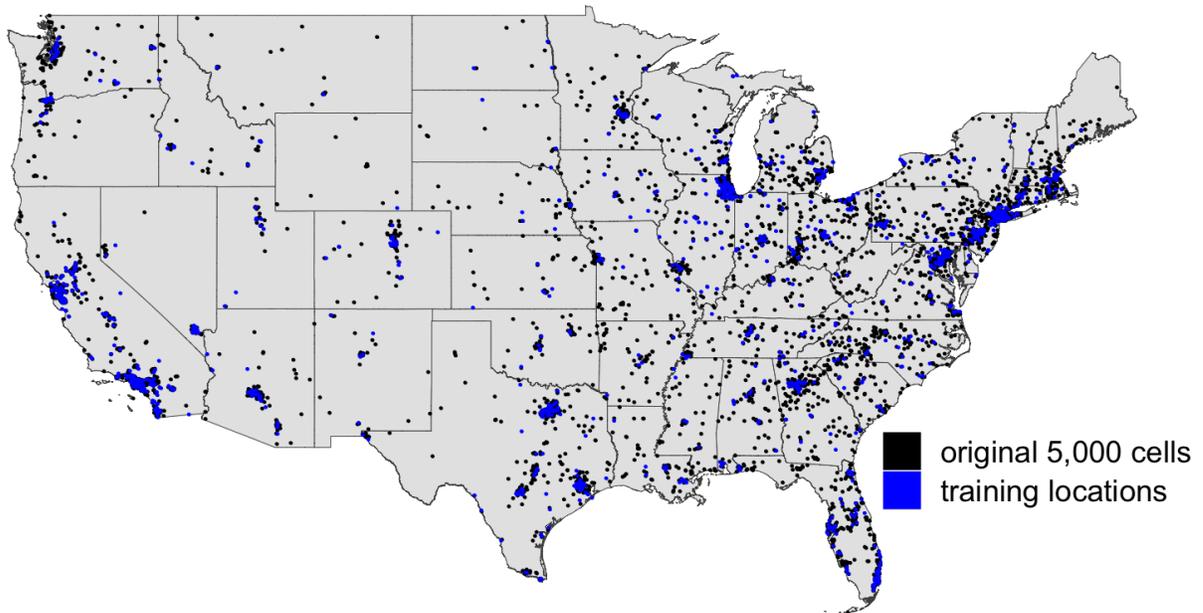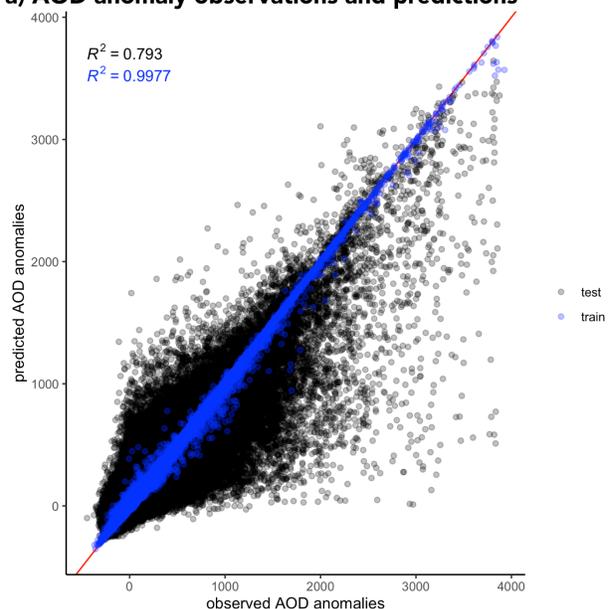| Feature	Source	Native resolution
Aerosol optical thickness anomalies on smoke-days (current, 1-day, 2-day, and 3-day lagged)	MERRA-2	50km
Elevation (mean and standard deviation in grid cells)	USGS NED	10m
Percent of area in each Level 1 land cover class (water, developed, barren, shrubland, herbaceous, cultivated, forest, wetlands)	USGS NLCD	30m
Distance to nearest fire cluster	HMS fire points	-
Size of nearest fire cluster (area and number of constituent fire points)	HMS fire points	-
Meteorology (daily mean, max, and min PBL, average sea level pressure)	ERA5 global	30km
Meteorology (total precipitation, average 2m air temperature, average eastward and northward wind speed, average surface pressure, 2m dewpoint temperature)	ERA5 land	11km
Latitude, Longitude, Month		-

Figure S7: Locations for AOD model training. Black cells are the original set of 5,000 grid cells, while blue cells are those ultimately used in model training that are a population-weighted sub-sample of the original set, at least 5km away from each other. The original set of cells not used in model training are used as a test set to check model performance. State outlines are reproduced with permission from the US Census Bureau TIGER/Line Shapefiles[6].

a) AOD anomaly observations and predictions

$R^2 = 0.793$
$R^2 = 0.9977$

predicted AOD anomalies

observed AOD anomalies

test
train

b) AOD anomaly model feature importance

AOT anomaly (current)
AOT anomaly (1-day lag)
Area of closest fire cluster
Latitude
Points in closest fire cluster
Sea level pressure
Longitude
Distance to closest fire cluster
AOT anomaly (2-day lag)
Wind speed (eastward)
Dewpoint temperature
Total precipitation
Temperature
Wind speed (westward)
Planetary Boundary Layer (mean)

Gain

aerosols
cross sectional
fire
meteorology

Figure S8: a) Model predicted AOD anomalies (y-axis) and observed AOD anomalies (x-axis) at training (blue) and test (black) locations. Test and training R^2 are labeled on the figure. b) Feature importance for the top 15 features (y-axis) from the final model, as measured by gain (x-axis) and colored by broad category of feature type.

Table S3: Smoke PM$_{2.5}$ Model inputs. NED = National Elevation Database, NLCD = National Land Cover Database

Feature	Source	Native resolution
Aerosol optical thickness anomalies (current, 1-day, 2-day, and 3-day lagged)	MERRA-2	50km
Percent of AOD observations missing	MODIS MAIAC	1km
Predicted aerosol optical depth anomalies (min, max, mean, 25th, 50th, and 75th percentiles)	predicted, trained on MODIS MAIAC	-
Elevation (mean and standard deviation in grid cells)	USGS NED	10m
Percent of area in each Level 1 land cover class (water, developed, barren, shrubland, herbaceous, cultivated, forest, wetlands)	USGS NLCD	30m
Distance to nearest fire cluster	HMS fire points	-
Size of nearest fire cluster (area and number of constituent fire points)	HMS fire points	-
Meteorology (daily mean, max, and min PBL, average sea level pressure)	ERA5 global	30km
Meteorology (total precipitation, average 2m air temperature, average eastward and northward wind speed, average surface pressure, 2m dewpoint temperature)	ERA5 land	11km
HYSPLIT trajectory points in 50 km buffer (by height quintiles: 0 - 1.1, 1.1-1.8, 1.8 - 2.7, 2.7 - 4.3, 4.3+ km AGL)	HYSPLIT simulations from HMS HYSPLIT points	-
Latitude, Longitude, Month		-

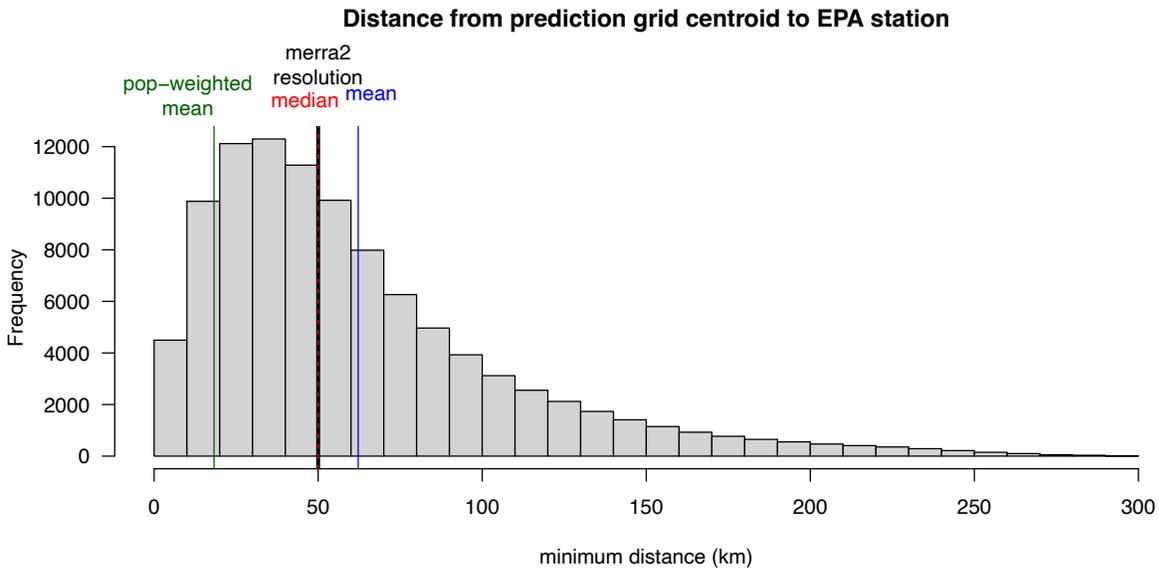**Distance from prediction grid centroid to EPA station**

Figure S9: Minimum distance from prediction grid cell centroid to EPA stations. We use average population density in each 10km prediction grid cell from WorldPop[25] and calculate distances between prediction grid cell centroids and all available EPA monitors in the study region and time period.

Cross-sectional land use and elevation   We use the U.S. Geological Survey (USGS) 1/3 Arc-Second National Elevation Dataset (NED) and calculate the mean and standard deviation of elevation within each grid cell[27]. For land cover, we use the USGS National Land Cover Dataset (NLCD) and calculate percentages of land area within each grid cell in land use Level 1 classes[28]. Datasets were accessed from and extractions were performed using the Google Earth Engine platform[22,23].

Smoke model tuning   For hyperparameter tuning, we use Bayesian optimization with 24 randomly chosen initial points and 16 iterations and searched over the following hyperparameter ranges: learning rate from 0.001 to 0.1, minimum split loss from 0 to 50, maximum tree depth from 2 to 25, fraction of columns to sample for each tree from 0.5 to 1, fraction of observations to subsample on each boosting iteration from 0.25 to 1, and minimum child weight from 1 to 50. For the spatial, nested cross-validation, for each of the five folds defined based on MERRA-2 grid cells, (1) we left the fold out for testing, (2) performed hyperparameter tuning on the remaining 4 folds using internal 4-fold cross-validation to identify the optimal number of trees for each parameter set, (3) trained a model using the four training folds of data with the identified hyperparameters and boosting iterations, and (4) predicted for the out-of-sample test fold. Repeating this

for all five folds resulted in out-of-sample predictions for all observations of the dataset, which we used to quantify model performance. While nested cross-validation provides a robust measure of out-of-sample model performance, for the final model used to predict over the entire 10km grid, we perform a similar procedure as steps (2 - 3) above, but using 5-fold cross validation and ultimately training the model over all 5 folds of data.

We also repeat the nested cross-validation for models without AOD anomaly predictions and without HYSPLIT trajectory features to get an out-of-sample quantification of the model's performance with and without those features.

Table S4: Out-of-sample smoke $PM_{2.5}$ model performance on different subset or aggregations of the data (rows) and different performance metrics (columns). Within R^2 controls for monitor and year fixed effects.

	within R^2	R^2	RMSE
smoke days	0.65	0.67	9.57
day $<50\mu g/m^3$	0.47	0.50	5.07
day $\geq 50\mu g/m^3$	0.36	0.44	70.99
month avg	0.67	0.67	3.75
year avg	0.68	0.72	2.89

Figure S10: Comparison of smoke $PM_{2.5}$ model performance without AOD anomaly predictions and HYSPLIT trajectory counts. a) Out-of-sample R^2 for all smoke days, days with observed smoke $PM_{2.5} < 50 \ \mu g/m^3$, and days with observed smoke $PM_{2.5} \geq 50 \ \mu g/m^3$ for the full model (red points), the model without HYSPLIT (orange points), and the model without AOD anomaly predictions (green points). b) Same as a) but for within R^2, which is measured as the predictive performance after controlling for monitor and year fixed effects. c) Same as a) but for root mean squared error (RMSE). Across all metrics, the full model performs best, followed by the model without HYSPLIT trajectories, and the model without AOD anomaly predictions performing worst, with the largest gap between best and worst models on days with extreme smoke (smoke $PM_{2.5} \geq 50 \ \mu g/m^3$).
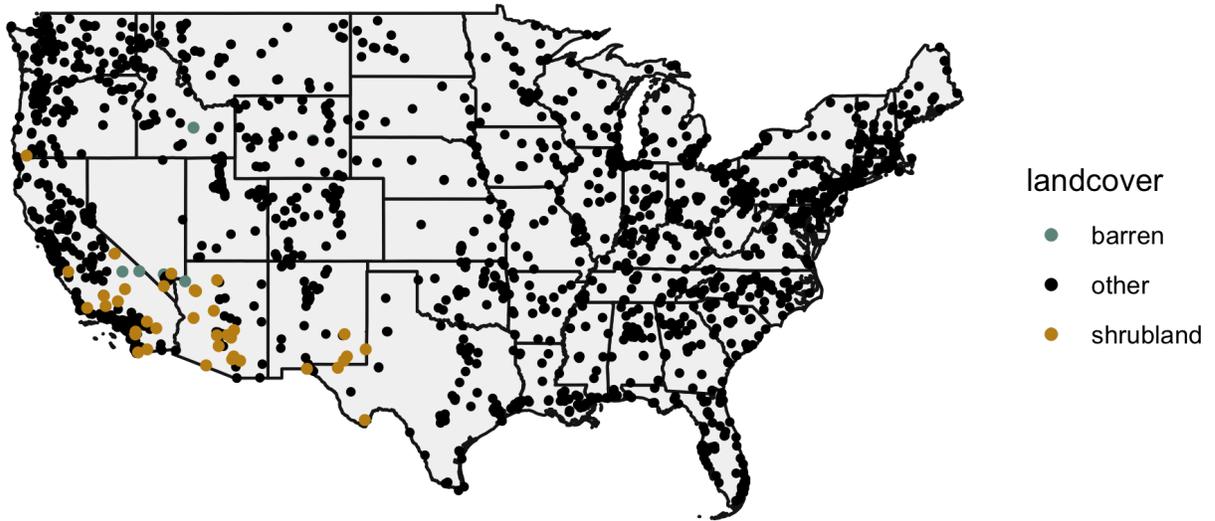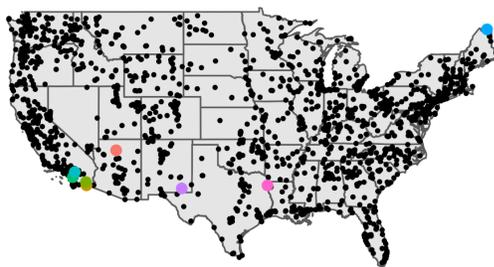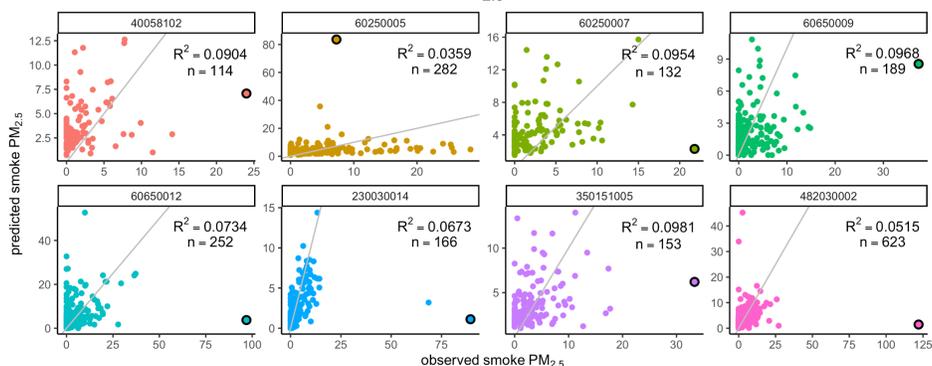
Figure S11: Landcover classification from MODIS Yearly Global Landcover Types[29], extracted at monitor locations from 2013 values. We used Annual Plant Functional Types classification, and grouped together all land covers (black) that are not barren (grey-blue) or shrubland (brown). State outlines are reproduced with permission from the US Census Bureau TIGER/Line Shape-files[6].

Figure S12: **a)** All EPA monitors are shown over the contiguous US, with 8 stations with $R^2 <$ 0.1 and over 100 smoke-day observations highlighted in colors. **b)** For each of the poor performing monitors, we plot observed (x-axis) and predicted (y-axis) smoke $PM_{2.5}$, with colors corresponding to the points on the map. These stations are often characterized by a small number of extreme outliers with high observed smoke $PM_{2.5}$ and under-predictions from the model. The grey line indicates the 1-1 line, and panel annotations state the monitor R^2 and the number of smoke-day observations. For each panel, the point in a black circle is the one with the greatest error between observed and predicted. **c)** For the observation with the worst performance in each station, we show the time series of $PM_{2.5}$ during the time window surrounding the observation. Black lines show observed monitor $PM_{2.5}$, red lines are inferred smoke $PM_{2.5}$, grey points at the bottom show days with smoke overhead, and blue points are the worst observation for each station. For many of these poorly predicted observations, an extreme spike in $PM_{2.5}$ occurred with no preceding or following days of high $PM_{2.5}$. State outlines are reproduced with permission from the US Census Bureau TIGER/Line Shapefiles[6].
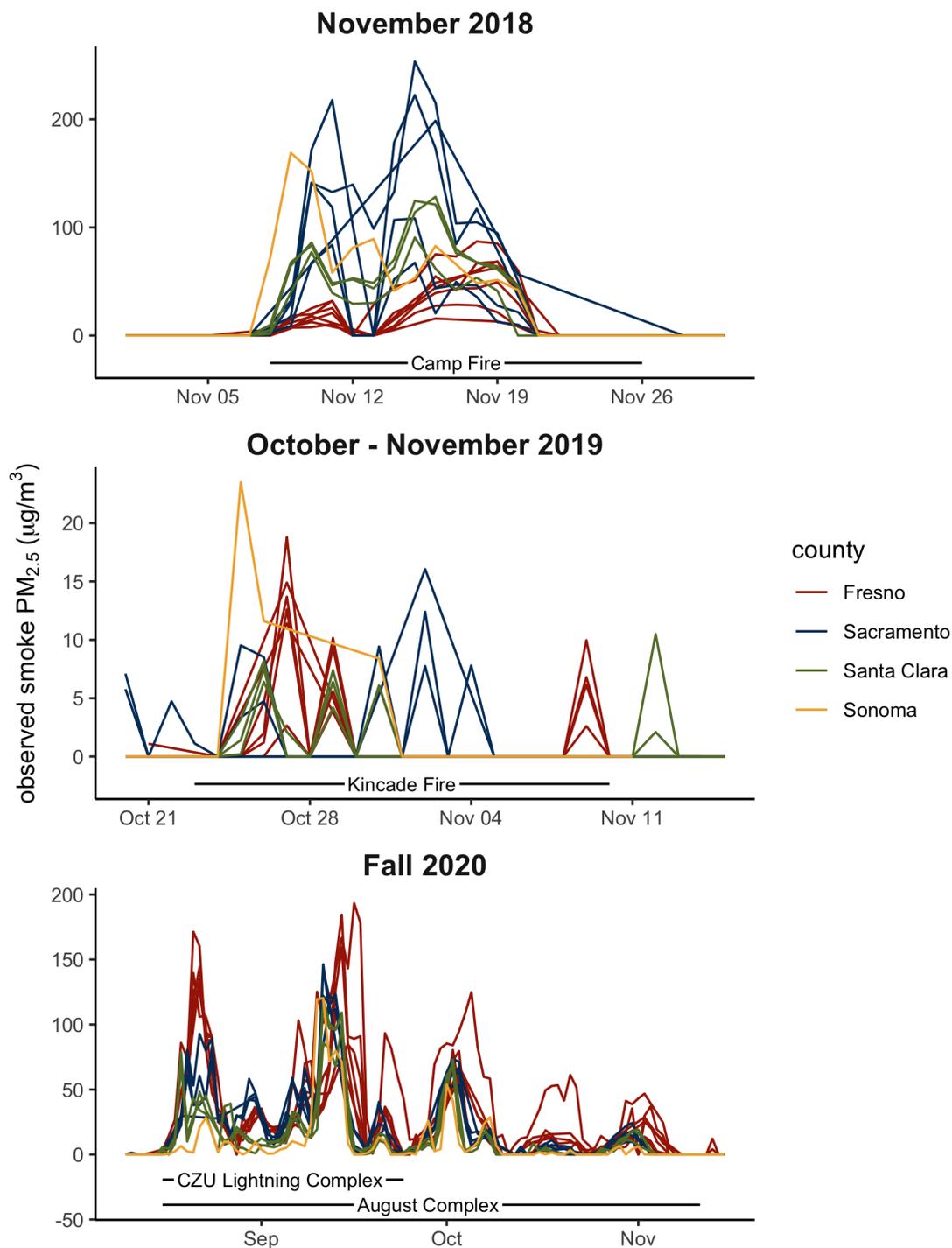
Figure S13: Observed smoke PM$_{2.5}$ during recent fires shows both spatial and temporal variation in concentrations. For select fires, time range are shown along the bottom axis (black lines and labels). Each line is a monitor, with lines colored by county. Colors, counties, and time periods correspond to those shown in Fig. 3. Straight line segments connect consecutive observations, so segments spanning multiple dates indicate no monitor observations on the intervening days.
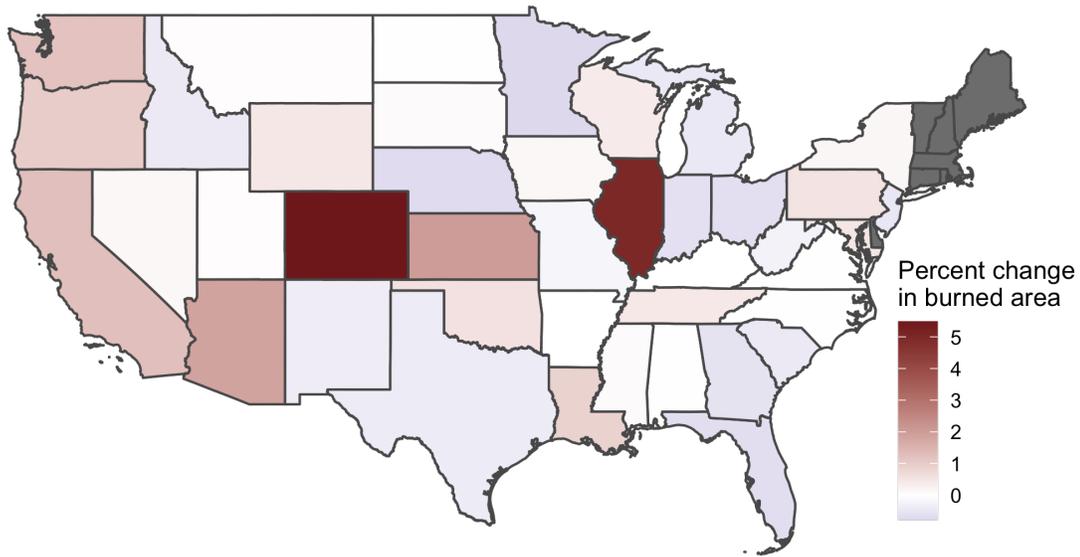
Figure S14: Decadal trend in burned area by state from 2006-2010 to 2016-2020. Annual burned area for each state was calculated as the area of intersection between burn perimeters with an ignition date in a given year[30]. As when calculating decadal trends in smoke $PM_{2.5}$, we then calculated average annual burned area during the 2006-2010 and 2016-2020 periods, then calculated the percent change from the early period to the late period to estimate the decadal percent change in burned area. States in grey intersected with no fire perimeters in at least one of the five year periods, so no trends were calculated. State outlines are reproduced with permission from the US Census Bureau TIGER/Line Shapefiles[6].
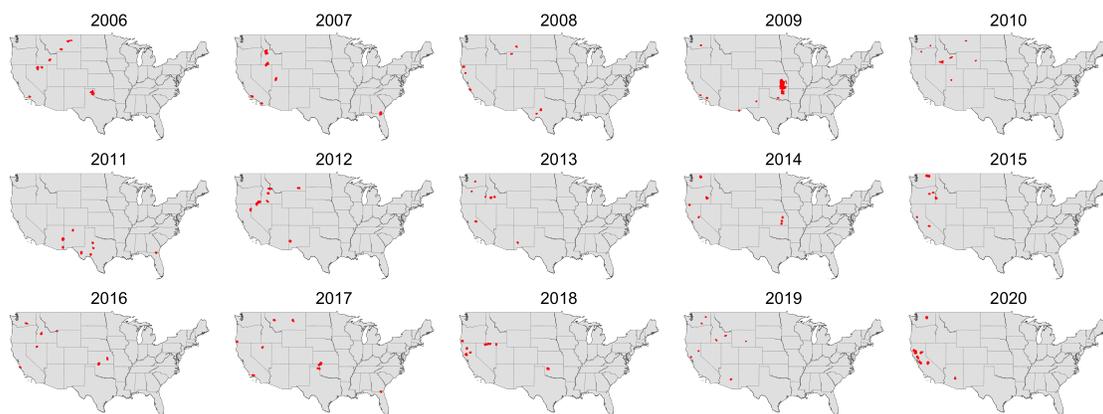

Figure S15: Eight largest fires by area for each year. Fires perimeters are reproduced with permission from the MTBS database of burned area boundaries[30]. State outlines are reproduced with permission from the US Census Bureau TIGER/Line Shapefiles[6].

Figure S16: Three largest fires by area for each state. Fires perimeters are reproduced with permission from the MTBS database of burned area boundaries[30]. State outlines are reproduced with permission from the US Census Bureau TIGER/Line Shapefiles[6].
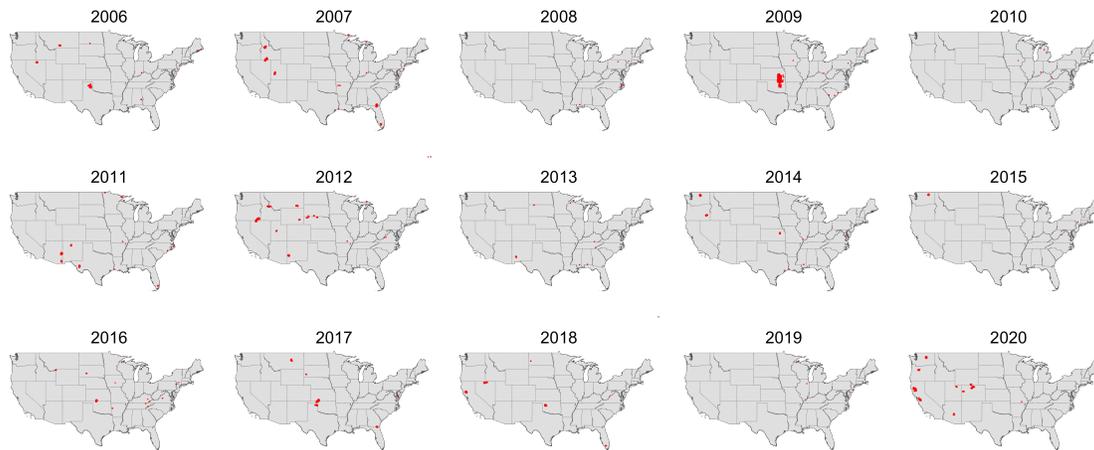

Figure S16: Three largest fires by area for each state. Fires perimeters are reproduced with permission from the MTBS database of burned area boundaries[30]. State outlines are reproduced with permission from the US Census Bureau TIGER/Line Shapefiles[6].
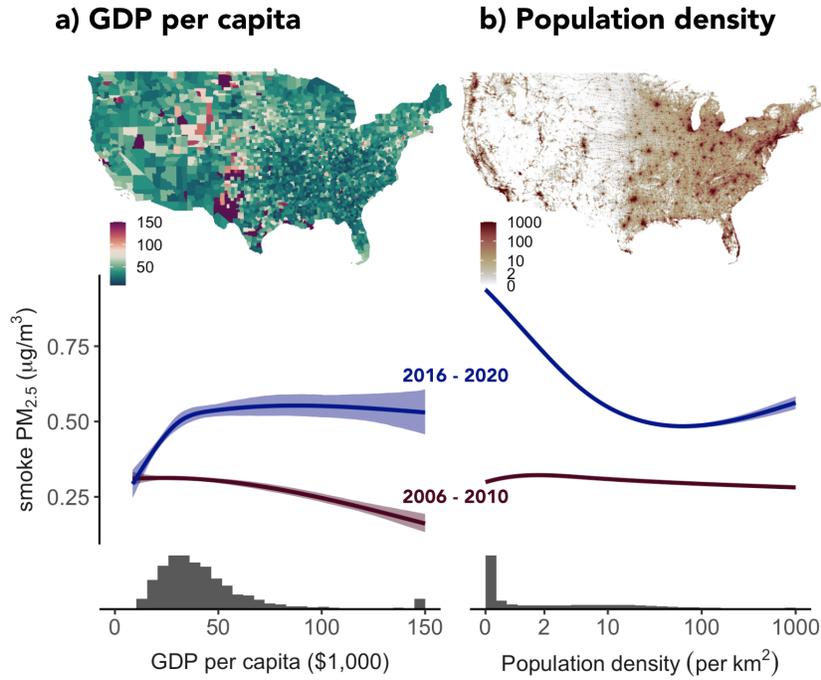
Figure S17: Comparison of smoke $PM_{2.5}$ to additional covariates. Gross domestic product (GDP) is in chained 2012 dollars from the Bureau of Economic Analysis[31], and divided by county ACS population estimates[32] to calculate GDP per capita using data from 2009 and 2019 for the 2006-2010 and 2016-2020 periods, respectively. We use average population density from WorldPop[25] from 2013, the midpoint year of our study, for each grid cell. Map and histogram show the distribution of real GDP per capita in 2019 and population density per km^2 in 2013. County outlines are reproduced with permission from the US Census Bureau TIGER/Line Shapefiles[6].
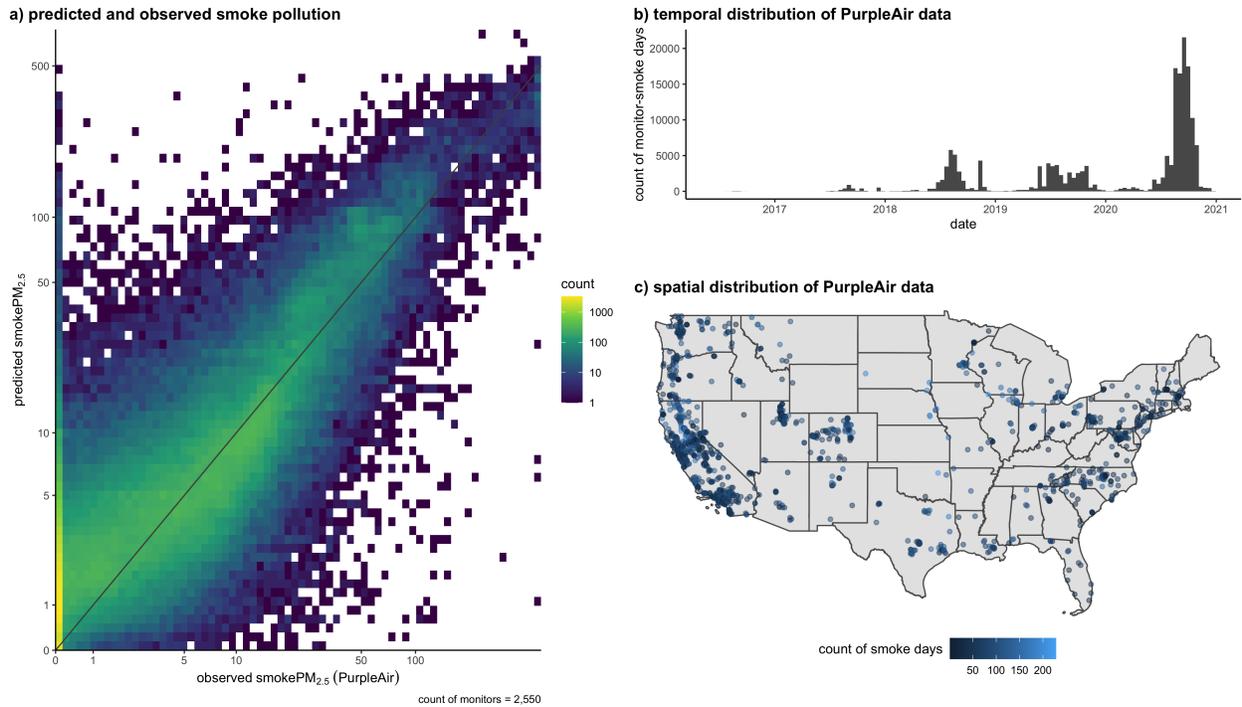
Figure S18: Comparison of smoke PM$_{2.5}$ predictions with smoke PM$_{2.5}$ observations at PurpleAir monitors. Raw PurpleAir PM$_{2.5}$ data are processed to hourly resolution as outlined in Burke et al.[33], calibrated according to the correction method developed in Barkjohn et al.[34], bottom-coded at 0 μg/m^3 , top-coded at 500 μg/m^3 for values between 500 and 1000 μg/m^3 , and discarded for values above 1000 μg/m^3 , prior to aggregation to the daily level and transformation into smoke PM$_{2.5}$. **a)** Count of PurpleAir monitor-smoke days and overlapping 10km grid cells on the same day within bins of smoke PM$_{2.5}$. Axes are pseudo-log transformed and color scale is log transformed. Black line indicates the 1-1 line. **b)** Distribution of PurpleAir monitor-smoke days over time. Bin width is approximately 2 weeks. **c)** Count of smoke days at each PurpleAir monitor. State outlines are reproduced with permission from the US Census Bureau TIGER/Line Shapefiles[6].

References

[1] National Oceanic and Atmospheric Administration. Hazard mapping system fire and smoke product. Accessed from `https://www.ospo.noaa.gov/Products/land/hms.html#about` on November 22, 2021.

[2] AF Stein, Roland R Draxler, Glenn D Rolph, Barbara JB Stunder, MD Cohen, and Fong Ngan. Noaa's hysplit atmospheric transport and dispersion modeling system. *Bulletin of the American Meteorological Society*, 96(12):2059–2077, 2015.

[3] Glenn D Rolph, Roland R Draxler, Ariel F Stein, Albion Taylor, Mark G Ruminski, Shobha Kondragunta, Jian Zeng, Ho-Chun Huang, Geoffrey Manikin, Jeffery T McQueen, and Paula M Davidson. Description and verification of the noaa smoke forecasting system: the 2007 fire season. *Weather and Forecasting*, 24(2):361–378, 2009.

[4] Steven J. Brey, Mark Ruminski, Samuel A. Atwood, and Emily V. Fischer. Connecting smoke plumes to sources using hazard mapping system (hms) smoke and fire location data over north america. *Atmospheric Chemistry and Physics*, 18(3):1745–1761, Feb 2018. ISSN 1680-7324. doi: 10.5194/acp-18-1745-2018.

[5] National Centers for Environmental Prediction. Global data assimilation system. Accessed from the NOAA ARL server `ftp://ftp.arl.noaa.gov/pub/archives/gdas1` on November 25, 2021.

[6] US Census Bureau. *2019 TIGER/Line Shapefiles*, 2019. Accessed using the R package *tigris*. Shapefiles are also available from `https://www.census.gov/cgi-bin/geo/shapefiles/index.php`.

[7] Earth Observing System Data and Information System. Nasa worldview snapshots. Accessed from `https://go.nasa.gov/3k1pI6N` on January 5, 2022.

[8] Katelyn O'Dell, Bonne Ford, Emily V Fischer, and Jeffrey R Pierce. Contribution of wildland-fire smoke to us pm2. 5 and its influence on recent trends. *Environmental science & technology*, 53(4):1797–1804, 2019.

[9] Marshall Burke, Anne Driscoll, Sam Heft-Neal, Jiani Xue, Jennifer Burney, and Michael Wara. The changing risk and burden of wildfire in the united states. *Proceedings of the National Academy of Sciences*, 118(2), 2021.

[10] US Environmental Protection Agency. Interagency monitoring of protected visual environments (improve). . Accessed from `http://vista.cira.colostate.edu/Improve/` in December 10, 2020.

[11] US Environmental Protection Agency. Chemical speciation network (csn). . `https://www.epa.gov/amtic/chemical-speciation-network-csn` Accessed in March 11, 2021.

[12] Katrine A Gorham, Sean M Raffuse, Nicole P Hyslop, and Warren H White. Comparison of recent speciated pm2.5 data from collocated csn and improve measurements. *Atmospheric Environment*, 244:117977, 2021.

[13] JL Hand, BA Schichtel, WC Malm, and NH Frank. Spatial and temporal trends in pm2. 5 organic and elemental carbon across the united states. *Advances in Meteorology*, 2013, 2013.

[14] Jenny L Hand, Anthony J Prenni, Bret A Schichtel, William Conrad Malm, and Judith C Chow. Trends in remote pm2. 5 residual mass across the united states: Implications for aerosol mass reconstruction in the improve network. *Atmospheric environment*, 203:141–152, 2019.

[15] H Hersbach, B Bell, P Berrisford, G Biavati, A Horányi, J Muñoz Sabater, J Nicolas, C Peubey, R Radu, I Rozum, D Schepers, A Simmons, C Soci, D Dee, and J-N Thépaut. Era5 hourly data on pressure levels from 1979 to present. *Copernicus climate change service (c3s) climate data store (cds)*, 10, 2018.

[16] J Muñoz Sabater. Era5-land hourly data from 1981 to present, copernicus climate change service (c3s) climate data store (cds)[data set], 2019.

[17] Global Modeling and Assimilation Office (GMAO). MERRA-2 tavg1_2d_aer_Nx: 2d,1-Hourly, Time-averaged, Single-Level, Assimilation, Aerosol Diagnostics V5.12.4. 2015. Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC). `https://doi.org/10.5067/KLICLTZ8EM9D`.

[18] Lianfa Li, Meredith Franklin, Mariam Girguis, Frederick Lurmann, Jun Wu, Nathan Pavlovic, Carrie Breton, Frank Gilliland, and Rima Habre. Spatiotemporal imputation of maiac aod using deep learning with downscaling. *Remote Sensing of Environment*, 237: 111584, Feb 2020. ISSN 00344257. doi: 10.1016/j.rse.2019.111584.

[19] Alexei Lyapustin, Yujie Wang, Sergey Korkin, and Dong Huang. Modis collection 6 maiac algorithm. *Atmospheric Measurement Techniques*, 11(10):5741–5765, 2018.

[20] Jianzhao Bi, Jessica H Belle, Yujie Wang, Alexei I Lyapustin, Avani Wildani, and Yang Liu. Impacts of snow and cloud covers on satellite-derived pm2. 5 levels. *Remote sensing of environment*, 221:665–674, 2019.

[21] Zhao-Yue Chen, Jie-Qi Jin, Rong Zhang, Tian-Hao Zhang, Jin-Jian Chen, Jun Yang, Chun-Quan Ou, and Yuming Guo. Comparison of different missing-imputation methods for maiac (multiangle implementation of atmospheric correction) aod in estimating daily pm2. 5 levels. *Remote Sensing*, 12(18):3008, 2020.

[22] Noel Gorelick, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment*, 202:18–27, 2017.

[23] Cesar Aybar, Quisheng Wu, Lesly Bautista, Roy Yali, and Antony Barja. rgee: An r package for interacting with google earth engine. *Journal of Open Source Software*, 5:2272, 2020.

[24] Xia Meng, Cong Liu, Lina Zhang, Weidong Wang, Jennifer Stowell, Haidong Kan, and Yang Liu. Estimating pm2. 5 concentrations in northeastern china with full spatiotemporal coverage, 2005–2016. *Remote sensing of environment*, 253:112203, 2021.

[25] Forrest R Stevens, Andrea E Gaughan, Catherine Linard, and Andrew J Tatem. Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PloS one*, 10(2):e0107042, 2015.

[26] Tianqi Chen and Carlos Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 785–794. ACM, 2016.

[27] US Geological Survey. National elevation dataset 1/3 arc-second (ned 1/3). courtesy of the us geological survey, 2009.

[28] John Dewitz. National land cover database (nlcd) 2016 products: Us geological survey data release. 2019.

[29] M Friedl and D Sulla-Menashe. Mcd12q1 modis/terra+ aqua land cover type yearly l3 global 500m sin grid v006 [data set](nasa eosdis land processes daac, 2019), 2020.

[30] Monitoring Trends in Burn Severity (MTBS) Project (USDA Forest Service/U.S. Geological Survey). MTBS data access: Burned areas boundaries dataset, 1984 - 2020. Accessed from `https://www.mtbs.gov/direct-download` on July 22, 2022.

[31] Regional Economic Accounts. Cagdp9: Real gdp in chained dollars by county and msa. *Bureau of economic analysis, US Department of Commerce.* Accessed from `https://apps.bea.gov/regional/downloadzip.cfm` on July 6, 2022.

[32] US Census Bureau. American community survey 5-year estimates. Accessed via R package tidycensus.

[33] Marshall Burke, Sam Heft-Neal, Jessica Li, Anne Driscoll, Patrick Baylis, Matthieu Stigler, Joakim A Weill, Jennifer A Burney, Jeff Wen, Marissa L Childs, and Carlos F Gould. Exposures and behavioural responses to wildfire smoke. *Nature Human Behaviour*, pages 1–11, 2022.

[34] Karoline K Barkjohn, Brett Gantt, and Andrea L Clements. Development and application of a United States-wide correction for PM2.5 data collected with the PurpleAir sensor. *Atmospheric Measurement Techniques*, 14(6):4617–4637, 2021.