



Status update: is smoke on your mind? Using social media to assess smoke exposure

Bonne Ford¹, Moira Burke², William Lassman¹, Gabriele Pfister³, and Jeffrey R. Pierce¹

¹Department of Atmospheric Science, Colorado State University, 1371 Campus Delivery, Fort Collins, CO 80523, USA

²Facebook, Menlo Park, CA 94025, USA

³National Center for Atmospheric Research, 3450 Mitchell Lane, Boulder, CO 80301, USA

Correspondence to: Bonne Ford (bonne@atmos.colostate.edu)

Received: 10 January 2017 – Discussion started: 19 January 2017

Revised: 8 May 2017 – Accepted: 15 May 2017 – Published: 22 June 2017

Abstract. Exposure to wildland fire smoke is associated with negative effects on human health. However, these effects are poorly quantified. Accurately attributing health endpoints to wildland fire smoke requires determining the locations, concentrations, and durations of smoke events. Most current methods for assessing these smoke events (ground-based measurements, satellite observations, and chemical transport modeling) are limited temporally, spatially, and/or by their level of accuracy. In this work, we explore using daily social media posts from Facebook regarding smoke, haze, and air quality to assess population-level exposure for the summer of 2015 in the western US. We compare this de-identified, aggregated Facebook dataset to several other datasets that are commonly used for estimating exposure, such as satellite observations (MODIS aerosol optical depth and Hazard Mapping System smoke plumes), daily (24 h) average surface particulate matter measurements, and model-simulated (WRF-Chem) surface concentrations. After adding population-weighted spatial smoothing to the Facebook data, this dataset is well correlated (R^2 generally above 0.5) with the other methods in smoke-impacted regions. The Facebook dataset is better correlated with surface measurements of $PM_{2.5}$ at a majority of monitoring sites (163 of 293 sites) than the satellite observations and our model simulation. We also present an example case for Washington state in 2015, for which we combine this Facebook dataset with MODIS observations and WRF-Chem-simulated $PM_{2.5}$ in a regression model. We show that the addition of the Facebook data improves the regression model's ability to predict surface concentrations. This high correlation of the Facebook data with surface monitors and our Washington state example

suggests that this social-media-based proxy can be used to estimate smoke exposure in locations without direct ground-based particulate matter measurements.

1 Introduction

Exposure to poor air quality is associated with negative impacts on human health (Dockery et al., 1993; Pope, 2007). As such, the Environmental Protection Agency (EPA) has set air quality standards to limit the concentration levels of pollutants in the United States, which has led to reductions in anthropogenic emissions. However, particulate matter (PM) also has natural and transboundary sources, which are more difficult to control. A large natural source of PM in the western US is from landscape fires, which are comprised of wildfires, prescribed burning on natural lands, and agricultural fires. Landscape fire smoke (LFS) drives much of the inter-annual variability in total $PM_{2.5}$ (PM with an aerodynamic diameter $< 2.5 \mu\text{m}$; Jaffe et al., 2008). The 2011 National Emissions Inventory (NEI2011; epa.gov) attributes $\sim 20\%$ of the primary $PM_{2.5}$ emissions in the US to wildfires, 15% to prescribed fires, and 1.5% to agricultural fires (epa.gov). Lelieveld et al. (2015) used concentration response functions derived from previous studies of total ambient PM (and smoking and household air pollution) to estimate that ~ 2500 premature mortalities are attributable to exposure to biomass burning (a broad category that includes wildland, prescribed, and agricultural fires) $PM_{2.5}$ per year in the US. However, the assumed toxicity and dose associated with LFS were as-

sumed to be the same as all other PM sources. Thus, it is important to determine the health responses specific to LFS.

Accurately attributing health outcomes to LFS requires a determination of the exposed population. Studies of health impacts often rely on (i) fixed site monitors (e.g., Pope et al., 2009), (ii) satellite products (e.g., Henderson et al., 2011; Rappold et al., 2011), or (iii) atmospheric model simulations (Alman et al., 2016; Fann et al., 2012; Johnston et al., 2012; Rappold et al., 2012). Each of these methods has limitations as an exposure metric. For example, fixed site monitors are sparse in much of the western US, and satellite products do not provide surface-level concentrations on their own. Atmospheric model simulations may be biased by their emission inventories (Davis et al., 2015; Zhang et al., 2014), spatial resolution (Misenis and Zhang, 2010; Pungert and West, 2013; Thompson et al., 2014; Thompson and Selin, 2012), or input meteorological fields (Cuchiara et al., 2014; Srinivas et al., 2015; Žabkar et al., 2013). Thus, there is a growing effort to include multiple datasets (e.g., Henderson et al., 2011; Yao et al., 2013) and create blended products that can exploit the strengths of each dataset (Brauer et al., 2015; van Donkelaar et al., 2015; Lassman et al., 2017; Gan et al., 2017; Reid et al., 2015; Yao and Henderson, 2014). However, these methods still only provide estimates of ambient concentration levels and not of actual exposure. Additionally, attributing health effects specifically to LFS exposure can be difficult as it requires separating the contribution of smoke from total $\text{PM}_{2.5}$ (Liu et al., 2015).

In this work, we propose the use of de-identified, aggregated Facebook data to determine population-level exposure for the summer of 2015, which was a particularly smoky year in the US (see Fig. S1 in the Supplement for the number of fire and smoke days). While there can be many different sources of poor air quality, the highest $\text{PM}_{2.5}$ concentrations measured during the study period were in regions and during time periods associated with wildfire smoke. We show that, region wide, this dataset is better correlated with surface measurements of $\text{PM}_{2.5}$ than other traditional means of estimating exposure, suggesting that it has the potential for use in estimating smoke exposure in locations without direct ground-based particulate matter measurements. We also present a test case for Washington state, in which we demonstrate that a regression model that includes our Facebook dataset is better able to predict surface $\text{PM}_{2.5}$ than a regression model that only has model-simulated $\text{PM}_{2.5}$ and satellite aerosol optical depth (AOD). We also compare our results to another measurement of internet behavior, Google Trends, as a proxy for air quality exposure.

The use of social media in risk and exposure assessment is a growing field. In the past decade, data mining of social media has provided a wealth of information to news outlets, marketing firms, and the social sciences (Burke and Kraut, 2016; Golder and Macy, 2011; Kosinski et al., 2013; Masedu et al., 2014; Youyou et al., 2015). Only recently have social media and internet behavior been used for research in both

the natural sciences and public health. Social media and internet behavior have been proposed to track epidemics and earthquakes (e.g., Broniatowski et al., 2013; Crooks et al., 2013; Ginsberg et al., 2009), fires (Abel et al., 2012; Bedo et al., 2015; De Longueville et al., 2009; Kent and Capello Jr., 2013), and poor air quality (Jiang et al., 2015; Mei et al., 2014; Tao et al., 2016), as well as to predict hospitalizations (Ram et al., 2015). A paper by Sachdeva et al. (2016) also proposed the use of Twitter content and geographic information to estimate LFS concentrations. In this paper, we show how daily Facebook posting trends “track” significant changes in air quality, such as those associated with dense smoke plumes from large wildfires. Furthermore, we show that Facebook posting trends could also improve estimates of $\text{PM}_{2.5}$ exposure by serving as an extra constraint on more traditional methods for estimating exposure.

2 Methods and datasets

2.1 Internet behavior datasets

2.1.1 Percent of Facebook posters

Our dataset is the percentage of distinct Facebook posters in each US city that used any of the following words in a post: “smoke”, “smoky”, “smokey”, “haze”, “hazy”, or “air quality”. References to cigarette smoking and other phrases not related to air quality were filtered out (see Supplement). The search generates de-identified and aggregated counts of posters each day divided by the number of people who used Facebook in that city. This method counts each person at most once per day, thus avoiding bias from a single person posting multiple times about air quality that day. Re-shares of news articles and friends’ posts were also excluded. No individual’s text was viewed by researchers. Our goal was to focus on wildfire smoke because wildfire smoke often leads to extreme air quality degradation over broad regions of the US in the summertime. However, because this list includes “air quality” and “haze” (and the results were aggregated), these search criteria can also highlight trends in Facebook posters discussing air quality degradation due to other emissions, such as fossil fuel combustion, and may better encompass more of the ways that people discuss their experiences of changes in the air from smoke or other particulate matter. Geographic location at the city level is determined by the IP address. Data were provided for 5 June through 27 October 2015.

We analyzed this dataset of the de-identified, aggregated percent of Facebook posters that matched our search criteria at the city, town, or other municipality level (See Fig. S2a for location centroids, referred to as “raw” throughout text). We translate the percent of Facebook posters in each region onto a standard latitude–longitude grid using an area smoothing procedure with data weighted by the population of the mu-

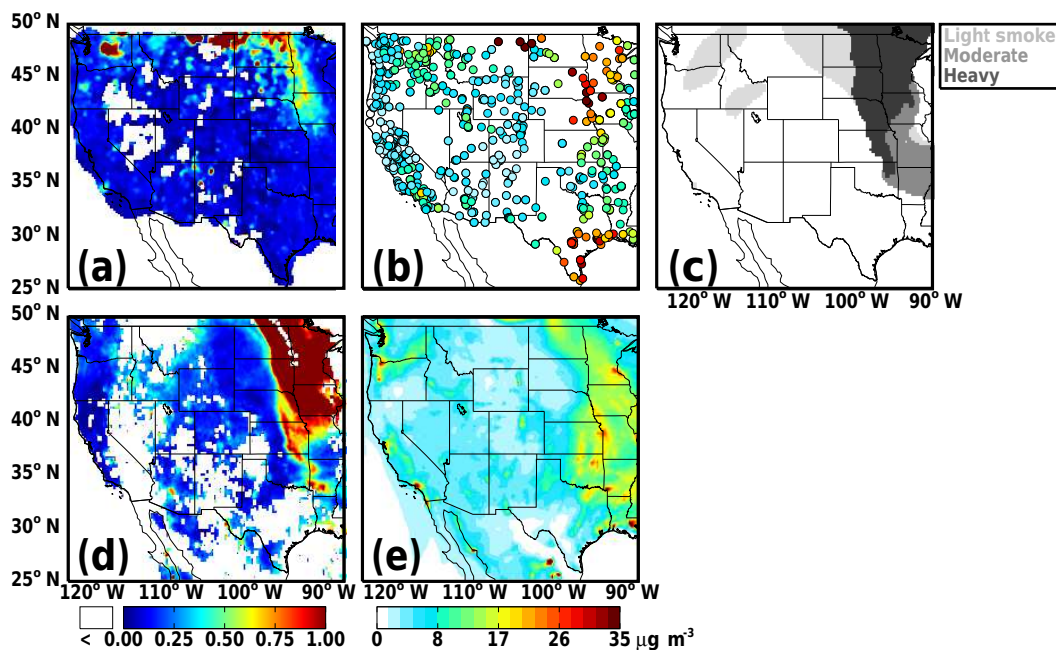


Figure 1. Example datasets for 29 June 2015. (a) Population-weighted (Eq. 1) percent of Facebook posters meeting the criterion (white signifies regions with weighted population < 10), (b) 24 h average surface $\text{PM}_{2.5}$ concentrations from surface measurement sites, (c) gridded HMS smoke product, (d) gridded, unfiltered MODIS Aqua and MODIS Terra AOD (white signifies no valid observation), and (e) WRF-Chem-simulated 24 h average surface $\text{PM}_{2.5}$ concentrations.

nicipality (See Fig. S2 for an example). The spatial interpolation allows us to estimate the magnitude of the response between the specific locations (centroids) and to compare to other gridded datasets. Additionally, we chose to weight the results by population because some of these locations are in areas with small populations (and potentially few posters on Facebook), which can skew results. We generated a fixed 0.25° grid using an inverse distance weighting to a power of 6 with a scale distance (or search neighborhood; d_s) of 20 km. The scale distance and power were set to sharply reduce the influence of more distant observations and chosen based on the grid resolution in order to maintain the regional variability of the Facebook posters. Our resulting gridded data are determined using the following formula:

$$f_i = \frac{\sum \left\{ f_c \times \frac{P_c}{1 + \left(\frac{d_{i,c}}{d_s} \right)^6} \right\}}{\sum \left\{ \frac{P_c}{1 + \left(\frac{d_{i,c}}{d_s} \right)^6} \right\}}, \quad (1)$$

where the percent of Facebook posters (f_i) at a grid location (i) is the sum of all of the products of the population (P_c) and the original percent of Facebook posters (f_c) at each “Facebook municipality” (c) weighted by the inverse of the distance (d) between location (i) and the Facebook municipality (c).

2.1.2 Google Trends

We analyzed Google Trends data (www.google.com/trends/) as a proxy for exposure and to evaluate the keywords used in our search criteria. Our reason for including this analysis is twofold: (1) to compare the results of our percent of Facebook posters to results using another internet behavior dataset and (2) to determine which keywords are most strongly correlated with $\text{PM}_{2.5}$ (as our “Percent of Facebook posters” dataset is an aggregated result for all search terms). We searched for “air quality”, “wildfire”, “smoke”, “pollution”, “haze”, “smog”, and “ozone” for 1 May to 31 October 2015 for every designated media area (DMA) in the western US. Google Trends results are determined from a random sample of searches with location determined by IP address and duplicates (when the same person searches for the same term multiple times) removed. Results for each search are aggregated and de-identified but limited to popular terms, with low values appearing as zero (the highest values are 100). Therefore, the popularity of a search term impacts the spatial resolution available of the aggregated results (country, DMA, or city). Because of the coarse resolution of the aggregated Google Trends data (DMA level), we chose to compare only to surface measurements and not the other gridded datasets. In order to determine the temporal correlation between the Google Trends and surface measurements, we identified the DMA in which each measurement site is located.

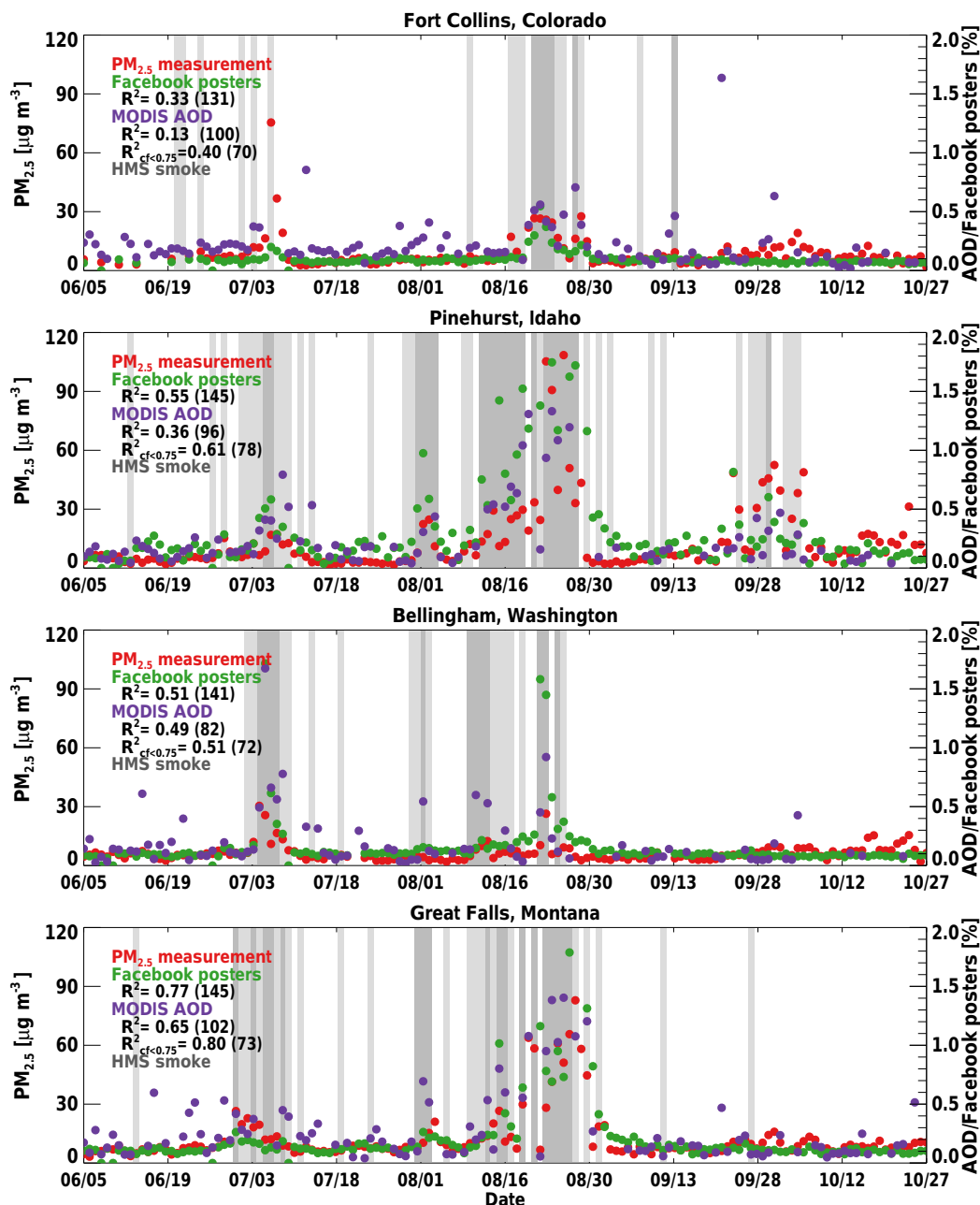


Figure 2. Time series of measured surface $\text{PM}_{2.5}$ concentrations (red), gridded and population-weighted percent of Facebook posters (green), MODIS AOD (purple), and days with HMS-denoted light (light gray) and moderate to thick (dark gray) smoke at the following locations for 5 June to 27 October 2015: (a) Fort Collins, CO; (b) Pinehurst, ID; (c) Bellingham, WA; and (d) Great Falls, MT. R^2 values for each dataset with the surface measurement are given along with the number of days available for the calculation noted in parentheses.

2.2 Surface measurements

We determined the temporal correlation of these datasets to several other datasets that are commonly used for estimating exposure to LFS on a daily timescale. We use 24 h average concentrations of total $\text{PM}_{2.5}$ mass from the EPA Air Quality System (AQS; data from www.epa.gov/aqs), which includes monitor data from different agencies, and sites from the

Interagency Monitoring of Protected Visual Environments (IMPROVE; data from <http://views.cira.colostate.edu/fed/>). At IMPROVE network sites, surface measurements of atmospheric composition are taken over a 24 h period every third day (Malm et al., 1994). Depending on the measurement method at the site, 24 h average concentrations are provided daily, every third day, or every sixth day at EPA-AQS sites. To maximize our data availability, we use measure-

ments from sites using the Federal Reference Method and the Federal Equivalent Method (FRM/FEM; 88101) as well as non-FRM/FEM (88502) sites (both are also used by the EPA for AQI summaries).

We determined the temporal correlations between the daily surface measurement and the internet behavior datasets at every site. However, in the “Results and discussion” section, we only show example time series for four of these locations. These four locations are shown because they were all impacted by wildfire smoke during the study period, but the response in the percent of Facebook posters varied among the sites, likely due to differences in surface concentrations, distance to fire, population, and cloud cover (discussed in “Results and discussion”).

2.3 Satellite products

2.3.1 Hazard Mapping System (HMS) smoke product

We use the Hazard Mapping System (HMS) fire and smoke analysis product, which is produced routinely by the National Oceanic and Atmospheric Administration (NOAA) National Environmental Satellite and Data Information Service (NESDIS) for the purpose of identifying fires and smoke emissions (<http://satepsanone.nesdis.noaa.gov>). The HMS smoke product uses observations from both geostationary and polar-orbiting satellites. Polygons determined from satellite visible image analysis are currently categorized as light, moderate, and heavy smoke and have assigned numerical values to estimate surface smoke concentrations (5, 16, 27 $\mu\text{g m}^{-3}$). This product is only available for daylight hours and each polygon is considered valid for a specific time period. We created a gridded surface from all the polygons valid for each day with the surface concentration values suggested at the same 0.25° grid resolution as our gridded percent of Facebook posters in order to calculate the temporal correlation between the two datasets for each grid. In grids with more than one polygon valid for a day, we take the maximum value in each grid location during that day. Data files were available for every day during our analysis period except 20 August 2015, although sub-daily smoke plume analysis periods could also be missing. To determine the correlation with surface measurements, we matched the site location to the corresponding grid box.

2.3.2 MODerate resolution Imaging Spectroradiometer (MODIS) AOD

For AOD from satellites, we use the Collection 6 MODerate resolution Imaging Spectroradiometer (MODIS) Level 2 10km aerosol optical depth (AOD) products from the Terra and Aqua platforms. Terra has a morning overpass ($\sim 10:30$ AMLT) and Aqua has an afternoon overpass ($\sim 1:30$ PMLT). With a swath width of 2330 km, the instruments provide almost daily coverage of the globe in cloud-free conditions. The MODIS algorithm can have difficulty

distinguishing thick smoke from cloud (van Donkelaar et al., 2011), causing some instances of heavy smoke to be erroneously filtered out (although Collection 6 has made improvements to the algorithm to minimize this misclassification; see Levy et al., 2013). We average the MODIS AOD observations from both instruments on the same 0.25° grid and use all quality levels for better coverage. We additionally use the MODIS cloud fraction (CF) products (“Cloud_Fraction_Land” and “Cloud_Fraction_Ocean”) in order to determine the presence of clouds and to determine whether cloudiness impacts Facebook postings on smoke. We calculate the temporal correlations between MODIS AOD and the “Percent of Facebook posters” dataset and the surface observations for the full dataset and excluding cloudy days.

2.4 Weather Research and Forecasting model with Chemistry (WRF-Chem) $\text{PM}_{2.5}$

Several models and model frameworks are also routinely used to estimate smoke exposure. Here, we use a chemical transport model, the Weather Research and Forecasting model with Chemistry (WRF-Chem). The simulation was completed for 5 June to 1 October 2015. We use Global Forecast System (GFS) meteorology, biogenic emissions from the Model of Emissions of Gases and Aerosols from Nature (MEGAN; Guenther et al., 2006), National Emissions Inventory 2011 (NEI) anthropogenic emissions, FINN biomass burning emissions (Wiedinmyer et al., 2011), MOZCART aerosol species and chemistry, and (MOZART) chemical boundaries (Emmons et al., 2010). Horizontal resolution is 15 km and there are 27 vertical levels. Concentrations are output for each model hour, which we then average to provide daily 24 h average surface concentrations in order to compare to the “Percent of Facebook posters” dataset and surface measurements.

2.5 Regression model

We present a test case to evaluate the feasibility and usefulness of including the “Percent of Facebook posters” dataset in a statistical model. We compare two geographically weighted regression (GWR) models that use MODIS AOD and WRF-Chem $\text{PM}_{2.5}$ with and without the “Percent of Facebook posters” dataset. GWR has previously been used in several different studies to predict surface air quality (Hu et al., 2013; Lassman et al., 2017; Song et al., 2014; You et al., 2016). For our test case, we focus on Washington state because of the extensive network of surface $\text{PM}_{2.5}$ measurements available for validating results. In our regression model, we determine the dependent variable (surface $\text{PM}_{2.5}$ at each measurement site) from a linear combination of these different predictor variables (MODIS AOD, WRF-Chem $\text{PM}_{2.5}$, and gridded percent of Facebook posters). A separate set of regression coefficients is determined at

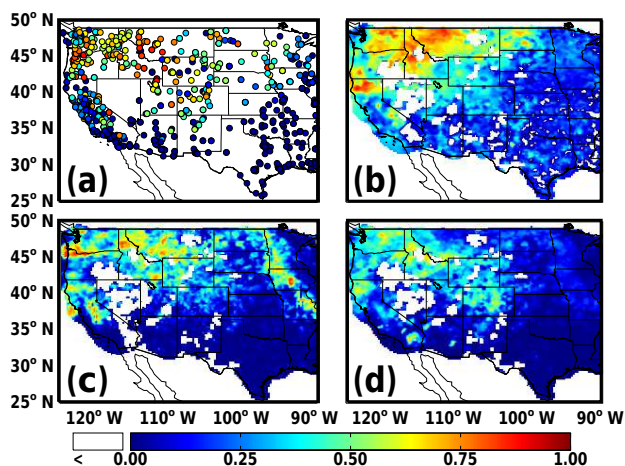


Figure 3. R^2 values for % of Facebook posters and (a) surface measurements of $\text{PM}_{2.5}$ (for sites with >35 days of measurements), (b) WRF-Chem $\text{PM}_{2.5}$, (c) MODIS AOD when cloud fraction was below 0.75, and (d) HMS smoke product for the period of 5 June to 30 September 2015.

each surface monitor location, which is then interpolated across the domain. We use the leave-one-out cross-validation (LOOCV) method to test our models, in which the regression coefficients determined at a single monitor are removed from the interpolation scheme. The resulting $\text{PM}_{2.5}$ predicted by the regression model is compared to the observed $\text{PM}_{2.5}$ concentrations. We calculate the temporal correlation, slope, and mean absolute error (MAE) for the two regression models.

3 Results and discussion

3.1 Comparison of percent of Facebook posters to conventional metrics

An example of the data used in this study is given in Fig. 1 for 29 June 2015, which shows a dense smoke plume from wildfires in Canada causing degraded air quality over the midwestern US and smoke from local fires in the northwest over Washington, Oregon, and Idaho. The impact of this smoke plume is evident in the HMS smoke product, the anomalously high surface $\text{PM}_{2.5}$ concentrations, the elevated MODIS AOD values, and the WRF-Chem $\text{PM}_{2.5}$. The spatial pattern in the percent of Facebook posters is somewhat consistent with regions of degraded surface air quality, suggesting that some people were aware of the degraded air quality. The extent of the “Facebook plume” does not extend as far east or south as the smoke plume observed by the satellite products (MODIS AOD and HMS smoke product), and hot spots in the percent of Facebook posters are centered around the border between eastern Montana and Canada. It should be noted that the surface measurements also do not

show a strong increase in surface concentrations as far south (Missouri and Arkansas), suggesting that the plume observed by the satellites might have been lofted above the surface. Additionally, while the HMS smoke product suggests only light smoke over northeastern Montana and MODIS AOD is only moderately higher than the surrounding region, surface $\text{PM}_{2.5}$ concentrations are elevated, which agrees with the spatial pattern of Facebook posters. In cases of a lofted plume or smoke concentrated at the surface, this new dataset might be more representative of surface air quality changes than these satellite products.

In Fig. 2, we also show example time series of the percent of Facebook posters and other datasets (surface $\text{PM}_{2.5}$ measurements, MODIS AOD, MODIS CF, HMS smoke product) used in this study for four different locations in the western US: Fort Collins, CO; Pinehurst, ID; Bellingham, WA; and Great Falls, MT. All four of these locations were impacted by wildfire smoke during the study period, but the response in the percent of Facebook posters varied among the sites, likely due to differences in surface concentrations, distance to fire, population, and cloud cover. From these time series, we see the two main fire event periods that impacted large areas of the US during the summer of 2015: (1) the Canadian wildfires in late June through early July and (2) the wildfires in the northwestern US (mainly Washington and Idaho) in August. The magnitude of impact on these different metrics for estimating air quality varies by location and event. For Pinehurst, ID, where the population was ~ 1600 in 2015, population weighting the Facebook poster time series improves the correlation with the 24 h average surface measurements ($R^2 = 0.55$ for gridded and $R^2 = 0.00$ for raw). In more populated regions, such as Fort Collins, CO (pop. $\sim 161\,000$), Bellingham, WA (pop. $\sim 85\,000$), and Great Falls, MT (pop. $\sim 60\,000$), population weighting the Facebook posters has little impact on the time series and resulting correlation with the surface measurements (as shown in Fig. S3). A further discussion of these time series is presented throughout this section.

In order to assess how well changes in the fraction of people posting about smoke and air quality on Facebook represent actual changes in surface air quality, we compare time series of the percentage of Facebook posts matching our criteria to time series of $\text{PM}_{2.5}$ measured at all of the different surface sites across the summer of 2015, such as shown in the example time series in Fig. 2. The coefficients of determination for all surface $\text{PM}_{2.5}$ measurement sites with the gridded, population-weighted Facebook posts are shown in Fig. 3a, which suggests that the best agreement between the two datasets is in regions that experienced heavy smoke and/or anomalously high $\text{PM}_{2.5}$ concentrations during the summer. This is to be expected based on our search criteria. For example, the Mt. Hood IMPROVE site in Oregon (Fig. 3) had 39 measurement days (5 June to 30 September) and 14 days on which the HMS smoke product suggested smoke over the location. This site provides the best R^2 be-

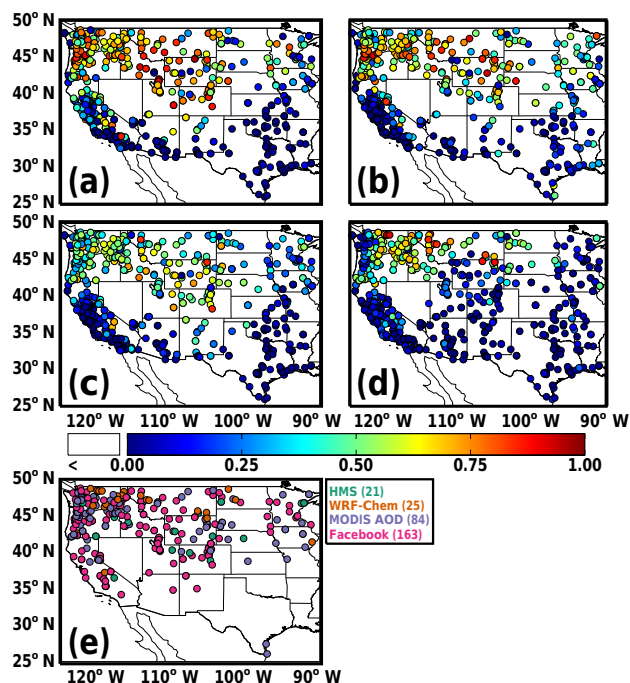


Figure 4. R^2 values for surface measurements of $\text{PM}_{2.5}$ with (a) percent of Facebook posters ($\text{CF} < 0.75$), (b) MODIS AOD ($\text{CF} < 0.75$), (c) HMS smoke, and (d) WRF-Chem-simulated $\text{PM}_{2.5}$ for the period of 5 June to 30 September 2015. (e) Product (HMS Smoke, WRF-Chem $\text{PM}_{2.5}$, MODIS AOD, or Facebook posters) that has the highest R^2 compared to surface measurements for the time period of 5 June to 30 September 2015 (sites are shown only if the resulting $R^2 > 0.5$). The number of sites in the western US (domain shown) where the product has the highest R^2 (and $R^2 > 0.5$) is given in parentheses.

tween the percent of Facebook posters and the measured surface $\text{PM}_{2.5}$ with a value of 0.97.

We also compare the agreement of the percent of Facebook posters against simulated concentrations from a chemical transport model simulation (WRF-Chem; Fig. 3b), which again shows the highest correlation in the northwestern US. The area was impacted by wildfire smoke for many days in the summer of 2015. We would expect this as our Facebook post search criteria are aimed at smoke and poor air quality and would likely only show changes in postings in regions where air quality was noticeably degraded.

Agreement between MODIS AOD and Facebook posting trends is shown in Fig. 3c, which also shows the best agreement in the northwestern US. Because thick smoke can occasionally be classified as cloud in the MODIS algorithm (van Donkelaar et al., 2011), we filter out MODIS AOD observations for which the cloud fraction was $> 75\%$. The impact of this filtering is shown in the time series in Fig. 2. The criterion reduced our number of useable observations but improved correlations at most sites (Fig. S4). Comparisons between Facebook posters and MODIS AOD are spatially sim-

ilar to WRF-Chem $\text{PM}_{2.5}$ and surface measurements, but the coefficients for MODIS AOD and Facebook posts are generally worse. However, this satellite product is derived for the full atmospheric column and is not necessarily directly related to surface concentrations. Smoke plumes (and transported pollution from other sources) can be lofted above the surface and may not impact surface-level exposure where astute Facebook posters would take notice.

Finally, we show R^2 for the values estimated by the HMS smoke product and the Facebook posters in Fig. 3d. Again, we see similar trends with the best agreement occurring in regions that experienced numerous smoke days. The correlation values are not as high as for MODIS AOD or WRF-Chem $\text{PM}_{2.5}$. The HMS smoke product only provides estimates for smoke, which is the primary focus of our search criteria, although it also includes phrases related to general air quality degradation. Additionally, as with MODIS AOD, the HMS smoke product may not be representative of actual surface-level exposure. Finally, the HMS smoke product only provides categorical estimates of “heavy,” “moderate,” or “light” smoke and likely cannot represent subtle changes in exposure concentration levels compared to MODIS AOD.

3.2 Evaluation of all metrics compared to surface measurements

While we have shown that our new dataset often correlates well with more traditional datasets that have been used to estimate smoke and/or $\text{PM}_{2.5}$ concentrations and exposure, we also investigate whether the percent of Facebook posters can be used to improve estimates when combined with the other datasets. In Fig. 4, we compare how well each dataset estimates $\text{PM}_{2.5}$. We show the coefficients of determination for Facebook posters (4a; similar to 3a but for days where $\text{CF} < 0.75$), MODIS AOD (with $\text{CF} < 0.75$; Fig. 4b), the HMS smoke product (Fig. 4c), and WRF-Chem $\text{PM}_{2.5}$ (Fig. 4d) with the surface monitors. From Fig. 4, we can evaluate which dataset best correlates with surface measurements in different regions of the western US.

We summarize these initial findings in Fig. 4e, which shows the dataset that was best correlated with the surface measurement at each site (and the R^2 had to be greater than 0.5). This figure shows that our “Percent of Facebook posters” dataset is better correlated with actual surface measurements at most sites in our domain for the given time period (5 June to 30 September 2015) compared to other datasets that are typically used to estimate exposure. We find that MODIS AOD and WRF-Chem $\text{PM}_{2.5}$ are better predictors in regions with low populations, such as North Dakota, eastern Montana, and eastern Washington. Additionally, WRF-Chem $\text{PM}_{2.5}$ and MODIS AOD are better predictors over much of the eastern US (not shown; R^2 values all less than 0.5), which is dominated by anthropogenic emissions during the time period. These “normal” day-to-day changes in anthropogenic pollution may be less likely to be

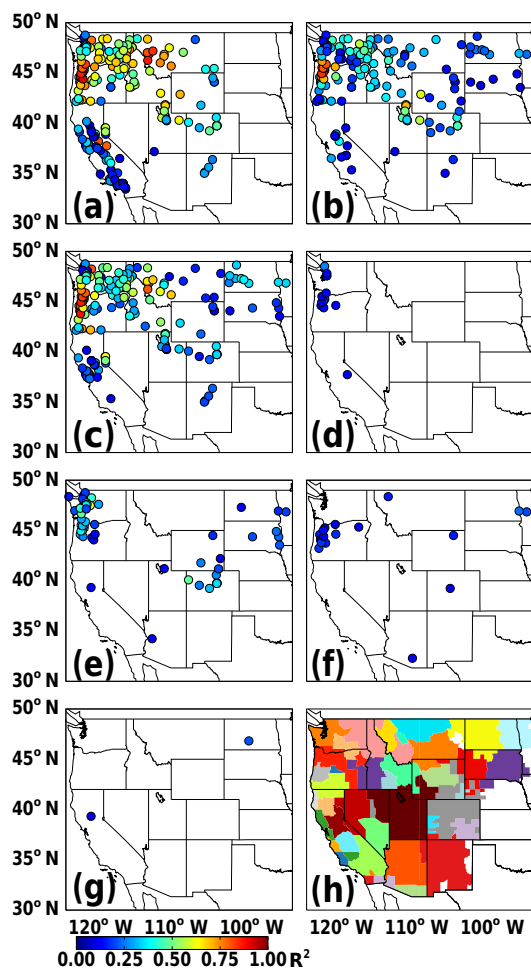


Figure 5. R^2 values at each measurement site for surface measurement and Google Trends search for (a) “air quality”, (b) “wild-fire”, (c) “smoke”, (d) “pollution”, (e) “haze”, (f) “smog”, and (g) “ozone.” Only sites where $R^2 > 0.1$ are shown. The 48 DMAs considered are shown in (h).

picked up by our Facebook post search criteria. We did not optimize the configuration of our WRF-Chem simulation to match surface observations. Changing emissions, meteorology, parameterization choices, grid resolution, or time steps may have improved surface concentration estimates, but the optimal configuration would likely differ by region and time period. However, our results shown in Fig. 4 suggest that Facebook posting could be used to help estimate exposure in conjunction with the other datasets.

However, if the aggregate percent of Facebook posters is used to estimate exposure, there may be a few limitations. While trends in Facebook posting seem to represent the variability in surface air quality over our study period at many sites, there is not a simple relationship between posting and $PM_{2.5}$. There did not appear to be a threshold $PM_{2.5}$ concentration at which it was guaranteed that people would start posting, region wide or in individual cities (e.g., there were

cases with high smoke but little posting, such as the July event in Fort Collins, CO). There are several potential reasons for this. (1) As noted, on cloudy days, people may not be able to distinguish poor air quality, especially if it is from long-range transport where residents are not aware of a nearby fire. (2) There could be a point of saturation or response fatigue; people experiencing multiple days of smoke may find it less interesting to post about, or they could experience a cognitive bias causing them to perceive improved air quality in comparison to previous air quality. To test this, we looked at the time series of the ratio of the percent of Facebook posters to surface concentrations, and this ratio does appear to decrease over time during smoke events lasting several days. A decrease throughout the season is only evident at a few sites, although this is difficult to compare because the major smoke event at most sites occurred in late August or early September with few to no smoke events occurring afterwards. (3) We noted that regions with a high Facebook posting percent were occasionally centered over areas where the population had experienced poor air quality on preceding days rather than the current regions of poor air quality. This time shift could suggest a lag in either individual awareness or in the time it takes to spread information among community-level social networks. Additionally, there could also be persistence in Facebook posts; air quality might improve in a location, but people are still posting about it. Conversely, awareness of events could spread through social networks more quickly than an air quality event (such as a smoke plume) is transported such that individuals discuss an event before it impacts them. Quantitatively, this is difficult to assess as it may be more event related than season specific. We compared ± 1 -day lag correlations between Facebook posts and surface measurements for all sites that had daily measurements (as opposed to every third day). Using the same day provided the best correlation at $\sim 90\%$ of sites. Slightly better correlations were found using the previous day’s measurement at several sites in Utah, and using the following day produced better estimates at several sites in Washington and Oregon, where there were broad regions and extended periods of degraded air quality due to local fires.

3.3 Cloudy day modification

We included the CF criterion for the above analysis for all datasets. We found that filtering out days with high CF improved the agreement of Facebook posts and MODIS AOD (Figs. 2 and S5). This led us to also hypothesize that people may have difficulty distinguishing poor air quality on cloudy days, especially farther downwind of a source. To test this, we also sampled the percent of Facebook posters and surface measurement time series at each site (with filtering) using the MODIS cloud fraction. Compared to correlations between surface measurements and Facebook posts for the full time period, using only the days with $CF < 0.75$ improved correlations most noticeably at sites that were gen-

erally more than 500 km downwind of fires (such as in Colorado, Wyoming, and Utah; Fig. S5) but had less impact at sites closer to the 2015 wildfires (Oregon, western Montana, Washington, and Idaho; see Fig. S1a for fire locations). Cloudiness as a possible impact on Facebook awareness is seen in the time series for Fort Collins, Colorado in Fig. 2a. Although concentrations were greater during the July event than the August event, the response in Facebook posts was much lower. Bellingham, WA was also impacted by smoke during the same period in July. Although lower surface concentrations were measured, the response in Facebook posts was greater. We noted that during the July event, however, the MODIS product reported a cloud cover of 100 % over Fort Collins. For the full time period, filtering out days with $CF > 0.75$ improved the R^2 between Facebook posts and surface measurements in Fort Collins from 0.33 to 0.54. Alternatively, in Great Falls, MT, which had many nearby fires, filtering only changed the R^2 from 0.77 to 0.79 even though roughly the same number of days met the 0.75 criteria for exclusion.

3.4 Google Trends comparison with surface measurements

We compared Google Trends data to surface measurements of $PM_{2.5}$. Our results are shown in Fig. 5 for each search term. As with the aggregate “Percent of Facebook posters” dataset, correlations are best in the northwestern US, specifically Washington, Montana, and Oregon, which are states that were heavily impacted by smoke in 2015. Although we compare to total $PM_{2.5}$, the best correlations were found not only for “air quality”, but also “wildfire” and “smoke”; as with the Facebook posters, we might expect this since wildfire smoke was the source of the most variability in surface $PM_{2.5}$ during this time period. The search terms that are more related to urban pollution (“pollution”, “smog”, “haze”, and “ozone”) have much lower correlations, and sites that do have $R^2 > 0.1$ are generally in urban areas or far downwind of smoke. “Ozone” in particular was not well correlated with $PM_{2.5}$ measurements (all $R^2 < 0.22$), which should be expected since ozone concentrations and $PM_{2.5}$ concentrations are not always well correlated (e.g., Reisen et al., 2011).

3.5 Google Trends search term comparison

We used the Google Trends data to analyze our Facebook search term criteria because we were not able to do this within the “Percent of Facebook posters” dataset. We chose several words that might be associated with “air quality” and determined the correlations between each word for each DMA as shown in Fig. 8. As with the actual concentrations of $PM_{2.5}$, we find that “air quality” is generally more associated with “smoke” and “wildfire” than words more commonly associated with urban sources like “smog”, “haze”, “pollution”, and “ozone”. Sachdeva et al. (2016) found that the distance

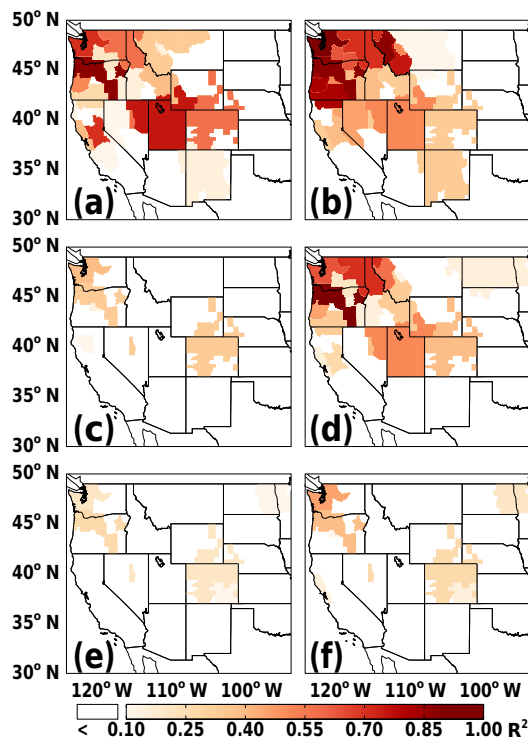


Figure 6. R^2 values for the following pairs of Google Trends search terms: (a) “air quality” and “wildfire”, (b) “air quality” and “smoke”, (c) “air quality” and “haze”, (d) “wildfire” and “smoke”, (e) “wildfire” and “haze”, and (f) “smoke” and “haze” for June–September 2015.

from the fires impacted the content of postings about the fire, and we also note some differences in our correlation maps based on distance. For example, closer to the fires (WA, OR, ID, MT), “air quality” is more associated with “smoke”. Farther away (CO, NV, UT, WY), “air quality” is more associated with “wildfire”. At these sites, “air quality” is also better correlated with “wildfire” than “smoke”, which may suggest that people are aware of the impact of the wildfires on air quality, but not able to see smoke. However, Google Trends is scaled by popularity in each region and data are only available on very popular terms. This could lead to a discrepancy in that the same number of people may be searching for these terms in different regions, but the relative popularity may be very different compared to other search terms, especially if there are other physical sources of “smoke” or impacts on “air quality” in a region. “Ozone”, “smog”, and “pollution” (terms that may be more associated with urban air pollution) are not well correlated with “air quality”, “smoke”, or “wildfires” over our study period; however, “haze” is moderately correlated in WA, OR, and CO (Fig. 6).

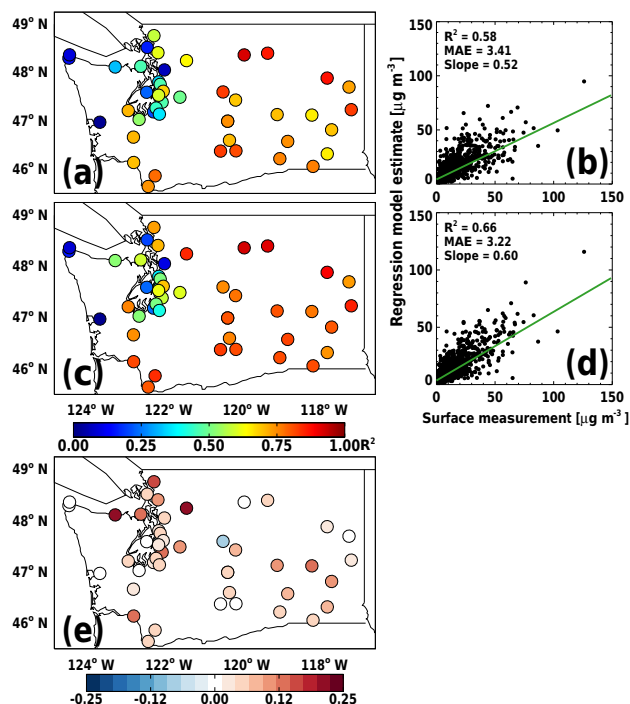


Figure 7. R^2 values at each measurement site for surface $\text{PM}_{2.5}$ and regression model estimate (a) using MODIS AOD and WRF-Chem $\text{PM}_{2.5}$ and (c) using MODIS AOD, WRF-Chem $\text{PM}_{2.5}$, and the percent of Facebook posters for 5 June to 30 September 2015. The difference in R^2 (e) between the two regression models (with Facebook posters and without Facebook posters). Also shown are scatterplots for all daily measured $\text{PM}_{2.5}$ and corresponding regression model estimates in the domain (b) using MODIS AOD and WRF-Chem $\text{PM}_{2.5}$ and (d) using MODIS AOD, WRF-Chem $\text{PM}_{2.5}$, and Facebook posters.

3.6 Geographically weighted regression test case for Washington state

As a first test case to evaluate the usefulness of this aggregate “Percent of Facebook posters” dataset in a statistical model, we compared two geographically weighted regression model estimates using MODIS AOD and WRF-Chem $\text{PM}_{2.5}$ with and without the Facebook posters. From Fig. 4, we see that WRF-Chem $\text{PM}_{2.5}$, MODIS AOD, and the “Percent of Facebook posters” dataset are all correlated with surface $\text{PM}_{2.5}$ in Washington state, and the best-correlated variable varies between surface sites. Therefore, a regression model could allow us to leverage the strengths from each dataset to create an improved estimate.

In Fig. 7, we show the results for our regression models with and without the Facebook posts. We see that including the Facebook posts in the regression model leads to improved R^2 values at many of the sites in Washington (only one site shows a decrease; Fig. 7e). Additionally, for the full dataset (of all sites and all days), there is an improved R^2 (0.66 compared to 0.58) and slope (0.60 compared to 0.52) with

a smaller error. While these improvements may be small, we find this is in part because the “Percent of Facebook posters” dataset explains much of the same variability as WRF-Chem $\text{PM}_{2.5}$ (and better explains variability in the urban region around Seattle, WA). We did not account for cloudy days in our regression analysis. Including information on cloud cover could potentially improve our regression model, which will be investigated further in ongoing work on this analysis.

4 Conclusions

In this paper, we introduced a novel concept of using the de-identified, aggregated percent of Facebook posters mentioning smoke, haze, or air quality to determine exposure by comparing to traditional datasets and in a regression model. We also looked at Google Trends data for the same time period and compared it to surface observations. The Facebook posts were useful in regions meeting two conditions: (1) the region was impacted by LFS, and (2) there was a large enough population posting to Facebook. The Google Trends data were also best correlated in regions impacted by smoke; however, it is aggregated at a much coarser resolution (DMA level), and therefore the impact of population density is unclear. For regions that meet these two criteria, the Facebook posts agreed well with more traditional datasets routinely used for estimating smoke concentrations. In fact, the dataset was often a better predictor of surface $\text{PM}_{2.5}$ than several of the other methods and/or datasets (MODIS AOD, HMS smoke product, WRF-Chem $\text{PM}_{2.5}$). Therefore, the percent of people in a region talking about air quality on Facebook could be useful in determining the spatial extent of exposure between surface monitors.

In further investigating regions and time periods of poor agreement, we noted that the cloud cover negatively impacted our correlations, suggesting that some environmental factors might impact awareness. We also found that in some regions, correlation improved when comparing to the previous or following day, possibly suggesting some influence of social media on awareness. Some of the disagreement could also be due to our search criteria, which could be further refined to reduce the number of false negatives (not recognizing that a post is about air quality) and false positives (including posts that are not about air quality) that likely occur with colloquial conversations. Other studies that have relied on Twitter messages have been able to optimize this process by examining subsets of individual posts (“Tweets”) to test for false positives. However, because this dataset does not provide information on individual posts, this is difficult solely within the dataset, but we do plan to test different search criteria in the future to aid in optimizing our dataset.

Even with some of these limitations, we demonstrated that the percentage of Facebook posters talking about air quality has strong potential for use in estimating exposure to poor air quality. Sachdeva et al. (2016) have shown similar results

with Twitter data, but only for a single fire in California. We believe that Facebook posts could provide some specific advantages over Twitter. Facebook is the most widely used social media site in the US, with 70 % of its participants active daily (Duggan et al., 2015) compared to 36 % for Twitter. Additionally, only 1 % of Twitter posts are georeferenced (Thom et al., 2013), and Google Trends relies on a subset of searches for a large region. In Sachdeva et al. (2016), the actual analysis only included 1297 tweets from a 45-day period covering a region of 40 000 km² in California and Nevada, and their statistical model was built from 705 tweets for a 37-day period covering a 7500 km² area. With a broader user base, Facebook posts could potentially provide better spatial resolution over a broader region. Therefore, this dataset of de-identified, aggregated counts of posters could be very useful for estimating population-level exposure. While we showed that Google Trends data were also moderately well correlated with surface PM_{2.5} in the northwest, results were only available for DMAs; there are only 210 in the US, leading to significantly less spatial information in the Google Trends data than with our percent of Facebook posters (which has results for >20 000 cities in the US). In 2015, there was a broad region of smoke over much of the US; therefore, correlations with Google Trends may be much higher than if we compared to years with only localized smoke events. Finally, we presented a first test case using the percent of Facebook posters in a statistical model to predict surface concentrations in Washington state for June–September 2015; this showed improvements in slope and R^2 values and a reduced error in predicted PM_{2.5}. We plan to extend this work in order to provide improved estimates of smoke exposure for the whole western US for the 2015 summer, which will then be used to quantify the health responses associated with exposure to wildfire smoke. Improving the understanding of these specific health effects can potentially aid the public and decision makers on when and how to take measures to reduce exposure. While social media will not be able to completely replace traditional methods of estimating exposure, social media datasets could improve estimates without the costly investment of additional surface monitors. Using social media datasets as a proxy for exposure also lends itself to an analysis of people's response to and understanding of smoke exposure (Sachdeva et al., 2016), which cannot be measured by traditional exposure methods.

Data availability. The 24 h average concentrations of total PM_{2.5} mass are available from the EPA Air Quality System at www.epa.gov/aqs, and the IMPROVE PM_{2.5} data are also available at <http://views.cira.colostate.edu/fed/>. The Collection 6, MODIS Level 2 10 km AOD products from the Terra and Aqua platforms are available at ladsweb.nascom.nasa.gov. The HMS fire and smoke analysis product is available at satepsanone.nesdis.noaa.gov. Google Trends data are available at www.google.com/trends. Our WRF-Chem model output (daily, 24 h average surface concentrations) is

available at <http://hdl.handle.net/10217/177042> (Ford et al., 2015a). The Facebook data retrieval was conducted internally at Facebook by a Facebook data scientist. To preserve the privacy of Facebook users and in accordance with the data use agreement, we are unable to provide the “Percent of Facebook posters” data. However, we do provide daily maps of the raw and gridded aggregate percent of Facebook posters at <http://hdl.handle.net/10217/177043> (Ford et al., 2015b).

The Supplement related to this article is available online at <https://doi.org/10.5194/acp-17-7541-2017-supplement>.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work was funded by NASA Applied Science grant NNX15AG35G.

Edited by: David Topping

Reviewed by: Sarah Henderson

References

- Abel, F., Hauff, C., Houben, G.-J., Stronkman, R., and Tao, K.: Twitcident: Fighting Fire with Information from Social Web Streams, in Proceedings of the 21st International Conference on World Wide Web, ACM, New York, NY, USA, 305–308, 2012.
- Alman, B., Pfister, G., Hao, H., Stowell, J., Hu, X., Liu, Y., and Strickland, M. J.: The association of wildfire smoke with respiratory and cardiovascular emergency department visits in Colorado in 2012: a case crossover study, *Environ. Health*, 15, 1–9, <https://doi.org/10.1186/s12940-016-0146-8>, 2016.
- Bedo, M., Blanco, G., Oliveira, W., Cazzolato, M., Costa, A., Rodrigues, J., Traina, A., and Traina Jr., C.: Techniques for effective and efficient fire detection from social media images, ArXiv150603844 Cs, available at: <http://arxiv.org/abs/1506.03844> (last Aaccess: 29 November 2016), 2015.
- Brauer, M., Freedman, G., Frostad, J., van Donkelaar, A., Martin, R. V., Dentener, F., Dingenen, R. van, Estep, K., Amini, H., Apte, J. S., Balakrishnan, K., Barregard, L., Broday, D., Feigin, V., Ghosh, S., Hopke, P. K., Knibbs, L. D., Kokubo, Y., Liu, Y., Ma, S., Morawska, L., Sangrador, J. L. T., Shaddick, G., Anderson, H. R., Vos, T., Forouzanfar, M. H., Burnett, R. T., and Cohen, A.: Ambient Air Pollution Exposure Estimation for the Global Burden of Disease 2013, *Environ. Sci. Technol.*, 50, 79–88, <https://doi.org/10.1021/acs.est.5b03709>, 2015.
- Broniatowski, D. A., Paul, M. J., and Dredze, M.: National and Local Influenza Surveillance through Twitter: An Analysis of the 2012–2013 Influenza Epidemic, *PLOS ONE*, 8, e83672, <https://doi.org/10.1371/journal.pone.0083672>, 2013.
- Burke, M. and Kraut, R. E.: The Relationship between Facebook Use and Well-Being depends on Communication Type and Tie Strength, *J. Comput.-Mediat. Commun.*, 21, 265–281, <https://doi.org/10.1111/jcc4.12162>, 2016.

- Crooks, A., Croitoru, A., Stefanidis, A., and Radzikowski, J.: #Earthquake: Twitter as a Distributed Sensor System, *Trans. GIS*, 17, 124–147, <https://doi.org/10.1111/j.1467-9671.2012.01359.x>, 2013.
- Cuchiara, G. C., Li, X., Carvalho, J., and Rappenglück, B.: Inter-comparison of planetary boundary layer parameterization and its impacts on surface ozone concentration in the WRF/Chem model for a case study in Houston/Texas, *Atmos. Environ.*, 96, 175–185, <https://doi.org/10.1016/j.atmosenv.2014.07.013>, 2014.
- Davis, A. Y., Ottmar, R., Liu, Y., Goodrick, S., Achtemeier, G., Gullett, B., Aurell, J., Stevens, W., Greenwald, R., Hu, Y., Russell, A., Hiers, J. K., and Odman, M. T.: Fire emission uncertainties and their effect on smoke dispersion predictions: a case study at Eglin Air Force Base, Florida, USA, *Int. J. Wildland Fire*, 24, 276–285, <https://doi.org/10.1071/WF13071>, 2015.
- De Longueville, B., Smith, R. S., and Luraschi, G.: “OMG, from Here, I Can See the Flames!”: A Use Case of Mining Location Based Social Networks to Acquire Spatio-temporal Data on Forest Fires, in *Proceedings of the 2009 International Workshop on Location Based Social Networks*, ACM, New York, NY, USA, 73–80, 2009.
- Dockery, D. W., Pope, C. A., Xu, X., Spengler, J. D., Ware, J. H., Fay, M. E., Ferris Jr., B. G., and Speizer, F. E.: An association between air pollution and mortality in six US cities, *N. Engl. J. Med.*, 329, 1753–1759, 1993.
- Duggan, M., Elison, N. B., Lampe, C., Lenhart, A., and Madden, M.: *Social Media Update 2014*, Pew Research Center, available at: <http://www.pewinternet.org/2015/01/09/social-media-update-2014/> (last access: 24 August 2016), 2015.
- Emmons, L. K., Walters, S., Hess, P. G., Lamarque, J.-F., Pfister, G. G., Fillmore, D., Granier, C., Guenther, A., Kinnison, D., Laepple, T., Orlando, J., Tie, X., Tyndall, G., Wiedinmyer, C., Baughcum, S. L., and Kloster, S.: Description and evaluation of the Model for Ozone and Related chemical Tracers, version 4 (MOZART-4), *Geosci. Model. Dev.*, 3, 43–67, <https://doi.org/10.5194/gmd-3-43-2010>, 2010.
- Fann, N., Lamson, A. D., Anenberg, S. C., Wesson, K., Risle, D., and Hubbell, B. J.: Estimating the National Public Health Burden Associated with Exposure to Ambient PM_{2.5} and Ozone, *Risk Anal.*, 32, 81–95, <https://doi.org/10.1111/j.1539-6924.2011.01630.x>, 2012.
- Ford, B., Lassman, W., and Pfister, G.: WRF-Chem simulated surface PM_{2.5}, available at: <http://hdl.handle.net/10217/177042> (last access: 16 June 2017), 2015a.
- Ford, B., Pierce, J. R., and Burke, M.: Maps of raw and gridded, population-weighted percent of Facebook Posters matching search criteria, available at: <http://hdl.handle.net/10217/177043> (last access: 16 June 2017), 2015b.
- Gan, R. W., Ford, B., Lassman, W., Pfister, G., Vaidyanathan, A., Fischer, E., Volckens, J., Pierce, J. R., and Magzamen, S.: A comparison of smoke estimation methods and their association with wildfire smoke and cardiopulmonary-related hospital admissions during the 2012 Washington wildfires, *GeoHealth*, 1, <https://doi.org/10.1002/2017GH000073>, 2017.
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., and Brilliant, L.: Detecting influenza epidemics using search engine query data, *Nature*, 457, 1012–1014, <https://doi.org/10.1038/nature07634>, 2009.
- Golder, S. A. and Macy, M. W.: Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures, *Science*, 333, 1878–1881, <https://doi.org/10.1126/science.1202775>, 2011.
- Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P. I., and Geron, C.: Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature), *Atmos. Chem. Phys.*, 6, 3181–3210, <https://doi.org/10.5194/acp-6-3181-2006>, 2006.
- Henderson, S. B., Brauer, M., MacNab, Y. C., and Kennedy, S. M.: Three measures of forest fire smoke exposure and their associations with respiratory and cardiovascular health outcomes in a population-based cohort, *Environ. Health Perspect.*, 119, 1266–1271, <https://doi.org/10.1289/ehp.1002288>, 2011.
- Hu, X., Waller, L. A., Al-Hamdan, M. Z., Crosson, W. L., Estes, M. G., Estes, S. M., Quattrochi, D. A., Sarnat, J. A., and Liu, Y.: Estimating ground-level PM(2.5) concentrations in the southeastern U.S. using geographically weighted regression, *Environ. Res.*, 121, 1–10, <https://doi.org/10.1016/j.envres.2012.11.003>, 2013.
- Jaffe, D., Hafner, W., Chand, D., Westerling, A., and Spracklen, D.: Interannual variations in PM_{2.5} due to wildfires in the Western United States, *Environ. Sci. Technol.*, 42, 2812–2818, <https://doi.org/10.1021/es702755v>, 2008.
- Jiang, W., Wang, Y., Tsou, M.-H., and Fu, X.: Using Social Media to Detect Outdoor Air Pollution and Monitor Air Quality Index (AQI): A Geo-Targeted Spatiotemporal Analysis Framework with Sina Weibo (Chinese Twitter), *PLOS ONE*, 10, e0141185, <https://doi.org/10.1371/journal.pone.0141185>, 2015.
- Johnston, F. H., Henderson, S. B., Chen, Y., Randerson, J. T., Marlier, M., DeFries, R. S., Kinney, P., Bowman, D. M. J. S., and Brauer, M.: Estimated Global Mortality Attributable to Smoke from Landscape Fires, *Environ. Health Perspect.*, 120, 695–701, <https://doi.org/10.1289/ehp.1104422>, 2012.
- Kent, J. D. and Capello Jr., H. T.: Spatial patterns and demographic indicators of effective social media content during the Horsethief Canyon fire of 2012, *Cartogr. Geogr. Inf. Sci.*, 40, 78–89, <https://doi.org/10.1080/15230406.2013.776727>, 2013.
- Kosinski, M., Stillwell, D., and Graepel, T.: Private traits and attributes are predictable from digital records of human behavior, *P. Natl. Acad. Sci. USA*, 110, 5802–5805, <https://doi.org/10.1073/pnas.1218772110>, 2013.
- Lassman, W., Ford, B., Gan, R. W., Pfister, G., Magzamen, S., Fischer, E. V., and Pierce, J. R.: Spatial and Temporal Estimates of Population Exposure to Wildfire Smoke during the Washington State 2012 Wildfire Season Using Blended Model, Satellite, and In-Situ Data, *GeoHealth*, 2017GH000049, <https://doi.org/10.1002/2017GH000049>, 2017.
- Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., and Pozzer, A.: The contribution of outdoor air pollution sources to premature mortality on a global scale, *Nature*, 525, 367–371, <https://doi.org/10.1038/nature15371>, 2015.
- Levy, R. C., Mattoo, S., Munchak, L. A., Remer, L. A., Sayer, A. M., Patadia, F., and Hsu, N. C.: The Collection 6 MODIS aerosol products over land and ocean, *Atmos. Meas. Tech.*, 6, 2989–3034, <https://doi.org/10.5194/amt-6-2989-2013>, 2013.
- Liu, J. C., Pereira, G., Uhl, S. A., Bravo, M. A., and Bell, M. L.: A systematic review of the physical health impacts from non-occupational exposure to wildfire smoke, *Environ. Res.*, 136, 120–132, <https://doi.org/10.1016/j.envres.2014.10.015>, 2015.

- Malm, W. C., Sisler, J. F., Huffman, D., Eldred, R. A., and Cahill, T. A.: Spatial and seasonal trends in particle concentration and optical extinction in the United States, *J. Geophys. Res.-Atmos.*, 99, 1347–1370, <https://doi.org/10.1029/93JD02916>, 1994.
- Masedu, F., Mazza, M., Di Giovanni, C., Calvarese, A., Tiberti, S., Sconci, V., and Valenti, M.: Facebook, quality of life, and mental health outcomes in post-disaster urban environments: the L'Aquila earthquake experience, *Front. Public Health*, 2, 286, <https://doi.org/10.3389/fpubh.2014.00286>, 2014.
- Mei, S., Li, H., Fan, J., Zhu, X., and Dyer, C. R.: Inferring air pollution by sniffing social media, in 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014), 534–539, 2014.
- Misenis, C. and Zhang, Y.: An examination of sensitivity of WRF/Chem predictions to physical parameterizations, horizontal grid spacing, and nesting options, *Atmos. Res.*, 97, 315–334, <https://doi.org/10.1016/j.atmosres.2010.04.005>, 2010.
- Pope, C. A.: Mortality Effects of Longer Term Exposures to Fine Particulate Air Pollution: Review of Recent Epidemiological Evidence, *Inhal. Toxicol.*, 19, 33–38, <https://doi.org/10.1080/08958370701492961>, 2007.
- Pope III, C. A., Ezzati, M., and Dockery, D. W.: Fine-particulate air pollution and life expectancy in the United States, *N. Engl. J. Med.*, 360, 376–386, 2009.
- Punger, E. M. and West, J. J.: The effect of grid resolution on estimates of the burden of ozone and fine particulate matter on premature mortality in the USA, *Air Qual. Atmosphere Health*, 6, 563–573, <https://doi.org/10.1007/s11869-013-0197-8>, 2013.
- Ram, S., Zhang, W., Williams, M., and Pengetnze, Y.: Predicting Asthma-Related Emergency Department Visits Using Big Data, *IEEE J. Biomed. Health Inform.*, 19, 1216–1223, <https://doi.org/10.1109/JBHI.2015.2404829>, 2015.
- Rappold, A. G., Stone, S. L., Cascio, W. E., Neas, L. M., Kilaru, V. J., Carraway, M. S., Szykman, J. J., Ising, A., Cleve, W. E., Meredith, J. T., Vaughan-Batten, H., Deyneka, L., and Devlin, R. B.: Peat bog wildfire smoke exposure in rural North Carolina is associated with cardiopulmonary emergency department visits assessed through syndromic surveillance, *Environ. Health Persp.*, 119, 1415–1420, <https://doi.org/10.1289/ehp.1003206>, 2011.
- Rappold, A. G., Cascio, W. E., Kilaru, V. J., Stone, S. L., Neas, L. M., Devlin, R. B., and Diaz-Sanchez, D.: Cardio-respiratory outcomes associated with exposure to wildfire smoke are modified by measures of community health, *Environ. Health*, 11, 71, <https://doi.org/10.1186/1476-069x-11-71>, 2012.
- Reid, C. E., Jerrett, M., Petersen, M. L., Pfister, G. G., Morefield, P. E., Tager, I. B., Raffuse, S. M., and Balmes, J. R.: Spatiotemporal Prediction of Fine Particulate Matter During the 2008 Northern California Wildfires Using Machine Learning, *Environ. Sci. Technol.*, 49, 3887–3896, <https://doi.org/10.1021/es505846r>, 2015.
- Reisen, F., Meyer, C. P. (Mick), McCaw, L., Powell, J. C., Tolhurst, K., Keywood, M. D., and Gras, J. L.: Impact of smoke from biomass burning on air quality in rural communities in southern Australia, *Atmos. Environ.*, 45, 3944–3953, <https://doi.org/10.1016/j.atmosenv.2011.04.060>, 2011.
- Sachdeva, S., McCaffrey, S., and Locke, D.: Social media approaches to modeling wildfire smoke dispersion: spatiotemporal and social scientific investigations, *Inf. Commun. Soc.*, 0, 1–16, <https://doi.org/10.1080/1369118X.2016.1218528>, 2016.
- Song, W., Jia, H., Huang, J., and Zhang, Y.: A satellite-based geographically weighted regression model for regional PM_{2.5} estimation over the Pearl River Delta region in China, *Remote Sens. Environ.*, 154, 1–7, <https://doi.org/10.1016/j.rse.2014.08.008>, 2014.
- Srinivas, C. V., Prasad, K. B. R. R. H., Naidu, C. V., Baskaran, R., and Venkatraman, B.: Sensitivity Analysis of Atmospheric Dispersion Simulations by FLEXPART to the WRF-Simulated Meteorological Predictions in a Coastal Environment, *Pure Appl. Geophys.*, 173, 675–700, <https://doi.org/10.1007/s00024-015-1104-z>, 2015.
- Tao, Z., Kokas, A., Zhang, R., Cohan, D. S., and Wallach, D.: Inferring Atmospheric Particulate Matter Concentrations from Chinese Social Media Data, *PLOS ONE*, 11, e0161389, <https://doi.org/10.1371/journal.pone.0161389>, 2016.
- Thom, D., Jankowski, P., Fuchs, G., Ertl, T., Bosch, H., Andrienko, N., and Andrienko, G.: Thematic Patterns in Georeferenced Tweets through Space-Time Visual Analytics, *Comput. Sci. Eng.*, 15, 72–82, 2013.
- Thompson, T. M. and Selin, N. E.: Influence of air quality model resolution on uncertainty associated with health impacts, *Atmos. Chem. Phys.*, 12, 9753–9762, <https://doi.org/10.5194/acp-12-9753-2012>, 2012.
- Thompson, T. M., Saari, R. K., and Selin, N. E.: Air quality resolution for health impact assessment: influence of regional characteristics, *Atmos. Chem. Phys.*, 14, 969–978, <https://doi.org/10.5194/acp-14-969-2014>, 2014.
- van Donkelaar, A., Martin, R. V., Levy, R. C., da Silva, A. M., Krzyzanowski, M., Chubarova, N. E., Semutnikova, E., and Cohen, A. J.: Satellite-based estimates of ground-level fine particulate matter during extreme events: A case study of the Moscow fires in 2010, *Atmos. Environ.*, 45, 6225–6232, <https://doi.org/10.1016/j.atmosenv.2011.07.068>, 2011.
- van Donkelaar, A., Martin, R. V., Spurr, R. J. D., and Burnett, R. T.: High-Resolution Satellite-Derived PM_{2.5} from Optimal Estimation and Geographically Weighted Regression over North America, *Environ. Sci. Technol.*, 49, 10482–10491, <https://doi.org/10.1021/acs.est.5b02076>, 2015.
- Wiedinmyer, C., Akagi, S. K., Yokelson, R. J., Emmons, L. K., Al-Saadi, J. A., Orlando, J. J., and Soja, A. J.: The Fire INventory from NCAR (FINN): a high resolution global model to estimate the emissions from open burning, *Geosci. Model Dev.*, 4, 625–641, <https://doi.org/10.5194/gmd-4-625-2011>, 2011.
- Yao, J. and Henderson, S. B.: An empirical model to estimate daily forest fire smoke exposure over a large geographic area using air quality, meteorological, and remote sensing data, *J. Expo. Sci. Environ. Epidemiol.*, 24, 328–335, <https://doi.org/10.1038/jes.2013.87>, 2014.
- Yao, J., Brauer, M., and Henderson, S. B.: Evaluation of a wildfire smoke forecasting system as a tool for public health protection, *Environ. Health Persp.*, 121, 1142–1147, <https://doi.org/10.1289/ehp.1306768>, 2013.
- You, W., Zang, Z., Zhang, L., Li, Y., and Wang, W.: Estimating national-scale ground-level PM_{2.5} concentration in China using geographically weighted regression based on MODIS and MISR AOD, *Environ. Sci. Pollut. Res.*, 23, 8327–8338, <https://doi.org/10.1007/s11356-015-6027-9>, 2016.
- Youyou, W., Kosinski, M., and Stillwell, D.: Computer-based personality judgments are more accurate than those made

- by humans, *P. Natl. Acad. Sci. USA*, 112, 1036–1040, <https://doi.org/10.1073/pnas.1418680112>, 2015.
- Žabkar, R., Koračin, D., and Rakovec, J.: A WRF/Chem sensitivity study using ensemble modelling for a high ozone episode in Slovenia and the Northern Adriatic area, *Atmos. Environ.*, 77, 990–1004, <https://doi.org/10.1016/j.atmosenv.2013.05.065>, 2013.
- Zhang, F., Wang, J., Ichoku, C., Hyer, E. J., Yang, Z., Ge, C., Su, S., Xiaoyang Zhang, Kondragunta, S., Kaiser, J. W., Wiedinmyer, C., and Silva, A. da: Sensitivity of mesoscale modeling of smoke direct radiative effect to the emission inventory: a case study in northern sub-Saharan African region, *Environ. Res. Lett.*, 9, 75002, <https://doi.org/10.1088/1748-9326/9/7/075002>, 2014.