



Article A Coupled Wildfire-Emission and Dispersion Framework for Probabilistic PM_{2.5} Estimation

David Melecio-Vázquez^{1,*}, Chris Lautenberger², Ho Hsieh¹, Michael Amodeo¹, Jeremy R. Porter^{1,3}, Bradley Wilson¹, Mariah Pope¹, Evelyn Shu¹, Valentin Waeselynck⁴ and Edward J. Kearns^{1,*}

- ¹ First Street Foundation, 32 Bridge St. Floor 3, Brooklyn, NY 11201, USA; ho@firststreet.org (H.H.); mamodeo@firststreet.org (M.A.); jeremy@firststreet.org (J.R.P.); bradley@firststreet.org (B.W.); mariah@firststreet.org (M.P.); evelyn@firststreet.org (E.S.)
- ² Reax Engineering Inc., 1921 University Avenue, Berkeley, CA 94704, USA; lautenberger@reaxengineering.com
- ³ Department of Sociology and Demography, City University of New York Graduate Center, 365 Fifth Avenue,
 - New York, NY 10016, USA
- ⁴ Spatial Informatics Group, 2529 Yolanda Ct., Pleasanton, CA 94566, USA; vwaeselynck@sig-gis.com
- Correspondence: dmeleciovazquez@firststreet.org (D.M.-V.); ed@firststreet.org (E.J.K.)

Abstract: Accurate representation of fire emissions and smoke transport is crucial for current and future wildfire-smoke projections. We present a flexible modeling framework for emissions sourced from the First Street Foundation Wildfire Model (FSF-WFM) to provide a national map for near-surface smoke conditions exceeding the threshold for unhealthy concentrations of particulate matter at or less than 2.5 μ m, or PM_{2.5}. Smoke yield from simulated fires is converted to emissions transported by the National Oceanic and Atmospheric Administration's HYSPLIT model. We present a strategy for sampling from a simulation of ~65 million individual fires, to depict the occurrence of "unhealthy smoke days" defined as 24-h average PM_{2.5} concentration greater than 35.4 μ g/m³ from HYSPLIT. The comparison with historical smoke simulations finds reasonable agreement using only a small subset of simulated fires. The total amount of PM_{2.5} mass-released threshold of 10¹⁵ μ g was found to be effective for simulating the occurrence of unhealthy days without significant computational burden.

Keywords: fire behavior model; smoke transport model; dispersion; air quality; particulate matter; fire emissions

1. Introduction

Wildfires are a significant contributor (15–30%) to atmospheric fine particulate matter (PM_{2.5}) pollution in the United States [1] with projected increases in the future under climatic changes that are favorable for wildfire activity [2,3]. In recent years, wildfire smoke has been a dominant contributor to adverse air quality events, particularly in the western US where large fires burning for extended periods of time have placed millions of people under hazardous air quality levels [4]. Particle emissions from wildfires are also known to travel long distances as evidenced by the fact that wildfires originating in the western US and southern Canada can drastically elevate PM_{2.5} levels in the northeastern US [5,6]. Globally, death estimates from particulate matter have been reported to be 3–4 million [7]. Increasing levels of wildfire smoke are expected to reach emission levels comparable to or greater than the improvements made in decades of anthropogenic emission reductions in the United States [8,9], which also carries significant health impacts. Exposure to fine particulate matter negatively impacts both the cardiovascular and respiratory systems, including causal associations with mortality (both short- and long-term), ER visits, and hospital admissions, among others [10–12]. In an analysis of six U.S. cities, a reduction in the concentration of 10 μ g/m³, on the order of concentration difference between cities, could result in about 36,000 fewer deaths per year [12]. Improved tools for the better characterization of wildfire smoke emissions, now and under future climate conditions, can aid in serving communities worldwide and is, therefore, a critical area of research.



Citation: Melecio-Vázquez, D.; Lautenberger, C.; Hsieh, H.; Amodeo, M.; Porter, J.R.; Wilson, B.; Pope, M.; Shu, E.; Waeselynck, V.; Kearns, E.J. A Coupled Wildfire-Emission and Dispersion Framework for Probabilistic PM_{2.5} Estimation. *Fire* **2023**, *6*, 220. https://doi.org/ 10.3390/fire6060220

Academic Editor: Kevin Tansey

Received: 4 May 2023 Revised: 28 May 2023 Accepted: 29 May 2023 Published: 31 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). Existing studies on wildfire smoke response under climate change projections for the 21st century are generally consistent in showing increased $PM_{2.5}$ or its constituent components (e.g., surface organic carbon) on the order of ~50% [13–16]. However, standard climate-scenario output from global climate models relies on fire emissions that do not adequately represent shifts in wildfire activity due to changes in climate and may consequently underestimate $PM_{2.5}$ concentrations into the future [8,13,17]. Alternative approaches to modeling current and or future wildfire smoke activity include fire-process models, statistical regressions, or offline chemical transport models, potentially coupled to historical or climate-projected meteorological fields. These approaches find largely similar trends of increasing wildfire smoke but have more limited representations of dynamic climate feedback and anthropogenic influences. Combining an accurate representation of fire emissions and smoke transport with a climate-sensitive wildfire behavior model is an active area of research (see examples in [18]).

Treating wildfires as emission sources with their distinctive heat- and mass-release variability is an important step for modeling the transport of smoke pollutants. To this end, NOAA's HYSPLIT model [19–21] is commonly used to model the transport and dispersion of particulate matter from wildfire smoke because it captures both particle trajectories and changes in pollutant concentrations. A widely available version of the software [22] can be used via the READY website https://www.ready.noaa.gov/HYSPLIT_traj.php (accessed on 30 May 2023) for trajectory calculations. HYSPLIT has proven to achieve good estimates of wildfire smoke transport, given appropriate input emission data sets, meteorological input data, and plume-rise and mixing layer parameterizations [23]. As is the case with HYSPLIT and other long-range transport models, simulating dispersion from wildfire emissions is sensitive to a variety of factors including (1) the source emission rates, (2) the smoke plume rise due to the energetics of the fire, (3) the transport of smoke by the mean wind, (4) and chemical interactions with the atmosphere after emission [24]. Although the last item is possible within HYSPLIT using linear mass conversion formulations, it is most useful for ozone studies [24], and therefore, not a component explored in this study. Throughout its development cycle, HYSPLIT has made significant improvements toward better representation of the plume-rise schemes [24–26].

Methods for estimating the emission rates from fires range from statistical methods [15,16,27] to process-based methods [28–30]. The ELMFIRE fire behavior model [31] is an open-source process-based type and can be adapted for either climate change impact studies [17] or for operational use [32], as seen in https://pyrecast.org (accessed on 30 May 2023). The advantages of using a process-based approach rather than a statistical approach include the ability to estimate impacts under a changing climate that is producing new environmental conditions for which we have no historical observations; it allows probabilistic estimates of future conditions to be built that can incorporate geographically and temporally variable wildfire suppression activities and fire ignition locations and enables statistical comparisons of future conditions with historical wildfire events and associated losses. Purely statistical estimates of future risks from wildfire and smoke may not be able to capture the changes in either the natural or the built environment at a resolution and fidelity that is adequate for informed decision-making. Additionally, using models in tandem with each other in a "modelling chain" [33,34] is a common tool where the strengths of multiple models can be combined in a computationally efficient approach. Towards this goal, we propose a modeling framework that integrates the outputs from the First Street Foundation Wildfire model (FSF-WFM) with deterministic smoke transport modeling via HYSPLIT transport and dispersion to estimate probabilistic smoke concentrations across the continental United States (CONUS).

2. Materials and Methods

We present a flexible framework for evaluating present and future air quality conditions from wildfire emissions across CONUS. The chemical transport framework brings together two main components, (1) a wildfire behavior model that is described in detail in [17], and (2) deterministic air dispersion modeling via HYSPLIT that estimates smoke transport and mixing for individual fires. In the following sections, we step through the development of the framework, discuss how emissions from the FSF-WFM are simulated, and how the emissions output is configured as input for the HYSPLIT air dispersion model. We then discuss how we sample from the FSF-WFM to estimate the number of days that have the potential to have adverse air quality conditions from the HYSPLIT outputs.

2.1. First Street Foundation Wildfire Model

In this research, we use a probabilistic estimate of current wildfire risk across the contiguous United States that was produced by the First Street Foundation and the Pyregence Consortium [17]. The property-specific and climate-resolving wildfire risk model was created at 30 m horizontal resolution using open government data such as the USFS' LANDFIRE fuels database and NOAA weather records to drive the ELMFIRE fire behavior model [31] in a series of Monte Carlo simulations. Over 100,000,000 simulations were conducted for each target year, based on historical ignition locations, to create a probabilistic estimate of the likelihood of wildfire, the mean and maximum expected flame lengths, and the likely exposure to flying embers for each 30 m across the landscape. Fuel types within the Wildland Urban Interface were approximated using 500 historical wildfires to allow the simulations to also provide risk estimates to properties within those inhabited areas. Of the 100 million fires that were ignited in the FSF-WFM, only 65 million fires reached the minimum fire-size threshold of 0.04 km² (10 acres) and were thus used to derive fire probability. From these fires, 20,000 randomly chosen fires became the starting point for the consideration of the impact of combustion and emission processes on air quality in CONUS.

2.2. Wildfire Smoke Emissions Output from FSF-WFM

We start with the underlying continuity balance for the combustion process simulated by the FSF-WFM wildfires. The relation between the generation rate of combustible gases $(\dot{m}, \text{kg/s}, \text{often referred to as "mass loss rate"})$ and heat release (\dot{Q}, W) is

$$Q = \dot{m} \,\Delta H_c,\tag{1}$$

where ΔH_c (J/kg) is the effective heat of combustion of biomass. In a bomb calorimeter which uses pure oxygen and high pressure to ensure complete combustion, the ΔH_c for biomass is approximately 18 MJ/kg, but for combustion of biomass in air, this value is closer to 12 MJ/kg.

The emission rate of major and minor species is usually estimated using a species yield. Smoke (or soot) yield is the fraction of a fire's mass loss rate that is converted to carbonaceous soot, and CO yield is the fraction of its mass loss rate that is converted to carbon monoxide. The negative health impacts of particulate matter, both fine ($\leq 2.5 \mu$ m) and coarse ($\leq 10 \mu$ m) are well-known in the literature, and in this study, we focus on the transport of fine particulate matter (PM_{2.5}) from wildfire smoke. The methods shown here can be applied to any constituent of wildfire smoke given that the appropriate (1) yield characteristics for the fire model are given and (2) the constituent characteristics are also captured in the HYSPLIT configurations (discussed later). The release rate of PM_{2.5} (m_s) is a function of PM_{2.5} smoke yield (Y_s) which is estimated from [35] by fuel type, as shown in Table 1, and defined as

$$\dot{m}_s = Y_s \dot{m} = Y_s \frac{Q}{\Delta H_C}.$$
(2)

It is seen that the smoke generation rate is a function of $Y_s/\Delta H_c$ and heat release rate, with the latter calculated by the fire spread model as described in [17].

Fuel Type	$PM_{2.5}$ Yield (Y_s)
Grass	0.54%
Grass/shrub	0.93%
Shrub	0.93%
Timber understory	1.3%
Timber litter	1.3%
Slash/blowdown	1.3%
Canopy	1.3%

Table 1. PM_{2.5} yield from the total smoke yield for different fuel types. Values from [35].

In the fire-spread model, the 30 m × 30 m pixel's time of ignition (t_{ig}) is determined by the spread model. For smoke generation purposes, each pixel burns for a duration t_{burn} , estimated as a pixel's edge length Δx divided by the fire's spread rate as it traverses that pixel (V, also calculated by the fire spread model) is

$$t_{burn} \approx \Delta x / V.$$
 (3)

We note that t_{burn} models the progressive heat release of a pixel, which should not be confused with the residence time of a fire, or how long a point location is covered by the fire front. A pixel burns and releases heat for times greater than t_{ig} and times less than $t_{ig} + t_{burn}$. A burning pixel's heat release rate per unit area (\dot{Q}'' , W/m²) is estimated from fireline intensity (w/m, calculated by the fire spread model) which includes a contribution from surface fuels (I_s) and canopy fuels (I_c) defined as,

$$\dot{Q}'' \approx \frac{I_s + I_c}{\Delta x}.$$
 (4)

Equation (4) captures the increase in heat (and therefore, smoke) production for rapidly spreading fires or for fires burning through areas of heavy fuel loading because fireline intensity increases with fuel loading and spread rate.

With Q'' determined for each burning pixel, a fire's total heat release rate is calculated by summing the heat release rate per unit area over all burning pixels and multiplying by the (fixed) area of each pixel. However, since smoke generation is typically output at hourly intervals but pixels generally burn only for seconds or minutes; the heat release rate is weighted by the fraction of the preceding hourly interval during which each pixel was burning. Therefore, each smoke generation rate represents the average over the preceding hour (or more generally, the smoke output interval). Note, that the preceding approach does not account for smoldering or post-frontal combustion; only emissions from the flame front are considered.

2.3. HYSPLIT Runtime Configurations

To model smoke dispersion, we use the NOAA ARL HYSPLIT model version 5.2.0 with input meteorological data spanning 2011–2020 over CONUS. From the list of publicly available meteorological input models in the NOAA ARL archive, the 12-km North American Mesoscale model (NAM) was chosen due to having the highest resolution in the archive for the period. The start dates of the 20,000 fires were selected from a much longer list of over 150 million ignitions that were created as part of the full-fire Monte Carlo simulation that was described in [17]. The 20,000 subset of fires used for smoke analysis were randomly taken from a subset of the 150 million possible fires that both achieved ignition in the model and also produced a fire size that was above a threshold value. For those 20,000 selected fires, the start dates were noted and the corresponding NAM time series beginning on those dates were used to drive the HYSPLIT model for the smoke simulation. This was conducted to ensure that the same fire weather that successfully drove the fire behavior simulation (using NOAA RMTA surface weather records) is also driving the smoke simulation (using the NAM multi-level weather records).

The HYSPLIT configuration details common to each fire emission event modeled are summarized in Table 2. The output concentration grid resolution was set to a $0.1^{\circ} \times 0.1^{\circ}$ grid centered at 38° N 97° W (approximately the center of CONUS), with a latitude extent of 40° and a longitude extent of 80°. Although not all fires will emit enough to cover all of CONUS, this setting ensures that all the individual output grids can be combined. Moreover, although the grid setting is large, HYSPLIT output is only saved for non-zero concentration levels, therefore, making the storage needs per fire low. The near-surface concentration layer that is analyzed is restricted to the 0–500 m above-ground-level (agl) so that concentrations are calculated on $0.1^{\circ} \times 0.1^{\circ} \times 500$ -m cells across CONUS. This region encompasses the full spectrum of heights where people in CONUS reside and where fine particles are known to be mostly uniformly distributed [36].

Details/Options	Title 3
Emission dataset	ELMFIRE Fire emissions; FINN emissions factors [35]
Plume rise scheme	Briggs [25]
Meteorology Inputs	NAM (12 km)
Mixing layer depth options	Met. model mixing layer height
Vertical Motion options	Met. model vertical velocity
Particle release mode	Full 3D vertical and horizontal
Number of Particles per cycle	10,000
Maximum number of particles	1,000,000
Horizontal resolution	0.1°
Surface concentration layer	0–500 m
Domain lat. center/extent, long. center/extent	38° N/40°, 97° W/80°
Model top	10 km (agl)
Average diameter and density of the particles	0.8 mm and 2 g/cm
Minimum mixing layer depth	150 m

Table 2. HYSPLIT runtime configurations for all fires.

HYSPLIT also contains options for different particle/puff release modes. Here, we use a full 3-D particle release mode, with a maximum allowable number of particles, MAXPAR, set at 1,000,000 which is large enough so that all particles being released are tracked throughout the life of the emission. The number of particles released in each cycle, NUMPAR, must also be chosen for each fire, and the final choice also directly impacts the computational cost (along with the resolution of the concentration domain. Figure 1 shows the 35 μ g/m³ (concentration for orange+ days) contour at a single time-step when NUMPAR values of 1000, 2500, 5000, and 10,000 are chosen. The patterns are similar across all four settings. The lower the number of particles the more mass each particle has, and therefore, the final concentration calculation may look noisier. As the NUMPAR is increased, the mass fraction per particle decreases and a smoother shape begins to appear. The particle characteristics of diameter and density are set the same as those used in [23], which are 0.8 mm and 2 g/cm, respectively, for PM_{2,5}. The chosen NUMPAR value was set at 10,000 for every HYSPLIT run.



Figure 1. Example showing 35 μ g/m³ contour from a HYSPLIT run. Shown here is a time step from a single fire event estimated at NUMPAR values of (**a**) 1000, (**b**) 2500, (**c**) 5000, and (**d**) 10,000.

2.4. HYSPLIT Emissions Input

The ELMFIRE output for each individual fire is a time-series of the emissions from the flame front. As the flame front moves, the emission characteristics change accordingly. As mentioned in Section 2.2, even though the focus of this study is on the transport of $PM_{2.5}$, appropriate changes to the configuration in Table 2 to the average diameter and density of the particles can adapt the application of this methodology to any constituent of wildfire smoke. The emission characteristics saved are the center latitude and longitude, the fire size in square meters, the rate of mass release in $\mu g/h$, and the rate of heat release in Watts. For the HYSPLIT input files, each fire characteristic is aggregated into an emission cycle. Here, an emission cycle is defined as a period of continuous emissions from the wildfire. In any given fire there are periods where the fire is still active but not emitting any smoke. These breaks in emissions occur when the fire spread rate is temporarily reduced due to increases in fine fuel moisture or during overnight hours when the use of a burn period precludes spread. These breaks are thus used to create the emissions cycle for each individual fire. Each emission cycle will, at most, be 12 h for a dry, daytime period. The lifetime of any fire was set to not exceed 7 days, the HYSPLIT runs were configured to continue 5 more days after the last emission release, for a max total model simulation time of 12 days.

The emission characteristics for each cycle in a single HYSPLIT run are defined using statistical summaries of the fire characteristics for that given cycle. For each emission cycle the location, area, mass released, and heat released is given, respectively, as the initial location of the cycle, the max area of the cycle, the mean of the mass-released (μ g/h) and the mean of the heat-released (W). The total duration of the HYSPLIT runs is set as the maximum duration of the fire (including when not emitting) and an additional 5 days to track the particle movements after the final emission cycle. Throughout the life of the fire (from burn start to the extinguishing of the flames) HYSPLIT tracks all the particles released since the start of the simulation.

2.5. Metrics and Sampling Methods

Ideally, HYSPLIT would be run on every single fire produced by FSF-WFM, but the computational expense of running HYSPLIT millions of times prohibits this. With the goal of reproducing an accurate distribution of smoke concentrations that is stable across our metric(s) of interest, we evaluated a subsampling methodology on randomly chosen fires. For this study, our target metric is $PM_{2.5}$ concentrations that reach an "unhealthy for sensitive groups" or above categorizations on the US EPA Air Quality Index (AQI), or a daily average $PM_{2.5}$ concentration greater than 35.4 µg/m³ [37]. This level is designated the color "Orange" on the AQI color scale. We use the term "orange+ days" to refer to when our model predicts daily conditions in which the daily $PM_{2.5}$ concentration is greater than 35.4 µg/m³. The AQI provides overall guidance on whether action is needed to curtail adverse conditions. The Air Quality Index (AQI) is also required to be reported daily in metropolitan areas that have a population of more than 350,000.

Concentration outputs were saved for each of the fires and the Python package MON-ETIO [38] was used to aggregate the hourly concentration ($\mu g/m^3$) data into 24-h mean concentrations to match the thresholds in the AQI formulation mentioned above. The start of the 24-h period begins at the first emission hour and ends 23 h later. For each individual simulation, the 24-h mean concentration is converted to an AQI value. All simulations are then aggregated to create the annualized expected number of days, at or above AQI "orange" levels, or "orange+ days" conditions.

Let *P* (i.e., A "point" or "pixel" in the HYSPLIT domain) be a location of interest; the quantity we aim to estimate is the Average Expected Oran ge+ Days (in days/year) defined as

$$AEOD_P := \frac{1}{S} \sum_{i=1}^{N} a_{Pi} = \frac{N}{S} \frac{1}{N} \sum_{i=1}^{N} a_{Pi},$$
(5)

where *S* is the number of simulated years, *N* is the number of simulated fires, *i* is an index that spans the set of simulated fires, and a_{Pi} is the number of orange+ days caused by fire

i at the location of interest (most often $a_{Pi} = 0$). For the results presented in this article, $S \approx 1.136 \times 10^6$ yr ($1/S \approx 8.8 \times 10^{-7}$). We see from Equation (5) that computing $AEOD_P$ reduces to computing the average

$$A_p := \frac{1}{N} \sum_{i=1}^{N} a_{Pi}.$$
 (6)

Computing A_P exactly would require running HYSPLIT on millions of fires, which would be prohibitively expensive. Instead, we estimate A_P using random subsampling as

$$A_P \approx A'_P := \frac{1}{n} \sum_{k=1}^n a_{Pi_k} \tag{7}$$

where *n* is the sub-sample size, and i_k is the index of the fire which was randomly selected as the k-th element of the sub-sample. Because the sub-sampling is random and treats all fires as exchangeable, basic probability theory tells us that A'_p can be viewed as a random variable, with an expected value equal to A_p and variance equal to v_P/n , where v_P is the variance of a_{Pi} when randomly choosing *i*. In particular, the variance of A'_p tends to zero as *n* is made large, justifying the approximation. The benefit of this is that we need to only run *n* fires through HYSPLIT instead of *N*. In this study, we demonstrate this approach with n = 20,000 fires or smaller. For computational convenience, the results shown in this study were derived only from fires simulated in the months (March, April, July, and August) which presents a bias; there would be no major difficulty in applying the computation to all 12 months, thus making the sub-sample truly representative. As shown later, these values of *n* were also empirically justified by observing the convergence of A'_p for increasingly large *n*.

Finally, we apply one last computation-saving approximation: thresholding. That is, we choose a predictor variable x_i which tends to correlate strongly with a_{Pi} , and estimate $a_{Pi} \approx 0$ when x_i is below a well-chosen threshold t_x , yielding

$$A'_{P} \approx A''_{P} := \frac{1}{n} \sum \{ a_{pi_{k}} | x_{i_{k}} > t_{x}, k = 1 \dots n \}.$$
(8)

This means that we will be able to run HYSPLIT on a number of fires that is much smaller than *n*. As will be explained in the detailed analysis later, we found that the total mass-released \dot{m}_{sum} (from the sum of every hourly rate of emission, \dot{m} , for each fire) to be a suitable predictor variable x_i , by observing the correlation of various candidate metrics x_i to the total count of orange+ days.

3. Results

Within the preliminary sample of 20,000 fires simulated in HYSPLIT, we first identified that the total mass released throughout the fire had the highest correlation (0.95) with the number of orange+ days over CONUS, as shown in Figure 2. We find that fires with a total mass-released less than 10¹⁵ µg do not substantially contribute to producing orange+ days throughout CONUS. As seen in Figure 3, approximately 75.8% of fires fell below this total mass-released threshold, with the remaining fires producing 92% of the pixels that produced orange+ days conditions. This suggests that the smoke that impacts AEOD across CONUS is driven primarily by larger fires that release more mass. We then examined how the AEOD changed as a function of the number of samples chosen, *n*. Figure 4 shows the distribution of AEOD(n) at two different locations to show how AEOD at any location of interest converges as the sample size is increased. In order to show the spread of AEOD for a fixed sample size, AEOD was calculated ten times for each sample size, n. Additionally, Figure 4 compares the convergence between a random selection of n fires to the fires that have the highest (green points) and lowest (orange points) values for AEOD. From the total number of 20,000 fires that were simulated, we find the variance of AEOD(n) stabilizes quickly as *n* increases, indicating that the final sample size can be less than the total number of fires and keep the error acceptably low, as will be shown next.



Figure 2. Distribution of fire emission characteristics of (**a**) total mass-released by the fire event, \dot{m}_{sum} , (**b**) maximum mass-release rate, (**c**) maximum fire area. Orange+ days pixels occur when the daily average PM_{2.5} concentration exceeds 35.4 µg/m. Correlation coefficients are 0.95, 0.89, and 0.83 for (**a**–**c**), respectively.



Figure 3. Empirical CDF of the total mass-released per fire. The red line is from the distribution of the total mass-released per fire and the orange line is the distribution total mass-released weighted by the fires that produce orange+ days.



Figure 4. Two representative distributions of AEOD(n) show the convergence of AEOD(n) as sample size increases, where (**a**) reaches convergence quicker than (**b**). The points in blue are estimations of AEOD where a fixed n is randomly sampled ten times to show the spread of values at each *n*. The green and orange points are the values of AEOD when n is chosen from the highest and lowest values of AEOD from the set of all HYSPLIT simulations ranked by the number of orange+ days generated at that location. The map in (**a**) shows the location of the pixel for (**a**,**b**).

To evaluate the impacts of different sub-samples across CONUS, we first applied the 10^{15} µg total mass-released threshold to the full set of 20,000 fires simulated in HYS-PLIT, leaving approximately 4000 fires. We then drew two additional samples based on a representative sampling calculation indicating that 3220 fires are representative of the larger population of fires based on highly restrictive parameters of an acceptable margin of error of 1% with a 99% confidence level of that margin of error [39], and 351 fires are representative under a 5% margin of error and a 95% confidence interval. The resulting maps of modeled AEOD for the 4000, 3220, and 351 number of fires and the differences between the 4000-count and 3220-count, and the 4000-count and 351-count are shown in Figure 5. We find that both sub-samples strongly reflect the distribution derived from the total number that meets the mass-threshold requirement. In general, the spatial patterns of Figure 5a are maintained in the results of the other representative samples in Figure 5b,c. The differences in the results shown in Figure $5d_{e}$ show the worsening of the errors when the sample size is decreased, as is to be expected. The northeast region is the least to show any occurrence of orange+ days conditions, and therefore, the difference plots in this region have very similar magnitudes. The largest absolute difference in performance between these two samples is regionally similar across sample sizes. On the lower end, the regions that show at least 1 day of orange-days seem to extend further with the more stringent sample-size selection criteria. Given the vastly smaller number of fires, there is also the potential for local maxima to appear more in the final output. The differences among the absolute maximum of these different options are very minor, with values of 11.8, 11.9, and 10.8 days for the total number of significant fires, the sample selected from the 99% CI, and 95% CI, respectively.

A comparison of the AEOD values estimated by our model against the historical data of wildfire smoke-only days from [40] is shown in Figure 6. The maximum number of days that can be expected to be experienced anywhere in CONUS is similar between the models 9 and 11 days for the historical (Figure 6b) and modeled (Figure 6a) data, respectively. The historical aggregate shows three different hotspots of orange+ days, one of which is not well-represented in our model. Spatially, the modeled AEOD against the year-to-year

variation in the Midwest region is comparable given that this region is always represented with at least some orange+ days conditions every year of the [40] dataset. The signal from the Idaho and Montana regions are consistent in both the model and the historical data. The feature in the northwest in Figure 6b is not captured by the model as seen in Figure 6a. This is due to the lack of representation of fires from outside of the CONUS region. This region is known for being heavily affected by smoke originating from fires outside of CONUS. In a climatology study of PM_{2.5} distributions over CONUS, Ref. [13] found that at least 50% of all PM_{2.5} may be attributed to fires outside of CONUS. The wide band of AEOD values greater than 5 days in the historical data is also consistent with the results from [13] regarding the effect of non-CONUS fires. The southeast region is a region where there is significantly less overlap between the historical and modeled data. The region has a small but still notable number of orange+ days in 2011, 2012 and 2016. The historical occurrences of orange+ days in the center region CONUS [41] may not always correspond with the modeled AEOD. However, both results, those in Figure 1 from [41] and from Figure 5a here, show a coherent picture of negative air quality impacts from wildfire smoke.



Figure 5. AEOD calculated from (**a**) the fires remaining after applying the threshold on the total mass-released, and from a random sample selection using a (**b**) 99% confidence interval (~3000 fires), and a (**c**) 95% confidence interval (~350). The difference between the (**a**,**b**) is shown in (**d**), and the difference between (**a**,**c**) is shown (**e**). (**a**–**c**) have maximum values of 11.81, 11.90, and 10.79 days, respectively. The distributions shown (**d**,**e**) range from ô0.64 to 0.64 and -4.35 and 3.55, respectively.



Figure 6. Comparison of the number of orange days estimated by the ELMFIRE-HYSPLIT model and historical data from [40]. (**a**) is the modeled AEOD (same values as Figure 5a), (**b**) average yearly number of days averaged over 2011–2020 period from [40], (**c**–**I**) are the individual years of the [40] data. The maximum number of days for each of the panels is (**a**) 11, (**b**) 9, (**c**) 17, (**d**) 57, (**e**) 25, (**f**) 43, (**g**) 36, (**h**) 33, (**i**) 48, (**j**) 53, (**k**) 10, (**l**) 63 days.

4. Conclusions

From a technical perspective, we have presented the methodology and preliminary results from ELMFIRE-HYSPLIT integration and from a practical perspective, this report introduces a novel conceptual approach to understanding the differential ways in which air quality will change in the future. Currently, the state-of-the-art in understanding the observed risk of smoke pollution is to use $PM_{2.5}$ metrics and systematically remove the influence of non-smoke contributors (similar to [40]) to then estimate past exposure and model future exposure. While this approach is certainly useful it relies on statistical modeling. In addition to what we can glean from the statistical analysis, the model produces a dynamically derived set of smoke results which are tied directly to a dynamically modeled set of wildfire exposure. In combination, the dynamic and statistical models give us an understanding of what we have been exposed to in the past and how that may change in the future. This is important for both individual and community level responses to poor air quality days, today and into the future.

Regarding the models produced in the current report, the coupling between the current FSF-WFM [17] and HYSPLIT allows for a probabilistic fire emissions product that can be used to quantify the risk of adverse air quality conditions anywhere in CONUS, flexible to a given climate scenario. The integration between ELMFIRE and HYSPLIT reflects the advantages of modularity seen in the implementation of the BlueSky framework [42] for operational use. It is the attempt of this paper to take that same idea of modularity and flexibility to tackle challenges for estimating the probability of smoke conditions that can affect the CONUS region from all possible scenarios of fire occurrence. Adverse conditions defined using the AQI values also give context to the impact of smoke concentration levels produced by wildfire smoke. In this research, "orange+ days" conditions, where the AQI value represents an exceedance of a daily average concentration of 35.4 µg/m^3 , is the first level at which behavior change is recommended due to the impact of the concentrations of particulate matter.

The evaluation of model results focused on the comparison to the data provided from the study conducted by [40]. While a wide network of $PM_{2.5}$ observations is available, an attribution problem occurs as the source of the $PM_{2.5}$ is difficult to distinguish. The efforts conducted in [40], therefore, create a dataset of the $PM_{2.5}$ concentrations due only to wildfire smoke and thus became key to our evaluation of the ELMFIRE+HYSPLIT integration. As better methods for source attribution of $PM_{2.5}$ station data develop, future efforts can improve the evaluation methods of this and similar smoke dispersion models.

ELMFIRE's deterministic approach to fire spread rate is currently being applied to help provide guidance on the impact of climate change on fire probability and how individual properties may be affected in the contiguous United States [17]; here we present the extension of those capabilities to smoke dispersion. Efficient use of available computing systems also revealed that focusing exclusively on the larger fires (i.e., total PM_{2.5} mass-released of 10^{15} is about 20–25% of all potential fires produced by ELMFIRE) can yield magnitudes of overall AEOD that are comparable to the historical period. The implications of this are that the sampled fires provide a generalizable set of data for inferring the relationship between fire incidence and smoke exposure. Furthermore, the reliance on statistically derived samples of the data allows for the structuring of future analyses in an environment in which constraints on time and compute resources may be present.

Both models in this integration have several sensitivities that can be addressed with further development. Fire behavior models can be largely impacted by the length of the fire-burn, the representation of geographically variable suppression, and calibration per pyrome. These things influence large fires such as the ones we see having the most impact on smoke [17]. Smoke dispersion models are very sensitive to input meteorological data and the plume-rise scheme which can have a large impact on the vertical distribution of the smoke. All HYSPLIT simulations used the 12-km NAM model results for the horizontal and vertical distributions of the mean wind fields. However, in this work, there was a careful effort to match the specific weather conditions between the surface fire-weather data used to drive ELMFIRE and the smoke dispersion meteorological data used to drive HYSPLIT.

Wildfires are increasing across the world in areas that are impacted by increasing temperatures, drying environments, and other conditions conducive to the onset of wildfires. With forecasts indicating that wildfire smoke will continue to be a major contributor to hazardous air quality levels across the United States, this work contributes to the larger body of research [13,18,40,43–45] aimed at developing accurate probabilistic estimates of smoke conditions. Furthermore, the modeling framework presented here also fully supports smoke estimation under climate change scenarios that include effects on wildfire behavior. The ELMFIRE-HYSPLIT integration aids in the exploration of model uncertainties and as new ideas and experiments continue to be developed to improve current forecasting and decision-making support. A similar framework is possible outside the US if a model such as ELMFIRE with comparable inputs to those of FSF-WFM in Appendix A of [17] can be found, in addition to the details outlined here for the smoke dispersion model. The modeling framework presented here can also take advantage of the advancements to come in the future [46,47].

Author Contributions: D.M.-V.—conceptualization, methodology, air dispersion coding, validation, visualization, supervision, writing—original draft, writing—review and editing; C.L.—conceptualization, methodology, fire behavior model coding, validation, writing—original draft, writing—review and editing; H.H.—air dispersion coding, methodology, validation, visualization, writing—original draft, writing—review and editing; M.A.—methodology, validation, visualization, writing—original draft, writing—review and editing; J.R.P.—methodology, validation, writing—original draft, writing—review and editing; B.W.—validation, writing—original draft, writing—review and editing; M.P.—validation, writing—original draft, writing—review and editing; E.S.—validation, writing—original draft, writing—review and editing; V.W.—validation, writing—original draft, writing—review and editing; E.J.K.—conceptualization, methodology, validation, visualization, writing—original draft, and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Software used for the fire behavior is called ELMFIRE, developed by Chris Lautenberger (lautenberger@reaxengineering.com). Source code is publicly available at https://github.com/lautenberger/elmfire/ and carries an EPLv2 license. Software used for smoke dispersion of fire emissions is called HYSPLIT and the license used is custom by agreement. HYSPLIT is available via https://www.ready.noaa.gov/HYSPLIT_linux.php. HYSPLIT meteorological input datasets are available via the portals via https://www.ready.noaa.gov/archives.php. ELMFIRE input dataset details can be found in Appendix A of [17]. The resulting particulate matter hazard layers will be freely and publicly available for noncommercial use at https://riskfactor.com, (accessed on 3 May 2023). The public availability of this climate information is meant to inform the public, enable new research efforts on air quality impacts from wildfire smoke, level the playing field with private commercial interests that already have access to this kind of information, and help address the privatization of climate impact information.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Urbanski, S.P.; Hao, W.M.; Nordgren, B. The Wildland Fire Emission Inventory: Western United States Emission Estimates and an Evaluation of Uncertainty. *Atmos. Chem. Phys.* 2011, 11, 12973–13000. [CrossRef]
- Brown, T.J.; Hall, B.L.; Westerling, A.L. The Impact of Twenty-First Century Climate Change on Wildland Fire Danger in the Western United States: An Applications Perspective. *Clim. Chang.* 2004, 62, 365–388. [CrossRef]
- 3. Westerling, A.L.; Bryant, B.P. Climate Change and Wildfire in California. *Clim. Chang.* 2008, 87, 231–249. [CrossRef]
- Xie, Y.; Lin, M.; Horowitz, L.W. Summer PM 2.5 Pollution Extremes Caused by Wildfires over the Western United States during 2017–2018. *Geophys. Res. Lett.* 2020, 47, e2020GL089429. [CrossRef]
- Dreessen, J.; Sullivan, J.; Delgado, R. Observations and Impacts of Transported Canadian Wildfire Smoke on Ozone and Aerosol Air Quality in the Maryland Region on June 9–12, 2015. J. Air Waste Manag. Assoc. 2016, 66, 842–862. [CrossRef]
- 6. Shrestha, B.; Brotzge, J.A.; Wang, J. Observations and Impacts of Long-Range Transported Wildfire Smoke on Air Quality Across New York State During July 2021. *Geophys. Res. Lett.* **2022**, *49*, e2022GL100216. [CrossRef]
- 7. Landrigan, P.J.; Fuller, R.; Acosta, N.J.R.; Adeyi, O.; Arnold, R.; Basu, N.; Baldé, A.B.; Bertollini, R.; Bose-O'Reilly, S.; Boufford, J.I.; et al. The Lancet Commission on Pollution and Health. *Lancet* **2018**, *391*, 462–512. [CrossRef]
- 8. Jacob, D.J.; Winner, D.A. Effect of Climate Change on Air Quality. Atmos. Environ. 2009, 43, 51–63. [CrossRef]

- Pienkosz, B.D.; Saari, R.K.; Monier, E.; Garcia-Menendez, F. Natural Variability in Projections of Climate Change Impacts on Fine Particulate Matter Pollution. *Earth's Future* 2019, 7, 762–770. [CrossRef]
- Karanasiou, A.; Alastuey, A.; Amato, F.; Renzi, M.; Stafoggia, M.; Tobias, A.; Reche, C.; Forastiere, F.; Gumy, S.; Mudu, P.; et al. Short-Term Health Effects from Outdoor Exposure to Biomass Burning Emissions: A Review. *Sci. Total Environ.* 2021, 781, 146739. [CrossRef]
- Reid, C.E.; Maestas, M.M. Wildfire Smoke Exposure under Climate Change: Impact on Respiratory Health of Affected Communities. Curr. Opin. Pulm. Med. 2019, 25, 179–187. [CrossRef] [PubMed]
- 12. Schwartz, J.; Laden, F.; Zanobetti, A. The Concentration-Response Relation between PM(2.5) and Daily Deaths. *Environ. Health Perspect.* **2002**, *110*, 1025–1029. [CrossRef] [PubMed]
- 13. Ford, B.; Val Martin, M.; Zelasky, S.E.; Fischer, E.V.; Anenberg, S.C.; Heald, C.L.; Pierce, J.R. Future Fire Impacts on Smoke Concentrations, Visibility, and Health in the Contiguous United States. *GeoHealth* **2018**, *2*, 229–247. [CrossRef]
- Liu, Y.; Liu, Y.; Fu, J.; Yang, C.-E.; Dong, X.; Tian, H.; Tao, B.; Yang, J.; Wang, Y.; Zou, Y.; et al. Projection of Future Wildfire Emissions in Western USA under Climate Change: Contributions from Changes in Wildfire, Fuel Loading and Fuel Moisture. *Int. J. Wildland Fire* 2021, *31*, 1–13. [CrossRef]
- Spracklen, D.V.; Mickley, L.J.; Logan, J.A.; Hudman, R.C.; Yevich, R.; Flannigan, M.D.; Westerling, A.L. Impacts of Climate Change from 2000 to 2050 on Wildfire Activity and Carbonaceous Aerosol Concentrations in the Western United States. *J. Geophys. Res.* 2009, 114, D20301. [CrossRef]
- 16. Yue, X.; Mickley, L.J.; Logan, J.A.; Kaplan, J.O. Ensemble Projections of Wildfire Activity and Carbonaceous Aerosol Concentrations over the Western United States in the Mid-21st Century. *Atmos. Environ.* **2013**, 77, 767–780. [CrossRef] [PubMed]
- 17. Kearns, E.J.; Saah, D.; Levine, C.R.; Lautenberger, C.; Doherty, O.M.; Porter, J.R.; Amodeo, M.; Rudeen, C.; Woodward, K.D.; Johnson, G.W.; et al. The Construction of Probabilistic Wildfire Risk Estimates for Individual Real Estate Parcels for the Contiguous United States. *Fire* 2022, *5*, 117. [CrossRef]
- Johnson, M.M.; Garcia-Menendez, F. Uncertainty in Health Impact Assessments of Smoke From a Wildfire Event. *GeoHealth* 2022, 6, e2021GH000526. [CrossRef]
- 19. Draxler, R. HYSPLIT4 User's Guide; NOAA Air Resources Laboratory: Silver Spring, MD, USA, 1997.
- Draxler, R.R.; Hess, G.D. An Overview of the HYSPLIT_4 Modeling System for Trajectories, Dispersion, and Deposition. *Aust. Meteor. Mag.* 1998, 47, 295–308.
- 21. 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. Meteorol. Soc.* 2015, *96*, 2059–2077. [CrossRef]
- 22. Rolph, G.; Stein, A.; Stunder, B. Real-Time Environmental Applications and Display SYstem: READY. *Environ. Model. Softw.* 2017, 95, 210–228. [CrossRef]
- Li, Y.; Tong, D.Q.; Ngan, F.; Cohen, M.D.; Stein, A.F.; Kondragunta, S.; Zhang, X.; Ichoku, C.; Hyer, E.J.; Kahn, R.A. Ensemble PM _{2.5} Forecasting During the 2018 Camp Fire Event Using the HYSPLIT Transport and Dispersion Model. *J. Geophys. Res. Atmos.* 2020, 125, e2020JD032768. [CrossRef]
- 24. Goodrick, S.L.; Achtemeier, G.L.; Larkin, N.K.; Liu, Y.; Strand, T.M. Modelling Smoke Transport from Wildland Fires: A Review. *Int. J. Wildland Fire* **2012**, 22, 83. [CrossRef]
- Briggs, G.A. *Plume Rise: A Critical Survey*; Air Resources Atmospheric Turbulence and Diffusion Lab.: Oak Ridge, TN, USA, 1969; p. 81 TID–25075, 4743102.
- 26. Sofiev, M.; Vankevich, R.; Ermakova, T.; Hakkarainen, J. Global Mapping of Maximum Emission Heights and Resulting Vertical Profiles of Wildfire Emissions. *Atmos. Chem. Phys.* **2013**, *13*, 7039–7052. [CrossRef]
- Val Martin, M.; Heald, C.L.; Lamarque, J.-F.; Tilmes, S.; Emmons, L.K.; Schichtel, B.A. How Emissions, Climate, and Land Use Change Will Impact Mid-Century Air Quality over the United States: A Focus on Effects at National Parks. *Atmos. Chem. Phys.* 2015, 15, 2805–2823. [CrossRef]
- Knorr, W.; Dentener, F.; Lamarque, J.-F.; Jiang, L.; Arneth, A. Wildfire Air Pollution Hazard during the 21st Century. *Atmos. Chem. Phys.* 2017, 17, 9223–9236. [CrossRef]
- Pierce, J.R.; Val Martin, M.; Heald, C. Estimating the Effects of Changing Climate on Fires and Consequences for U.S. Air Quality, Using a Set of Global and Regional Climate Models; Joint Fire Science Project 13-1-01-4; Colorado State University: Fort Collins, CO, USA, 2017.
- Thonicke, K.; Spessa, A.; Prentice, I.C.; Harrison, S.P.; Dong, L.; Carmona-Moreno, C. The Influence of Vegetation, Fire Spread and Fire Behaviour on Biomass Burning and Trace Gas Emissions: Results from a Process-Based Model. *Biogeosciences* 2010, 7, 1991–2011. [CrossRef]
- Lautenberger, C. Mapping Areas at Elevated Risk of Large-Scale Structure Loss Using Monte Carlo Simulation and Wildland Fire Modeling. *Fire Saf. J.* 2017, 91, 768–775. [CrossRef]
- Lautenberger, C. Wildland Fire Modeling with an Eulerian Level Set Method and Automated Calibration. *Fire Saf. J.* 2013, 62, 289–298. [CrossRef]
- Bates, J.T.; Pennington, A.F.; Zhai, X.; Friberg, M.D.; Metcalf, F.; Darrow, L.; Strickland, M.; Mulholland, J.; Russell, A. Application and Evaluation of Two Model Fusion Approaches to Obtain Ambient Air Pollutant Concentrations at a Fine Spatial Resolution (250m) in Atlanta. *Environ. Model. Softw.* 2018, 109, 182–190. [CrossRef]

- 34. Fallah Shorshani, M.; André, M.; Bonhomme, C.; Seigneur, C. Modelling Chain for the Effect of Road Traffic on Air and Water Quality: Techniques, Current Status and Future Prospects. *Environ. Model. Softw.* **2015**, *64*, 102–123. [CrossRef]
- Wiedinmyer, C.; Akagi, S.K.; Yokelson, R.J.; Emmons, L.K.; Al-Saadi, J.A.; Orlando, J.J.; Soja, A.J. The Fire INventory from NCAR (FINN): A High Resolution Global Model to Estimate the Emissions from Open Burning. *Geosci. Model Dev.* 2011, 4, 625–641. [CrossRef]
- 36. Deng, X.; Li, F.; Li, Y.; Li, J.; Huang, H.; Liu, X. Vertical Distribution Characteristics of PM in the Surface Layer of Guangzhou. *Particuology* **2015**, *20*, 3–9. [CrossRef]
- U.S. Environmental Protection Agency. Office of Air Quality Planning and Standards. *EPA* 454/B-18-007; Technical Assistance Document for the Reporting of Daily Air Quality—The Air Quality Index (AQI); U.S. Environmental Protection Agency: Washington, DC, USA, 2018.
- Baker, B.; Pan, L. Overview of the Model and Observation Evaluation Toolkit (MONET) Version 1.0 for Evaluating Atmospheric Transport Models. *Atmosphere* 2017, 8, 210. [CrossRef]
- Krejcie, R.V.; Morgan, D.W. Determining Sample Size for Research Activities. *Educ. Psychol. Meas.* 1970, *30*, 607–610. [CrossRef]
 Childs, M.L.; Li, J.; Wen, J.; Heft-Neal, S.; Driscoll, A.; Wang, S.; Gould, C.F.; Qiu, M.; Burney, J.; Burke, M. Daily Local-Level
- Estimates of Ambient Wildfire Smoke PM _{2.5} for the Contiguous US. *Environ. Sci. Technol.* **2022**, *56*, 13607–13621. [CrossRef] 41. Kaulfus, A.S.; Nair, U.; Jaffe, D.; Christopher, S.A.; Goodrick, S. Biomass Burning Smoke Climatology of the United States:
- Implications for Particulate Matter Air Quality. *Environ. Sci. Technol.* **2017**, *51*, 11731–11741. [CrossRef] [PubMed]
- 42. Larkin, N.K.; O'Neill, S.M.; Solomon, R.; Raffuse, S.; Strand, T.; Sullivan, D.C.; Krull, C.; Rorig, M.; Peterson, J.; Ferguson, S.A. The BlueSky Smoke Modeling Framework. *Int. J. Wildland Fire* **2009**, *18*, 906. [CrossRef]
- Brown, T.; Clements, C.; Larkin, N.; Anderson, K.; Butler, B.; Goodrick, S.; Ichoku, C.; Lamb, B.; Mell, R.; Ottmar, R.; et al. Validating the Next Generation of Wildland Fire and Smoke Models for Operational and Research Use—A National Plan; Joint Fire Science Program: Boise, ID, USA, 2014; p. 70.
- 44. Oglesby, R.J.; Marshall, S.; Taylor, J.A. The Climatic Effects of Biomass Burning: Investigations with a Global Climate Model. *Environ. Model. Softw.* **1999**, *14*, 253–259. [CrossRef]
- 45. Oliveri Conti, G.; Heibati, B.; Kloog, I.; Fiore, M.; Ferrante, M. A Review of AirQ Models and Their Applications for Forecasting the Air Pollution Health Outcomes. *Environ. Sci. Pollut. Res.* **2017**, *24*, 6426–6445. [CrossRef]
- Liu, Y.; Kochanski, A.; Baker, K.R.; Mell, W.; Linn, R.; Paugam, R.; Mandel, J.; Fournier, A.; Jenkins, M.A.; Goodrick, S.; et al. Fire Behaviour and Smoke Modelling: Model Improvement and Measurement Needs for next-Generation Smoke Research and Forecasting Systems. *Int. J. Wildland Fire* 2019, *28*, 570. [CrossRef] [PubMed]
- Prichard, S.; Larkin, N.; Ottmar, R.; French, N.; Baker, K.; Brown, T.; Clements, C.; Dickinson, M.; Hudak, A.; Kochanski, A.; et al. The Fire and Smoke Model Evaluation Experiment—A Plan for Integrated, Large Fire–Atmosphere Field Campaigns. *Atmosphere* 2019, 10, 66. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.