NBER WORKING PAPER SERIES

THE CHANGING RISK AND BURDEN OF WILDFIRE IN THE US

Marshall Burke
Anne Driscoll
Jenny Xue
Sam Heft-Neal
Jennifer Burney
Michael Wara

Working Paper 27423
http://www.nber.org/papers/w27423

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2020

The authors thank the Robert Wood Johnson Foundation, the Hewlett Foundation, and National Science Foundation (CNH-L #1715557) for funding. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w27423.ack

The Changing Risk and Burden of Wildfire in the US
Marshall Burke, Anne Driscoll, Jenny Xue, Sam Heft-Neal, Jennifer Burney, and Michael Wara
NBER Working Paper No. 27423
June 2020
JEL No. I14,Q53

ABSTRACT

Recent dramatic and deadly increases in global wildfire activity have increased attention on the causes of wildfires, their consequences, and how risk from fire might be mitigated. Here we bring together data on the changing risk and societal burden of wildfire in the US. We estimate that nearly 50 million homes are currently in the wildland-urban interface in the US, a number increasing by 1 million houses every 3 years. Using a statistical model that links satellite-based fire and smoke data to pollution monitoring stations, we estimate that wildfires have accounted for up to 25% of $PM_{2.5}$ in recent years across the US, and up to half in some Western regions. We then show that ambient exposure to smoke-based $PM_{2.5}$ does not follow traditional socioeconomic exposure gradients. Finally, using stylized scenarios, we show that fuels management interventions have large but uncertain impacts on health outcomes, and that future health impacts from climate-change-induced wildfire smoke could approach projected overall increases in temperature-related mortality from climate change. We draw lessons for research and policy.

Marshall Burke
Department of Earth System Science
Stanford University
Stanford, CA 94305
and NBER
mburke@stanford.edu

Anne Driscoll
Stanford University
616 Serra St
Stanford, CA 94305
anne.driscoll@stanford.edu

Jenny Xue
473 Via Ortega
Stanford, CA 94305
United States
jennyxue@stanford.edu

Sam Heft-Neal
Center on Food Security and the Environment
Stanford University
473 Via Ortega
Stanford, CA 94305
sheftneal@stanford.edu

Jennifer Burney
University of California San Diego
9500 Gilman Dr. #0519
La Jolla, CA 92093
jburney@ucsd.edu

Michael Wara
473 Via Ortega
Stanford, CA 94305
United States
mwara@stanford.edu

Over the past four decades, burned area from wildfires has roughly quadrupled in the US (Fig 1a).[1] This rapid growth has been driven by a number of factors, including the accumulation of fuels due to a legacy of fire suppression over the last century,[2] and a more recent increase in fuel aridity (Fig 1b, shown for the western US), a trend which is expected to continue as the climate warms.[3,4] These increases have happened parallel to a substantial rise in the number of houses in the wildland-urban interface (WUI). Using data on the universe of home locations across the US and updated national land cover maps, we update earlier studies[5,6] and estimate that there are now ∼49 million residential homes in the WUI, a number that has been increasing by roughly 350k houses per year over the last two decades (Fig 1c; see Supplement). As firefighting effort focuses substantially on the protection of private homes,[7] these factors have contributed to a steady rise in spending on wildfire suppression by the US government (Fig 1d), which in recent years has totaled ∼$3 billion per year in federal expenditure.[1] Total prescribed burn acreage has increased in the southeastern US but has remained largely flat elsewhere (Fig 1e), suggesting to many that there is under-investment in this risk-mitigation strategy, given the massive overall growth in wildfire risk.[8]

What are the consequences of this change in fire activity for overall air quality and for health outcomes, and how should policy respond? Large increases in wildfire activity have been accompanied by substantial increases in the number of days with any smoke in the air across the US (Fig 1f), as estimated from satellite data.[9] Such increases have been observed throughout the continental US, not just in the West, and threaten to undo the substantial improvements in air quality observed across the US over the last two decades (Fig 1g). The fingerprints of wildfire are already visible in upward-trending spring- and summertime organic carbon concentrations observed in rural areas in the US south and west (Fig S1), respectively, and studies find that having any smoke in the air can increase morbidity and mortality among exposed populations.[10,11]

A challenge in understanding the broader contribution of changing wildfire activity to air quality is the difficulty in accurately linking fire activity to related pollutant exposures in oftendistant population centers. Satellite-based measures of smoke exposure are increasingly available and are appealing because plume monitoring intuitively links source and receptor regions. Such data, however, cannot yet be used to precisely measure smoke density or to separate surface-level smoke from smoke higher in the atmospheric column, and thus are difficult to link to existing exposure-health response relationships.[12,13] Chemical transport models (CTMs), which can directly model the movement and evolution of wildfire emissions, offer an alternate approach for linking local pollution concentrations to specific fire activity. However, generating accurate exposure estimates from CTMs requires surmounting several major uncertainties in the pathway between source and receptor. First, large uncertainties in wildfire emissions inventories have been shown to lead to many-fold differences in wildfire-attributed $PM_{2.5}$ concentrations across the US (and >20x regional differences in high fire years) when different inventories are used as input to the same CTM,[14,15] and integration of satellite observations only slightly improves performance.[16] Second, the detailed conditions surrounding emissions – such as the height of emissions injections and very localized meteorology – and their transport may not be captured by models, and can dramatically impact downstream exposure estimates.[17,18] Finally, CTM representations of atmospheric chemistry may not accurately capture the evolution of wildfire smoke.[19–23] In addition to model-related uncertainty, the computational expense of running CTMs over large spatial and temporal scales means that models are rarely validated against the long time series of concentration measurements available from hundreds of ground stations across the US.

To further understand the changing contribution of wildfire to particulate matter exposure across the US, we train and validate a simple statistical model that relates changes in satellite-estimated smoke plume exposure and fire activity to ground-measured $PM_{2.5}$ concentrations across regions in the US (Supplemental Information; Fig S2). Our approach does not rely on uncertain emissions inventories and alleviates difficulties in modeling plume dispersion, and results can easily be validated against over a decade of ground data on which the model was not trained. Model estimates are robust to alternate ways of incorporating fire and plume data (Fig S3, and Tables S1-

S3) and performance in predicting variation in overall $PM_{2.5}$ is on par with benchmark remote-sensing based approaches (Fig S4) and exceeds reported performance of CTMs (see Supplemental Information). We compare estimates from this reduced-form approach to other region-specific estimates of smoke concentrations in the literature, finding that our approach provides similar estimates of the share of overall $PM_{2.5}$ from smoke as recent studies covering smaller regions or periods (Fig S5).

Our results show that the contribution of wildfire smoke to $PM_{2.5}$ concentrations in the US has grown substantially since the mid 2000s, and in recent years has accounted for up to half of the overall $PM_{2.5}$ exposure in western regions as compared to $<20\%$ a decade ago (Fig 1h). While increases in contribution of smoke to $PM_{2.5}$ are concentrated in the Western US, they can also be seen in other regions (Fig 2a-b), a result of long-distance transport of smoke from large fires. Indeed, in midwestern and eastern regions of the US, a growing share of smoke is estimated to originate from fires in the western US or from outside the US[12] (Fig 2c-d), mirroring recent findings on the substantial transboundary movement of overall $PM_{2.5}$ within the US.[24] Exposure patterns are also pertinent to environmental justice debates: we find that while counties with higher proportions of non-Hispanic whites in the population are less exposed to total $PM_{2.5}$, as has long been recognized in the environmental justice community, they are actually more exposed on average to ambient $PM_{2.5}$ from wildfire smoke (Fig 2e-f). How these differences in ambient smoke-based $PM_{2.5}$ exposure translate to actual individual exposures will depend on a variety of individual factors, including disparities in time spent outdoors and home characteristics, many of which could correlate with socioeconomic factors.

These trends and patterns highlight important points of tension between existing air quality regulation and the growing threat from wildfire smoke, and raise important unanswered research questions that will be critical to informing policy choice. Current approaches to regulation in the US treat air quality primarily as a local problem, wherein counties are penalized if pollutant concentrations exceed designated short- or long-term thresholds. Current regulation under the Clear

Air Act also potentially exempts wildfire smoke – but not smoke from prescribed burns – from attainment designation. These approaches appear at odds with the transboundary nature and growing contribution of wildfire smoke to air quality.

To better guide policy, a first key scientific contribution will be a better quantification of smoke exposures and agreed upon methods for validating these exposures. Both statistical and transport-based approaches to exposure assessment have their strengths and shortcomings, and the performance of both should be evaluated based on metrics relevant to the measurement of downstream health responses. In particular, to isolate smoke exposure from potential confounds, most statistical approaches in recent health impact studies use variation over time in pollution exposure to estimate health effects. This implies that smoke models used for estimating health impacts should be evaluated in their ability to predict temporal variation in $PM_{2.5}$ at relevant locations, not just spatial patterns in $PM_{2.5}$ levels; most existing validation efforts focus on the latter. To guard against overfitting, these evaluations must be done on ground data not used in model training.

A second key scientific question is the nature of health responses to wildfire smoke. Growing evidence indicates a range of negative health consequence associated with wildfire smoke exposure,[10,25] consistent with a vast literature on the broader health consequences of polluted air. Most recent evidence suggests that there is no "safe" level of exposure to key pollutants such as $PM_{2.5}$,[26,27] but differences in the shape of the pollution-health response function at low levels of exposure can have large implications for the benefits of pollution reduction.

To illustrate this sensitivity, we use three such recently-published response functions[26,28,29] and simulate changes in elderly mortality predicted by various changes in $PM_{2.5}$ exposure induced by mitigating wildfire smoke. Guided by existing estimates of how prescribed burning reduces subsequent wildfire activity[30] (see Supplemental Information), we combine these three response functions with our statistical $PM_{2.5}$ model to evaluate stylized scenarios in which the use of prescribed burning changes the interannual distribution and overall amount of $PM_{2.5}$ from smoke. Estimates of the annual number of elderly lives saved for a given change in smoke differ by a factor of three across published response functions, implying large average differences in the benefits of smoke mitigation (Fig 3). Evidence on whether certain populations are more susceptible to smoke exposure is also lacking.[10,25]

The large potential health benefits of smoke mitigation also raise key questions about wildfire management strategies. For instance, existing evidence does not provide a comprehensive understanding of how a given prescribed burning intervention will change the timing, amount, and spatial distribution of smoke, and we find that alternate estimates of the efficacy of prescribed burning in reducing the subsequent size of wildfires[30] can lead to more than two-fold differences in estimated health benefits of prescribed burns (Fig 3). Similarly, current fire suppression efforts understandably focus on protecting homes and structures, but the overall population health impact of a heavily polluting wildfire that does not threaten structures could be much worse than that of a smaller fire that does threaten structures. In addition, fuels management activities are targeted at local community protection and ecosystem benefits and do not consider likely downstream impacts of wildfire on large populations. Additional quantitative work is needed to help navigate these difficult trade-offs.

A third key question is whether source-agnostic $PM_{2.5}$-health response functions are appropriate for estimating wildfire-smoke-specific health impacts. Although it is commonly hypothesized, existing literature is mixed on whether exposure to wildfire smoke has different health impacts than exposure to other sources of $PM_{2.5}$.[31] Improved science on this topic – including necessary investments in speciated monitoring to distinguish wildfire-specific pollutants – will be critical for understanding wildfire impacts.

Fourth, how might the interaction of climate change and wildfire risk shape policy priorities? A warming climate is responsible for roughly half of the increase in burned area in the US,[4] and future climate change could lead to up to an additional doubling of wildfire related particulate emissions in fire-prone areas[32] or a many-fold increase in burned area.[33,34] Costs from these increases include both the downstream economic and health costs of smoke exposure, as well as the cost of suppression activities, direct loss of life and property, and other adaptive measure (e.g. power shutoffs) that have widespread economic consequences. Whether accounting for these wildfire-related costs meaningfully increases the estimated overall economic damages from climate change is unknown, but stylized calculations suggest that increased mortality from climate-change-induced wildfire smoke is roughly on par with projected overall increases in temperature-related mortality – itself the largest estimated contributor to economic damages in the US[35] (see Supplemental Information). A key policy question will be whether and to what degree to modify current exceptions to the Clean Air Act granted to states for pollution impacts from wildfire smoke, as these erode gains from efforts aimed at reducing other $PM_{2.5}$ pollution sources.

Finally, wildfires could interact with the COVID-19 pandemic in important but unknown ways. COVID-19 is likely to impede the ability of government at all scales to respond to wildfire risk, both before and after fires occur. This could present specific challenges in many parts of the West, where record drought in the 2019/20 rainy season followed an accumulation of fuels during a relatively wet 2018/19 season. Wildland firefighter trainings have been delayed or cancelled, government agencies acknowledge that COVID-19 could degrade their firefighting capabilities, and traditional approaches to wildfire evacuation are likely to prove far more challenging due to lack of social distancing at evacuation centers. If large fires do occur and burn uncontained in 2020, this might in turn worsen COVID-related health outcomes, as early evidence suggests that exposure to air pollution increases both COVID cases and deaths in the US[36,37] (a finding consistent with the relationship between pollution and other viral respiratory illness[38,39]). A better causal understanding of the impact of air pollution on COVID outcomes, including that from wildfires, is a critically urgent research priority and could be important in guiding labor- and finance-constrained firefighting effort.

References

[1] National Interagency Fire Center, Wildland fire statistics (2019).

[2] J. R. Marlon, *et al.*, *Proceedings of the National Academy of Sciences* **109**, E535 (2012).

[3] A. L. Westerling, H. G. Hidalgo, D. R. Cayan, T. W. Swetnam, *science* **313**, 940 (2006).

[4] J. T. Abatzoglou, A. P. Williams, *Proceedings of the National Academy of Sciences* **113**, 11770 (2016).

[5] A. Bar-Massada, S. I. Stewart, R. B. Hammer, M. H. Mockrin, V. C. Radeloff, *Journal of environmental management* **128**, 540 (2013).

[6] V. C. Radeloff, *et al.*, *Proceedings of the National Academy of Sciences* **115**, 3314 (2018).

[7] P. Baylis, J. Boomhower, Moral hazard, wildfires, and the economic incidence of natural disasters, *Tech. rep.*, National Bureau of Economic Research (2019).

[8] C. A. Kolden, *Fire* **2**, 30 (2019).

[9] W. Schroeder, *et al.*, *International Journal of Remote Sensing* **29**, 6059 (2008).

[10] W. E. Cascio, *Science of the total environment* **624**, 586 (2018).

[11] N. Miller, D. Molitor, E. Zou, *working paper* (2017).

[12] S. J. Brey, M. Ruminski, S. A. Atwood, E. V. Fischer, *Atmospheric Chemistry and Physics* **18**, 1745 (2018).

[13] K. O'Dell, B. Ford, E. V. Fischer, J. R. Pierce, *Environmental science & technology* **53**, 1797 (2019).

[14] S. N. Koplitz, C. G. Nolte, G. A. Pouliot, J. M. Vukovich, J. Beidler, *Atmospheric environment* **191**, 328 (2018).

[15] T. S. Carter, *et al.*, *Atmospheric Chemistry & Physics* **20** (2020).

[16] F. H. Johnston, *et al.*, *Environmental health perspectives* **120**, 695 (2012).

[17] R. A. Kahn, *et al.*, *Geophysical Research Letters* **35** (2008).

[18] Y. Rastigejev, R. Park, M. P. Brenner, D. J. Jacob, *Journal of Geophysical Research: Atmospheres* **115** (2010).

[19] L. Zhang, *et al.*, *Atmospheric Chemistry and Physics* (2014).

[20] J. C. Liu, *et al.*, *Epidemiology (Cambridge, Mass.)* **28**, 77 (2017).

[21] H. Forrister, *et al.*, *Geophysical Research Letters* **42**, 4623 (2015).

[22] S. Tomaz, *et al.*, *Environmental science & technology* **52**, 11027 (2018).

[23] M. J. Gunsch, *et al.*, *Atmospheric Chemistry and Physics* **18** (2018).

[24] I. C. Dedoussi, S. D. Eastham, E. Monier, S. R. Barrett, *Nature* **578**, 261 (2020).

[25] C. E. Reid, *et al.*, *Environmental health perspectives* **124**, 1334 (2016).

[26] Q. Di, *et al.*, *New England Journal of Medicine* **376**, 2513 (2017).

[27] T. Deryugina, G. Heutel, N. H. Miller, D. Molitor, J. Reif, *American Economic Review* **109**, 4178 (2019).

[28] R. Burnett, *et al.*, *Proceedings of the National Academy of Sciences* **115**, 9592 (2018).

[29] J. D. Sacks, *et al.*, *Environmental Modelling & Software* **104**, 118 (2018).

[30] M. Cochrane, *et al.*, *International Journal of Wildland Fire* **21**, 357 (2012).

[31] C. Black, Y. Tesfaigzi, J. A. Bassein, L. A. Miller, *Environmental toxicology and pharmacology* **55**, 186 (2017).

[32] M. D. Hurteau, A. L. Westerling, C. Wiedinmyer, B. P. Bryant, *Environmental science & technology* **48**, 2298 (2014).

[33] X. Yue, L. J. Mickley, J. A. Logan, J. O. Kaplan, *Atmospheric Environment* **77**, 767 (2013).

[34] D. J. Wuebbles, D. W. Fahey, K. A. Hibbard (2017).

[35] S. Hsiang, *et al.*, *Science* **356**, 1362 (2017).

[36] X. Wu, R. C. Nethery, B. M. Sabath, D. Braun, F. Dominici, *medRxiv* (2020).

[37] C. Persico, K. R. Johnson, *et al.*, Deregulation in a time of pandemic: Does pollution increase coronavirus cases or deaths?, *Tech. rep.*, Institute of Labor Economics (IZA) (2020).

[38] Y. Cui, *et al.*, *Environmental Health* **2**, 15 (2003).

[39] G. Singer, J. G. Zivin, M. Neidell, N. Sanders, *medRxiv* (2020).

[40] D. Glickman, B. Babbitt, *Federal Register* **66**, 751 (2001).

[41] V. C. Radeloff, *et al.*, *Ecological applications* **15**, 799 (2005).

[42] A. Van Donkelaar, *et al.*, *Environmental science & technology* **50**, 3762 (2016).

[43] A. Van Donkelaar, *et al.*, *Environmental health perspectives* **118**, 847 (2010).

[44] N. Fann, *et al.*, *Science of the total environment* **610**, 802 (2018).

[45] D. Jaffe, W. Hafner, D. Chand, A. Westerling, D. Spracklen, *Environmental science & technology* **42**, 2812 (2008).

[46] X. Liu, *et al.*, *Journal of Geophysical Research: Atmospheres* **122**, 6108 (2017).

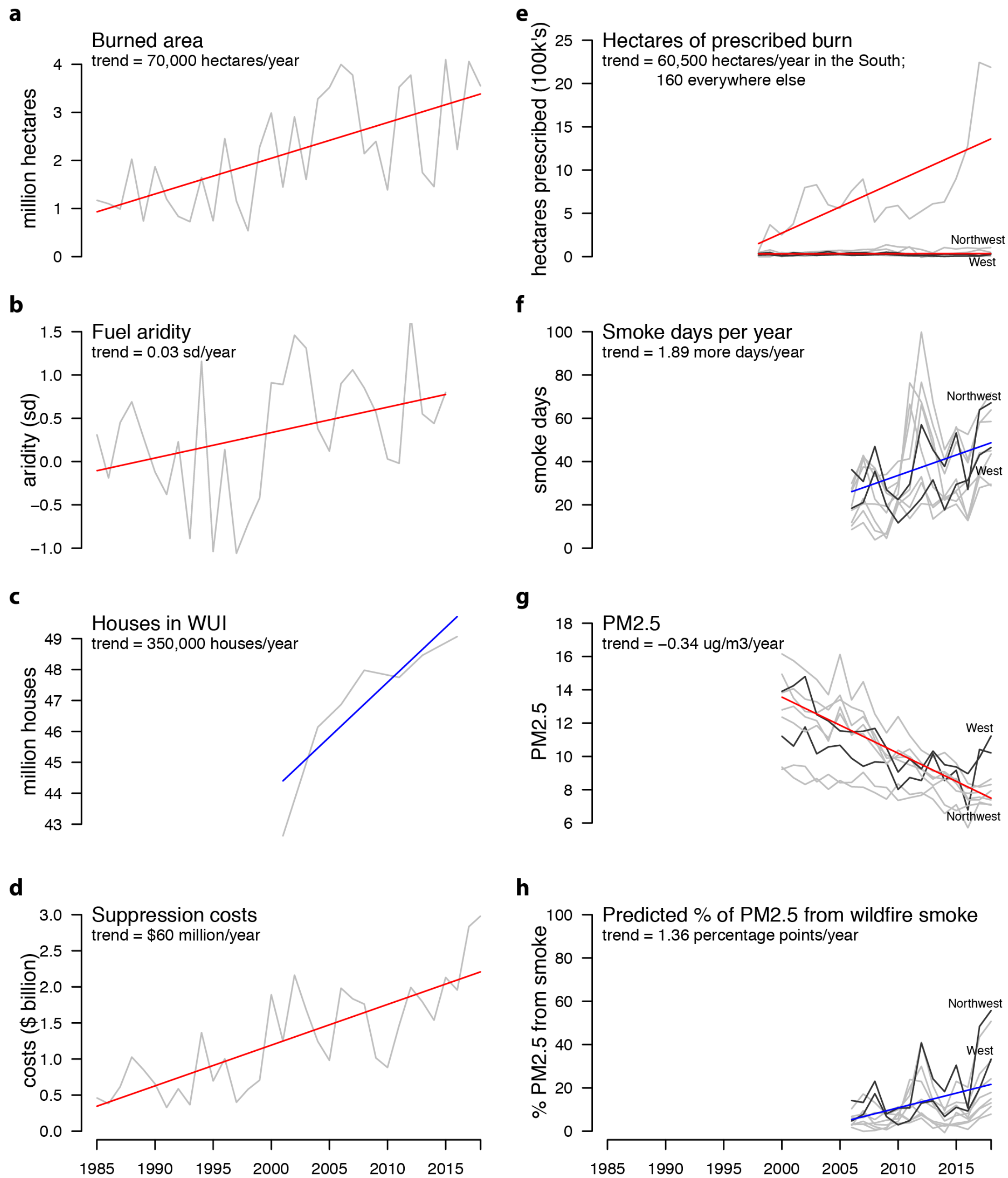

Figure 1: **Trends in the drivers and consequences of wildfire**. **a** Increases in burned area in public and private US lands[1] have been driven in part by rising fuel aridity, shown here over the western US[4] (**b**). The number of homes in the Wildland Urban Interface (WUI) has also risen quickly (**c**, our calculations, see Supplemental Information), which has contributed to rising suppression costs (**d**) incurred by the federal government. **e** Prescribed burn area has increased substantially in the South but is flat in all other regions.[1] **f** Smoke days have increased throughout the US, perhaps undermining decadal improvements in air quality across the US (**g**). **h** We calculate an increasing proportion of overall $PM_{2.5}$ attributable to wildfire smoke, particularly in the West. Red and blue lines in each plot indicate linear fits to the historical data, with slopes reported in the upper left of each panel; all are significantly different from zero ($p < 0.01$ for each), except for prescribed burn in regions outside the South. Red lines indicate underlying data are from published studies or government data, blue lines indicate novel estimates from this paper.

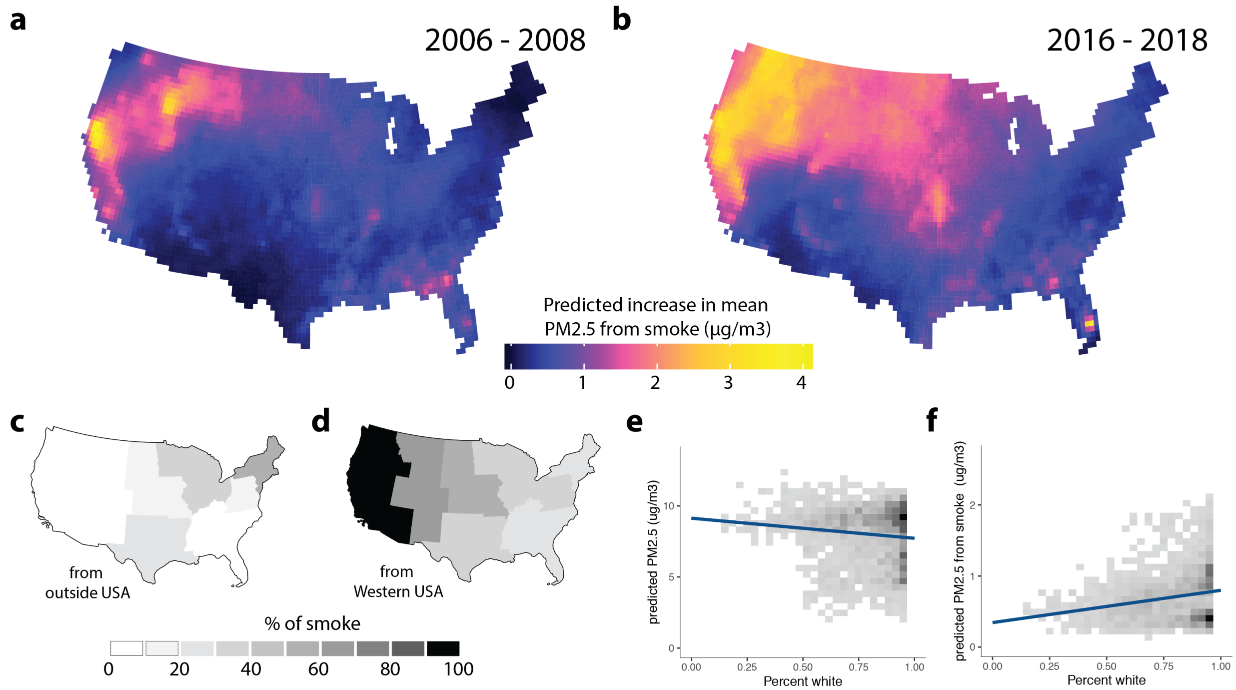

Figure 2: **The quantity, source, and incidence of wildfire smoke**. **a-b** Average predicted ug/m3 of $PM_{2.5}$ attributable to wildfire smoke in 2006-08 and 2016-28, as calculated from a statistical model fitting satellite-derived smoke plume data. **c** Share of smoke originating outside US June-Sept 2007-2014, (calculated from[12]), with a substantial amount of smoke the northeast and midwest originating from Canadian fires, and about 60% of smoke in the north east originating outside the country; nationally, ∼11% of smoke is estimated to originate outside the country. **d** The share of smoke originating in the western US June-Sept 2007-2014. Smoke originating in the western US accounts for 54% of the smoke experienced in the rest of the US. **e-f** Racial exposure gradients are opposite for particulate matter from smoke as compared to total particulate matter: across the coterminous US, counties with higher population proportion of non-Hispanic whites have lower average particulate matter exposure but higher average ambient exposure to particulate matter from smoke ($p < 0.01$ for both relationships).

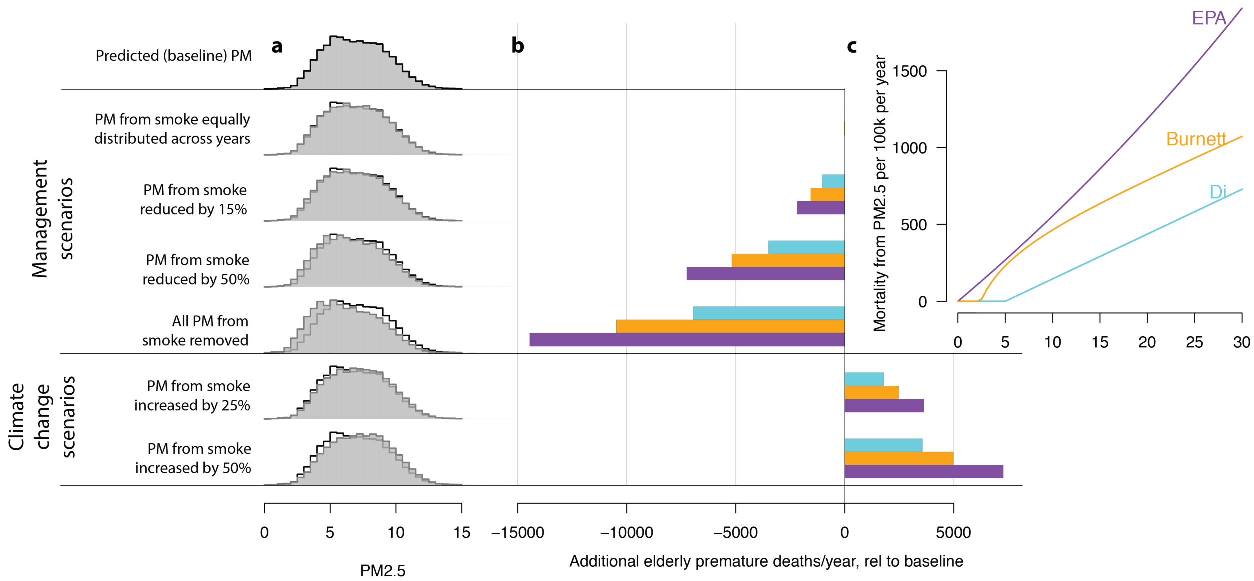

Figure 3: **Health consequences of changes in smoke exposure depend on the assumed dose-response function and on the magnitude of management- or climate-driven changes in smoke. a** Distributions of $PM_{2.5}$ for all grid-cell years in the contiguous US 2006-2018 under several stylized wildfire management strategies and climate change scenarios (see Supplemental Information for details). The baseline distribution of total predicted $PM_{2.5}$ from all sources is in black. Grey distributions show alternative scenarios in which the timing and/or amount of overall smoke-related $PM_{2.5}$ is altered through management interventions or increased due to climate, including the (hypothetical) full elimination of smoke $PM_{2.5}$. **b** Annual number of avoided premature deaths in the US population age 65+ for each management strategy, calculated by combining the $PM_{2.5}$ distributions in **(a)** with published long-run $PM_{2.5}$ exposure-response functions depicted in **c**.[26,28,29]

Supplemental Information

Estimating houses in WUI

The wildland urban interface (WUI) is the area where houses and other man-made structures are near or overlap wildland vegetation. The U.S. Federal Register defines two types of WUI.[40] Intermix WUI is defined as an area with more than one structure per 40 acres (6.17 houses/km^2) and with over 50% wildland vegetation. Interface WUI is defined as an area with less than 50% vegetation but that lies within 1.5 miles (2.4km) of a 5km^2 or larger area that is at least 75% wildland vegetation.

To map WUI in the conterminous U.S., we use an approach similar to Bar-Masada et al,[5] where we create two rasters: 1) a household density map of the conterminous U.S.; and 2) a wildland vegetation density map identifying areas with high levels of vegetation. The household density raster was created using CoreLogic household data,, which proprietary dataset which includes detailed information on all existing housing structures (including the year the structure was built, longitude and latitude, household value, and so on). We convert the dataset into a spatial points layer, and filter for residential houses only. For households without a recorded year built we assume that the household was built on or before the year 2000 and code the missing year as 2000. The vegetation density raster was computed using the National Land Cover Dataset, a 30m resolution raster of land type across the U.S. We use the same vegetation classification as Radeloff et al.[41],[6] where we classify forests (classes 41-43), shrublands (classes 51 and 52), grasslands (class 71), and woody wetlands (class 90) as wildland vegetation.

We use the vegetation and housing density rasters to compute intermix WUI. To speed up computation time, instead of calculating household and vegetation density using a cell-level moving window approach,[5] we aggregate the 30m resolution NLCD dataset up to 1km cells. To obtain vegetation density, we sum the area that is wildland vegetation within each 1km cell and code

"1" for each cell where wildland vegetation is over 50% and "0" otherwise. To obtain housing density, we compute the number of houses per 1km cell and code "1" for each cell where there are more than 6.17 houses, and "0" otherwise. Intermix WUI is then calculated by multiplying the household density and vegetation density rasters. The resulting intermix raster has a cell value of "1" if household density is over 6.17 houses and wildland vegetation is over 50% and "0" otherwise.

To calculate interface WUI, we first identify continuous patches of wildland vegetation by converting the vegetation density raster into spatial polygons. We keep polygons that are over $5km^2$, and create a buffer of 2.4km around each polygon. The polygons are then converted into a vegetation buffer raster, where cells that fall within the 2.4km buffer around each polygon are labelled "1", and "0" otherwise. Interface WUI is then calculated by multiplying the household density raster and the vegetation buffer raster. The resulting interface raster has a value of "1" if household density is over 6.17 and if cells fall within the buffered area.

Finally, we calculate the number of households that fall within either the interface or the intermix WUI by using the CoreLogic spatial points layer which contains the individual household locations. For each household, we identify whether the household falls into either WUI category by extracting the intermix and interface raster cell values. For each year the NLCD is available, we create new household density and vegetation rasters and classify WUI status for all households. Finally, we count the number of households inside the interface and intermix WUI areas for each year. Annual estimates are plotted in Fig. 1c.

Estimating $PM_{2.5}$ attributable to wildfire smoke

To quantify the variation in $PM_{2.5}$ attributable to wildfire smoke, we estimate a statistical model that relates satellite-derived data on wildfire smoke and fire activity along to $PM_{2.5}$ data collected by ground monitors. Our goal is to provide a tractable way to measure the contribution of smoke to overall $PM_{2.5}$ at a local level and across time in the US using a model that can be directly and easily validated across wide geographies against ground data on which the model was not trained.

The advantage of this "reduced-form" statistical approach, relative to chemical transport model (CTM) based approaches, is that it does not rely directly on uncertain emissions inventories or on an ability to correctly model atmospheric chemistry or plume dispersion, both of which have been shown to challenge CTM-based approaches. Furthermore, the model can easily be run at large spatial and temporal scale, and estimates of overall $PM_{2.5}$ can be readily validated against ground data on which the model was not trained using cross-validation. Generating temporal data across many locations is important, as such variation is what is typically exploited in recent studies to estimate the impacts of air pollution on health outcomes.[11,27] The disadvantage of the statistical approach is an inability to easily constrain where the smoke is in the atmospheric column or a way to directly measure smoke density in each plume, and the lack of foolproof way of linking particular smoke plumes to source fires, although methods for the latter are being explored.[12] We describe and validate our approach to indirectly constraining these factors below.

We train a statistical model that relates variation in nearby emissions sources and both near and distant fires to ground-measured concentrations of $PM_{2.5}$ at 100km grid cells throughout the US:

$$PM_{irt} = f(plumes_{irt}) + Z_{irt} + \alpha_r + \delta_t \tag{1}$$

where PM_{irt} is annual average $PM_{2.5}$ concentration at grid cell i in region r in year t, calculated as the average of any EPA monitoring stations reporting in that grid cell. The variable $plumes_{irt}$ includes the count, size, and source of wildfire plumes that overlay the grid cell in a given year, as derived from remotely-sensed smoke and fire variables, and $\mathbf{Z_{irt}}$ is a vector of time-varying sources of particulate emissions (direct or precursor) derived from a range of data sources. Any other sources of average difference in $PM_{2.5}$ between regions (e.g. unmodeled sources of stable regional emissions) is picked up by the vector of regional intercepts α_r, and any additional time-varying sources of $PM_{2.5}$ that are common across all grid cells are accounted for by the vector of year intercepts δ_t.

Regions are defined starting from the USGS physiographic divisions but are adapted to ensure that they are similarly sized and contain at least two reporting EPA monitors. We initially define our regions following USGS's second level divisions ("provinces"). However, for the five largest provinces (each of which covers 11-25% of the lower 48) we use the next level USGS division referred to as "sections". The 9 sections that did not have the required two EPA stations are manually merged with their smallest neighbor, leaving a final count of 42 regions (Fig S2).

Our *plumes* measure includes both fire and smoke variables constructed from satellite-based estimates of smoke plumes and of fire activity compiled as part of NOAA's Hazard Mapping System (HMS) Fire and Smoke Product.[9] Both sources are generated by trained analysts who analyze sub-daily animated imagery primarily from the geostationary operational environmental satellite system (GOES). The system has provided a measure of smoke and fire activity every few hours throughout the daytime across the entire US for more than a decade, and we utilize the daily lists of active fires and smoke plumes. Our *plumes* measure is constructed as follows:

1. Beginning with HMS daily fire location data, which is provided as active fire points on a given day, we group contiguous fire points into a single fire by creating a buffer of 2.9 km around each point and merging points whose buffers overlap. We use 2.9km as each cell is 4km across, so from the centroid to the edge of the cell is $\sqrt{(2^2 + 2^2)} = 2.83 \sim 2.9$km. In order to capture large fires that are burning over several days, and have produced large plumes over that time, we use fire points from the day in question as well as the three days prior.

2. For each plume, we identify the largest intersecting fire active on that day and associate the plume with that "origin fire".

3. We then find the distance from each grid cell the plume overlaps to the origin fire, also noting the size of the origin fire.

4. Finally, for each grid-cell-year we generate counts of plumes overlapping that grid cell in each of nine classes, where the classes are a combination of origin fire size tercile (small = <140 acres, medium == 140-800 acres, large == >800 acres) and distance to origin fire tercile (close = <280km, med = 280-1000km, far>1000km). In our baseline model, these nine covariates are used as regressors in Equation 1.

The goal of our binning approach is to provide the model information that can proxy for the density of smoke and where it is in the atmospheric column: fires further away are likely to be higher in the column,[12] and larger and more nearby fires are likely to generate more smoke near the ground. Even though our method of matching to origin fire is likely imperfect, our binning approach does appear to improve on alternative specifications that do not use this information: models that simply count overhead plumes and do not use information on distance or size of origin fire perform less well and over-attribute high smoke $PM_{2.5}$ to the Midwest, which experiences many plumes from distant fires (Fig S3 and Table S3). Our binned model attributes less smoke $PM_{2.5}$ to the Midwest, and overall attributes a somewhat smaller share of overall $PM_{2.5}$ to wildfire smoke (Fig S3) – although the temporal and spatial patterns are largely similar.

Non-wildfire-related time-varying sources of $PM_{2.5}$ are collected from various sources and informed by the EPA's National Emissions Inventory (NEI), which identifies key sources of particulate emissions across US counties. We assemble annual information on activity related to industrial processes, aircraft traffic, natural gas production, mining, power plants, transportation, and agriculture. In particular, we assemble county-year-level data on population, construction payroll, mining payroll, airport locations and taxi times, natural gas production, and cattle head count. In addition, we assemble zip code-year-level data on power plant emissions as well as 1km grid cell-year-level data on traffic. Each covariate is harmonized to our 100km grid by overlaying the relevant administrative boundary at which the data are available with our grid. For missing data in grid cells that have 2 or more non-missings, we linearly interpolate between years. For grid cells that are entirely missing a variable, we impute using all other variables. We chose not to include weather and climate variables as they are likely to capture some of the variation in $PM_{2.5}$ that is attributable to smoke, i.e. locations or years that are especially arid are likely to have higher $PM_{2.5}$ concentrations given the role of fuel aridity in fire risk.

The last important components of Equation 1 are the region and year fixed effects which account for average differences between regions and common changes over time in both smoke exposure and overall $PM_{2.5}$. The fixed effects offer a parsimonious way of dealing with remaining uncertainties in non-smoke sources of $PM_{2.5}$. For instance, they help ensure that the model learns that the co-occurance of lower average smoke exposure and higher average $PM_{2.5}$ (e.g. in the Central and Northeast regions) does not indicate that reducing smoke exposure will increase average $PM_{2.5}$. Similarly, the year fixed effects help the model learn that the overall increase in smoke exposure over our study period and the average decline in overall $PM_{2.5}$ again does not imply that increasing smoke exposure decreases average $PM_{2.5}$. The fixed effects will be particularly important in this task if our time-varying data on non-smoke sources of $PM_{2.5}$ are either poorly measured or missing key contributors to $PM_{2.5}$.

Once we have built the model with these variables, we use it to predict $PM_{2.5}$ across the nation at observed smoke levels, then to predict the counterfactual by setting smoke exposure to 0. Predictions are bottom coded at 0, and the amount of $PM_{2.5}$ attributed to smoke is the difference between the two predictions.

To train and validate the model, we focus on grid cells with at least one EPA station reporting at any point during our sample. We then partition these cells into disjoint sets for 10-fold cross validation, where all years in a given cell are in one fold. To be able to estimate the region and year fixed effects, all folds contain cells with $PM_{2.5}$ observations in all regions and all years, and no cell (and thus no monitoring station) contributes information to more than one fold.

Model validation and robustness Our model validates well against data on which it was not trained, explaining 60% of the variation in overall $PM_{2.5}$ at these held out locations (Fig S4a). This performance is very close to performance of benchmark satellite-based $PM_{2.5}$ data generated by van Donkelaar et al[42] in the US, which we aggregate to our 100km grids for comparison (Fig S4b). The van Donkelaar data have been used widely for US and global pollution-health studies, although we emphasize that van Donkelaar are using satellite information from multiple sensors to predict overall $PM_{2.5}$ variation, and are not attempting to estimate the wildfire component explicitly. Our temporal and spatial predictive performance in the continental US also mirrors that of van Donkelaar et al (Fig S4c-d).

Our performance appears to exceed that of many CTM-based studies in predicting overall variation in $PM_{2.5}$, although many smoke-based CTM studies do not report validation performance or are not directly validated on the same monitoring stations or over time, hampering comparison. To our knowledge the most comprehensive comparison of CTM and ground-based PM observations reports an $r^2 = 0.16$ for GEOS-Chem (a leading CTM) over western North America.[43] Comparatively (and excluding the part of Canada in their estimates) our model has an $R^2 = 0.44$ over that region.

Our baseline model that uses binned counts of plumes by fire size and distance outperforms other ways of using the smoke and fire data (Table S3), but both temporal and spatial patterns of wildfire-attributed $PM_{2.5}$ are generally consistent using these alternate approaches (Fig S3).

A remaining concern with our attributable smoke share estimates is that our simplified statistical model could be missing key non-smoke sources of $PM_{2.5}$, which would lead us to underestimate total $PM_{2.5}$ (the denominator in the shares calculation) and thus perhaps lead to upward-biased estimates of attributable smoke share reported in Figs 1-3. However, for reasons stated above, non-smoke sources of $PM_{2.5}$ tend to have opposite temporal and spatial relationships with overall $PM_{2.5}$ as compared to smoke: overall $PM_{2.5}$ has been trending down while smoke has trended up, and regions with fewer fires have historically had higher average $PM_{2.5}$. This means that exclusion of key non-smoke sources of $PM_{2.5}$ could bias estimates of the impact of smoke in Equation 1 downward, resulting in smaller attributable shares.

To understand the direction of bias in attributable shares, we systematically dropped key non-smoke sources of $PM_{2.5}$ in our data individually and jointly re-calculated the attributable shares. Dropping individual sources appeared to have minimal effect on overall estimated smoke shares (Table S1), but dropping them jointly or dropping the fixed effects as well substantially *reduced* the estimated smoke shares – again consistent with expected downward bias in coefficients in Equation 1 if spatial and temporal patterns in non-smoke $PM_{2.5}$ are not accounted for. This exercise suggests that our overall estimates of attributable smoke shares are, if anything, conservative.

Comparison to other estimates We compare our results to three studies that provided temporal estimates of wildfire-attributable $PM_{2.5}$ at either regional or national scale.

Fann et al[44] run a CTM with and without wildfire smoke to estimate the contribution of wildfire smoke to total $PM_{2.5}$ across the continental US between the years 2008-2012. They do not provide validation of model runs against ground data. We compare spatial and temporal patterns in smoke-attributed $PM_{2.5}$ finding general agreement in overall spatial patterns, and similarity in the level and trend in mean smoke-attributed $PM_{2.5}$ (Fig S5). Overall, we have higher estimates in the midwest, which is partially attributable to the fact that we also include international fires, which contribute significantly to the midwest.

O'Dell et al[13] use both a statistical approach and a CTM to estimate the contribution of wildfire smoke to $PM_{2.5}$ in the Pacific Northwest. They do not provide validation of CTM results against ground data. They report percent of $PM_{2.5}$ attributable to smoke during the summer months (July, August, and September) from 2006-2016. As our estimates are annual, we calculate a comparable version of their estimates by assuming there is no $PM_{2.5}$ from smoke in other months (as summer is the main fire season - see sinusoidal seasonal pattern across IMPROVE sites in Figure S1) and the $PM_{2.5}$ in other months is equivalent to the summer non-smoke $PM_{2.5}$.

$$\text{Comparable yearly O'Dell estimate} = \frac{\text{summer smoke PM}_{2.5}}{\text{total summer PM}_{2.5} + (\text{summer non-smoke PM}_{2.5}) * 3}$$

These assumptions are likely to cause an underestimate of O'Dell's full year estimate of the percent of $PM_{2.5}$ from smoke, as the assumption that there is no $PM_{2.5}$ from smoke in other months is typically incorrect; 33% of smoke days in our data are outside that period. Nonetheless the trend in our estimates matches those of O'Dell , and our estimates are only slightly higher than theirs, as approximated from their figure data (Fig S5d).

Jaffe et al.[45] estimate statistical models that relate station-measured $PM_{2.5}$ to nearby burned area, and use these models to estimate the contribution of wildfire smoke to overall $PM_{2.5}$ across five regions in the Western US for the years 1988-2004. Jaffe et al use data from the IMPROVE network, which are mainly located in rural environments in which the share of anthropogenic $PM_{2.5}$ will likely be lower than the EPA stations that are the focus of our analysis. The difference in time periods between their analysis and ours (1988-2004 versus 2006-2018) also makes comparison difficult, but we proceed as it is one of the few other papers that provides comparable estimates. We aggregate our gridded data to the five regions used in Jaffe et al and compare regional $PM_{2.5}$ smoke shares (Fig S5e). Despite the differences in time periods, we see similar overall regional patterns, with our estimates smaller than Jaffe et al in Rockies and larger in coastal western states (Jaffe et al regions 4 and 5).

Finally, we compare our results to estimates of wildfire $PM_{2.5}$ contributions from EPA National Emissions Inventory (NEI) data in 2008, 2011, 2014 and 2017. To estimate wildfire emissions, EPA uses data on burned area collected by national and local agencies with information on fuel type and fuel-specific emissions factors to convert burned area to local emissions. Importantly, these estimates will represent the smoke emitted from a given location, not the overall contribution of smoke to $PM_{2.5}$ in that location (as much of this smoke will originate from fires elsewhere), and so differences between NEI and our estimates could in large part reflect the importance of non-local smoke to local $PM_{2.5}$ concentrations. For comparison, we aggregate both our and NEI results to the state-year level, comparing estimates of wildfire-attributed $PM_{2.5}$, where NEI estimates are calculated as emissions per acre to make them comparable. Estimates are positively related (Fig S5f) but the differences in some years are large, again likely reflecting the importance of non-local sources of smoke.

Estimating health impacts of changes in smoke exposure We use three published exposure-response curves that relate changes in overall $PM_{2.5}$ to elderly (65+) mortality to investigate how changing forest management strategies and climate scenarios might change the overall health impacts of wildfire smoke.[26,28,29] This exercise requires a number of assumptions. First, we assume that changes in exposure to wildfire-derived $PM_{2.5}$ have the same effect on mortality as changes in overall exposure to $PM_{2.5}$. Existing work has clearly shown that smoke exposure increases mortality risk,[10,11] but does not yet conclude whether these impacts are different from other sources of $PM_{2.5}$ exposure. A second key assumption is that these response curves reflect the causal effect of changes in $PM_{2.5}$ on mortality. This is also an area where further research could improve estimates, as the studies we use (which are also those used in current policymaking and research on air quality impacts) use cohort-based or cross-sectional study designs that likely cannot fully isolate variation in $PM_{2.5}$ from other factors that affect mortality risk. More recent quasi-experimental research designs (e.g.[11,27]) will likely prove useful in this regard. Finally, our scenario analyses are highly stylized, and are meant to explore the rough magnitude of changes in smoke and health impacts and the sensitivity of overall estimates to input parameters. Our approach could be expanded to incorporate a more detailed assessment of how specific management interventions would shape pollution exposures and downstream health outcomes.

We create several different stylized scenarios for changes in smoke-related $PM_{2.5}$ exposure under different management regimes or changes in climate. These scenarios use our statistical model to predict grid-level changes in total $PM_{2.5}$ as smoke variables are perturbed. The changing overall distribution of $PM_{2.5}$ across grid cells are shown in Fig 3. The specific scenarios are as follows:

- **Original Model/Baseline**. Use the total modeled $PM_{2.5}$ for each grid-cell year.
- **Redistribute $PM_{2.5}$ from smoke equally across years**. For each grid cell, sum up all $PM_{2.5}$ from smoke over the 13 year study period, divide it equally and add it to the predicted $PM_{2.5}$ with no smoke for each year. This results in a slight shrinkage of the distribution, years with high smoke have reduced $PM_{2.5}$, and years with low $PM_{2.5}$ from smoke have increased $PM_{2.5}$. This scenario is meant to mimic a scenario in which prescribed burning increases smoke in low-wildfire years and decreases it in otherwise high-wildfire years, but the total amount of smoke-derived $PM_{2.5}$ does not change.
- **Redistribute $PM_{2.5}$ from smoke and reduce the total amount of $PM_{2.5}$ from smoke**. Here we assume that fuels management strategies are able to reduce the overall amount of wildfire smoke by 25%, and that these strategies redistribute the smoke evenly across years. This stylized scenario is roughly consistent with a tripling of prescribed burn area, given existing research that suggests that each hectare burned under a prescribed burn leads to an average of 1.8 fewer hectares burned in wildfires,[30] and that an acre of prescribed burn produces half the $PM_{2.5}$ as an acre of wildfire burned[46] (as wildfires are estimated in recent work to have higher particulate emissions factors and consume more fuel). In historical data, prescribed burned acreage is roughly 5% of total wildfire acreage in western US states (this ratio is higher in the Southern US, see Fig 1e). In a scenario with A acres of total burned area (wildfire + prescribed burn), $A * 0.05$ of those would be from prescribed burn and the remaining $A * 0.95$ from wildfire.

$$\text{original abstracted total emissions} = 0.95A + (0.05A) * 0.5 = 0.975A$$

$$\text{burned area w/ X extra Rx burning} = A + X - (1.8X) = A - 0.8X$$

$$\begin{aligned}
\text{emissions w/ X extra Rx burning} &= \frac{\text{new wildfire acreage} + (\text{new prescribed acreage} * 0.5)}{\text{original total acreage}} \\
&= \frac{(0.95A - 1.8X) + (0.05A + X) * 0.5}{0.95A + (0.05A) * 0.5} \\
&= \frac{0.975A - 1.3X}{0.975A} \\
&= \frac{1,702,144 - 249,648}{1,702,144} = 0.85
\end{aligned}$$

Using A equal to the average total burned area in the western US states and assuming that prescribed burn area triples ($X = A * 0.11$; $A = 1,745,789$) leads to a reduction in smoke of about 15%. We reduce our predicted smoke (as summed over all 13 years and spread equally within each grid cell) by 15%. We note that prescribed burn area in the Southeastern US has more than tripled in the last decade, suggesting such increases are in principle possible.

- **Redistribute all $PM_{2.5}$ from smoke and dramatically reduce the total amount of $PM_{2.5}$ from smoke**. Similarly, research shows in area where prescribed burning is successful each hectare burned under a prescribed burn leads to 4.9 fewer hectares burned in wildfires.[30] Using the same method as above, we substitute this improved efficacy of prescribed burning, leading to a reduction in smoke of about 50%.

$$\begin{aligned}
\text{emissions w/ X extra Rx burning} &= \frac{(0.95A - 4.9X) + (0.05A + X) * 0.5}{0.95A + (0.05A) * 0.5} \\
&= \frac{0.975A - 4.5X}{0.975A} \\
&= \frac{1,702,144 - 864,165}{1,702,144} = 0.49
\end{aligned}$$

- **Remove all $\mathbf{PM_{2.5}}$ from smoke**. Use the predictions for each grid-cell year when smoke is set to 0. This functions as an estimate of the overall mortality attributable to $PM_{2.5}$ from smoke.

- **Climate change increases smoke $\mathbf{PM_{2.5}}$ by 25%**. Climate change is likely to substantially increase wildfire burned area, with estimates ranging from a 25% increase in burned area in parts of the US to many-fold increases.[32–34] Here we simulate the impact of the lower range of those estimate – a 25% increase in burned area – assuming that smoke $PM_{2.5}$ increases linearly with burned area and that all portions of the US experience this 25% increase.

- **Climate change increases smoke $\mathbf{PM_{2.5}}$ by 50%**. This estimate is near the median estimate for changes in wildfire emissions in California,[32] but remains lower than other estimates.[33,34]

Changes in the cell-level distribution of smoke are shown in Fig 3a. These changes are then applied to the three exposure-response functions in Fig 3c, with changes in the mortality rate calculated as the response-function-specific difference in mortality rate between a given scenario s and base scenario b, i.e. $f_r(PM_s) - f_r(PM_b)$. Changes in mortality rate are calculated at each grid cell, applied to grid-level population estimates, and then summed across grids to get the change in elderly premature deaths shown in Fig 3b. For simplicity, we assume that changes in $PM_{2.5}$ exposure are applied to the current population of elderly, and thus do not account for future population growth in this demographic or changes in the underlying elderly mortality rate.

Comparison of temperature versus smoke mortality impacts under climate change Existing studies on the impact of climate change in the US find that the impacts of future temperature change on mortality are by far the largest component of total monetized impacts of climate change.[35]

We calculate that a 50% increase in smoke would lead to an average 0.35 μg/m^3 change in $PM_{2.5}$, which existing dose/response functions indicate would lead to an increase in elderly deaths of between 9-20 per 100,000 people. Estimates in Hsiang et al[35] suggest that each additional degree C increase in temperature leads to roughly 24 additional deaths per 100,000 elderly people (using linear function reported in Hsiang et al Table S12).

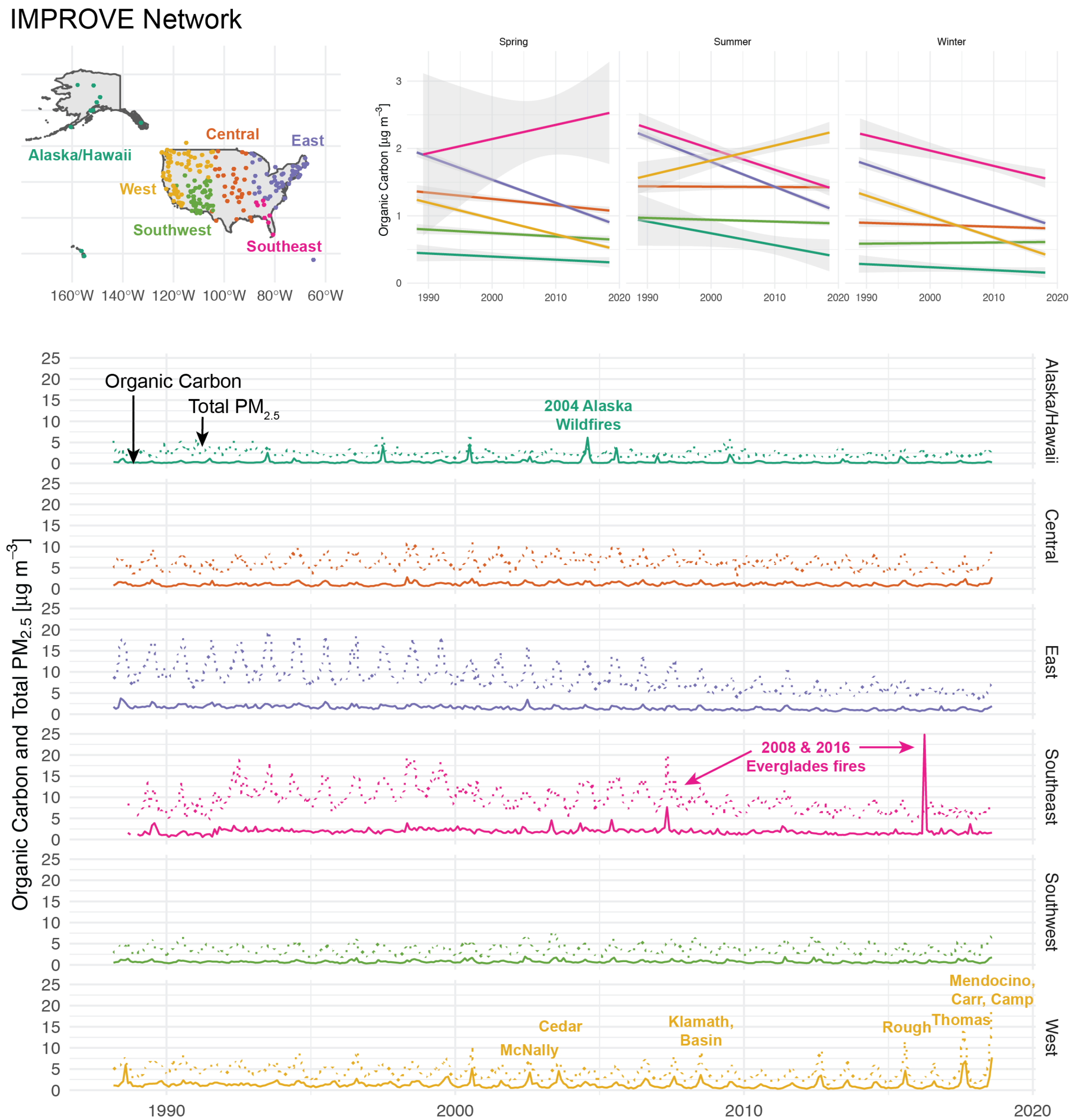

Figure S1: **Regional trends in organic carbon (OC) particulate concentrations**. **a** Locations of IMPROVE stations measuring organic carbon and total $PM_{2.5}$. **b**. Top right panels show seasonal trends by region show increasing OC in the Southeast in the spring, and in the West in the summer, consistent with observed fire activity. **c** Bottom plots show time series of OC and total $PM_{2.5}$ by region; spikes in OC are coincident with large fire events in most regions.

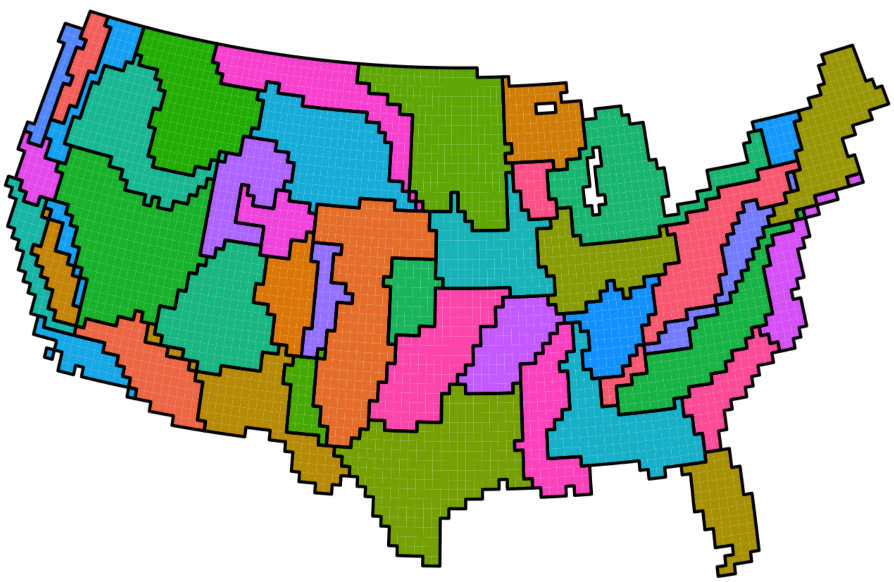

Figure S2: **The adapted USGS physiographic sections used in the statistical smoke model.**

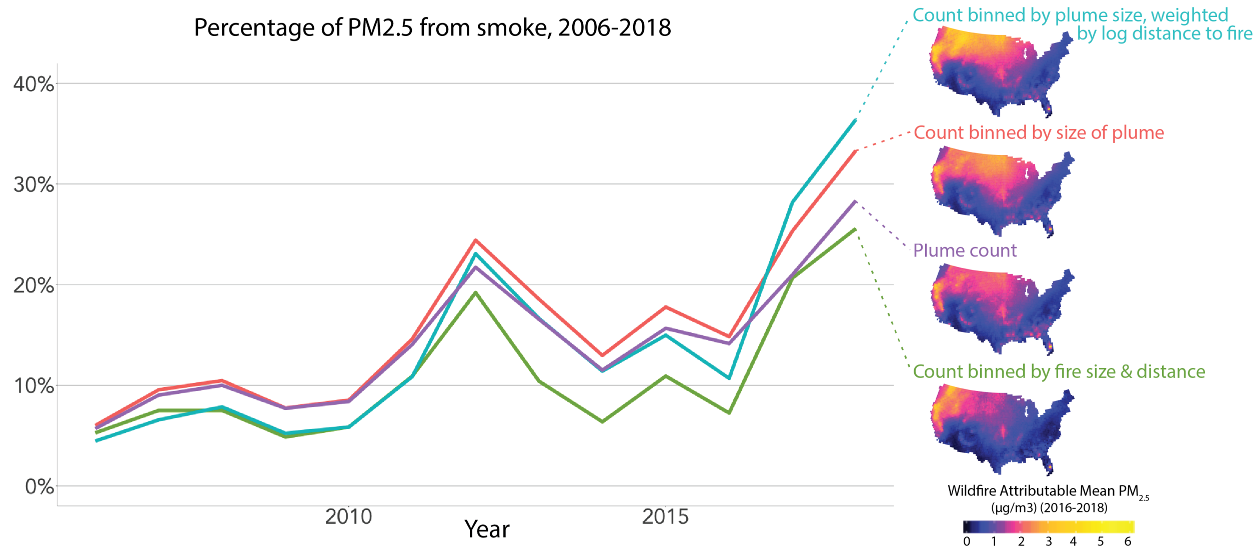

Figure S3: **Statistical model is robust to alternate approaches to incorporating fire and smoke data**. Our baseline model associates each plume with its likely source fire, and allows the impact of that plume on station $PM_{2.5}$ to differ as a function of the size of the source fire and the distance of the source fire from the location of interest. Estimates from this baseline model are similar but slightly more conservative than estimates from simpler models that model smoke simply as a function of the count of overhead plumes (purple line) or the counts of overhead plumes by size (red line), and more conservative than a model that uses plume size and allows distance to fire to enter non-linearly (blue line). Spatial patterns across models are also similar, although our baseline model attributes less smoke to the US midwest (despite that region having a high number of overhead plumes), consistent with these plumes being from distant fires and their smoke being higher in the column.[12]

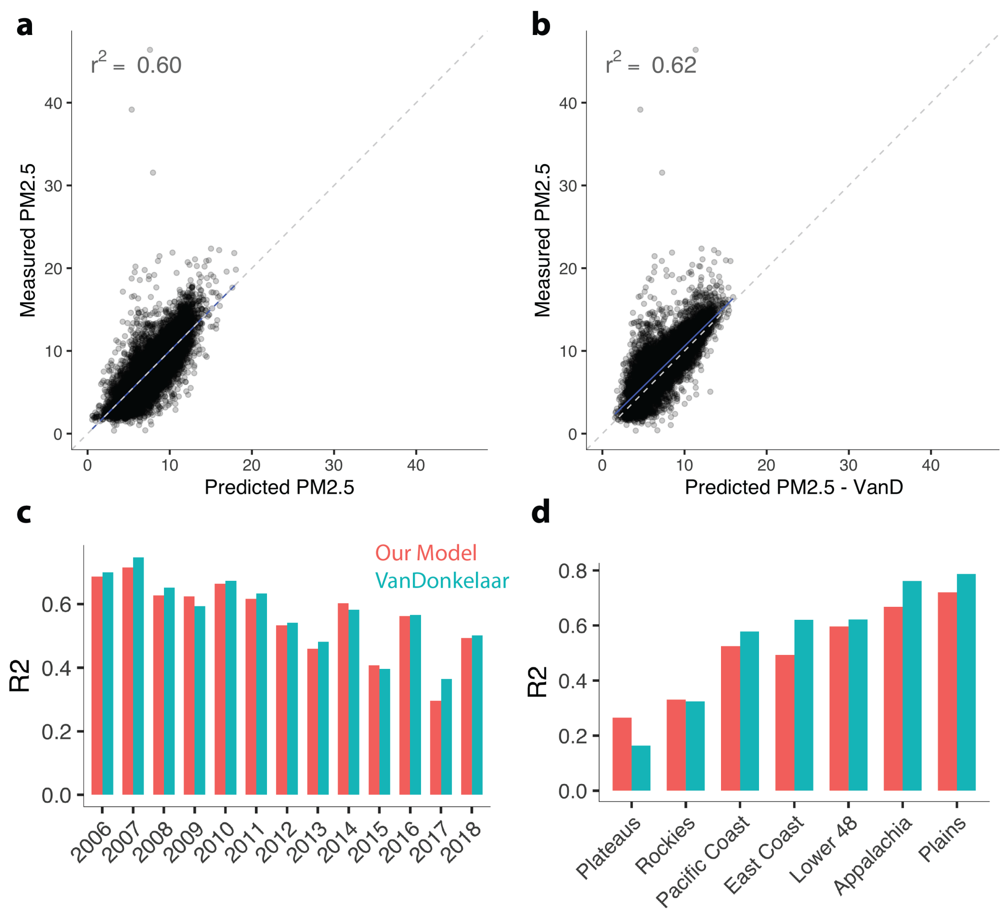

Figure S4: **Comparison of model performance in predicting overall $PM_{2.5}$ to existing benchmarks**. Our statistical model shows similar predictive performance to benchmark remote-sensing/statistical approaches ("van Donkelaar",[42]) in predicting variation in $PM_{2.5}$ from remote data. **a** Comparison of predicted versus observed $PM_{2.5}$ from our model, evaluated on US monitoring stations on which model was not trained; each dot is a station-month. **a** Comparison of observed versus predicted $PM_{2.5}$ from van Donkelaar, evaluated on US monitoring stations; these stations contributed to model development in van Donkelaar.[42] **c-d** Predictive performance of our model and the van Donkelaar model across years and regions in the US show very similar patterns of performance.

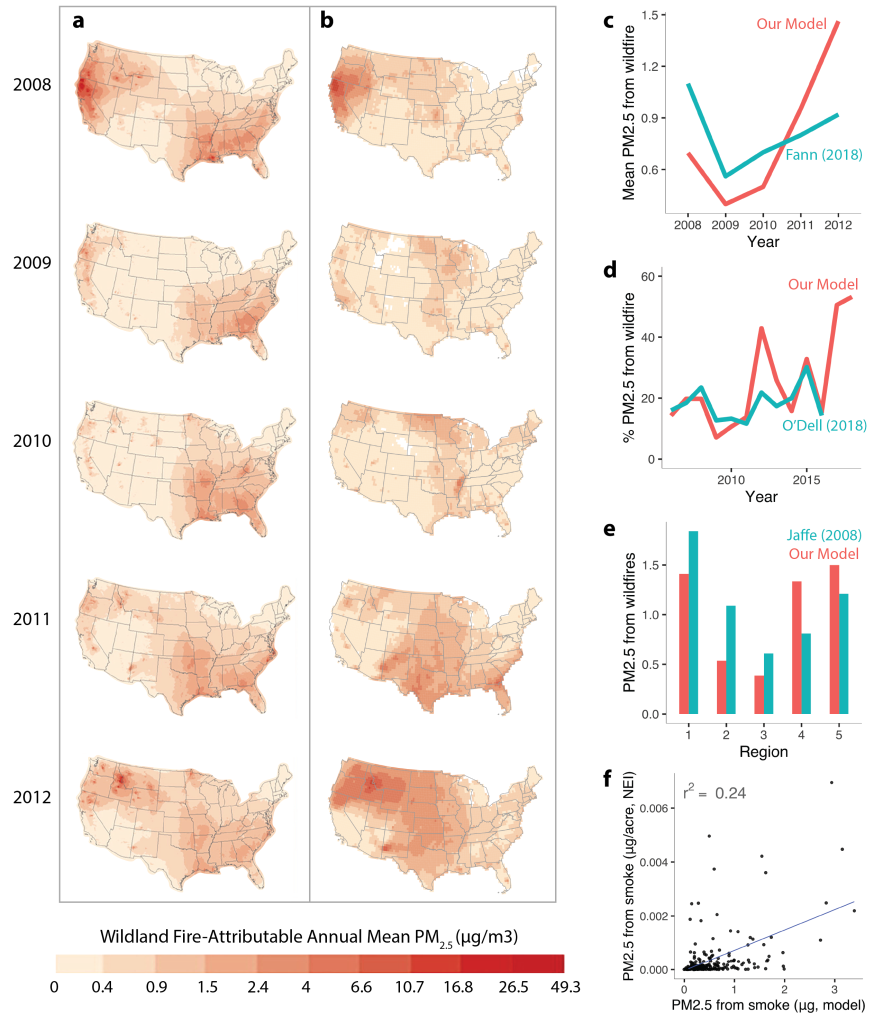

Figure S5: **Comparison of our estimates of attributable $PM_{2.5}$ from smoke to other estimates. a-b** Comparison with estimates in ref[44] for 2008-2012 show similarities in both spatial and temporal patterns of wildfire-attributed annual mean $PM_{2.5}$. Our estimates for the Midwest appear higher, perhaps because we include plumes originating from fires outside the US, which is important in the upper Midwest which is estimated to receive between 30 and 60 percent of smoke from outside the US.[12] **c** Comparison with estimates in ref[13] for the Pacific Northwest. O'Dell reports percentages for the July, August, September season, while our estimates are annual. To create a comparable number we assume that there is 0 smoke $PM_{2.5}$ in other seasons, and calculate (smoke PM)/(total summer PM + 3*summer non-smoke pm). **e** Comparison with ref,[45] who generate estimates of 5 regions in the Western US over the years 1988-2004. Comparison is imperfect as our estimates are for 2006-2018, but estimates follow a similar regional pattern. Even though burned area has grown substantially, our estimates are more conservative than Jaffe et al's estimates in three regions; our estimates are higher in Regions 4 and 5 (California, Oregon, Washington) where burned area has grown most rapidly . **f** Comparison with National Emissions Inventory estimates of primary $PM_{2.5}$ emitted from wildfires, aggregated to the state-year level for 2014 and 2017.

Table S1: **Sensitivity of mean estimated % $PM_{2.5}$ from wildland fire smoke upon removal of different covariates from statistical model.** Overall estimates are not sensitive to removal of individual co-variates, but change when all co-variates are removed and change substantially when fixed effects are also removed, as expected.

None	Coal	Cattle	Mining	Traffic	All	All+FE
13.5%	13.7%	13.3%	13.4%	13.5%	10.1%	2.1%

Table S2: **Sensitivity of mean estimated % $PM_{2.5}$ from wildland fire smoke using observed total $PM_{2.5}$ instead of predicted total $PM_{2.5}$ in denominator.** "Observed" total $PM_{2.5}$ is the average computed from monitoring data each monitoring station. The sample is thus limited in this comparison to the subset of locations in our data that have monitoring stations

Share using modeled $PM_{2.5}$	Share using observed $PM_{2.5}$
10.2%	8.5%

Table S3: **Model performance across different methods of quantifying $PM_{2.5}$ from smoke**. We compare overall model performance (as measured by r^2 between observed $PM_{2.5}$ and predicted $PM_{2.5}$) between a model with no smoke or fire variables but all other inputs, and a model that includes those inputs plus the smoke and fire variables, coded in different ways. Performance is evaluated on held-out stations on which the model was not trained. Our baseline model that uses counts of plumes by fire size and distance explains an additional 6% of variation in observed $PM_{2.5}$ relative to a model with no smoke variables. Alternate models that also use fire distance variables perform nearly as well, and simpler models that use only counts of plumes perform less well.

Binned by fire & distance	Weighted by log distance to fire	Binned by size of plume	Raw count
7.1%	7.0%	5.7%	5.5%